Using Information Entropy to Model a Wide Variety of Complex Behaviors within Organizations

by Rodney Fuller
Department of Psychology
Claremont Graduate School
Claremont, CA 91711 USA
<fullerr@cgsvax.claremont.edu>

Abstract

This paper deals with issues surrounding ***not*** codifying the semantic possibilities within a problem space as they relate to explaining variability in behavior or structure. This paper proposes that given a sufficient amount of complexity in the problem space (either high uncertainty in the probability of an events occurrence; ambiguity due to the subjective nature of an individual's decision space; or high uncertainty in the role of individuals within the problem space) that many of the problems involved in explaining variability in behavior are negligible, and a computational model of the problem space can be created.

There are several problems with computational models of individuals doing group work on ambiguous tasks. These problems include: 1. The difficulty in statistically modeling the uncertain or ambiguous environment. 2. the subjective nature of the individual models within the group, and 3. the multiple levels of meaning and analysis that exist both within group processes and between uncertain processes. One possible solution for dealing with computational models of the above problems is to create an adaptive model of coordination behaviors related to group interaction on basic decision making processes and activities like categorization, consensus making, and creative elicitation. This paper proposes that in situations of sufficient uncertainty (either at an organizational, individual, or cognitive level) some of the problems involved in explaining the shared understanding of groups of individuals can be largely ignored.

A well known Indian fable about four blind men and an elephant can illustrate the types of problems and questions this paper hopes to deal with: All four of the blind men are brought to an object which they are told is an elephant, and they are asked to determine what an elephant is (Thompson, 1957). One of the blind men describes it as like a snake, another like a tree, still another describes the experience as something without a beginning or an end, and finally the last blind man describes the elephant as a rope.

Using any of the accepted techniques in analyzing group decision making today it would be difficult (if not impossible) to say that the four blind men described the same experience because 1) the probabilistic model of the above shared decision space is unknown, 2) that the individual cognitive models used by the participants are too different, or 3) that the mechanisms for managing the dynamics of group decision making behaviors required to reach consensus are unknown. I will propose that if sufficient complexity and uncertainty exists within the problem space that modelers of problem spaces should be able to ask and answer the following two questions: A) could the group ever decide that they had shared the same experience? and B) given that this group does differ what is the most complicated explanation they can derive regarding the above experience?

I. The Complexity of Problem Spaces:

Traditional categorization models based on Bayesian probability (Pearl, 1988) are modeled after the classic statement of probabilistic logic: "given that I know Y what is the probability of X?" This can be displayed as: $P(X|Y) = P(X,Y)/P(Y)$. In this equation the probability of event X is explained in terms of the probability of X and Y occurring together divided by the probability of Y occurring by itself. The formula reduces mathematically down to the statement that X occurs a certain percentage of the time given that Y has already occurred.

This foundation of Bayesian logic is based on the assumption that 1) the total number of possible events is known, and 2) that the overall probability of selecting an event is known. These assumptions do not hold up in situations of uncertainty or ambiguity.

One of the first problems one encounters when modeling uncertain or ambiguous behaviors is

that the sample space (the realm of all possible outcomes) is unknown. A well defined problem sets out its "problem space" by defining those events that are impossible, and by defining or measuring a given a rate of success and by knowing the total number of trials required for "certainty." In the example of the tossing of a fair coin one can assume that the events "odd," "blue," and "none" would all be illegal (or impossible) events. The probability of any future event can be predicted if the probability of either "heads" or "tails" is known for a sufficiently large number of trials. This model of prediction does not work if the uncertainty present in the environment is sufficiently high. High uncertainty creates the following two problems for any model that utilizes probabilistic reasoning: 1) while the null set is assumed to exist, the assignment of the null value to any one outcome can not be known *a priori* (everything is possible—or at least what is impossible is not known), and 2) there is no assurance that the model of behavior selected will try and explain the variability at a level at which the events are mutually exclusive ("the" level of analysis problem). The violation of these two assumptions frees the ultimate model of the "uncertain" behavior from two of the three main assumptions of the linear statistical model (Hays, 1988).

In an uncertain situations you do not know how many possible occurrences there are, and this creates a unique problem in categorization: You can't determine how many events there are within a given category until you know what the category is, and you can't determine what the category is until you know all of the possible events. This assumption is partially based on the argument that no categorization will ever be complete until time stops because at any time a new event can anticipate a entire new structure of cognitive classification. This problem can be shown theoretically by the concept of information entropy in Claude Shannon's information theory (Shannon and Weaver, 1963).

Shannon showed that every symbol, or sequence of symbols, can be described in terms of how much information it contained—or how much uncertainty it explained—and he called this measure "information entropy." The basis of Shannon's information theory is the question of how many decision points does it take to exactly specify an experience? "That is, information is a measure of one's freedom of choice when one selects that message. . . . The concept of information applies not to the individual messages (as the concept of meaning would), but rather to the situation as a whole, the unit of information indicating that in this situation one has an amount of freedom of choice, in selecting a message" (Shannon and Weaver, 1963, p. 9). In this way Shannon created a mathematical theory of communication that focused on the "degrees of freedom" that one has in communicating an experience, rather than the qualitative significance of any one message. "To be somewhat more definite, the amount of information is defined, in the simplest cases, to be measured by the logarithm of the number of available choices" (Shannon and Weaver, 1963, p. 9).

> The concept of entropy in information theory is closely connected with the concept of entropy in statistical mechanics. If we draw a sequence of n independent and identically distributed ... random variables, we will show that the probability of such a 'typical' sequence is $2^{-nH(X)}$ and that there are about $2^{nH(X)}$ such 'typical' sequences .... The notion of descriptive complexity of a random variable can be extended to define the descriptive complexity of a single string. The Kolmogorov complexity of a binary string is defined as the length of the shortest computer program that prints out the string. It will turn out that if the string is indeed random, the Kolmogorov complexity is close to the entropy. Kolmogorov complexity is a natural framework in which to consider problems of statistical inference and modeling and leads to a clearer understanding of Occam's Razor 'The simplest explanation is the best.' .... Entropy is the uncertainty of a single random variable. We can define conditional entropy, which is the entropy of a random variable, given another random variable. The reduction in entropy due to another random variable is called the mutual information" (Cover and Thomas, 1991, p. 6).

The implications of how information theory can provide a better model for models of cognitive categorization is because it provides a measure of the amount of uncertainty that exists in any one instance or sequence of instances. A second reason is that, according to Kolmogorov, the complexity of any one given model or explanation is independent of the representation used and essentially equal to the length of the shortest explanation that can produce that sequence. Kolmogorov's ideas are important because it uses the uncertainty within a problem space to search

for and select the "best" level of analysis or representation for explaining the variance of that space.

II. Individual Cognitive Models and Complexity:

A second category of problems related to modeling the decision making behavior of individuals is that knowledge representations are subjective. For the sake of simplicity the "problem" of subjectivity will be reduced to that of verifying that different individuals are talking about the same "cognitive category," and the more "procedural" (script or role based) implications of subjectivity will be talked about as the problems involved in modeling multiple levels of group behavior (Section III). To further simplify the problem of categorization I will only use examples of individuals categorizing color terms:

Categorization and Subjectivity

An individual's capacity for making decisions is partially based on their past experiences and on their ability to categorize these experiences in a meaningful way. According the Mervis and Rosch (1981) "a category exists whenever two or more distinguishable objects or events are treated equivalently. This equivalent treatment may take any number of forms, such as labeling distinct objects or events with the same name, or performing the same action on different objects" (p. 89). In this sense, categorization is considered one of the "most basic [of all] cognitive functions" (Corter and Gluck, 1992, p. 291, see also Mervis and Rosch, 1981). George Lakoff writes that "without the ability to categorize, we could not function at all, either in the physical world or in our social and intellectual lives" (Lakoff, 1987, p. 5-6).

In situations of high complexity and uncertainty the issues related to how one forms and finds categories is directly related to how one makes decisions. Eleanor Rosch, one of the pioneers in the field of categorization and cognitive-linguistic architecture has stated: "Empirical findings have established that: (*a*) categories are internally structured by gradients of representativeness; (*b*) category boundaries are not necessarily definite; (*c*) there is a close relation between attribute clusters and the structure and formation of categories" (Mervis and Rosch, 1981, p. 109). In the following three sections these aspects of categorization will be addressed:

A. Category Structure:

According to the classical theory of categorization "attributes are combined arbitrarily to form items" (Mervis and Rosch, 1981, p. 91). According to learning theory the appropriateness of these categories is then refined over time. But "the contention that the division of real world objects into categories is originally arbitrary would make sense *only if the attributes in the world formed a total set*: that is, if all combinations of attribute values were equally likely to occur" (Mervis and Rosch, 1981, p. 91, emphasis added). It is important to realize that the division (or categorization) of objects in real and mental space is not arbitrary: 1) most combinations of "features" do not occur in naturally occurring categories, and 2) that some combinations of features are thought to be more basic in determining membership within a category than others. The importance of these two points is explained below:

1. Categories are not simply combinations of features.

It should be obvious that not every combination of features will produce a category that will attract members. For example, if animals were categorized by the following features: coat (fur or feathers), oral opening (beak or mouth), and primary mode of locomotion (walk or fly) then it should be obvious that some of the possible combinations do not describe existing natural categories (there are no animals which have fur, mouths and fly, nor are there any with feathers, a mouth and which walk—see Mervis and Rosch, 1981 for more examples).

This issue can also be approached on a much larger scale. Davies (1990) in an article titled "why is the physical world so comprehensible?" explains: "The most striking feature of many complex systems is their non-random nature. The universe is populated by distinct classes of recognizable things: galaxies, stars, crystals, clouds, bacteria, people. Given the limitless variety of ways in which matter and energy can arrange themselves, almost all of which would be "random," the fact that the physical world is a coherent collection of mutually tolerant, quasi-stable entities is surely a key scientific fact in need of explanation. The non random nature of cosmic complexity is captured by the concept of organization, or, to use a more fashionable word, depth" (p. 61).

Davies goes on to show that this "non random nature of cosmic complexity" can be explained

"if we view the world algorithmically" and that "the existence of regularities may be expressed by saying that the world is algorithmically compressible" (1990, p. 63). At this point it is important to realize that the world (both mental and physical) can be explained in a way that does not require the random or arbitrary distribution of all events.

2. Some category members are more basic than others.

In addition to the above evidence that not all categories (or combinations of features) are created equal there is also empirical evidence that not all members within a category are considered equal. Not only are some levels of description more basic than others (people remember them better and faster) but also some of the members within a category are considered "better" (or more basic) members than others. The evidence for basic categories (or levels of description) comes from anthropologists (Berlin and Kay, 1968; but see also Corter and Gluck, 1992) who propose that those categories that transcend linguistic and cultural levels should be considered "basic." Rosch (1976; Mervis and Rosch, 1981, Rosch, 1975a and b; Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976) argues that those levels of categories which are the most cognitively efficient are the levels of categorization where the information value of the attributes is maximized or "prototypical." The research on "basic categories" has shown, for example, that the word "chair" is recognized faster than either its subordinate "lazy chair" or its super-ordinate "furniture." And the research on "basic" membership within a category (or prototypicality) has shown that people consider a canary to be a better bird than a penguin.

B. Category Boundaries:

Not only do subjects disagree about the within category membership of colors (which "red" is the best red?) but they also disagree about the between category membership of colors. In our example of color categorization the issues related to category boundaries can be posed as the following question: "in the continuum of color between two colors, for example, blue and green, what determines linguistically and cognitively the separation between bluish-green and greenish-blue?"

Brown and Lenneburg (1954) proposed that the codability of a color (a composite measure of agreement in naming, length of name, and response latency in naming) correlated with a subjects memory accuracy in a recall task. It was proposed that the improved memory for color chips in the delayed recall task occurred because subjects were able to better linguistically code the color stimuli. This correlation between codability and recall failed to appear when several researchers replicated the original study using Farnsworth-Munsell colors in an array of uniform low saturation (Burnham and Clark, 1955; Lenneberg, 1961). In 1968 Berlin and Kay showed that there were some regions of the color spectrum that were more psychologically salient than others, and that these areas corresponded with the focal points of the primary color terms used in the language. The discovery of focal colors is considered important because for the first time scientists had discovered a set of cognitive terms that seem to be universal (constant over effects of experience, language and culture).

Eleanor Rosch (Heider 1972) followed up Berlin and Kay's (1968) work with a study that tested the old theory of linguistic relevance in light of the new focal color research. She tested 23 non-English speaking subjects on the number of words in their response, the number of letters, and the latency of response. She found that the focal colors were described with a significantly fewer number of words (and letters) and in a shorter amount of time. She also found that there were **no** significant differences between the number of words used, the number of letters used and the response time for the two types of non-focal colors tested: Internominal (those areas of the color chart where Berlin and Kay found no focal colors) and boundary colors (colors adjacent—2 or more spaces away—to the focal colors and the internominal color). Rosch (Heider 1972) also found that the non-English speaking subjects performed the same as her English speaking subjects in that the "length difference between focal and nonfocal colors was not simply the result of subjects naming focal colors with basic names and nonfocal colors with secondary names. This lack of significance between the boundary and "non-categorical" is surprising, and goes against what most researchers and theorists would predict. Berlin and Kay (1968) found that while the identity of the focal colors was stable individuals disagreed (both within and between subject disagreement) on those colors that lie on the color category boundaries.

C. Category Structure and Attributes:

"If one believes that categories consist of determinate necessary and sufficient criteria, one can develop a model which attempts to explain representativeness and indeterminate boundary effects by means of processes operating on a deterministic representation" (Mervis and Rosch, 1981, p. 101). The deterministic representations that have been used to date represent both rule-based (Miller & Johnson-Laird, 1976; Johnson-Laird, 1988) and fuzzy logic based models (Kay & McDanials, 1978; Kay, 1981). Both of these representations can not handle the effects of context on the categorization.

Paul Kay and Chad McDaniel (1978, Kay 1981) next proposed that basic color terms were categorized based on a set of fuzzy logic rules (identity, union and intersection) of patterns of firing in the retina. Relying on the recent discovery of the opponent processes research (De Valois, Abramov and Jacobs, 1966; De Valois and Jacobs, 1968) in the color visual system of primates Kay and McDaniel proposed that the basic colors black, white, red, yellow, green and blue were determined by the actual retinal response of either the rods, cones, or the opponent process summation of the horizontal cells. They also proposed that the remaining basic colors resulted from either the union of several of these more basic colors (dark-cool = black or green or blue, light-warm = white or red or yellow, warm = red or yellow, cool; grue = green or blue), or from the intersection of two of these colors (brown = black and yellow, purple = red and blue, pink = red and white, orange = red and yellow, and gray = white and black). This model of basic color categorization explains how some cultures demonstrate "transitional" color terms (like "grue"—a combination of green and blue) as well as explaining why there are no basic color terms for combinations of the two "opponent pairs" red and green, and yellow and blue.

Mervis and Roth (1981) empirically tested Kay and McDanials (1978) theory and found that "in the case of color categories, the best examples (foci) appear to be strongly influenced by the psychophysical properties of the visual system. Nevertheless, the degree to which a color precept is representative of a color category should be affected by how similar it is perceived to be to other members of the category" (p. 401). In one experiment Mervis and Roth asked subjects how good of an example the sample color was compared to three other colors. According to the Kay and McDanials theory a non-focal color (like lime) should be constructed from the separate inputs of green and yellow. Mervis and Roth (1981) found that subjects answered that the color "lime" was a better example of the color lime than either green and yellow. They conclude that "every category contains at least one member—the best example—that is a better example (or, in the limiting sense, as good an example) of its own category than any other category" (Mervis and Roth, 1981, p. 405).

In summary the problems related to the modeling decision making behavior of an individuals cognitive structures and subjectivity are 1) the structures are different, 2) that not every combination of features results in a naturally occurring category, and 3) that the prototypicality of category members on those features changes within the category. These problems may be resolved by selecting a domain where the representation that accounts for the form, structure, and context of the events behavior is universal—as is the case with color terms.

III. Modeling Behavior at Multiple Levels: Roles, Groups and Complexity:

A third problem related to the modeling of uncertain group decision making processes is the problem that all decision processes are ambiguous or uncertain at one or more levels, otherwise there would be no need to "make" a decision. It is also important to realize that while every cognitive and behavioral decision space is uncertain (incomplete) at some level there are also many other levels where the decision space that are not ambiguous. The problem of finding one perspective that captures all of the necessary information without creating a un-resolvable uncertainty at one or more levels is not as easy as one might think. Jim Crutchfield explains why:

> "The epistemological problem of nonlinear modeling is: Have we discovered something on our data or have we projected the new found structure onto it? This was the main lesson of attempting to reconstruct equations of motion from a time series: When it works, it works; When it doesn't, you don't know what to do; and in both cases it is ambiguous what you have learned. Even though data was generated by well-behaved, smooth dynamically systems, there was an extreme sensitivity to the assumed model class that completely swamped 'model order estimation.' Worse still there

was no *a priori* way to select the class appropriate to the process. This should be contrasted with what is probably one of the more important practical results in statistical modeling: within a model class a procedure exists to find, given a finite amount of data, an optimal model that balances prediction error against model complexity. Despite representations to the contrary, this 'model order estimation' does not address issues of class inappropriateness and what to do when confronted with failure" (Crutchfield 1992a, p. 68).

It is important to distinguish between simple deterministic systems and complex deterministic systems, because according to Crutchfield (1992a) "there appears to be a way out of the model class discovery dilemma. The answer that hierarchical machine reconstruction gives is to start at the lowest level of representation, the given discrete data, and to build an adaptive series of models within a series of model classes of increasing computational capability until a finite causal model is found. Within each level there is a model-order-estimation inference of optimal models, just as indicated. And there is an induction from a series of approximate models within a lower 'inappropriate' class to the next higher model class" (p. 68)

The logarithm (base 2) of any one statement or outcome is also known as the information entropy of the statement. This means that in situations where the total number of possible outcomes is known, and each outcome has an equally likely chance of selection, the minimum number of choices that must be made to decide on one outcome is equal to the $\log_2$ of the number of possible outcomes. The maximum number of decisions will be one less than the total number of outcomes. For example, if someone wanted to communicate to someone else that they should choose one out of 8 outcomes then the shortest message explaining the decision process would contain information about the three decisions ($\log_2 8 = 3$) that must be made to select the correct outcome. Shannon's model is based on the amount of information needed to explain or differentiate between the number of choices available. "That information be measured by entropy is, after all, natural when we remember that information, in communication theory, is associated with the amount of freedom of choice we have in constructing messages. Thus for a communication source one can say, ...'This situation is highly organized, it is not characterized by a large degree of randomness or of choice–that is to say, the information (or the entropy) is low" (Shannon and Weaver, 1963, p. 13).

If the probability of each outcome is equal then we can say that the selection of any one outcome explains an equal amount of the uncertainty within the decision making process. You could also say that each outcome contributes exactly the same amount of information (or entropy) to the communication. The amount of entropy contained in a communication becomes important when the total number of outcomes is unknown (uncertainty) or when determining if the information is specific enough (the ambiguity of the message). Thus the information content of an infinite non-repeating string, like that found in the number π (pi), according to Shannon's theory would be as long as the string itself (because no information repeats within the string), and the entropy (or information content) of an infinitely long string of repeating numbers would be trivial. This example also shows that what may be close to impossible to explain at one level (the infinitely long string of the number π), can be easily explained at another level (the written symbol "π"), and that the information entropy explained at these various levels of symbols varies widely. Information entropy, or the information content of any communication that can be transmitted, "is ... approximately the logarithm of the reciprocal probability of a typical long sequence divided by the number of symbols in the sequence" (Shannon and Weaver, 1963, p. 54).

The complexity of decision making behavior can be explained in the same way: "The success at each stage in hierarchical reconstruction is controlled by the amount of given data, since this puts an upper bound on statistical accuracy, and an error threshold, which is largely determined by the observers available computational resources. The goal is to find a finite causal model of minimal size and prediction error while maximizing the extraction of information from the given data" (Crutchfield, 1992a, p. 68).

Any process which behaves in a simple predictable manner can be modeled with one knowledge representation. Once a data model has been selected one can calculate the upper and lower limits on the entropy of this model. The area between these bounds is known as the bandwidth of the data model: and these bounds

set limitations on both what is the simplest interaction that model can explain and what is the most complex explanation that can be shared. The essence of Kolmogorov's model is that he reversed the whole problem of having to know what the right answer was before one could create the appropriate data representation—he showed that given the fact that any stream of data can be randomly and completely "chunked," and more importantly that one of these chunks *must* be the right data structure to explain the uncertainty within the system.

Kolmogorov's process in creating meaning for complex data streams is to calculate the information entropy (the degree of uncertainty) in every possible chunk from the data stream. In this way Kolmogorov also frees the nonlinear modeling process from a pre-selected data representation, as well as the many problems that *a priori* selection brings. The "chunk" with the lowest entropy will be that data representation that explains the most uncertainty in the given model at that level of analysis.

Kolmogorov's theorems of complexity resolve some of the problems of uncertainty in decision making (having to determine *a priori* which data structure will explain the complexity of the cognitive/social model) but it does not resolve the problem of ambiguity of the process (non-stochastic and groups). Some of the earliest theorists in "communication and organization theory" concluded that it is necessary to study communication at multiple levels within the organization (Thayer, 1967), but that models of small group behavior must include "organizational", "interpersonal" and "intra-personal" levels. An example of work done involving the organizational and interpersonal levels that captures the dynamics of small group behavior are recent studies by Bernardo Huberman and Natalie Glance (1993; Glance and Huberman, 1993a and b).

Huberman and Glance (1993; Glance and Huberman, 1993a and b) have shown that the degree of cooperation present in a group decision making task can be predicted by using nonlinear models. These models are sensitive to the past history of the groups behavior, but also on their expectations about the future behavior of the other members of the group. Glance and Huberman (1993a) found that the "results reveal several different dynamical regimes that should be observable as the size of the group changes. In one regime, although cooperation may persist for very long times even for groups exceeding a critical size. group behavior eventually decays to overall defection. In another situation a system can be stuck in a non-cooperative state even though its size is well below that guaranteeing long term cooperation. *These effects are shown to depend strongly on the degree of uncertainty pervading the system, as well as on the length of the individual's horizon*" (p. 283, emphasis added).

Glance and Huberman (1993a) proposed that while no individual participant could "directly observe the effort of another, each member observes instead the collective output and can deduce overall group participation using knowledge of individual and group production functions" (p. 283). They found that an individuals performance could be best predicted in a model that considered the past (in terms of the most recent "average" cooperation rate) and by considering either an optimistic or pessimistic view of the near future (the individual tendency for cooperation to foster cooperation, and defection to foster more defection). In an earlier study Ceccatto and Huberman (1989) found that, as long as the group is neither too large or too small, that short fluctuations away from the local minimum (the most recent average behavior) would eventually return to the previous equilibrium, but that longer deviations resulted in giant fluctuations during which a large portion of the participants can switch strategies (from cooperation to defection). "Once this critical mass is reached, the remaining agents rapidly switch into the new strategy that corresponds to the optimal . . . equilibrium and the system slides into a new optimal state" (Glance and Huberman 1993a, p. 290).

These sudden mass shifts in strategy also depend on "the amount of imperfect knowledge that individuals have about the state of the system, or in other words, on what [the other participants are doing]" (Glance and Huberman 1993a, p. 290). When the participants in the system decide to seek a new equilibrium these changes happen very quickly. Glance and Huberman (1993a) report that the time required to make this shift is equal to the logarithm of the number of participants (if there are 8 participants then the change can occur in 3 turns, if there are 246 participants then the transformation will occur in 8 turns). "The theory predicts that nothing much happens for long times, but when it does, it happens very fast" (Glance and Huberman 1993a,

p. 290). But the above statement is clarified as "it is only in the case of imperfect knowledge that many individuals can change their behavior. This is because, in evaluating the number of members cooperating, imperfect knowledge amounts to occasional large errors in the individual's estimation of the actual number cooperating" (Glance and Huberman 1993a, p. 290).

In many situations the expectation that organizations of individuals work as "bounded rational" agents is unrealistic: Eisenberg (1984) has said that "people in organizations confront multiple situational requirements, develop multiple and often conflicting goals, and respond with communicative strategies which do not always minimize ambiguity, but may nonetheless be effective" (p. 228). This echoes many of the writings of Karl Weick (1979) and James March (1990; March and Olsen, 1976). One way to study differences between the intra- and interpersonal levels of organizations is to look at organizational behaviors as equivocal communications. One of the first theorists to study incongruence in communication noted that all communication is, in principle, based on four characteristics: Communication always involves a sender, some content, a receiver and a context (Haley, 1959). Bavelas, Black, Chovil and Mullett (1990) propose that "the data of communication can be the messages themselves and that the explanation of a message can be sought in the immediate, observable interpersonal situation in which it occurs" (p. 28). John Maynard Smith (1972) has proposed several possible algorithms for modeling communication related to conflict behavior (see Figure 1). One of the interesting things about these strategies is that they are not effective until a critical percentage of the population has adopted them—Maynard Smith (1972) found that once the "retaliator" program was predominant within the population none of the other strategies could surpass it.

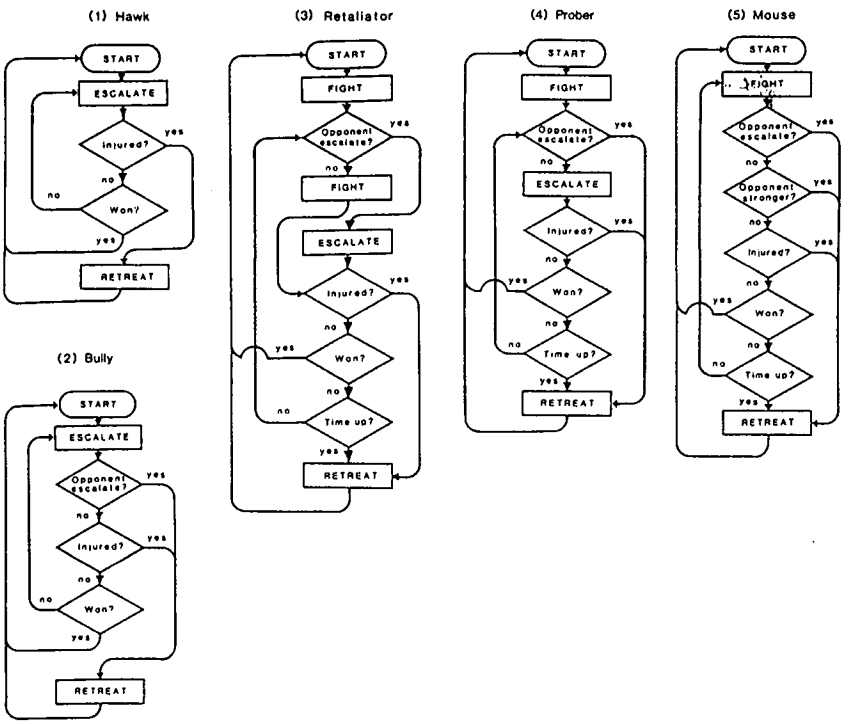

Figure .1. Flow charts of algorithms for five alternative strategies for conflict behavior. (Strategies adapted from Maynard Smith 1972, p. 19, and Maynard Smith and Price 1973, p. 26.)

Figure 1 is taken from Beniger (1986). p. 81.

Uncertainty and Multiple Levels of Meaning:
"One of the principal problems with previous models of the communication process is that they universally neglect the way in which multiple levels of information are exchanged among sender, receiver, and environment" (Targowski and Bowman, 1988, p. 10). In recent years several theories have proposed solutions to the problem of modeling communication between multiple levels. These theories have proposed models for solving the multiple level communication problems in dynamic cognitive environments (Crutchfield, 1992a and b) and for collaborative human-computer communication (Hale, Hurd and Kasper, 1992). This section will explain how information entropy can be used to resolve the problems surrounding multiple levels of meaning in communication and how uncertainty is a crucial part of the process to resolve this problem.

Shannon's theory allows for the modeling of multiple levels of communication through the idea of calculating the entropy of a string at different levels of analysis, or "morphs" and then calculating the minimum entropy for the string. This allows for a single stream of information to be used in manner that provides adaptive between level representation of the semantic content of a message. An example of this process that shows how the differences in information entropy for strings based on letter probability and word probability account for meaning is:. if we assume that the total number of possible outcomes (or in this case symbols) are all equally likely, and that these 27 outcomes will be labeled as the letters 'a' through 'z' and space. The probability of a trigram (a string of three symbols) such as "and" given that each outcome is equally likely is 3 times the probability of any one single event (or 3(1/27) = .11). If we look at the probability that the word "and" would occur independent of the letter probabilities then we find that the probability of that word occurring is .028 (27873 occurrences out of a possible 1,000,000 words, Johansson and Hofland, 1989). Thus the lowest entropy we could calculate for the trigram "and" is also the most semantically relevant "morph" or representation.

Within organizational behavior recent studies have shown that the decision making processes (both social and cognitive) in uncertain tasks are recursive (Cohen, March and Olsen, 1972; Morgan, 1986; Masuch and LaPotin, 1989; March 1990; Beach 1991): participants return to the same statements and processes until they have been refined or clarified. This recursive process also allows for a model of the outcome likelihood to be developed for group decision making processes. This model can then be used to calculate the information entropy of the event for the decision process.

An example of how information entropy might model group decision making behavior in situations of sufficient "uncertainty" deals with the behaviors of "brainstorming" and "consensus making." In traditional decision theory the alternatives that will be considered are generated near the start of the processes. These alternatives are then considered and narrowed and some means of consensus making behavior is applied (usually by vote). In uncertain environments researches found that groups would keep returning to the "brainstorming" stage after they had started evaluation and consensus making (Mintzberg, Rasinghani, and Théorêt 1976; Nutt, 1984)—participants found that their problem space was too confining or that they ran out of alternatives. Using measures of mutual information and information distance (Shannon and Weaver, 1963) the decision making process can be modeled so that participants can monitor the scope of the problem space through the stages of alternative generation and consensus making.

In decision making tasks ambiguity is where one level of knowledge or description seems more accurate than a lower level even though that level may have multiple meanings. In information theory ambiguity of this sort is handled by using Kolmogorov's theorems of complexity to model the categorization process. "Kolmogorov [defined] the intrinsic descriptive complexity of an object ... [also known as] the algorithmic (descriptive) complexity of an object to be the length of the shortest binary computer program that describes that object" (Cover and Thomas, 1991, p. 144). According to the classic Shannon argument the descriptive complexity of this event is equal to its information entropy, which in turn depends on the probability of occurrence of the event. "Kolmogorov went further .... The Kolmogorov complexity of an object dispenses with the probability distribution. Kolmogorov made the crucial observation that the definition of complexity is essentially [independent of the representation and] that the expected length of the shortest binary computer description of a random variable is approximately equal to its entropy" (Cover and Thomas, 1991, p. 144).

According to Cover and Thomas (1991) an "important technique for thinking about Kolmogorov complexity is... –if one person can describe a sequence to another person in such a manner as to lead unambiguously to a computation of that sequence in a finite amount of time, then the number of bits in that communication is an upper bound on the Kolmogorov com-plexity" (p. 147). There are two important implications to be drawn from Kolmogorov's theorems of complexity: The first is that by using his theorems we can model uncertain decision making process (because the measures of complexity are representation independent), and second that a means of semantically categorizing events without the *a priori* engineering of the knowledge domain is possible.

It is important to realize that the *a priori* selection of Bayesian probabilities to rep-resent uncertain and ambiguous events places an upper and lower bound on the entropy of the possible model. But by using Kolmogorov's models of complexity we have successfully reversed this: And as Kolmogorov said himself "Information theory must precede probability theory, and not be based on it" (quoted in Cover, Gacs, and Gray, 1989, p. 840).

In summary modelers of decision making processes can avoid many of the problems related to designing systems to support decision making behaviors when sufficient complexity exists. Uncertainty can be used to provide designers and computational modelers with problem environments that reflect the cognitive subjectivity and social biases of both human and cognitive agents.

References:

Bavelas, J. B., Black, A., Chovil, N. and Mullett, J. (1990). *Equivocal Communication.* London: Sage Publications.

Beach, L. R. (1991). Four revolutions in behavioral decision theory. Conference presentation at the Claremont Applied Leadership Conference. Claremont, CA. Feb, 26, 1991.

Beniger, J. R. (1986). *The control revolution: technological and economic origins of the information society.* Cambridge, Massachusetts: Harvard University Press.

Berlin, B. and Kay, P. (1968). *Basic Color Terms: Their Universality and Origin.* Berkeley and Los Angeles: University of California Press.

Brown, R. W. and Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology,* vol 49, 454-462.

Burnham, R. and Clark, J. (1955). A test of hue memory. *Journal of Applied Psychology,* 39. 164-172.

Burns, B. and Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: the representation of hue, saturation and brightness. *Perception and Psychophysics.* Vol 43(5). p. 494-507.

Ceccatto, H. A. and Huberman, B. A. (1989). Persistence of nonoptimal strategies. *Proceedings of the National Academy of Sciences of the United States of America,* 86. 3443-3446.

Cohen, M. D.; March, J. G and Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly, 17.* P. 1-25.

Corter, J. E. and Gluck, M. A. (1992). Explaining basic categories: feature predictability and information. *Psychological Bulletin,* Vol. 111(2). p. 291-303.

Cover, T. M.; Gacs, P. and Gray, R. M. (1989). Kolmogorov's contributions to information theory and algorithmic complexity. *The Annals of Probability.* Vol. 17(3). pp. 840-865.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory.* New York: John Wiley and Sons, Inc.

Crutchfield, J. P. (1992a). Knowledge and meaning: chaos and complexity. in Lam, L. and Naroditsky, V. (Eds.) *Modeling Complex Phenomena.* New York: Springer-Verlag.

Crutchfield, J. P. (1992b). Semantics and thermodynamics. In Casdagli, M. and Eubank, S. (Eds.) *Nonlinear Modeling and Forecasting.* Redwood City, California: Addison-Wesley Publishing Company.

Crutchfield, J. P. (1990a). Reconstructing language hierarchies. In H. A. Atanspracher,& H. Scheingraber (Eds.) *Information Dynamics.* New York: Plenum Press.

Crutchfield, J. P. (1990b). Information and its metric. In L. Lam and H. C. Morris, (Eds.) *Nonlinear Structures in Physical Systems: Pattern Formation, Chaos and Waves.* San Jose State University, San Jose California. November 17-18, 1989.

Davies, P. C. W. (1990). Why is the physical world so comprehensible? in Zurek, W. H.

(Ed.) *Complexity, Entropy and the Physics of Information.* Redwood City, CA: Addison-Wesley Publishing Company.

De Valois, R. L., Abramov, I. & Jacobs, G.H. (1966). Analysis of response patterns on LGN cells. *Journal of the Optic Society of America.* Vol 56. pp. 966-977.

De Valois, R. L., Jacobs, G. H. (1968). Primate Color Vision. *Science.* Vol. 162. pp. 242-254.

Eisenberg, E. M. (1984). Ambiguity as strategy in organizational communication. *Communication Monographs*, 51, 227-242.

Glance, N. S. and Huberman, B. A. (1993a). The outbreak of cooperation. Journal of Mathematical Sociology. Vol. 17(4). pp. 281-302.

Glance, N. S. and Huberman, B. A. (1993b). Organizational fluidity and sustainable cooperation. In K. Carley and M. Prietula (Eds.) *Computational Organizational Theory.* Hilldale, New Jersey: Lawrence Earlbaum Associates.

Hale, D. P.; Hurd, J. E. & Kasper, G. M. (1991). A knowledge exchange architecture for collaborative human-computer communication. *IEEE Transactions on Systems, Man, and Cybernetics.* 21(3). p. 555-564.

Haley, J. (1959). An interactional description of schizophrenia. *Psychiatry*, 22, 321-332.

Harrison, J. R. and March J. G. (1984). Decision making and postdecision surprises. *Administrative Science Quarterly, 29.* p. 26-42.

Heider, E. Rosch. (1972). Universals in color naming and memory. *Journal of Experimental Psychology.* Vol. 93(1). pp. 10-20.

Heider, E. Rosch and Olivier, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology.* Vol. 3. pp. 337-354.

Hays, W. M. (1988). *Statistics*, 4th Ed. New York: Holt Rinehart and Winston, Inc.

Huberman, B. A. and Glance, N. S. (1993). Diversity and collective action. *Interdisciplinary Approaches to Nonlinear Complex Systems.* Berlin: Springer-Verlag.

Johansson, S. and Hofland, K. (1989). *Frequency analysis of english vocabulary and grammer: based on the LOB corpus.* New York: Oxford Univeristy Press.

Jones, G. V. (1983). Identifying basic categories. *Psychological Bulletin.* Vol. 94(3). pp. 423-428.

Kasperson, C. J. (1985). An explanation of the relationship between performance, decision making, and structure. *Human Relations, 38*(5). 441-456.

Kay, P. (1981). Color perception and the meaning of color words. In the proceedings of *The Third Annual Conference of the Cognitive Science Society.* Berkeley, California.

Kay, P. and McDaniel, C. K. (1978). The linguistic significance of the meaning of basic color terms. *Language.* Vol. 54(3). p. 610-646.

Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind.* University of Chicago Press.

Lakoff, G. (1973). Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic.* Vol. 2. p. 458-508.

Lenneberg, E. (1961). Color naming, color recognition, color discrimination: a re-appraisal. *Perception and Moter Skills*, 69. 472-481.

Lutz, J. H. (1990). Category and measure in complexity classes. *1990 SIAM Journal of Computing.* Vol. 19(6). p. 1100-1131.

March, J. G. (1990). How decisions are made. Keynote Address at the 1990 conference on Computer-Supported Cooperative Work. Unpublished document.

March, J. G. and Olsen, J. P. (1976). *Ambiguity and choice in organizations.* Bergen, Norway: Universitetsforlaget.

Masuch, M. and LaPotin, P. (1989). Beyond garbage cans: An AI model of organizational choice. *Administrative Science Quarterly*, 34. p. 38 - 67.

Maynard Smith, J. (1972). *On evolution.* Edinburgh University Press.

Mervis, C. B. and Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology.* Vol. 32. p. 89-115.

Mervis, C. B. and Roth, E. M. (1981). The internal structure of basic and non-basic color categories. *Language.* Vol. 57(2). p. 384-405.

Miller, G. A. and Johnson-Laird, P. N. (1976). *Language and Perception.* Cambridge, Massachusetts: The Belkamp Press.

Mintzberg, H.; Rasinghani, D. and Théorêt, A. (1976). The structure of "unstructured" decision processes. *Administrative Science Quarterly, 21.* p. 246-275.

Morgan, G. (1986). *Images of Organizations.* Newberry Park, Califor-nia: Sage press.

Morgan, M. G. & Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in qualitative risk and policy analysis.* Cambridge University Press.

Nutt, P.C. (1984). Types of organizational decision processes. *Administrative Science Quarterly, 29.* p. 414-450.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, California: Morgan Kaufmann Publishers.

Rosch, E. (1988). Coherences and categorization: a historical view. In F. S. Kessel (Ed.) *The Development of Language and Language Research-ers: Essays in Honor of Roger Brown.* Hilldale, New Jersey. Lawerence Earlbaum and Associates.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8. 382-439.

Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology*. Vol. 7. p. 532-547.

Rosch, E. (1975b). The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance,* 1. 303-322.

Shannon, C. E. and Weaver, W. (1963). *The Mathematical Theory of Communication.* Urbana, Illinois: University of Illinois Press.

Targowski, A. S. and Bowman, J. P. (1988). The layer-based, pragmatic model of the communication process. *The Journal of Business Communication.* Vol. 25. p. 5-24.

Thayer, L. (1967). Communication and organizational theory. In F. E. X. Dance (Ed.) *Human Communication Theory: Orginal Essays.* New York: Holt, Rinehart and Winston, Inc.

Thompson, S. (1957). *Motif-index of folk-literature.* Indiana University Press: Bloomington, Indiana.

Weick, K. E. (1979). *The social psychology of organizations.* Second edition. New York: Random House.