A note to our readers:

Except for the use of brief excerpts for the purposes of review, all requests for permission to reprint from this text should be addressed to the University of Chicago Press, 5801 South Ellis Avenue, Chicago, Illinois, 60637

~ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences-Permanence of Paper for Printed Library Materials, ANSI Z39.48- 1992.

Cybernetic Explanation*

It may be useful to describe some of the peculiarities of cybernetic explanation.

Causal explanation is usually positive. We say that billiard ball B moved in such and such a direction because billiard ball A hit it at such and such an angle. In contrast to this, cybernetic explanation is always negative. We consider what alternative possibilities could conceivably have occurred and then ask why many of the alternatives were not followed, so that the particular event was one of those few which could, in fact, occur. The classical example of this type of explanation is the theory of evolution under natural selection. According to this theory, those organisms which were not both physiologically and environmentally viable could not possibly have lived to reproduce. Therefore, evolution always followed the pathways of viability. As Lewis Carroll has pointed out, the theory explains quite satisfactorily why there are no bread-and-butter-flies today.

In cybernetic language, the course of events is said to be subject to restraints, and it is assumed that, apart from such restraints, the pathways of change would be governed only by equality of probability. In fact, the "restraints" upon which cybernetic explanation depends can in all cases be regarded as factors which determine inequality of probability. If we find a monkey striking a typewriter apparently at random but in fact writing meaningful prose, we shall look for restraints, either inside the monkey or inside the typewriter. Perhaps the monkey could not strike inappropriate letters; perhaps the type bars could not move if improperly struck; perhaps incorrect letters could not survive on the paper. Somewhere there must have been a circuit which could identify error and eliminate it.

Ideally—and commonly—the actual event in any sequence or aggregate is uniquely determined within the terms of the cybernetic explanation. Restraints of many different kinds may combine to generate this unique determination. For example, the selection of a piece for a given position in a jigsaw puzzle is "restrained" by many factors. Its shape must conform to that of its several neighbors and possibly that of the boundary of the puzzle; its color must conform to the color pattern of its region; the orientation of its edges must obey the topological regularities set by the cutting machine in which the

puzzle was made; and so on. From the point of view of the man who is trying to solve the puzzle, these are all clues, i.e., sources of information which will guide him in his selection. From the point of view of the cybernetic observer, they are *restraints*.

Similarly, from the cybernetic point of view, a word in a sentence, or a letter within the word, or the anatomy of some part within an organism, or the role of a species in an ecosystem, or the behavior of a member within a family— these are all to be (negatively) explained by an analysis of restraints.

The negative form of these explanations is precisely comparable to the form of logical proof by *reductio ad absurdum*. In this species of proof, a sufficient set of mutually exclusive alternative propositions is enumerated, e.g., "P" and "not P," and the process of proof proceeds by demonstrating that all but one of this set are untenable or "absurd." It follows that the surviving member of the set must be tenable within the terms of the logical system. This is a form of proof which the nonmathematical sometimes find unconvincing and, no doubt, the theory of natural selection sometimes seems unconvincing to nonmathematical persons for similar reasons—whatever those reasons may be.

Another tactic of mathematical proof which has its counterpart in the construction of cybernetic explanations is the use of "mapping" or rigorous metaphor. An algebraic proposition may, for example, be mapped onto a system of geometric coordinates and there proven by geometric methods. In cybernetics, mapping appears as a technique of explanation whenever a conceptual "model" is invoked or, more concretely, when a computer is used to simulate a complex communicational process. But this is not the only appearance of mapping in this science. Formal processes of mapping, translation, or transformation are, in principle, imputed to *every* step of any sequence of phenomena which the cyberneticist is attempting to explain. These *mappings* or transformations may be very complex, *e.g.*, where the output of some machine is regarded as a transform of the input; or they may be very simple, *e.g.*, where the rotation of a shaft at a given point along its length is regarded as a transform (albeit identical) of its rotation at some previous point.

The relations which remain constant under such transformation may be of any conceivable kind.

This parallel, between cybernetic explanation and the tactics of logical or mathematical proof, is of more than trivial interest. Outside of cybernetics, we look for explanation, but not for anything which would simulate logical proof. This simulation of proof is something new. We can say, however, with hindsight wisdom, that explanation by simulation of logical or mathematical proof was expectable. After all, the subject matter of cybernetics is not events and objects but the *information* "carried" by events and objects. We consider the objects or events only as proposing facts, propositions, messages, percepts, and the like. The subject matter being propositional, it is expectable that explanation would simulate the logical.

Cyberneticians have specialized in those explanations which simulate *reductio ad absurdum* and "mapping." There are perhaps whole realms of explanation awaiting discovery by some mathematician who will recognize, in the informational aspects of nature, sequences which simulate other types of proof.

Because the subject matter of cybernetics is the propositional or informational aspect of the events and objects in the natural world, this science is forced to procedures rather different from those of the other sciences. The differentiation, for example, between map and territory, which the semanticists insist that scientists shall respect in their writings, must, in cybernetics, be watched for in the very phenomena about which the scientist writes. Expectably, communicating organisms and badly programmed computers will mistake map for territory; and the language of the scientist must be able to cope with such anomalies. In human behavioral systems, especially in religion and ritual and wherever primary process dominates the scene, the name often *is* the thing named. The bread *is* the Body, and the wine *is* the Blood.

Similarly, the whole matter of induction and deduction —and our doctrinaire preferences for one or the other— will take on a new significance when we recognize inductive and deductive steps not only in our own argument but also in the relationships among data.

Of special interest in this connection is the relationship between *context* and its content. A phoneme exists as such only in combination with other phonemes which make up a word. The word is the *context* of the phoneme. But the word only exists as such—only has "meaning"—in the larger context of the utterance, which again has meaning only in a relationship.

This hierarchy of contexts within contexts is universal for the communicational (or "emic") aspect of phenomena and drives the scientist always to seek for explanation in the ever-larger units. It may (perhaps) be true in physics that the explanation of the macroscopic is to be sought in the microscopic. The opposite is usually true in cybernetics: without context, there is no communication.

In accord with the negative character of cybernetic explanation, "information" is quantified in negative terms. An event or-object such as the letter K in a given position in the text of a message *might* have been any other of the limited set of twenty-six letters in the English language. The actual letter excludes (*i.e.*, eliminates by restraint) twenty-five alternatives. In comparison with an English letter, a Chinese ideograph would have excluded several thousand alternatives. We say, therefore, that the Chinese ideograph carries more information than the letter. The quantity of information is conventionally expressed as the log to base 2 of the improbability of the actual event or object.

Probability, being a ratio between quantities which have similar dimensions, is itself of zero dimensions. That is, the central explanatory quantity, information, is of zero dimensions. Quantities of real dimensions (mass, length, time) and their derivatives (force, energy, etc.) have no place in cybernetic explanation.

The status of energy is of special interest. In general in communicational systems, we deal with sequences which resemble stimulus-and-response rather than cause-and-effect. When one billiard ball strikes another, there is an energy transfer such that the motion of the second ball is energized by the impact of the first. In communicational systems, on the other hand, the energy of the response is usually provided by the respondent. If I kick a dog, his immediately sequential behavior is energized by his metabolism, not by my kick. Similarly, when one neuron fires another, or an impulse from a microphone activates a circuit, the sequent event has its own energy sources.

Of course, everything that happens is still within the limits defined by the law of energy conservation. The dog's metabolism might in the end limit his response, but, in general, in the systems with which we deal, the energy supplies are large compared with the demands upon them; and, long before the supplies are exhausted, "economic" limitations are imposed by the finite number of available alternatives, there is an economics of probability. This economics differs from an economics of energy or money in that probability—being a ratio—is not subject to addition or subtraction but only to multiplicative processes, such as fractionation. A telephone exchange at a time of emergency may be "jammed" when a large fraction of its alternative pathways are busy. There is, then, a low probability of any given message getting through.

In addition to the restraints due to the limited economics of alternatives, two other categories of restraint must be discussed: restraints related to "feedback" and restraints related to "redundancy."

We consider first the concept of feedback: When the phenomena of the universe are seen as linked together by cause-and-effect and energy transfer, the resulting picture is of complexly branching and interconnecting chains of causation. In certain regions of this universe (notably organisms in environments, ecosystems, thermostats, steam engines with governors, societies, computers, and the like), these chains of causation form circuits which are *closed* in the sense that causal interconnection can be traced around the circuit and back through whatever position was (arbitrarily) chosen as the starting point of the description. In such a circuit, evidently, events at any position in the circuit may be expected to have effect at *all* positions on the circuit at later times.

Such systems are, however, always *open:* (a) in the sense that the circuit is energized from some external source and loses energy usually in the form of heat to the outside; and (b) in the sense that events within the circuit may be influenced from the outside or may influence outside events.

A very large and important part of cybernetic theory is concerned with the formal characteristics of such causal circuits, and the conditions of their stability. Here I shall consider such systems only as sources of *restraint*.

Consider a variable in the circuit at any position and sup-pose this variable subject to random change in value (the change perhaps being imposed by impact of some event external to the circuit). We now ask how this change will affect the value of this variable

at that later time when the sequence of effects has come around the circuit. Clearly the answer to this last question will depend upon the characteristics of the circuit and will, therefore, be *not random*.

In principle, then, a causal circuit will generate a non-random response to a random event *at that position in the circuit at which the random event occurred*.

This is the general requisite for the creation of cybernetic restraint in any variable at any given position. The particular restraint created in any given instance will, of course, depend upon the characteristics of the particular circuit— whether its overall gain be positive or negative, its time characteristics, its thresholds of activity, etc. These will together determine the restraints which it will exert at any given position.

For purposes of cybernetic explanation, when a machine is observed to be (improbably) moving at a constant rate, even under varying load, we shall look for restraints —*e.g.*, for a circuit which will be activated by changes in rate and which, when activated, will operate upon some variable (*e.g.*, the fuel supply) in such a way as to diminish the change in rate.

When the monkey is observed to be (improbably) typing prose, we shall look for some circuit which is activated whenever he makes a "mistake" and which, when activated, will delete the evidence of that mistake at the position where it occurred.

The cybernetic method of negative explanation raises the question: Is there a difference between "being right" and "not being wrong"? Should we say of the rat in a maze that he has "learned the right path" or should we say only that he has learned "to avoid the wrong paths"?

Subjectively, I feel that I know how to spell a number of English words, and I am certainly not aware of discarding as unrewarding the letter K when I have to spell the word "many." Yet, in the first level cybernetic explanation, I should be viewed as actively discarding the alternative K when I spell "many."

The question is not trivial and the answer is both subtle and fundamental: choices are not all at the same level. I may have to avoid error in my choice of the word "many" in a given context, discarding the alternatives, "few," "several" "frequent," etc. But if I can achieve this higher-level choice on a negative base, it follows that the word many and its alternatives somehow must be conceivable to me— must exist as distinguishable and possibly labeled or coded patterns in my neural processes. If they do, in some sense, exist, then it follows that, after making the higher-level choice of what word to use, I shall not necessarily be faced with alternatives at the lower level. It may become unnecessary for me to exclude the letter K from the word "many." It will be correct to say that I know positively how to spell "many"; not merely that I know how to avoid making mistakes in spelling that word.

It follows that Lewis Carroll's joke about the theory of natural selection is not entirely

cogent. If, in the communicational and organizational processes of biological evolution, there be something like levels—items, patterns, and possibly patterns of patterns—then it is logically possible for the evolutionary system to make something like positive choices. Such levels and patterning might conceivably be in or among genes or elsewhere.

The circuitry of the above mentioned monkey would be required to recognize deviations from "prose," and prose is characterized by pattern or—as the engineers call it—by redundancy.

The occurrence of the letter K in a given location in an English prose message is not a purely random event in the sense that there was ever an equal probability that any other of the twenty-five letters might have occurred in that location. Some letters are more common in English than others, and certain combinations of letters are more common than others. There is, thus, a species of patterning which partly determines which letters shall occur in which slots. As a result: if the receiver of the message had received the entire rest of the message but had not received the particular letter K which we are discussing, he might have been able, with better than random success, to guess that the missing letter was, in fact, K. To the extent that this was so, the letter K did not, for that receiver, exclude the other twenty-five letters because these were already partly excluded by information which the recipient received from the rest of the message. This patterning or predictability of particular events within a larger aggregate of events is technically called "redundancy.

The concept of redundancy is usually derived, as I have derived it, by considering first the maximum of information which might be carried by the given item and then considering how this total might be reduced by knowledge of the surrounding patterns of which the given item is a component part. There is, however, a case for looking at the whole matter the other way round. We might regard patterning or predictability as the very essence and *raison d'être* of communication, and see the single letter unaccompanied by collateral clues as a peculiar and special case.

The idea that communication is the creation of redundancy or patterning can be applied to the simplest engineering examples. Let us consider an *observer* who is watching A send a message to B. The purpose of the transaction (from the point of view of A and B) is to create in B's message pad a sequence of letters identical with the sequence which formerly occurred in A's pad. But from the point of view of the observer this is the creation of redundancy. If he has seen what A had on his pad, he will not get any new information about the message itself from inspecting B's pad.

Evidently, the nature of "meaning," pattern, redundancy, information and the like, depends upon where we sit. In the usual engineers' discussion of a message sent from A to B, it is customary to omit the observer and to say that B received information from A which was measurable in terms of the number of letters transmitted, reduced by such redundancy in the text as might have permitted B to do some guessing. But in a wider universe, *i.e.*, that defined by the point of view of the observer, this no longer appears as a

"transmission" of information but rather as a spreading of redundancy. The activities of A and B have combined to make the universe of the observer more predictable, more ordered, and more redundant. We may say that the rules of the "game" played by A and B explain (as "restraints") what would otherwise be a puzzling and improbable coincidence in the observer's universe, namely the conformity between what is written on the two message pads.

To guess, in essence, is to face a cut or slash in the sequence of items and to predict across that slash what items might be on the other side. The slash may be spatial or temporal (or both) and the guessing may be either predictive or retrospective. A pattern, in fact, is definable as an aggregate of events or objects which will permit in some degree such guesses when the entire aggregate is not available for inspection.

But this sort of patterning is also a very general phenomenon, outside the realm of communication *between* organisms. The reception of message material by *one* organism is not fundamentally different from any other case of perception. If I see the top part of a tree standing up, I can predict —with better than random success—that the tree has roots in the ground. The percept of the tree top is redundant with (*i.e.*, contains "information" about) parts of the system which I cannot perceive owing to the slash provided by the opacity of the ground.

If then we say that a message has "meaning" or is "about" some referent, what we mean is that there is a larger uni-verse of relevance consisting of message-plus-referent, and that redundancy or pattern or predictability is introduced into this universe by the message. If I say to you "It is raining," this message introduces redundancy into the universe, message-plus-raindrops, so that from the message alone you could have guessed—with better than random success—something of what you would see if you looked out of the window. The universe, message-plus-referent, is given pattern or form —in the Shakespearean sense, the universe is *informed* by the message; and the "form" of which we are speaking is not in the message nor is it in the referent. It is a correspondence between message and referent.

In loose talk, it seems simple to locate information. The letter K in a given slot proposes that the letter in that particular slot is a K. And, so long as all information is of this very direct kind, the information can be "located": the information about the letter K is seemingly in that slot.

The matter is not quite so simple if the text of the message is redundant but, if we are lucky and the redundancy is of low order, we may still be able to point to parts of the text which indicate (carry some of the information) that the letter K is expectable in that particular slot.

But if we are asked: Where are such items of information as that: (a) "This message is in English"; and (b) "In English, a letter K often follows a letter C, except when the C begins a word"; we can only say that such information is *not* localized in any part of the text but is rather a statistical induction from the text as a whole (or perhaps from an

aggregate of "similar" texts). This, after all, is meta-information and is of a basically different order—of different logical type—from the information that "the letter in this slot is K."

This matter of the localization of information has bedeviled communication theory and especially neurophysiology for many years and it is, therefore, interesting to consider how the matter looks if we start from redundancy, pattern or form as the basic concept.

It is flatly obvious that no variable of zero dimensions can be truly located. "Information" and "form" resemble contrast, frequency, symmetry, correspondence, congruence, conformity, and the like in being of zero dimensions and, therefore, are not to be located. The contrast between this white paper and that black coffee is not somewhere between the paper and the coffee and, even if we bring the paper and coffee into close juxtaposition, the contrast between them is not thereby located or pinched between them. Nor is that contrast located between the two objects and my eye. It is not even in my head; or, if it be, then it must also be in your head. But you, the reader, have not seen the paper and the coffee to which I was referring. I have in my head an image or transform or name of the contrast between them; and you have in your head a transform of what I have in mine. But the conformity between us is not localizable. In fact, information and form are not items which can be localized.

It is, however, possible to begin (but perhaps not com-plete) a sort of mapping of formal relations within a system containing redundancy. Consider a finite aggregate of objects or events (say a sequence of letters, or a tree) and an observer who is already informed about all the redundancy rules which are recognizable (*i.e.,* which have statistical significance) within the aggregate. It is then possible to delimit regions of the aggregate within which the observer can achieve better than random guessing. A further step toward localization is accomplished by cutting across these regions with slash marks, such that it is across these that the educated observer can guess, from what is on one side of the slash, something of what is on the other side.

Such a mapping of the distribution of patterns is, however, in principle, incomplete because we have not considered the sources of the observer's prior knowledge of the redundancy rules. If, now, we consider an observer with *no* prior knowledge, it is clear that he might discover some of the relevant rules from his perception of *less* than the whole aggregate. He could then use his discovery in predicting *rules* for the remainder— rules which would be correct even though not exemplified. He might discover that "H often follows T" even though the remainder of the aggregate contained no example of this combination. For this order of phenomenon a different order of slash mark—meta-slashes will be necessary.

It is interesting to note that meta-slashes which demarcate what is necessary for the naive observer to discover a rule are, in principle, displaced relative to the slashes which would have appeared on the map prepared by an observer totally informed as to the rules of redundancy for that aggregate. (This principle is of some importance in aesthetics. To the aesthetic eye, the form of a crab with one claw bigger than the other is not simply

asymmetrical. It first proposes a rule of symmetry and then subtly denies the rule by proposing a more complex combination of rules.)

When we exclude all things and all real dimensions from our explanatory system, we are left regarding each step in a communicational sequence as a *transform* of the previous step. If we consider the passage of an impulse along an axon, we shall regard the events at each point along the pathway as a transform (albeit identical or similar) of events at any previous point. Or if we consider a series of neurons, each firing the next, then the firing of each neuron is a transform of the firing of its predecessor. We deal with event sequences which do not necessarily imply a passing on of the same energy.

Similarly, we can consider any network of neurons and arbitrarily transect the whole network at a series of different positions; then, we shall regard the events at each transsection as a transform of events at some previous transsection.

In considering perception, we shall not say, for example, "I see a tree," because the tree is not within our explanatory system. At best, it is only possible to see an image which is a complex but systematic transform of the tree. This image, of course, is energized by my metabolism and the nature of the transform is, in part, determined by factors within my neural circuits: "I" make the image, under various restraints, some of which are imposed by my neural circuits, while others are imposed by the external tree. An hallucination or dream would be more truly "mine" insofar as it is produced without immediate external restraints.

All that is not information, not redundancy, not form and not restraints—is noise, the only possible source of *new patterns*.