

INFORMATION, CONSTRAINT, AND MEANING

Martin Zwick

Systems Science PhD Program
Portland State University
Portland, Oregon 97207

ABSTRACT

Despite the familiar and correct disclaimer that information theory (Shannon and Weaver, 1949) does not concern the semantic level of communication, the technical definition of information nonetheless bears directly and importantly on the subject of meaning. Meaning, at least in one sense of the word, is the recognition of constraint and is based on isomorphism of structure. Constraint reduces information, yet information is also the very substrate of meaning. Meaning is thus the union of the informative and the intelligible (Moles, 1958), the reconciliation of this dialectical opposition being achievable in several different ways.

- - -

It is commonly understood that "information" in the mathematical theory of communication (Shannon and Weaver, 1949) refers only to the technical aspects of signal transmission and not at all to questions of meaning. To deal with such questions, Bar-Hillel and Carnap (1953, 1964) developed a semantic theory of information formally similar to Shannon's theory but based upon a logical, as opposed to statistical, notion of probability. The Shannon entropy expression, $H = -\sum p_j \log p_j$, in the semantic theory is used for the average uncertainty associated with a set of possible meanings, as compared, in the technical theory, to the average uncertainty associated with a set of (emitted or received) symbols. Both the technical and semantic theories are thus essentially theories of selection.

But by virtue of this formal similarity, the technical and semantic levels remain fully distinct and unbridgeable. This complete separability of levels precludes any quantitative linkage between information and meaning, though meaning depends upon the receipt of information, and if meaning is to be quantified, one might expect that its amount should be a function of (and possibly vary monotonically with) the amount of information. The relationship, however, must be more complex since from a random sequence of symbols one gains maximum information, but no meaning. Indeed, Weaver remarked on his vague feeling that

information and meaning may prove to be something like a pair of canonically conjugate variables in quantum theory . . . subject to some joint restriction that condemns a person to the sacrifice of the one as he insists on having much of the other

(Shannon & Weaver, 1949, p 28)

but gave no specific basis for such a complementarity. The purpose of this note is to pursue this suggestion further and by doing so demonstrate a complex relationship between information, constraint, and meaning.

Information, for Shannon and Weaver, is reduction in uncertainty $i = -\Delta u = u_{\text{initial}} - u_{\text{final}}$. If a definite message is received, i.e., the final uncertainty is zero, then the information gained is equal to the initial uncertainty. Messages from a source about which there is complete initial uncertainty, i.e., a random source, carry maximum information. Thus, for example, if a sequence of symbols s_1, s_2, \dots, s_n is received, one gains maximal information when receipt of s_j in no way reduces the uncertainty of what s_{j+1} (and succeeding symbols) will be. To the extent that the source is nonrandom and there exist constraints between symbols at different times, the information provided by the message is reduced (Shannon and Weaver, 1949; Miller, 1963).

Constraint reduces information. Yet strong constraint is present in natural language, e.g., in the sequence of letters in words and of words in sentences. This constraint is normally referred to as redundancy; and for the English language, redundancy is approximately 50% (Shannon and Weaver, 1949). This redundancy must be performing some useful function, since it reduces the information content of messages from what is theoretically achievable. The function is normally assumed to be error correction: if part of the message is lost, redundancy enhances the possibility that the receiver can reconstruct what was lost or ignore it.

It is unlikely, however, that the function of error correction accounts

sufficiently for redundancy in language. Consider an ideal communication situation where the message received is exactly the message sent. Error correction would not be necessary, so a redundancy-free language would seem to be optimal, since a minimal number of symbols would be required for any message.

Yet such a language is hardly conceivable. Because of the complete absence of constraint between sequential symbols, a message will appear to be coming from a random source. How could it be understood? The message will be perceived as a unit without parts since there can be no basis to demarcate substrings in a string of symbols emitted from a random source. For parts to exist in some relation, some constraint must exist between them, which has by assumption been precluded.

The meaning of such a message would have to be looked up in a (vast) reference table. This follows from the unitary character of every message: each will be mappable onto a unique number with a unique meaning. (E.g., message #2,092,154 might be translated as "The dog chased the cat.") Such a table would have to encompass all possible linguistic utterances, whose number is extremely large, indeed noncomputable. So a table of this sort could not in fact be prepared. One may object: why should it be necessary to translate the original message into a natural language utterance? Why cannot each message (say, "2,092,154") "point" directly to its meaning, i.e., to our experience of the world received via the senses?

This raises the central question: What is required for experience to have meaning, which can be the semantic content of a message? The answer has been clearly articulated by Ashby (1956). For experience to be comprehensible, some constraint must exist in the elements of experience. (Mathematically, constraint between two variables, say x and y , is defined set-theoretically (Wiener, 1917), as the joint occurrence of (x, y) in only a subset of the Cartesian product of the sets of x and y values). What can be understood is constraint, i.e., the order which we see or (perhaps) project onto the world.

If what is intelligible in experience is constraint, it is natural that the means by which we represent or express experience should exhibit a corresponding constraint. Constraint implies redundancy. Here, then, is a second explanation for redundancy in language. Constraint is a basis for meaning. The meaning of an utterance

is based on a mapping between the constraint embodied in experience and the constraint represented in language. To be more precise: What is mapped are elements and the constraining relations between them. In language, the words are the elements; the nonarbitrary sequence of words in sentences, the constraints. (Words themselves are constrained forms of smaller units, and sentences are elements in larger units, but we shall here ignore these other levels of language.) Words map onto different "chunked" aspects of our experience; sentences map onto the relations we perceive between these chunked aspects. There are thus at least two senses of "meaning": meaning₁, the mapping of one element onto another, and meaning₂, the mapping of a relation (constraint) onto a relation (in addition to a mapping of elements onto elements).

The first sense of "meaning" is fairly simple, namely the relation of symbol and referent, e.g., the relation of the word "dog" to the familiar four-legged creature referred to. It is in this sense of meaning that "2,092,154" can mean "The dog chased the cat." Only this type of meaning is possible in a language with no redundancy; this also resembles Bar-Hillel and Carnap's notion of meaning as involving a selection from a set of possibilities.

The second sense of "meaning" includes the first as a component, but is more complex. Meaning₂ is the isomorphism of two or more structures, where "structure" is a set of elements organized by some constraint. If the constraint in a received message is sufficiently isomorphic to some reference constraint, the former is taken to "mean" the latter. The reference constraint can be internal, e.g., some mental representation of past experience, or external; and the two structures can be given either in the same or in different modes of representation. Meaning₂ arises from the recognition of constraint.

Constraint can be quantitatively assessed as information-theoretic "transmission," which we will write as "c." For a message consisting of n symbols, s_1, s_2, \dots, s_n ,

$$c = \sum_j u(s_j) - u(s_1, s_2, \dots, s_n)$$

The u 's represent initial uncertainties; the final uncertainties are zero. $u(s_1, s_2, \dots, s_n)$ is thus the information gained, i , upon receipt of the message. $\sum_j u(s_j)$ is the maximum information which could possibly have been

gained, were there no redundancy in the message, and is a constant, call it K , for all fixed length messages. Thus,

$$c + i = K$$

i.e., the sum of the amount of constraint and the amount of information is constant. Thus redundancy, defined as $\frac{c}{K}$, is necessary, not merely for error-correction, but more fundamentally for structure, from which there arises the possibility of meaning₂.

One could then conceivably express the dependence of meaning upon constraint in terms of a monotonic dependence of the quantity of the former upon the quantity of the latter, i.e., by $m \propto c$. Given $i + c = K$, here is a clear basis for Weaver's idea of information and meaning being "subject to the sacrifice of one . . . {from} . . . having much of the other," quoted earlier.

But the matter is more complex and perhaps even paradoxical. Meaning₂, being based on an isomorphism of structure, requires that a message embody constraint. Constraint reduces information. Yet meaning presupposes and depends upon the receipt of information; and if the meaning of a message could be quantified, it might be expected also to vary monotonically with the message's information content. A message with little information cannot support much structure and thus much meaning. Hence constraint is necessary for meaning, yet also, as it were, undermines (reduces) it.

This analysis is not restricted to language, but applies to all forms of communication. Information and constraint must similarly be balanced in the visual arts, in music, etc. Too much information becomes chaos, which has no meaning. Too much constraint is boring because, however "meaningful" the communication is, it conveys too little information.

This dialectical tension has been eloquently and extensively discussed by Moles (1958) who notes the fundamental opposition in all forms of communication between what is intelligible and what is informative. This idea serves as the foundation for Moles' rich and deep analysis of the relationship between information theory and aesthetic perception, this analysis showing conclusively that, despite the oft-cited and necessary disclaimer that Shannon's theory does not explicitly deal with meaning (which, strictly speaking, is correct), the technical theory nonetheless bears very considerably--and directly--on this

important subject.

There are at least three ways by which tension between variety (information) and constraint can be satisfactorily resolved. In the first, a compromise or balance is reached between the two conflicting needs. In the second, the extremes are not subjected to compromise but are superposed. The message is divided into two parts, one embodying extreme constraint, the other embodying extreme variety, so that the goals are met separately. In the third, there is also no compromise but the conflicting needs for variety and constraint are satisfied sequentially in time, either once or repeatedly.

One might represent quantitatively the strategy of compromise. Let m equal the "amount" of meaning₂ in a message.

If, from one point of view, m should vary directly with the amount of constraint in a message and yet, from another perspective, it should vary directly with the amount of information in a message, one might consider the possibility of writing $m = \beta ic$, where β is some constant. Since $i + c = K$, maximum m is obtained when $i = c = \frac{K}{2}$, i.e., when redundancy is 50%. This, we recall, is the redundancy of English, and the comparable figure for French is 45% (Moles, 1958); presumably the redundancies of other languages tend to be closer to this value rather than to either extreme of 0 or 100%. This analysis is suggestive as a starting point for more intensive study, but should not yet be taken too seriously.

Another quantitative illustration can be given for this idea of compromise. The analysis (Voss and Clark, 1975, 1978; see also Gardner, 1978) of the sound spectra of musical works (of greatly varying styles--from Bach and Beethoven to the Beatles) reveals correlations of strength inversely proportional to frequency; that is, music resembles what physicists call $1/f$ noise. Such correlations lie intermediate between white noise, whose frequency spectrum is flat and which corresponds to the complete absence of order, and Brownian noise, which is more ordered and has correlations varying inversely with the square of the frequency. The $1/f^1$ spectrum is "intermediate" between $1/f^0$ and $1/f^2$ spectra: music represents a compromise between high information (white noise) and moderate constraint (Brownian noise).

To avoid the implication that there is some specific optimal balance of information and constraint, it might be preferable to speak of a range

of compromise values, or a "plateau of meaning" between the extremes of order (maximum constraint) and chaos (maximum information) as shown in Figure 1. (The plateau imagery is adapted from Miller, 1971; Hardin, 1963.)

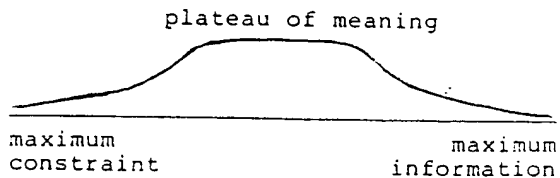
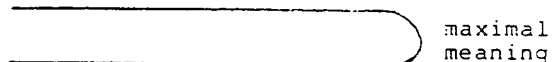


Figure 1.

One might also fold the horizontal axis over, giving the representation of Figure 2; meaning now increases monotonically in one direction.

minimal meaning
due to
insufficient
constraint



minimal meaning
due to
insufficient
information

Figure 2.

Or, one might employ catastrophe-theoretic imagery (Thom, 1974). Using the dual of the cusp catastrophe (where the region inside the bifurcation set corresponds to a stable equilibrium surface), we have the representation of Figure 3.

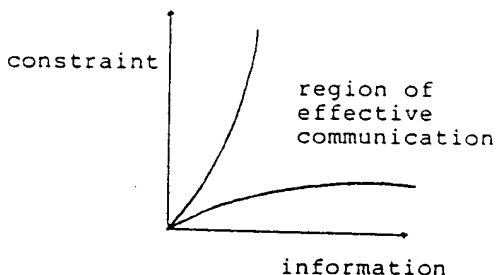


Figure 3. Constraint and information treated as independent variables (K is allowed to vary).

With the simple schema of Figure 1, one might interpret the aim of certain movements in the arts as the widening of this plateau in the direction of one or the other extremes. Minimal art (typified, say, by some monochromatic painting, in its most extreme manifestation by a white on white canvas) is an extension in the direction of maximal constraint. The "randomness" of some of the creations of Jackson Pollock or John Cage are attempts to extend the domain of effective communication in the opposite direction, towards maximal variety. Cage has also experimented at the other extreme, i.e., with minimal forms having extreme order.

In the face of overly chaotic or constrained forms, a viewer/listener may be forced to become aware of more--or new forms of--constraint or variety than was initially apparent. The first impression of a work of art or music as being too chaotic or too ordered may really be the product of artistic tradition, social convention, and/or personal habit.

Artistic experiments probing the meaningful limits of order or chaos may also depend for their success upon a different mechanism. Such works operate at a meta-level of communication, saying, in effect: this is art despite its extreme randomness, or despite its extreme order. The "frame message" of the work (Hofstadter, 1978), the primary communication of any message that it is in fact a message, guarantees at least a certain kind of meaning, irrespective of content. What is exhibited in museums, performed on stages, etc or what is done by "artists," is defined and usually accepted (sometimes reluctantly) as art.

As Hofstadter has pointed out, such communications require, for their appreciation, familiarity with the history and current manifestations of the particular art form. This meaning is not internal, i.e., deriving from some constraint upon variety, but external. It is neither meaning₁ (a label or mapping invoked by some act of selection), nor meaning₂ (an isomorphism of constraint), but a third alternative: the communication in its entirety is an element in a larger system, which consists of other elements constrained by a higher level relation. Meaning₃ is defined in terms of context. To put it simply: meaning₂ is based on "structure," the internal order present in a system, while meaning₃ is based on "function" or "history," the external order in which it participates, whether synchronic or diachronic. The three senses of meaning are summarized diagram-

matically in Figure 4.

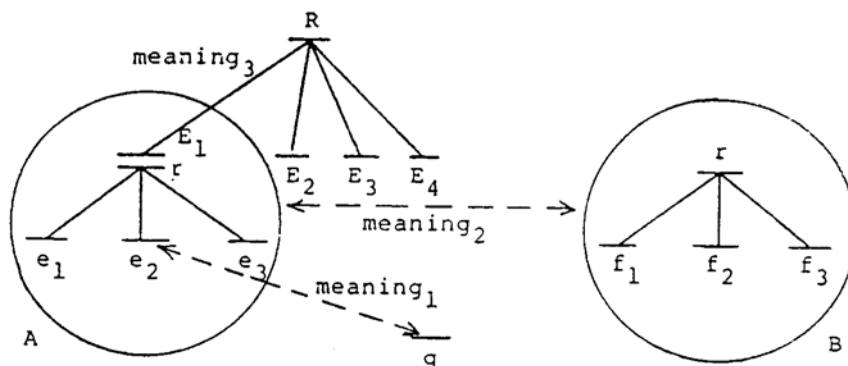


Figure 4. "Three Senses of Meaning." Elements $e_1, e_2,$ and e_3 are organized (constrained) by relation r , to form system A . Meaning_2 is the isomorphism of system A with some other system, B . Meaning_1 is a simpler concept, referring to the mapping of a structureless element, say e_2 , onto some other element, say g . System A can also act as an element, E_1 , in a higher level system constrained by relation R . The entire system can have a meaning_3 by virtue of the relation R of E_1 to other elements, $E_2, E_3,$ and E_4 .

In terms of either meaning_1 or meaning_3 , a message is treated as a unit. (For meaning_1 the message is structureless, while for meaning_3 it may have some internal order which underlies its relations to other units.) For both meaning_1 and meaning_3 , therefore, the preceding discussion of complementarity of information and constraint is irrelevant; such considerations apply only to meaning_2 .

The second means by which the need for both variety and constraint may be reconciled, namely the partitioning of a message in two, is illustrated by the DNA molecule which fulfills Schrodinger's scientific prophecy that the physical substrate of genetic memory should be an "aperiodic crystal." DNA is such a union of opposites: maximal constraint is embodied in the crystalline, i.e., perfectly periodic, sugar-phosphate backbone of each strand of the molecule; maximal information is embodied in the aperiodicity, i.e., irregular sequence, of bases attached to the sugars (Figure 5).

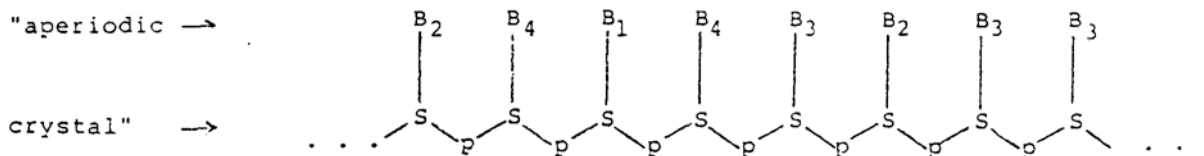


Figure 5. Symbolic representation of one (stretched-out) strand of DNA illustrating Schrodinger's idea of an aperiodic crystal. (S = sugar, P = phosphate, B_{1,2,3,4} = bases = adenine, thymine, cytosine, guanine.)

Normally the DNA message is said to reside in the base sequence, that is, only the aperiodic part, which determines the amino acid sequence in proteins. (This is what Hofstadter terms an "inner message.") Yet DNA must also "communicate" how it is to be read, its decoding mechanism; that is, it must provide for its interaction with a variety of enzymes. (Hofstadter calls the specification of decoding procedures the "outer message.") A great proportion (but not all) of this message is located in the spatial structure and chemical properties of the sugar-phosphate backbone. In this case, then, the inner message embodies high information and the outer message embodies high constraint, though both components of the message are read simultaneously.

The third way by which the tension between variety and constraint may be resolved is to communicate a message which is first perceived as having high information but whose structure is soon thereafter cognized. This is succinctly summarized in the expressive statement of Gray (1974): "Meaning is the digestion of newness into sameness." Newness is information; sameness is isomorphism of structure.

For example, up to the end of a good detective or mystery story there will appear to be no order (in the explicit clues, in the characters of the protagonists, etc), but when the mystery is finally revealed, the pieces fall into place. What seemed to be a situation of maximum chaos is revealed to be a situation of maximum order. It is not a compromise that is reached between information and constraint. Rather, both extremes are superposed, but sequentially. Here, too, the frame message (include in this the specification of genre) is essential. The reader can bear the multiplicity of yet unrelated facts because the frame message guarantees the existence of some unifying order. Either it can be figured out by the reader--this is the challenge posed by this genre--or it will finally be revealed in the end.

This means of reconciling the contrary needs of information and constraint is not restricted to the detective or mystery story, but is a general feature of the novel and applies to masterpieces as well as potboilers. All this refers to the single level of plot, but such dialectical tensions and reconciliations exist at multiple structural levels in literary works, as well as in other artistic media. (See, e.g., Moles, 1958, Chapter VI, and the discussion of Gardner, 1978, of the work of Voss and the theory of fractals of Benoît Mandelbrot.)

Another example of this general schema is the placing of the verb in

German at the end of sentences; anticipation and suspense enhance the informational richness of the message when the key to its meaning is finally delivered. Here, too, the frame message ("This is a sentence in German") sustains the listener by providing a guarantee of the eventual comprehensibility of the utterance.

This approach of reconciling the conflicting needs of information and constraint can also be varied by proceeding, not from information to constraint, but in the reverse direction. Rather than being derived from the recognition of order in apparent disorder, meaning may be based upon the introduction of new variety, and hence imperfection, into existing constraint.

In imperfection there is novelty and thus information. This information has meaning since it occurs in the context of a recognized constraint. Moreover, the imperfection not only creates novelty but actual tension, which sustains interest in the unfolding, spatial and/or temporal, of the work. But the matter does not stop here. The tension must be resolved, i.e., the imperfection removed, typically by being perceived as lawful, i.e., the reflection of a "higher" (more encompassing) ordering principle. This means of reconciling the two needs of communication can be linked to the earlier approach discussed above: One proceeds to order either from disorder or from the distortion or imperfection in some prior order. Or there may be some genuine, if irregular, alternation of banality and originality (Moles, 1958).

In summary, then, while Shannon's information theory does not explicitly speak of meaning, the quantitative measure of information introduced by that theory cannot be dissociated from discussions of the semantic level of communication. "Information" and "meaning" are, rather, ultimately linked, via the universal dialectical complementarity of information and constraint. The tension produced by this complementarity can be resolved in a number of alternative ways, the consideration of which gives rise to several different meanings of "meaning."

REFERENCES

- Ashby, W Ross
 1976 An Introduction to Cybernetics. London: Methuen (first published, 1956).
- Bar-Hillel, Y
 1964 Language and Information. Reading, Mass: Addison-Wesley and Jerusalem: Jerusalem Academic Press.
- Bar-Hillel, Y and R Carnap
 1953 "Semantic Information." British Journal for the Philosophy of Science 4: 147-157.
- Gardner, Martin
 1978 "Mathematical Games." Scientific American (April): 16 ff.
- Gray, William
 1974 "Current Issues in General Systems Theory and Psychiatry." General Systems XIX; reprinted in: Chen, G et al The General Theory of Systems Applied to Management and Organizations, vol II, (1980) Seaside, Cal: Intersystems: 519-522.
- Hardin, Garrett
 1963 "The Cybernetics of Competition: A Biologist's View of Society." Perspectives in Biology and Medicine VII (Autumn): 61-84; reprinted in: Buckley, W (ed) Modern System Research for the Behavioral Scientist (1968) Chicago: Aldine: 449-459.
- Hofstadter, Douglas R
 1980 Gödel, Escher, Bach. New York: Vintage.
- Miller, G Tyler
 1971 Energetics, Kinetics, and Life. Belmont, Cal: Wadsworth Publishing Co.
- Miller, George A
 1963 "What is Information Measurement?" American Psychologist 3-11. reprinted in: Buckley, W (ed) Modern System Research for the Behavioral Scientist (1968) Chicago: Aldine: 123-128.
- Moles, Abraham
 1966 Information Theory and Aesthetic Perception. trans by Joel E Cohen. Urbana: University of Illinois Press. (published in French, 1958).
- Schrödinger, Erwin
 1967 What is Life? London: Cambridge University Press. (first published 1944).
- Shannon, Claude E and Warren Weaver
 1975 The Mathematical Theory of Information. Urbana, Ill: University of Illinois Press. (first published 1949)
- Thom, Rene
 1975 Structural Stability and Morphogenesis. New York: Benjamin Press. (translated by D H Fowler).
- Voss, Richard and John Clark
 1975 "'1/f Noise' in Music and Speech." Nature 258: 317-318.
- Voss, Richard and John Clark
 1978 "'1/f Noise' in Music; Music from 1/f Noise." Journal of Acoustical Society of America 53: 258-263.
- Wiener, Norbert
 1917 "A Simplification of the Logic of Relations." Proceedings of the Cambridge Philosophical Society 17: 387-390.