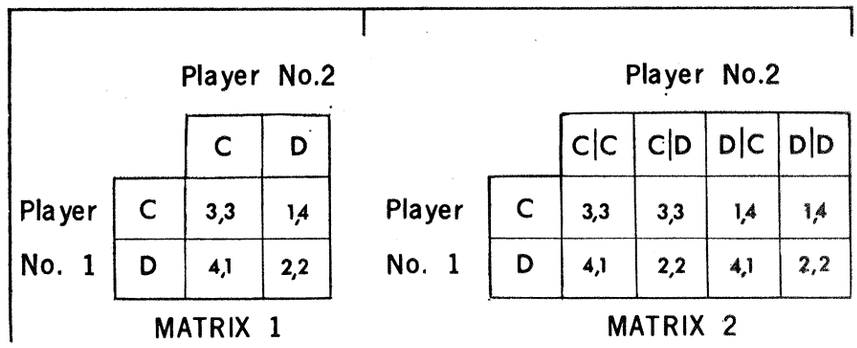



Matrix 1 represents the game known as Prisoner's Dilemma, in which the use of "individual rationality" leads to an outcome that is worse for both players than another outcome they could have attained (see text). The name comes from a classic example in which two suspects are taken into custody and separated. They are told by the district attorney that he does not have adequate evidence to convict them in a trial; that if neither confesses he will have them booked on a trumped-up minor charge with a one-year sentence; that if both confess they will both be prosecuted but he will recommend leniency (a five-year sentence); and that if one confesses and the other does not, the confessor will be released for turning state's evidence and the other will have the book thrown at him (ten-year sentence). (Note that strategy D is to confess.) Matrix 2 represents the "2-metagame" taken from Matrix 1.

life participants in a situation modeled by a game—must ultimately make decisions on the level of the basic game, not some abstract construction formulated by a nonplayer (nonparticipant) called a game theorist. This position has been argued by Harris (*4*).

In answer to this question, Howard presents three interpretations of what it means to consider progressively higher meta-levels. He says that the sequence game, 2-metagame, 1-2-metagame, and so on "may be traversed: (i) in the mental processes of a single player, as he reasons 'if I do this, he will do that; but if he is going to react in that way, I should do this . . . '; (ii) in the process of bargaining and negotiation between players; (iii) in the process of making physical moves in a crisis situation (e.g., a post-war 'Berlin crisis') with the object of conveying one's intentions to the other players" (p. 101). Real players, he argues, actually negotiate in terms of essentially metagame concepts. An example of second-level thinking is " 'If it were the case (I do not say it is) that my client could pay such and such damages, would you be willing to settle out of court?' " (p. 99).

Rapoport (*5*) calls this work a genuine "escape from paradox." To resolve a paradox, says Rapoport, one must find the logical error that leads to the apparent contradiction which constitutes that paradox. In this case one might simply "recognize that 'rationality' has two different, irreconcilable meanings in this context. Howard's metagame model goes farther. There the collectively rational strategy becomes also *individually* rational."

This interest in the mathematical transcending of a paradox is not Howard's central theme. In a reply (*6*) to both Harris and Rapoport, he stresses two points: that stability of outcomes is the crux of the theory, and that the theory is predictive or empirical, not normative. Though predictive, the theory "does not predict what players will do, but only what they will do under conditions of *stability*, i.e., when each somehow succeeds in predicting the other's choice."

The book treats many other provocative and significant points, which can only be touched fleetingly here: metagame analysis for individuals in a coalition in a game with more than two players; the relationship of "free will" to the concept of choosing a metastrategy; and perhaps most important the third, previously unnoted, "breakdown of rationality."

This third breakdown hinges on the notion of a "sure-thing strategy," one which is best no matter what you think the other player will do. Use of such a strategy might seem unimpeachably "rational." Now if one of two players has a strategy, then the second has an "induced" strategy, best for him under the assumption that the first plays his sure-thing. One might think that the resulting outcome would benefit the sure-thing player. This is true in the second meta-level of Prisoner's Dilemma but false at that level in other equally simple situations. In fact, says Howard, to be "rational" in the sense of choosing a sure-thing strategy "is to be a 'sucker' that capitulates entirely to the other side" (p. 181). I suggest that no one act on this advice without fully digesting Howard's arguments.

Mathematical and nonmathematical readers alike may want to know that rigorous proofs of theorems are given roughly "equal time" with examples and interpretative comments. Though parts of Howard's work have appeared earlier, in *General Systems* and in *Papers* of the Peace Research Society (International), much here is new and it is most helpful to have such a unified presentation of the subject.

HENRY HAMBURGER
*School of Social Sciences,*
*University of California, Irvine*

**References**

1. G. Hardin, *Science* **162**, 1243 (1968).
2. For additional examples see T. C. Schelling's fascinating article "On the ecology of micromotives," *Public Interest* No. 25, p. 61 (1971); M. Olson, *The Logic of Collective Action* (Harvard Univ. Press, Cambridge, Mass., 1965); R. Hardin, "Collective action as an agreeable n-prisoners' dilemma," *Behav. Sci.* **16**, 472 (1971).
3. L. F. Richardson, *Arms and Insecurity* (Quadrangle, Chicago, 1960).
4. R. J. Harris, "Note on Howard's theory of metagames," *Psychol. Rep.* **24**, 849 (1969).
5. A. Rapoport, "Comments on 'Paradox regained,' " *ibid.* **26**, 272 (1970).
6. N. Howard, "Note on the Harris-Rapoport controversy," *ibid.*, p. 316.

## Understanding Sentences

**Psychology of Reasoning.** Structure and Content. P. C. WASON and P. N. JOHNSON-LAIRD. Harvard University Press, Cambridge, Mass., 1972. viii, 264 pp., illus. $10.

There are two distinct but related aspects to the meaning of the term "rationality." One has to do with overt behavior, the other with mental processes. Behavior is judged to be rational when it seems temperate or measured, as opposed to instinctive or impulsive. The first sense of rationality, while the more common, is actually derived from the second. It is based on the not always accurate assumption that complex mental processes lead to behavior that is temperate in nature.

The notion of rationality as it applies to mental processes is sometimes used quite broadly to cover all thinking and decision making, and some-

times more specifically to designate deductive processes akin to those treated by formal logic. Wason and Johnson-Laird are concerned with rationality in the latter, narrow sense. They examine people's ability to trace the logical consequences of certain types of assertions and interrelated sequences of assertions—for example, how people interpret negative and conditional statements, how they solve syllogisms, and how they deal with the deductive aspects of hypothesis testing. The authors report their own research in detail, and although they also examine other, closely related research, the book presents neither a thorough review of the literature nor a broad overview of the issues. Instead, it provides insight into selected aspects of the reasoning processes, and thus clarifies certain elements in an emerging picture of how human beings derive information from language.

The major recurrent theme of this research is noted in the book's title in the distinction between "structure" and "content." The studies repeatedly show that content affects how people interpret particular logical forms. While similar findings have been noted in a variety of studies by other researchers, Wason and Johnson-Laird have explored some particularly salient instances of this principle in their research. For example, consider a person who understands a conditional like "If John wins the lottery he will have a lot of money." He knows that "John does not win the lottery" does not imply "John will not have a lot of money" and that "John has a lot of money" does not imply "John won the lottery." But if given the abstract formulation "If $p$ then $q$," that same person may believe that if $q$ occurs $p$ must have occurred, or that if $p$ did not occur then $q$ would not occur. Content also affects interpretation of sentences that describe relations between classes, for example whether "All A are B" is taken to imply that "All B are A." An important factor in interpretation is whether the relations described correspond to people's knowledge and beliefs. Ambiguities in the interpretation of such sentence forms naturally contribute to the difficulty of syllogisms, which involve sequences of such sentences. Content even affects the difficulty of understanding negative sentences. Typically it takes longer to verify negative than positive sentences, for example "Three is not an even number" versus "Three is an odd number." However, when negatives deny some expected state of affairs they are easy to interpret.

In explaining these phenomena, the authors point out certain vagaries in the use of particular logical forms in discourse. For example, sentence negation does not specify what is being denied. A person might say "John did not run down the street yesterday" because no such occurrence took place, or because the time or place of the occurrence, or the actor or type of action, was wrongly stated. "Or" may be used to indicate at *least* one ("John or Bill can carry it") or at *most* one ("choice of vegetable or salad"). The ambiguity of the conditional is perhaps even more striking. Conditionals are used in discourse to describe various types of relations between events, and may even be used without asserting a relation between events at all. The authors' example, "If you want the money, it is in the teapot," simply asserts the existence of money in the teapot.

One could argue that because these logical operations are encoded ambiguously in the language people must rely on content to interpret sentences which include them. One could also argue in the opposite way, namely, that it is because people cannot think about logical operations in any general way that there do not exist straightforward unambiguous terms to specify these logical operations independently of particular content. That is, if it were important to people to distinguish between class inclusion and classes that completely overlap, there might be no form "All A are B" which can be used to express either relation.

In the case of the conditional, at any rate, it seems clear that people's difficulty does not stem simply from its ambiguous usage, but rather from conceptual difficulties in dealing with this type of relation. This is demonstrated in a memorable set of experiments where the authors ask sophisticated adults to test an apparently simple hypothesis. The task involves a set of cards each of which has a numeral printed on one side and a letter on the other. Four cards are presented showing, respectively, E, K, 4, and 7. Subjects are told to verify the claim that "If a card has a vowel on one side then it has an even number on the other." Most subjects know that one must turn over the card with the E to make sure there is an even number on the other side. Only "the rare individual" realizes that it is also necessary to turn the 7 because if the letter on the other side were a vowel that would falsify the conditional. Even professional logicians sometimes fail to realize that they must check "not $q$" to make sure that $p$ did not occur.

The conclusion the authors draw from their studies is that people do not carry out a fixed set of logical operations when they are presented with particular logical terms in sentences. They do not possess a straightforward, computer-like program that always executes a particular logical operation given a particular linguistic input. Rather, they consider particular types of information they are presented with and then interpret the logical terms in a way that seems consistent with a type of relation that could hold for that content.

The picture of mental processes that emerges from these studies is broadened by a demonstration of what the authors designate as "the logical fallacy of induction by simple enumeration." Subjects were given the task of discovering a simple numerical rule. This rule was exemplified in the series 2, 4, 6. The subject was to generate triads of numbers, and to be told in each case whether the triad fit the rule. He was to report when he knew the rule. Many subjects generated only ascending series of successive even numbers and after a few such trials announced that they knew the rule. Even after they were told that the rule was *not* ascending series of successive even numbers, many gave another such triad on the next trial. Unless one attempts to falsify one's hypothesis by producing sequences that violate it, one cannot discover the answer, which in this case was *any* ascending series of numbers.

Certainly the use of abstract materials is not the source of this particular limitation of the thought process. If the material concerned people's actual beliefs, they would surely be even less likely to generate counterhypotheses. The authors' demonstration shows in bold relief a human limitation which is closely related to the tendency toward superstition and prejudice. We might wish that people had a general procedure for evaluating incoming information by attempts at falsification

before accepting it as fact. Since this does not seem to come naturally, one hopes such a general procedure can be taught more successfully in the future than it has been in the past.

JANELLEN HUTTENLOCHER
*Department of Psychology, Teachers College of Columbia University, New York City*

## Computer-Age Art

**Science and Technology in Art Today.** JONATHAN BENTHALL. Praeger, New York, 1972. 180 pp., illus. Cloth, $8.95; paper, $4.95. Praeger World of Art Series.

During the last five or six years nearly a dozen books have appeared about the growing enthusiasm of artists for scientific and technological phenomena. Some are historical résumés, others tend to be overenthusiastic apologetics for future-oriented art. Because of its author's breadth of perspective and consistently fair critical sense, this book is in many ways the best of the lot.

Benthall possessed a sound technical background when he began to write a monthly column on "art and technology" for *Studio International* four years ago. Two lecture series which he planned for the Institute of Contemporary Arts in London remain the most provocative of recent attempts to span the gap between human mythic systems and scientific rationalism. Among the topics covered by his book are: the significance of "media studies" (studies of techniques of communication) to the fine arts, photography as an art medium, the esthetics of computer systems and laser holography, recent evaluations of kinetic art, art and the ecological trend, the analysis of art through linguistic models, and finally a summation of the cultural attitudes—and illusions—that tend to distinguish the goals of art and science.

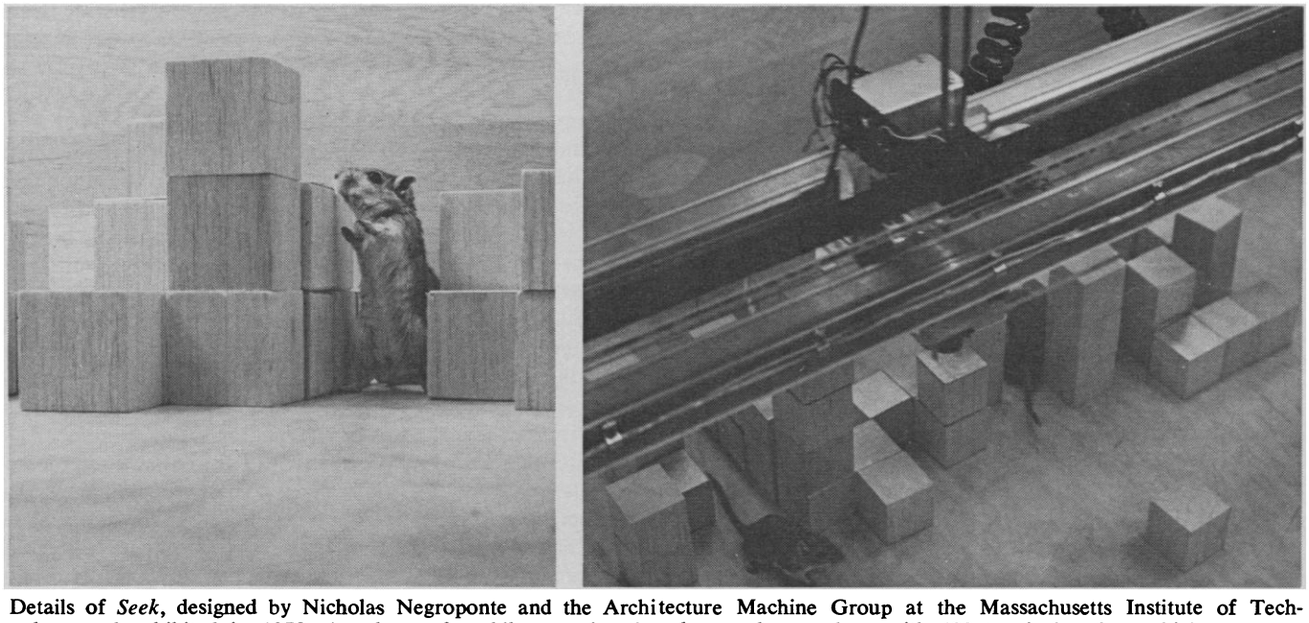
Most recent attempts to fuse sophisticated technology with art have met with bitter hostility within the art world itself. Benthall is very much aware of this, and his own cautious, hardheaded criticism of much recent "Teck-Art" is the best response to the emotionalism generated elsewhere, both pro and con. For instance, his clarification of the question of creativity within "computer graphics" is helpful. He writes in part,

> The term 'computer art' is itself a provocation (even more than 'artificial intelligence') because the very terms in which we often characterize art—'humanity', 'warmth', 'spontaneity', 'sincerity', 'originality' and so forth—are laden with implicit prejudice against the values of which the machine is a symbol.

To an art critic who dismisses an example of the ecologically oriented art of Alan Sonfist—one in which crystals vaporize and condense in a container—as a "high-school physics project," Benthall replies,

> Looked at in this way, all of Sonfist's work would be quite meaningless. . . . One cannot really argue with those who refuse to respond to such work as legitimate forms of art. But one can insist that a painting by, say, Rothko is not meant to be taken as an experiment in the drying properties of paints, nor did Duchamp exhibit his famous urinal to measure the viscosity of urine. Conversely, the technical sophistication or novelty of a work of art is no guarantee of its *artistic* interest.

Benthall does not bog down in elaborate technical detail, nor does he take sides. His book differs in two other respects from its predecessors. He devotes a good deal of space to various research projects which neither by intention nor by popular definition qualify as art but which either have strong esthetic appeal or parallel research conducted by artists. In dealing with Joel Weizenbaum's conversational computer programs or Nicholas Negroponte's computer-controlled gerbil environment, Benthall is dealing with questions raised by artists and esthetic-ians—rightly seeing that if the art-



Details of *Seek*, designed by Nicholas Negroponte and the Architecture Machine Group at the Massachusetts Institute of Technology and exhibited in 1970. A colony of gerbils occupies the glass enclosure along with 480 two-inch cubes, which a computer transports, stacks, and aligns by means of an overhead electromagnet. The computer "configures the original assemblage of blocks by using a random number generator which has been programmed to have tendencies to provide enclosures . . . nooks, crannies, and mazes in which gerbils can play." The gerbils' movements continually disrupt the constructions, and the computer continually realigns displaced blocks or, where the dislocation is of sufficient magnitude to be interpreted as "a gerbil-desired move," attempts a new arrangement. A spectator who perceives the display as a metaphor for life in the Machine Age may identify "now with the freedom of the gerbils to consume and excrete, scurry, court and squabble, now with the responsibilities of a lumbering bureaucracy to keep the environment orderly." [From *Science and Technology in Art Today*]