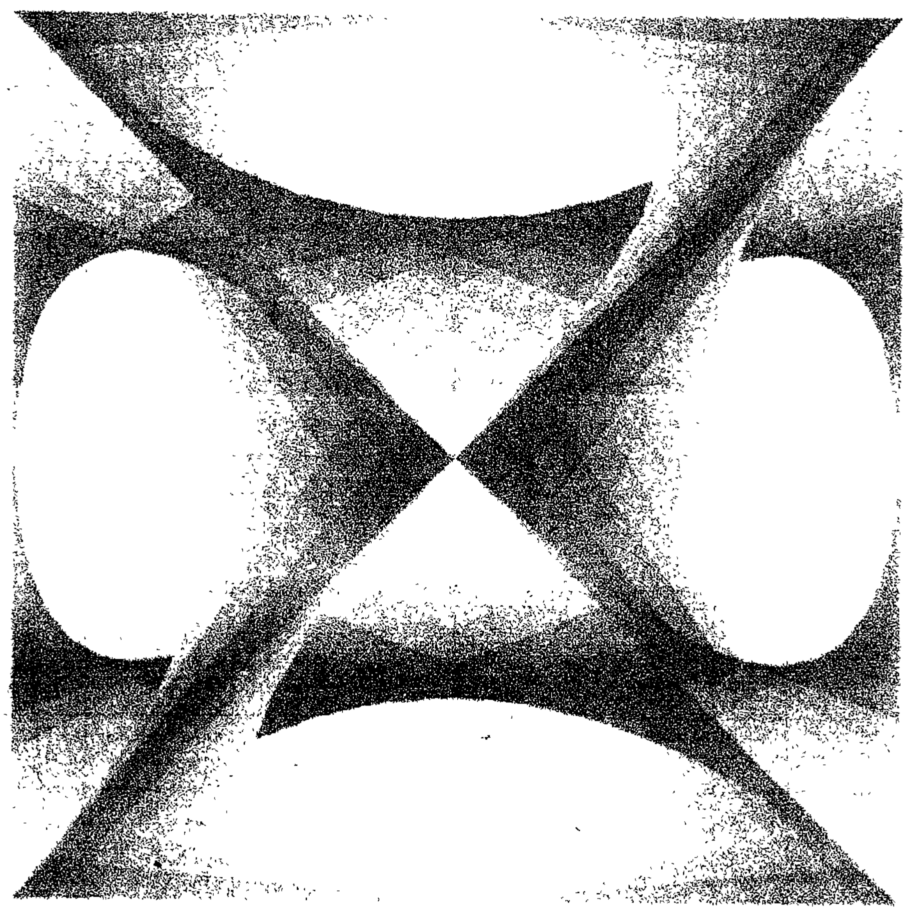
Symmetry: Culture and Science

Symmetry and Information

The Quarterly of the
International Society for the
Interdisciplinary Study of Symmetry
(ISIS-Symmetry)

Volume 8, Number 2, 1997



Symmetry: Culture and Science
Vol. 8, No. 2, 207-214, 1997

SYMMETRY: CULTURE AND SCIENCE

IS SYMMETRY INFORMATIVE?

John E. Gray and Andrew Vogt

address: Code N92, Department of Mathematics, Naval Surface Warfare Center/DD Georgetown University, Dahlgren, VA 22448 Washington, DC 20057-0996, U.S.A., E-mail: jgray@nswc.navy mil, andy@gumath1 math georgetown edu

1. INTRODUCTION

Is symmetry informative? The answer is both yes and no. We examine what information and symmetry are and how they are related. Our approach is primarily mathematical — not because mathematics provides the final word, but because it provides an insightful and relatively precise starting point.

Information theory treats transformations that messages undergo from source to destination. Symmetries are transformations that leave some property of interest unchanged. In this respect the studies of information and symmetry can both be regarded as a Quest for the identity transformation.

2. WHAT IS INFORMATION?

Shannon and Weaver (the latter wrote an introduction to Shannon's papers (Shannon and Weaver, 1964)) explicitly called Shannon's work communication theory. In fact, even more narrowly they called it "a mathematical theory of communication." Despite popular usage, no comprehensive theory of information exists. The common notion of "information" — i.e., facts, knowledge, data, structure — is probably too broad for scientific and mathematical characterization. Information theory is as applicable to the communication of a stream of nonsense as it is to the most profound emanations of the human psyche. Shannon (1964), in a fruitful oversimplification, declared that

> "... the semantic aspects of communication are irrelevant to the engineering aspects."

Every form of communication has three components: a message from a source (e.g., words spoken into a telephone), a signal representing this message that travels along a channel (e.g., a series of electrical pulses traveling along a telephone wire), and the message actually received by the destination (e.g., the reconstituted words heard by a listener on another telephone). The message is encoded as a signal, transmitted along the channel, and received and decoded at the destination. Enroute noise (e.g., the static on a telephone line) may alter the signal, so that the message received differs from the message sent.

The problem that information theory addresses is how best to encode and decode the message, taking into account the characteristics of source, channel, noise, and destination. "Best" is equated with maximum accuracy and speed subject to cost constraints. On the assumption that one-time-only messages are of no importance for design, the theory concerns itself with average accuracy, average speed, and average cost.

A message is a succession of elements (e.g., letters, words, sentences, digits, images, musical notes, etc.). These elements — call them symbols — are drawn from a preexisting set of possible symbols (e.g., an alphabet, dictionary, etc.), and over the course of many communications each symbol is assumed to occur with a definite relative frequency. (This is sometimes referred to as the stationarity assumption. It is false in most applications, but as frequencies usually change slowly with time, it is a useful approximation.)

Information theory associates with each symbol the quantity

$$\log_2(1/p)$$

where p is the relative frequency or probability of the symbol. This quantity came to be called the "information"[1] of the symbol. It is a measure of infrequency: the more infrequent the symbol, the greater the "information." This quantity has no direct connection with the particular information represented by the symbol. Indeed, since it depends only on the probability of the symbol, not its content or meaning, no such connection is possible. We call this quantity the "infrequency." If a symbol has a fifty percent probability of occurring, its infrequency is $\log_2(1/(1/2)) = \log_2(2) = 1$ and the symbol is said to consist of one "bit," a term coined by John Tukey. If it has probability $1/(2^n)$, the infrequency is $\log_2(2^n) = n$ bits.

The infrequency is related to binary coding. Encoding replaces one symbol set by another (e.g., alphanumeric computer keyboard input by the 0's and 1's of machine language). A binary code has two symbols (say 0 and 1) and encodes a message as a finite string of successive 0's and 1's. Binary codes are convenient in electronics and computer science since 0 and 1 can be translated into on/off states of electric circuits. If only three source symbols are to be sent — one of them of probability 1/2, and the other two of probability 1/4 each, the first symbol might be encoded as 0 and the other two as 1 and 01. Another choice is to use 0, 10, and 11 respectively. The latter code, called a Huffman code, allows us to interpret the sequence as soon as it is sent rather than waiting to see if more symbols will be sent. A "good" binary code, as the examples suggest, assigns shorter strings (fewer digits) to more frequent symbols and longer strings to less frequent symbols. The infrequency of a symbol is approximately the number of 0's and 1's needed for the symbol in a good binary code.

Parenthetically we note that the number of 0's and 1's in a binary signal is called the "bit count" and is itself spoken of as a measure of information. The bit count has nothing directly to do with either frequency or information content. It is just the total raw binary data actually transmitted between sites or stored at a site. Curiously, bit count treats 0's and 1's as if they are equal, but depending on the implementation the two digits may have

[1] There is only one function $f: p \to f(p)$ taking real numbers between 0 and 1 to real numbers and satisfying: (i) $f(1/2) = 1$ (binary normalization); (ii) $f(p_1p_2) = f(p_1) + f(p_2)$ (additivity), and (iii) if $p_1 \geq p_2$, then $f(p_1) \leq f(p_2)$ (decrease). That function is given by $f(p) = \log_2(1/p)$. Additivity ensures that when independently generated symbols are combined in series for successive transmission along a single channel or in parallel for simultaneous transmission along multiple channels, the "information" of the compound symbol will be the sum of the "information" of the individual symbols

different time and storage constraints.

To summarize the behavior of the message source, Shannon introduced a quantity that, at John von Neumann's suggestion, he called "entropy." A readable review of different notions going by the name of entropy is contained in Çambel (1993). The entropy of a source, denoted by H, is:

$$H = p_1 \log_2(1/p_1) + \ldots + p_n \log_2(1/p_n)\,,$$

where this sum is over all source symbols. Shannon explicitly regarded this quantity as a measure of the amount of information produced by the source. Rather, H is the average infrequency of the source symbols, without regard to their information content. It can be shown that H is (approximately) the average bit count of the source symbols in a good binary code. Low entropy sources are more compressible than high entropy ones: their symbols can be expressed in fewer bits on average.

Given a noiseless channel able to handle C bits per second (i.e., the channel can transmit up to C 0's and 1's each second), and given a source with entropy H, Shannon proved that source messages can be encoded so that the rate at which the source symbols are transmitted on average is approximately C/H symbols per second. However, no encoding has a rate that exceeds C/H. Huffman codes and other codes that compress symbols of maximum frequency and expand those of minimum frequency can be used to achieve near optimum transmission rates.

The presence of noise in a channel complicates the communication problem. Environmental noise can append itself to the signal, replace or distort part of the signal, or delete part of the signal. In such a situation redundancy in the signal is appropriate to minimize the probability that the message is lost. A trade-off occurs between the benefits of compression (high speed and low transmission cost) and the benefits of redundancy (high fidelity). Codes that detect and/or reduce error become desirable: a familiar example is the parity check code that appends a 0 or 1 to each sequence of, say, seven 0's and 1's so that the sum of the digits is even. If one digit is changed during transmission, the sum will no longer be even and the receiver who computes the sum will detect the error. Shannon, Feinstein, and others showed that a noisy channel has a characteristic transmission capacity (say a net rate of C bits per second error-free), and that when a source generates symbols at a rate of R symbols per second and has entropy H bits per symbol, then provided $HR \leq C$, source messages can be encoded to achieve a transmission rate approximately equal to HR with error rate as close to zero as we like. If the transmission rate is required to exceed C, some portion of the message will be lost; but at any rate below the channel capacity C, codings exist that ensure negligible loss.

Shannon's theory also treats continuous sources (sources having a continuum of symbols rather than just n symbols, for example, waveforms), p-ary codes, continuous codes, properties of noise, and other characterizations of entropy. A word of warning: the near optimum codes for noisy channels, whose existence is guaranteed by the theory, are not explicitly constructed. Finding codes that are appropriate to a particular channel and achieve low error rates has become the central problem of modern coding theory. As Weaver (1964) noted, the expense of finding a good code and the effort of translating the source message into that code can sometimes offset the savings the code is designed to realize. If message characteristics stay constant over a reasonable time, modern digital technology lessens this concern.

Concerned about careless usage of his ideas, the founder (Shannon, 1956) wrote an editorial entitled "The Bandwagon:"

> "In the first place, workers in other fields should realize that the basic results of the subject are aimed in a very specific direction, a direction that is not necessarily relevant to other fields .... Indeed, the hard core of information theory is, essentially, a branch of mathematics, a strictly deductive system. A through understanding of the mathematical foundation and its application is surely a prerequisite to other applications.... — but the establishing of such applications is not a trivial matter of translating words to a new domain, but rather the slow tedious process of hypothesis and experimental verification."

The ordinary usage of the word "information" should be mentioned. That usage is, roughly, that information is communicated structure[2]. For a discussion of communication (and of the meaning and value of communications), see Gray and Vogt (1994). The structure of things, of entities arising out of matter, is how the raw materials or building blocks are assembled into a whole. The structure is the manner in which the parts relate to each other and to the whole; the way that the parts are organized into a whole, achieving the harmony that makes them into a whole. Information is a presentation of some structure by a source to a destination, usually in language or graphically. Whether the original structured object itself qualifies as a presentation, or whether the source and the destination can be construed as merely factories (and not as knowing subjects) are questions we do not answer.

Some structures, both man-made and natural, can be encoded, but there is no universal language for such encoding. Thus context is important. Algorithmic information theory (Chaitin, 1987) proposes to measure information content by the bitcount of the needed code. To date the domain to which this theory applies is uncertain because it has not been established that context may be eliminated in favor of an absolute standard. In this respect, information has a nonquantitative and nonmathematical sense that transcends the theories of Shannon, Chaitin, and others.

3. WHAT IS SYMMETRY?

The ordinary notion of symmetry is readily expressed in mathematical terms. A set is a collection of objects. A symmetry is a one-to-one correspondence between the members of the set that preserves some property or properties. A rectangle can be rotated 180 degrees about its center. Each point is transformed into another point, except for the center which remains fixed. This transformation preserves the distance between points. So it is a symmetry of the rectangle with respect to distance. The collection of all symmetries of a set with respect to a property is a group: the mathematical composition of two symmetries is a symmetry, the inverse of a symmetry is a symmetry, and the identity transformation, which leaves each point fixed, is a symmetry.

Symmetry is usually associated with geometry (size and shape), but other kinds of properties also give rise to symmetries. For example, the sibling relation is symmetric: if a and b are siblings, then so are b and a. The transformation taking a to b and b to a preserves

[2] The term "instruction" is also used, emphasizing the practical, action-oriented aspect of information.

this form of kinship. Indeed, any transformation of a set of animals that permutes the siblings within each family is a symmetry for the sibling relation. In the theory of relativity, to take a more sophisticated example, the Lorentz transformations are symmetries of space-time preserving the Minkowski distance

$$(t_2 - t_1)^2 - (1/c^2)(x_2 - x_1)^2$$

between events. Einstein proposed a new kind of geometry based on space-time rather than space, and determination of the symmetries of this geometry was one of the first tasks he undertook. Another symmetry group in space-time is the orthochronous Poincare'group consisting of transformations T of space-time that preserve "causality": an event y is in the future cone of the event x if and only if the event $T(y)$ is in the future cone of the event $T(x)$. The future cone of an event x is the set of all events that might potentially be influenced by x (for example, by the arrival of a signal coming from x at a speed below that of light).

In scientific fields symmetry arguments are sometimes used to deduce the equations of a theory. A symmetry is proposed and the equations preserved by it, or compatible with it in an appropriate sense, are enumerated. These are the candidates for the proper equations. Additional symmetries may narrow down the field of candidates even more. Alternatively, an empirical equation is discovered. Then theoreticians study it to see what symmetries are associated with it. But these scientific uses of symmetry are not always straightforward. Decisions must be made about the sets on which the symmetries act and what properties are to be preserved. Scientists speak, sometimes cryptically, of invariance, covariance, and contravariance. The technicalities involved are cumbersome enough so that details are often omitted with the result that only experts know exactly what type of symmetry is meant or what exactly is being preserved.

Other examples of symmetries in science include bilateral symmetry in biology, phase symmetries in chemistry, and the well-known symmetry groups of mathematical physics. With regard to phase, a solid or crystal has symmetries that preserve shape and volume, a liquid has symmetries that preserve volume but not shape[3], and a gas has symmetries that preserve neither. The phase transition from gas to solid is an instance of symmetry-breaking: a symmetry applies at one time but not another. In mathematical physics, in addition to the Lorentz transformations mentioned above, symmetries include particle interchange, parity, charge conjugation, time reversal, the eight-fold way, and so-called supersymmetries. An industry has grown up in which a symmetry group is proposed (typically a group of linear transformations on an n-dimensional real or complex vector space preserving some metric), and then an attempt is made to relate the group, the space, and the metric, to known physical phenomena.

Symmetry is central to art. In painting, sculpture, and architecture, symmetry is used to harmonize the elements of a composition. The same is true in dance, music, costume, and spectacle. A precision marching band or a military formation derives much of its spectator appeal from symmetry. Even in science aesthetics and symmetry are linked: if an equation or physical principle is considered beautiful, then probably an underlying symmetry is the source of the appeal.

From the mathematical notion of symmetry one rather radical conclusion can be drawn. Any transformation of a set preserves some property — perhaps a familiar one or perhaps

---

[3] Strictly speaking, a solid or liquid of infinite extent should be considered But if we are studying small-scale properties, we can ignore boundary conditions

one of no particular interest, obscure or complicated[4]. So the transformation is a symmetry with respect to that property. Thus, any transformation is a symmetry! If this conclusion seems bizarre, still it has the virtue of reminding the student of symmetry that symmetry cannot be properly discussed without singling out certain properties as important and worthy of preservation.

Felix Klein's Erlangen Programm in geometry is an important source of the mathematical notion of symmetry discussed here. In Klein's view every transformation group can be regarded as a symmetry group, and the properties preserved by these symmetries constitute a geometry.

4. HOW INFORMATION AND SYMMETRY ARE RELATED

The concern of information theory is communication between a source and destination. A communication channel performs a transformation on the source message, transforming it into the received message. What is the desired transformation? Naively, it is the identity transformation! Thus, there is an immediate and fundamental connection with symmetry: a communication is an attempt to "mirror" the source message at the receiver. In the 'real world' of communication, the mirror is fractured, and it may be difficult to maintain the identity of source message and received message, but one may hope that the transformation preserves essentials.

A set of possible messages exists. The source selects one of these messages, encodes it, and transmits it, and the receiver decodes it. What is received is also a member of the set of possible messages. Ideally it is the same member. For this to make sense the message must be considered abstractly — not as a piece of paper with some scratchings on it, transformed into another piece of paper, but as a statement, pure and simple, without regard to medium, typesize, or other irrelevant characteristics of the material presentation.

Granted, this viewpoint has drawbacks. Apart from the issue of delineating exactly what messages are to be included in the set and what characteristics are relevant, there is the fact that communication channels have noise and error in them. These errors may not be the same on each transmission. In the spirit of Shannon's approach, we can assume that the pattern of errors/noise is stationary, that distributions[5] of source messages are transformed into distributions of received messages, and once again that the ideal transformation is the identity transformation.

A more sophisticated view is that certain messages are equivalent and there is no need to insist that precisely the same message is received as is sent. In this case any transformation that permutes equivalent messages is acceptable. Such a transformation is a symmetry on the set of messages, a symmetry that preserves the property of equivalence. What is desirable is a channel that transforms each equivalence class to itself, i.e., a channel that performs the identity transformation on the set of equivalence classes. The issue then is how to select a representative of each equivalence class to be the source message so as to

---

[4] Given a one-to-one transformation $T. S \rightarrow S$, we can, for example, define an equivalence relation on members of the set S by $x \sim y$ if and only if $x = T^{(n)}(y)$ for some integer $n = 0, \pm 1, \pm 2, \ldots$. Then T preserves equivalence: $x \sim y$ if and only if $T(x) \sim T(y)$.

[5] This is a subtlety. As in statistical mechanics transformations are conceived as applying not to the individual message but to the collective.

maximize the probability that the received message is in the same equivalence class.

Symmetry has another role in information theory. We add symmetry (redundancy) to a message so that it can be decoded in the presence of noise. For example, a message can be sent more than once, or check bits can be appended to it. Messages are encoded so that they have some special property known to the destination. If the received message has this property, then the channel has executed a transformation that preserves the property, i.e., a symmetry. The property is chosen so that for transformations likely to be executed by the channel (including symmetries) the original message can usually be recovered. Symmetries can also be applied by the receiver[6] or the source to verify that the message has the required property. The message is expanded by application of symmetry to prevent transmission failure. Obversely, a message can be compressed by removal of symmetry to achieve economy. For example, if a video signal has a constant background, this background need not be retransmitted. The transformation from past images to future images preserves background, and hence need not be fully implemented. However, a transmission protocol must be established to permit partial implementation. This protocol is dependent on the nature of the medium and the traffic.

Symmetry or the lack of it can also guide the choice of codes. If the source traffic has high entropy, then the symbols occur with approximately equal probability. In this case there are many good codes: arbitrary permutations of good codes are also good codes.

If, however, the source traffic has low entropy, some source symbols occur with much more frequency than others. A good code will assign these symbols fewer bits, while less frequent symbols will receive more bits. Arbitrary permutations of good codes will no longer yield good codes since short strings may get permuted into long ones. The source traffic is asymmetric and good codes will be rarer because of the symmetry-breaking.

When we consider information in the wider sense, as structure, we find that symmetry is an important structural principle. The relationship of part to part may be symmetric (e.g., bilateral symmetry), and hierarchical symmetries may be present in the relation of subpart to part versus part to whole (e.g., approximate "fractal" self-similarity). Substructures are often repetitive. A platform requires several supports: the principle of parsimony solves this repetitive problem in a repetitive way with equally spaced legs of similar design. New solutions are reserved for new problems. The harmony, integrity, and even beauty of structures are aliases for a symmetry that is both functional and economical.

A fundamental fact of biology is the replication of structures. Replication (asexual reproduction) is an attempt to duplicate an existing form, to implement the identity transformation on the set of possible forms. Just as in information theory, the environment may alter the transformation. If the identity transformation (cloning?) cannot be achieved, the objective is at least a transformation that preserves important properties (first of all, viability), i.e., a symmetry. Replication to the extent that it is the identity transformation is immortality.

5. FINAL THOUGHTS

Replication can be viewed as the attempt to maintain identity across time. Mutation is a

[6] The message a is sent as $(a, f(a))$ where $f(a)$ is some quantity that is calculated from a. If the received message is (b, c), the recipient in effect expands it out to $(b, f(b), c)$, interchanges $f(b)$ and c, and expects to find $f(b) = c$ since the interchange is a symmetry in the absence of error

breaking of the replication symmetry which leads to the equations: replication with error = symmetry-breaking = diversity as a descriptive hypothesis for evolution.

Symmetry-breaking is violation of symmetry. If a transformation does not preserve some expected or desired property, it is not a symmetry. The observer may have been mistaken in assuming that the symmetry was present, or the symmetry may be present for a time and then disappear. The message with error, the mutant life form, the anisotropic physical phenomenon all illustrate the violation of symmetry. In a sense each represents a deviation from the identity transformation: some property deemed critical to identity does not survive the transformation.

With more complicated structures more properties are deemed critical and hence there are fewer symmetries. Symmetry-breaking is equated with the "higher" forms encountered in biology (although it can also refer to a statistical anomaly such as an extremely hot spot in a plasma, cf. chapter 9 of Morowitz (1992)). Matsuno (1985) argues that protobiological information is due to symmetry-breaking in a hypothesized interaction Hamiltonian. Certain systems are degenerate (have the same energy and are not physically distinct). When the degeneracy is broken through an interaction, distinct physical states are revealed, and new structure is produced. However, symmetry-breaking is not the whole story since in biology symmetries occur that are not exhibited in inanimate matter.

Why do people find appeal in symmetry? One answer is that just as we seek knowledge of structures, we are also symmetry-seeking animals. Abstraction, our method of knowing, is finding the sameness in different objects, attending to invariants and ignoring differences. Is symmetry informative? No, because redundancy and repetition and sameness do not add to the store of information. Yes, because symmetry allows us to understand identity — to combat error in communication and replication, to solve similar problems by similar means, and to discern the common threads in the multiplicity that surrounds us.

REFERENCES

Çambel, A B. (1993) *Applied Chaos Theory. A Paradigm for Complexity*, New York· Academic Press

Chaitin, G. J. (1987) *Algorithmic Information Theory*, Cambridge: Cambridge University Press

Gray, J. E. and Vogt, A. (1994) Information and meaning, *WESScomm*, **4** (1), 2-10

Matsuno, K. (1985) How can quantum mechanics of material evolution be possible?· Symmetry and symmetry-breaking in protobiological evolution, *BioSystems*, **17**, 179-192.

Morowitz, H. J. (1992) *Beginnings of Cellular Life. Metabolism Recapitulates Biogenesis*, New Haven· Yale University Press

Shannon, C. E. (1956) The bandwagon, *IEEE Trans Information Theory*, **2** March.

Shannon, C E. and Weaver, W. (1964) *The Mathematical Theory of Communication*, Urbana· University of Illinois Press. (Originally published in 1949, this contains a reprint of Shannon's 1948 papers that appeared in the *Bell System Technical Journal.*)