Econometrica, Vol. 64, No. 2 (March, 1996), 343–355

WHY ARE CERTAIN PROPERTIES OF BINARY RELATIONS RELATIVELY MORE COMMON IN NATURAL LANGUAGE?

BY ARIEL RUBINSTEIN[1]

The aim of this paper is to explain the fact that certain properties of binary relations are frequently observed in natural language while others do not appear at all. Three features of binary relation are studied: (i) the ability to use the relation to indicate nameless elements; (ii) the accuracy with which the vocabulary spanned by the relation can be used to approximate the actual terms to which a user of the language wishes to refer; (iii) the ease with which the relation can be described by means of examples. It is argued that linear orderings are optimal according to the first criteria while asymmetric relations are optimal according to the second. From among complete and asymmetric relations (tournaments), those which are transitive are optimal according to the third criterion.

KEYWORDS: Binary relations, natural language, linear orderings, optimality.

1. INTRODUCTION

ECONOMIC THEORY ATTEMPTS TO EXPLAIN REGULARITIES in human interaction. This paper is based on methods rooted in economic theory and applies formal approaches to deal with questions regarding what is probably the most fundamental nonphysical regularity in human interaction: the natural language.

The object of the investigation is binary relations defined within a *finite* set Ω. A binary relation within a set Ω, is specified as a connection between pairs of elements within Ω. A binary relation on a set Ω (or its extension) is taken to be a subset of $\Omega \times \Omega$-$\{(\omega, \omega) | \omega \in \Omega\}$ (the set of all pairs of members of Ω with the exception of the "diagonal"). Thus no element relates to itself. The reason for this condition is that typically the exclusion of self-reference is a matter of analytical convenience and not of substance.

Binary relations are common in natural language. For example, "Person x knows person y," "tree x is to the right of tree y," "picture x is similar to picture y," "chair x and chair y are both brown," and so on.

The nature of many binary relations requires them to satisfy certain properties. We refer to a property of the relation R as a sentence in the language of calculus of predicates which includes a name for the binary relation R, variable names, connectives, and qualifiers, but does not include any individual names of the set of objects Ω. For example, the relation "x is a neighbor of y" must, in

[1]This project was begun during the academic year 1982/3, while I was a Fellow at the Institute for Advanced Studies at the Hebrew University. The second phase, during the year 1992/3, was partially supported by the United States-Israel Binational Science Foundation, Grant No. 1011-341.

I am grateful to Tzachi Gilboa, Asa Kasher, David Lewis, Isaac Levy, Benni Moldevano, Avishai Margalit, Joseph Raz, Amos Tversky, Ruth Weintraub, Asher Wolinsky, and the editor and a referee of this journal for their comments and encouragement. Special thanks to Noga Alon (from the Mathematics Department at Tel Aviv University) for his key contribution to Proposition 3.

any acceptable use of this relation, satisfy the symmetry property (if x is a neighbor of y, then y is a neighbor of x). The relation "x is to the right of y" must satisfy the properties of a *linear ordering*: completeness (for every $x \neq y$, either x relates to y or y to x); asymmetry (for every x and y, if x relates to y, y does not relate to x); and transitivity (for every x, y, and z, if x relates to y and y to z, then x relates to z).

On the other hand, the nature of many other binary relations, such as the relation "x loves y," does not imply any specific properties that the relation must satisfy. It may be true that in a particular reference to this relation whenever "x loves y" then also "y loves x." However, there is nothing in our understanding of the relation "x loves y" which makes that symmetry necessary; all binary relations are feasible with regard to this relation.

In fact, the real object of our investigation is to examine the *properties* of binary relations which appear in natural language. The number of these properties is small although in principle one could define an infinite number of them. It is difficult to find natural examples of binary relations which must satisfy even simple properties such as the following:

P1: *For every x and y the number of z for which xRz is equal to the number of z for which yRz.*

P2: *If xRy and xRz* ($y \neq z$), *then for every a such that both yRa and zRa also xRa.*

Is it just a coincidence that human perception uses certain properties of binary relations and not others?

The argument presented in this paper is based on the following premises: (i) In order to function, a natural language can include only a small number of structured binary relations; (ii) binary relations have several functions in natural language; and (iii) evolutionary forces make it more likely that the "optimal" structures are observed in natural language.

The first premise states that language is inherently equipped with few of the properties of binary relations. If this were not the case, it would be difficult to understand how human beings are able to deduce from a small number of observations such as "a is better than b" that the relation "being better than" is a linear ordering. Only if the number of possible structures is limited, can a small number of observations enable the user of a language to identify the appropriate type of a relation.

This premise has been the subject of classical philosophical discussions. The riddle of induction is a fundamental philosophical inquiry into the factors which confirm an induction (see, for example, Goodman (1972), Quine (1969), and Watanabe (1969)). Watanabe concludes: "...we are attaching non-uniform importance to various predicates, and that this weighting has an extra-logical origin" (Watanabe (1969, p. 376) and Quine state that there is no escape from assuming that a certain kind of predicate (like "green" and not like "grue" in Goodman's paradox) has a preferred status called "natural kind" (see Quine

(1969)). From a philosophical perspective, the current paper is an attempt to examine the origins of special natural kinds of binary predicates.

This thesis is also rooted in modern linguistics, a field which attempts to explain the fact that speakers of a language are able to follow its grammatical rules after listening to only a small number of sentences (see, for example, Piattelli-Palmarini (1970)). The underlying thesis is that the brain's hardware comes equipped with universal rules that are compatible with a small number of grammatical structures and which enable the speaker to identify grammatical rules from only a few examples.

The second premise is the central one of this paper. It states that binary relations have certain useful functions in human interaction. Three criteria of the usefulness of a binary relation are examined. It is argued that the fact that certain properties are more common in natural language can be credited to their superior performance according to these criteria. Let us briefly introduce the criteria and the derived conclusions.

Indication-friendliness: Binary relations often enable the user to indicate "nameless" elements in a set. Consider, for example, the phrase "the second tree from the right." The comprehensibility of the phrase depends on the familiarity of the users with the binary relation "being on the left." The phrase is used in circumstances in which the trees within a well-defined group do not have distinguishable names familiar to both the speaker and the listener. The relation "being to the left of" is a linear ordering. It will be shown that only linear orderings allow the user to construct, for every subset and for every element in the subset, a sentence which can single out an element from among the members of the subset.

Informativeness: A binary relation may be used to convey information regarding the connection between objects. For example, if Ω is a set of cities and a manager of a taxi company wishes to instruct a driver what requests for going from a city x to the city y he should accept. For the description of the relation he can use only existing vocabulary which may be too poor. Using an "imprecise" relation causes "loss" which he will try to minimize. Our investigation is in the level of the "planner of the vocabulary:" a preferable structure is one which spans a vocabulary which minimizes the expected "loss" inflicted on the user. It will be shown that any structure which includes the asymmetry requirement is "nearly" optimal in this respect.

Ease of Describability: Assume that a "father" wants to pass on to his son the priorities necessary on a hunting trip in the forest. Let Ω be the set of types of animals. Given that the properties of linear ordering are understood by the son, the father can provide him with a list of observations, that is, statements of the type "animal a is more desirable than animal b." Given that the number of elements in Ω is n, the minimal number of observations he has to make is $n-1$ (the $n-1$ observations a_iRa_{i+1} where $a_1Ra_2Ra_3 \dots Ra_{n-1}Ra_n$). In general, we may ask what is the minimum number of observations which are necessary to describe the relation?

Note that this is not the same problem as that of discovering the content of a binary relation by a user who is informed about the structure of the relation but is not instructed on how to expose its content. For example, if all orderings on a three-element set are equally likely, the user will need an average of 2.5 observations to discover a linear ordering at the same time that an instructor can describe that linear ordering using only two observations.

It is conjectured that for sets with more than five elements, a linear relation is an "efficient" structure (or at least "almost efficient") in the sense that the process of describing it requires the least number of observations from among the complete asymmetric relations. That is, for $n > 5$, there is no way to define any complete asymmetric binary relation with less (or at least significantly less) than $n - 1$ observations.

The third premise, which links the first two, is that evolutionary forces select structures which are optimal or nearly optimal with respect to the functions they fulfill. Even if a linguistic engineer does not exist, evolutionary forces "prefer" human beings who are equipped with structures which permit better forms of communication. This idea, which is popular in Economics, was also noted by philosophers. For example, Quine states: "If people's innate spacing of qualities is a gene-linked trait, then the spacing that has made for the most successful inductions will have tended to predominate through natural selection" (Quine (1969, p. 126)).

To summarize, the goal of this paper is to spell out several functions for binary relations in natural language and to derive the properties which an optimally designed binary relation would require. There are other criteria to evaluate types of binary relations, such as the usefulness for simplifying problems (a function which is carried out by similarity relations). The choice of the three functions presented in this paper was motivated by the availability of analytical results and does not express an evaluation of their relative importance.

2. INDICATION-FRIENDLINESS

The user of a language may wish to refer to elements in a set Ω in order to point out a certain element to a second party. If the element has a name, he may use it. If there is no mutually recognized "name" for the element and the two parties understand the meaning of a certain binary relation defined on Ω, the user may utilize the relation to define the element. "The third tree on the right" is a way of indicating one object out of many by using the relation "x stands to the left of y." The phrase "the seventh floor" is a way of indicating a location in a building given the linear ordering "floor x is above floor y." On the other hand, the relation "the symbol a on the clock is clockwise to the symbol b (in the smallest angle)" does not enable the user to indicate a certain line on the clock without having at least one designated element (such as "12 o'clock").

In this section, a binary relation is evaluated as a tool for indicating elements in a set in which the objects do not have names. In light of this function, a relation may be evaluated according to whether it enables the user to refer unambiguously to all elements in the set or in any subset of it. We are led to the following definition:

DEFINITION: A binary relation R on a set Ω is *indication-friendly* if for every subset A of Ω and every element $a \in A$, there is a formula $f_{a,A}(x)$ in the language of the calculus of predicates with the name of the binary relation (and without individual constants) so that a is the only element in X satisfying it.

Any linear ordering is indication-friendly. For every subset X of Ω the formula $P_1(x) \equiv \forall y(x \neq y \rightarrow xRy)$ defines the "maximal" element in the set. The formula $P_2(x) \equiv \forall y(x \neq y$ and $-P_1(y) \rightarrow xRy)$ defines the "second to the maximal" element and so on.

On the other hand, consider the set $\Omega = \{a, b, c, d\}$ and the nonlinear binary relation depicted in Diagram 1 (aRb, aRc, dRa, bRd, bRc and cRd). In the set Ω the element a is defined by "x relates to two elements, one of which also beats two elements." The element b is defined by "x beats two elements, which beat one element each." And so on. However, whereas the relation R allows the user to define any element in the set Ω, the relation is not effective in defining elements in the subset $\{a, b, d\}$.

PROPOSITION 1: *Let Ω be a finite set. A binary relation R is indication-friendly if and only if R is a linear ordering.*

PROOF: We have already indicated that if R is a linear ordering on Ω, then for every subset A and every $a \in A$ there is a formula indicating a. On the other hand, assume that R is indication-friendly. In order to indicate the two elements in a two member set $A = \{a, b\}$, it must be that either aRb or bRa but not both. In order to indicate the three elements in a three element set $A = \{a, b, c\}$ it must be that in addition there is no cycle. *Q.E.D.*

3. INFORMATIVENESS

A binary relation on Ω is sometimes used as a means to transfer or store information concerning a certain relationship between elements of Ω. Consider for example, a case in which the user of the language is interested in describing the relation "student x read the paper of student y" within a group of Ph.D. students. This relation by nature does not have any particular structure. The

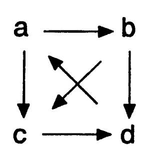

DIAGRAM 1

user of the language may describe the relation by listing the pairs of students who satisfy the relation. Alternatively, he could use a known binary relation which is "close" enough to the relation he wants to describe. He may use the sentence "any x who is older than y" read his paper. This may not be entirely correct and he may add that "the exceptions are a who did not read b's paper (though he is older) and c who read d's paper (though he is younger)." Thus the relation "older than" helped the user to describe the relation he really wished to describe and in some sense we can say that the distance between the two relations is 2. Note that the relation "older than" allows the user to also infer its negation ("x is younger than"), the universal relation "every x read the paper of every y" and the empty relation ("no student read any of the other students' papers").

Thus, in general, when the user wishes to refer to a specific relation, he may find the language insufficient to describe the relation and will have to use the existing vocabulary to approximate the relation. In this section, binary relations are evaluated according to the effectiveness of the vocabulary which they span.

Consider the designer of a language who is able to design only one binary relation at the "initial stage of the world," "beyond the veil of ignorance." Of course, real life language includes many relations and each one's effectiveness depends on the entire set; the assumption that the designer is planning only one binary relation is made for analytical convenience.

The user forms a vocabulary spanned by the basic binary relation. It is assumed in this analysis that the vocabulary includes four relations: the relation, its negation, the universal relation ("everything") and the empty relation ("nothing"). (Benni Moldevano pointed out to me that one can think of other spanned vocabularies, such as the one which will also include the relation S defined by xSy if "there is z such that xRz and zRy.") When the "world" is realized, the user may wish to refer to some relation which does not necessarily belong to the vocabulary spanned by the basic binary relation. It is assumed that the user uses the relation in the vocabulary which is the "closest" to the one he really intends. However, we have yet to define what is meant by "closeness." The distance between any two binary relations R' and R'' is taken to be the number of pairs (a, b) for which it is not true that $aR'b$ iff $aR''b$. Note that by this measure any pair for which R' and R'' disagree gets the same weight. "Beyond the veil of ignorance," the relative importance of these mistakes is not known and thus giving equal weights to all mistakes seems a reasonable approach.

Finally, in designing the binary relation, the planner wishes to minimize the expected "loss" inflicted on the user by optimal use of inaccurate relations. Again, "beyond the veil of ignorance," it seems proper to assume that he assigns equal probabilities a priori to all possible binary relations that the user may wish to refer to.

For now we will abstract from the terminology of choosing a binary relation: Let X be a (finite) set. For any S, a subset of X, let $V(S) = \{X, S, -S, \phi\}$ be the vocabulary spanned by S. When the user of the language wishes to refer to a subset $Z \subseteq X$, he uses an element in $V(S)$ which is the "closest" to Z. We define the loss function of adhering to A rather than to B as the cardinality of the asymmetric difference between A and B (the set of all elements which are in A and not in B or in B and not in A), that is $d(A,B) = |(A-B) \cup (B-A)|$. One interpretation of this distance function corresponds to the case in which the user who wishes to refer to a set Z, and uses a term A in the vocabulary, lists elements in Z which are excluded from or appended to the term A (the sentence "you may eat bread or any fruit with the exception of apples and bananas" has three exceptions), and bears a "cost" proportional to the number of elements he has to exclude or append.

For a set B and a vocabulary V, define $\delta(B,V) = \min_{A \in V} d(A,B)$, the distance of the set B from the closest set in the vocabulary V. Assigning equal probabilities a priori to all possible sets to which the user may wish to refer, the problem to be solved by the designer is $\min_S \Sigma_{Z \subseteq X} \delta(Z, V(S))/2^{|X|}$. Clearly, the problem boils down to the choice of the *number* of elements in the optimal set S.

To clarify the nature of the problem, verify the detailed calculation for $|X| = 4$. Consider a set S which contains 2 elements. The user can refer to a set Z, which is 1 of the sets X, ϕ, S or $-S$, without bearing any loss. He can approximate a set Z, which contains 1 or 3 elements, by using the set ϕ or X, respectively, and thus bears a loss of 1. If he wishes to refer to 1 of 4 2-element sets which are not S or $-S$, he bears a loss of 2. Thus, the average loss is $[4(0) + 8(1) + 4(2)]/16 = 1$. Similar calculations lead to the conclusion that the choice of S as an empty or 4-element set leads to the expected loss of $5/4$ (no loss for ϕ and X, a loss of 1 for the 8 sets of size 1 and 3, and a loss of 2 for the 6 sets of size 2, i.e., $[2(0) + 8(1) + 6(2)]/16 = 5/4$). If S is taken to be a 1- (or 3-) element set, the average loss is only $[4(0) + 6(1) + 6(1)]/16 = 3/4$. Thus, in this case, splitting X into 2 *unequal* categories of sizes 1 and 3 yields smaller expected imprecision.

Despite this example, the intuition that choosing S's size to be half that of X is quite correct.

PROPOSITION 2: *Let X be an n-element set. Assume that a set Z is drawn with uniform probability from the set of all subsets of X. Denote by $L_n(k)$ the expected loss from optimal use of the vocabulary spanned by a k-member subset of X. Then, $n/2$ "almost" minimizes $L_n(k)$ (the difference between $L_n(n/2)$ and the minimum loss is in the magnitude of $1/\sqrt{n}$).*

PROOF: See Appendix.

We now come to the final stage of the argument. Let Ω be a finite set. Let $X = \Omega \times \Omega - \{(\omega, \omega) | \omega \in \Omega\}$ be the set of all pairs of distinct elements of Ω. Recall that we identify a binary relation as a subset of X. The designer's aim is to reduce the expected number of "conflicts" between the relation the user may have in mind and the one he will be able to refer to by using the vocabulary spanned by the relation which the designer constructs. Following Proposition 2, an optimal binary relation will include (approximately) half of the pairs of X. If the designer wants the relation R to also be effective when the reference set is any subset A of Ω, any induced relation $R_{|A}$ should include (approximately) half of the pairs (a, b). This is equivalent to the condition that the relation R has to be both complete and asymmetric.

We can now wrap up the argument. Our fictitious planner wishes to design a binary relation which will span the user's vocabulary. His goal is to decrease the expected inaccuracy of the term which the user will use. Viewing a relation as a set of pairs of elements, the problem was linked to the design of a category in a set of elements where choosing half the members of the set is nearly optimal. Requiring that the relation induces half of the possible pairs of elements in any subset of Ω is equivalent to requirements of completeness and asymmetry.

4. EASE OF DESCRIBABILITY

A user may wish to communicate the content of a binary relation to someone else. Different types of binary relations may be transferred with varying degrees of difficulty. In this section, it is argued that from among the complete and asymmetric binary relations (called *tournaments*), those which are transitive (and thus are linear orderings) are (nearly) optimal in this respect.

To motivate the definitions consider the linear ordering $aRbRcRd$ on the set $\Omega = \{a, b, c, d\}$ of size $n = 4$. There are 64 tournaments, on this set, 24 of which are linear orderings. Assuming that R is a linear ordering, only $n - 1 = 3$ observations (aRb, bRc, and cRd) are needed to characterize it. In contrast, if the only property known to the listener is that the relation is a tournament, $n(n-1)/2 = 6$ observations are needed to characterize the relation.

In general, we will analyze a situation in which a new user of the language is aware of a set Ω and of the structure of a binary relation (a list of properties it satisfies). He becomes acquainted with the structure of the relation either because it is embedded in his "hardware" or because a veteran user informs him of the properties of the relation. An *observation* of the relation R is a statement of the type $\alpha R \beta$ where α and β are names of objects in Ω. The veteran user makes a number of observations which allow the listener to complete the acquisition of the relation (using logical operations) as a unique complete and asymmetric binary relation on Ω. The complexity of the process of acquiring a relation is measured by the minimum number of observations that are required. The complexity of making the inferences is ignored, which is hardly realistic.

Formally, we say that $\langle f, \{a_i R b_i\}_{i \in I} \rangle$ defines the binary relation R on Ω if: (i) f is a sentence in the language of predicate calculus with one binary relation name, R and for all $i, a_i, b_i \in \Omega$, and (ii) R is the unique binary relation on Ω satisfying the sentence f and the set of observations $\{a_i R b_i\}_{i \in I}$.

We will measure the complexity of the definition by the number of elements in the set I and we will measure the complexity of a binary relation R, $l(R)$ by the minimal complexity of any of its definitions.

EXAMPLE 1: Consider the tournament R on the set $\Omega = \{a, b, c\}$ where aRb, bRc, and cRa. The sentence f, which states that R is complete, asymmetric and anti-transitive $(\forall x, y, z(xRy \text{ and } yRz \rightarrow -xRz))$ with the unique observation aRb, is a definition of R. Thus $l(R) = 1$.

EXAMPLE 2: Let $\Omega = \{a_1, a_2, \ldots, a_n\}$ and let R be a linear relation on Ω. The sentence f, expressing completeness, asymmetry, and transitivity with the $n-1$ observations $\{a_i R a_{i+1}\}_{i=1\ldots n-1}$, consists of a definition. Obviously, there is no definition of a linear relation with less than $n-1$ observations. Thus, $l(R) = n-1$.

EXAMPLE 3: Let $\Omega = \{a, b, c, d\}$ $(n = 4)$ and let R be a tournament so that aRx for all x and $bRcRdRb$. That is, a "beats" the other 3 members which are cyclically related one to the other. The sentence f, $\exists wxyz[wRx, wRy, wRz, xRy, yRz \text{ and } zRx]$, with the 3 observations aRb, aRc, and bRc, is a definition of the relation: a must stand for w in the sentence f as he "beats" more than 1 element. The observation bRc then determines the "direction of the cycle" in $\{b, c, d\}$.

EXAMPLE 4: Consider the relation R on the set $\Omega = \{a, b, c, d\}$ described in Diagram 1. Choose the sentence f to be $\exists x_1, \ldots, x_4(\Lambda_{a_i R a_j} x_i R x_j)$. There are 24 different binary relations on Ω which satisfy f. The 3 observations aRb, bRc, and cRd are consistent with 4 different relations which satisfy f, and the 3 observations bRc, bRd, and dRc are consistent with 2 relations which satisfy the structure. It is easy to verify that $l(R) = 4$.

The mathematical problem we want to impose in this section is

$$\min_{R \text{ is a tournament}} l(R).$$

Note that for a given (finite) set of objects $\Omega = \{a_1, \ldots, a_n\}$, the "structure" of any binary relation R can be fully expressed by the sentence $\phi_R(x_1, \ldots, x_n) = \exists x_1, \ldots, x_n(\Lambda_{a_i R a_j} x_i R x_j)$. Therefore, the above optimization problem can be restated as follows: Start with a tournament in which the names of the vertices have been erased and find the minimum number of observations required to discover the names of the vertices (up to isomorphism).

In the first version of this paper I presented the above mathematical question and conjectured that when $n > 3$, $l(R) \geq n-1$. Fishburn, Kim, and Tetali (1994) investigated the problem. They found one additional counterexample for $n = 5$ and showed that these two relations are the only exceptions for $n \leq 7$.

EXAMPLE 5 (Fishburn, Kim, and Tetali (1994)): Refer to Diagram 2. This relation, R, defined on the 5-element set $\Omega = \{a, b, c, d, e\}$ satisfies $l(R)$. It is nicely characterized by the fact that for every x there are precisely 2 elements

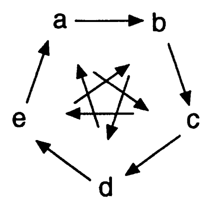

DIAGRAM 2

"beaten" by x. The 3 observations aRb, aRc, and eRb, define the relation through the chain of conclusions $\{dRa, eRa\}$, $\{cRe, dRe\}$, $\{bRd, cRd\}$, and finally bRc.

Examples 2 and 5 show that for $n = 3$ and $n = 5$, linear ordering is not the most "economical" structure. Are there any other binary relations, where $n > 7$, which are defined with less than $n - 1$ observations? I am not able to prove the conjecture that for $n > 5$, $l(R) \geq n - 1$. The following weaker proposition has been suggested and proved by Noga Alon:

PROPOSITION 3: *For any ε there is $n(\varepsilon)$ such that for any $n > n(\varepsilon)$, $l(R) > (1 - \varepsilon)n$ for any asymmetric complete relation R.*

PROOF: See Appendix.

Thus, at least for large sets, linear orderings are almost optimal with respect to the criterion of minimizing the number of observations which are required for their definition.

Notice that in the above discussion, we allowed the relation to be defined by a formula which depends on the number of elements in the set Ω. Linear orderings, are defined by a formula which does not depend on the number of elements in the set Ω. This leads to another weaker conjecture:

CONJECTURE: *Let ϕ be a sentence in the predicate calculus language which includes a single name of binary relation, R. Then, there exists n^* such that for any $|\Omega| \geq n^*$ and any tournament R which is defined by the sentence ϕ, $l(R) \geq |\Omega| - 1$.*

I conclude with one comment on the use of the term "describability" and not "learnability." In the mechanism discussed in this section, the instructor chooses the observations to be observed. The number $l(R)$ is the minimum number of observations which are *chosen* by the instructor to be revealed. When he chooses the set of observations the instructor knows the relation which he wishes to describe. The "student," on the other hand, if he wishes to acquire the content of the relation, has to "search in the dark." An alternative related measure of the complexity of the relation is the expected number of inquiries that the student, knowing only the structure of the relation, will have to make in order to conclude the content of the relation. In this respect cycle (one observation required to learn the relation) is preferred over linear ordering (an average of 2.5 observations are required).

5. CONCLUDING REMARKS

This paper focused on three functions which may be associated with binary relations. The optimality of certain properties of binary relations with respect to these functions is taken as a potential explanation of the fact that these properties are relatively common in natural language. Of course, the list of functions is partial and there are others which could provide alternative explanations for the commonality of these and other properties (such as those which define the equivalence and similarity relations).

This paper dealt only with binary relations. Similar questions could be asked regarding the principles which guide our classification of objects (unary relations). Why is the class of furniture divided into the set of chairs and the set of tables and not into another partition in which some of the chairs and some of the tables are classified together in a class of "chables" and others classified into a separate class of "tairs"? Why is the class of chairs a more basic category than the class of furniture or the class of armchairs (see, for example, E. Rosch and B. B. Lloyd (1978))?

The part of the paper which deals with "informativeness" is connected with several theoretical economic issues concerning the design of information to be stored or transferred. One aspect of bounded rationality is that decision makers are constrained in the information structure they can use. Given his constraints, the decision maker (or evolutionary force) selects what he will know. Note, in particular, Dow (1991), which analyzes a simple search model in which a searcher travels between two stores searching for a cheap indivisible unit of some good. At the first store, he observes the price but can retain only one out of a given finite number of potential symbols in his memory. At the second store, based on his memory and the price at the second store, he has to make a final decision as to where to buy the good. Dow then analyzes the consumer ex-ante problem of designing the optimal classification of the set of possible prices in the first store.

A most exciting research challenge is the search for an economic-theoretical explanation of the principles of universal grammar which according to modern linguistics, is part of human's hardware.

Dept. of Economics, Tel Aviv University, Tel Aviv, Israel 69978
and
Dept. of Economics, Princeton University, Princeton, NJ 08544, U.S.A.

Manuscript received January, 1995; final revision received April, 1995.

APPENDIX

PROOF OF PROPOSITION 2: (a) We first show that if S is a set of size k, then the expected distance of Z from $V(S)$ is

$$L_n(k) = n/2 - k/2^k C([(k-1)/2], k-1)$$
$$-(n-k)/2^{n-k} C([(n-k-1)/2], n-k-1)$$

where $C(\alpha, \beta) = (\beta!/\alpha!(\beta-\alpha)!)$ and $C(m/2, m) = 0$ where m is odd.

The proof is combinatorial. First, notice that

$$\delta(A, V(S)) = \delta(A \cap S, \{\phi, S\}) + \delta(A \cap (N-S), \{\phi, N-S\}).$$

This follows from the fact that the closest set to A in $V(S)$ is the union of the set closest to $A \cap S$ from among $\{\phi, S\}$ and the set closest to $A \cap (N-S)$ from among $\{\phi, N-S\}$.

The user's expected loss, multiplied by 2^n, is:

$$\Sigma_A \delta(A, V(S)) = \Sigma_A \delta(A \cap S, \{\phi, S\}) + \delta(A \cap (N-S), \{\phi, N-S\})$$
$$= 2^{n-k} \Sigma_{B \subset S} \delta(B, \{\phi, S\}) + 2^k \Sigma_{B \subset N-S} \delta(B, \{\phi, N-S\})$$
$$= 2^{n-k} U_k + 2^k U_{m-k}$$

where $U_m = \Sigma_{B \subseteq M} \delta(B, \{\phi, M\})$ and M is a set of m elements.

It is claimed that $U_m = m[2^{m-1} - C([(m-1)/2], m-1)]$. To see this, take any element $b \in M$. The number of pairs $(B, M-B)$ in which b appears in the smallest set is the number of pairs (B_1, B_2) where $B_1 \cap B_2 = \phi$ and $B_1 \cup B_2 = M - \{b\}$, with the exception of those in which $|B_1| = |B_2|$.

By algebraic manipulation we now obtain that the expected loss is:

$$L_n(k) = n/2 - k/2^k C([(k-1)/2], k-1)$$
$$-(n-k)/2^{n-k} C([(n-k-1)/2], n-k-1).$$

(b) When n is large, $k = n/2$ "almost" minimizes L_n (that is, the difference between $L_n(n/2)$ and the minimum loss is in the magnitude of $1/\sqrt{n}$). To prove (b), define $\pi^* = 1/\sqrt{2\pi}$. Use Sterling's formula to replace $l!$ with $l^{l+1/2} e^{-l}/\pi^*$ and $C(l/2, l)$ with $(2/\pi^*) 2^l/\sqrt{l}$. Then $L_n(k)$ behaves asymptotically as

$$L'_n(k) = n/2 - \pi^* k/\sqrt{k-1} - (n-k)\pi^*/\sqrt{n-k-1}.$$

The function L'_n satisfies

$$n/2 - \pi^* \sqrt{k-1} - \pi^* \sqrt{n-k-1} \geq L'_n(k) \geq n/2 - \pi^* \sqrt{k} - \pi^* \sqrt{n-k}.$$

Therefore the value of $\max_k L_n(k) - L_n(n/2)$ is in the magnitude of $1/\sqrt{n}$.

Finally, by manipulation of $L'_n(n/2)$ we obtain that the minimal value of L_n is approximately $n/2 - \sqrt{n}/\sqrt{\pi}$. *Q.E.D.*

PROOF OF PROPOSITION 3: Let $C(\alpha, \beta) = (\beta!/\alpha!(\beta-\alpha)!)$ be the number of subsets of size α in a β-element set). Let r be the largest integer for which $n > r3^{C(2,r)}$. It will be proved that $l(R)$ is at least $n - 2n/r = n(1-2/r)$. This is sufficient for the proof since r is in the order of $(\log n)^{1/2}$. Assume that there is a set H of less than $n - 2n/r$ observations which defines the tournament R. The set H has at least $2n/r$ connected components. At most, n/r of them have up to r vertices. The total number of tournaments on r vertices is $3^{C(2,r)}$. Therefore, by the choice of r, it must be that two of the connected components are isomorphic. Consider any two vertices, v_1 and v_2, in the two components. If $v_1 R v_2$, then there is a tournament R' which extends H, is isomorphic to R and in which $v_2 R' v_1$. Thus, H does not define the relation R. *Q.E.D.*

REFERENCES

DOW, J. (1991): "Search Decisions with Limited Memory," *Review of Economic Studies*, 58, 1–14.

FISHBURN, P., J. H. KIM, AND P. TETALI (1994): "Tournament Certificates," AT&T Bell Labs, mimeo.

GOODMAN, N. (1972): *Problems and Projects*. Indianapolis and New York: Bobbs-Merrill.

PIATTELLI-PALMARINI, M., ED. (1970): *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge: Harvard University Press.

QUINE, W. V. (1969): *Ontological Relativity and Other Essays*. New York: Columbia University Press.

ROSCH, E., AND B. B. LLOYD, EDS. (1978): *Cognition and Categorization*. Hillsdale: Lawrence Erlbaum Associates.

WATANABE, S. (1969): *Knowing and Guessing*. New York: John Wiley.