Manuscript published in
International Journal of General Systems 38,2: 189-212 and a correction *38*,6: 667-668, 2009.

Ross Ashby's information theory: a bit of history, some solutions to problems, and what we face today

Klaus Krippendorff
University of Pennsylvania, Philadelphia, PA

Starting with my acquaintance with W. Ross Ashby, this paper offers an account of one aspect of his work, information theory. It shows his motivation to embrace information theory, how he extended it, where it failed while fertilizing reconstructability analysis, and what it took for a solution to emerge. Based on Ashby's recognition of the material limits of computation, the paper continues his concern for transcomputational phenomena and proposes a paradigm shift to cope with what probably is one of the most complex system we face today, cyberspace. Replacing the epistemological stance of discovery with that of design, the paper discusses how the enormous information quantities of cyberspace implicate human agency, a concept necessary to explain the ecology of electronic artifacts that populate cyberspace. It suggests a new direction for constructing theories of complex systems involving their human constituents not recognized by Ashby but derivable from his ideas.

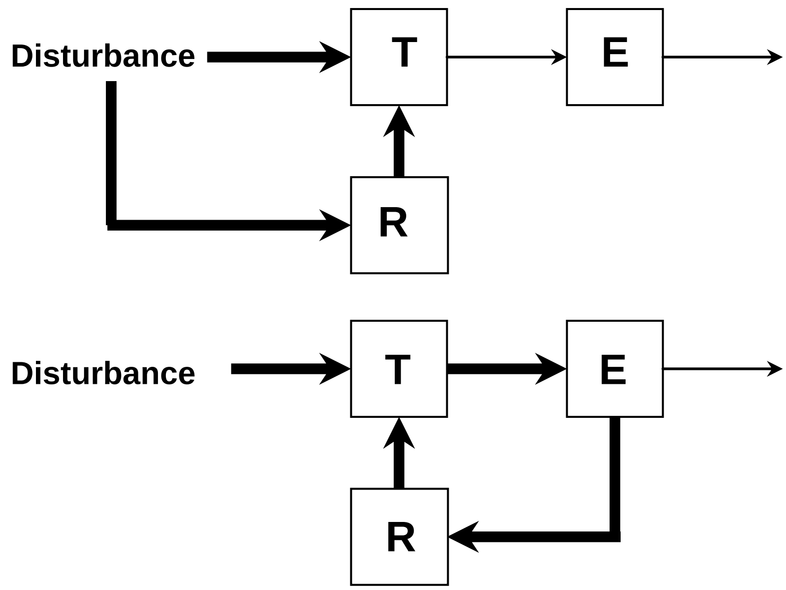
Keywords: communication theory, complexity, cybernetics, cyberspace, decomposition, electronic artifacts

1. Personal preliminaries

In 1959, I spent a summer in Oxford England to learn English. I was then a student at the legendary Hochschule für Gestaltung in Ulm, now closed which is typical of avant-garde institutions. There, we heard about cybernetics, information theory, and other exciting intellectual developments. Norbert Wiener had visited the Ulm school before my time. At the famous Oxford bookstore Blackwell I bought two books, W. Ross Ashby's (1956) *An Introduction to Cybernetics* and Ludwig Wittgenstein's (1922) *Tractatus Logicus Philosophicus*. I can't say I fully understood either of them at that time and I had no idea that both authors would have a profound effect on my academic future.

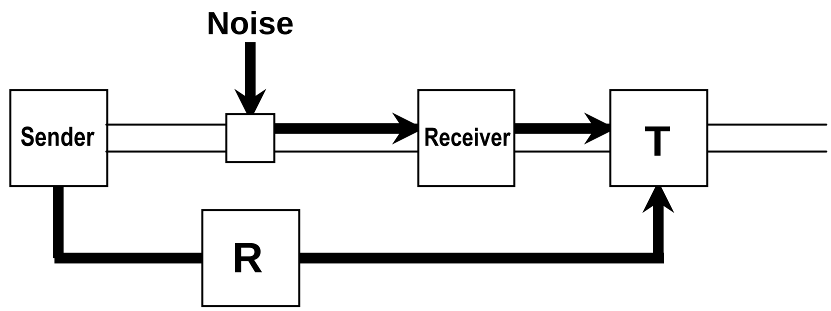
Three years later I visited the University of Illinois in Urbana in search of a university to study. I met Heinz von Foerster at his Biological Computer Laboratory. He mentioned that Ross Ashby was teaching a course on cybernetics. I had no idea that Ashby was in Urbana and the prospect of studying with him was decisive in becoming a student at U of I. I enrolled in Ashby's two-semester course in 1962-1963. He became an important member on my dissertation committee. The dissertation reconceptualized content analysis but in one chapter, I developed an information calculus for what may and what cannot be inferred by this methodology (Krippendorff, 1967).

Part of Ashby's *Introduction to Cybernetics* concerned variety and constraints in systems. Shannon's (Shannon and Weaver, 1949) entropy measures did not play an important role in this introduction except in arguing for his famous Law of Requisite Variety (1956:206-218). This law concerns the limit of successful regulation. It states that disturbances D that affect the essential variables E of a system, which are to remain within limits of a system's viability, may be counteracted by a regulator R provided the variety R has at its disposal equals or exceeds the variety in the disturbances D. In short, only variety can restrain variety. He discussed two kinds of regulation, when regulators pick up the disturbances before they affect the essential variables, anticipatory regulation, and when regulators pick up the effects of the disturbances on the essential variables, error-controlled regulation, which involved a feedback loop. Figure 1 shows these two kinds, T denoting what he called "table," a variable that responds to two effects, the solid lines representing the variety of disturbances and the variety that a perfect regulator would require.



Ashby's Regulators R and Two Versions of the Law of Requisite Variety
Figure 1

Shannon's theorem states that communication through a channel that is corrupted by noise may be restored by adding a correction channel with a capacity equal to or larger than the noise corrupting that channel. This may be seen in Figure 2, with solid lines representing the amount of noise that enters a communication channel, reduces what the receiver gets from the sender, and the required capacity of the correction channel R. This is how far the *Introduction to Cybernetics* went. Today, Shannon's 10th Theorem is considered a special case of Ashby's Law of Requisite Variety.



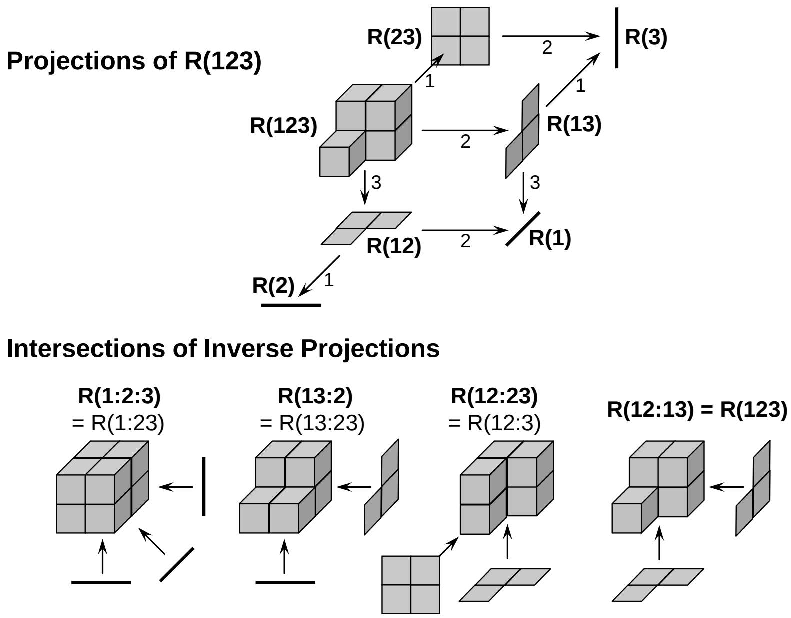
Noisy Communication Channel and Correction Channel R
Figure 2

2. Ashby's information theory

By the time I became his student, Ashby had developed many interpretations of his Law of Requisite Variety, including that the human ability to understand systems is limited by the variety available to the cybernetician's brain relative to the complexity of the system being experimented with. Although the concept of second-order cybernetics (Foerster, et al. 1974; Foerster, 1979) was not known at this time, Ashby always included himself as experimenter or designer of systems he was investigating as observer.

It is important to stress that Ashby defined a system not as something that exists in nature, which underlies Bertalanffy's (1968) *General Systems Theory* and fuelled much of the general systems movement. He did not distinguish systems from their environment and generalize what makes such systems viable. Ashby always insisted that anything can afford multiple descriptions and what we know of a system always is what he called an "observer's digest." For him a system consisted first of all of a set of variables chosen for attention and second of relationships between these variables, established by observation, experimentation, or design. He built many mechanical devices and then explored their properties. One was a black box – Heinz von Foerster later called it the "Ashby Box" – which had two switches and two lights, each either on or off. He asked his students of a preceding cohort to figure out its input-output relationships. This must have been a most frustrating assignment because every hypothesis advanced to explain the input-output relations seemed to eventually fail in subsequent trials. It turned out that – while the system was strictly determinate – the combinatorial number of possibilities that would have had to be explored far exceeded human capabilities. There was a true answer, but one that could not be found by systematic explorations. Pushing the limits of analyzing complex systems became an important part of Ashby's work.

Before fully embracing information theory, Ashby had developed the idea of decomposing complex multivariate relations into simpler constituents, using set theory. This culminated in his influential Constraint Analysis of Many-valued Relations (1964). It defined a process for systematically testing whether a seemingly complex constraint (within many variables) could be decomposed into several simpler constraints (involving co-occurrences in fewer variables) and be recomposed to the original constraint without loss. Figure 3, taken from Roger Conant's (1981a) account of constraint analysis, demonstrates in part (a) how the result of a constraint, the relation R(123), is projected onto three planes in two variables and what this implies about the whole when the missing variable is added. (b) Shows three aggregations of pairs of two-variable projections, and the aggregation of the three one-variable projections, all in three variables. Visually obvious is that R(12)×R(13) preserves the original relation, = R(123), whereas R(1)×R(2)×R(3) fails completely and the other two aggregations fail partially, in either cases ≠R(123). This graphical illustration suggests that the original relation is not as complex is it may have seemed but not as simple to allow the three variables to be regarded as independent.



Geometrical Interpretation of the Constraint Analysis of a Three-Variable Relation
Figure 3

Ashby was attracted to information theory, not only because of his Law of Requisite Variety, but also because it promised to generalize his constraint analysis to probabilistic systems and finding an elegant algebra of relations. Shannon's theory had distinguished signals from noise and patterns from random variation, and raised the hope of separating the defining properties of a system from accidental or irrelevant variations – all of which to find hidden simplicities in apparently complex systems, a theme that guided much of Ashby's work.

Shannon's entropies largely served to quantify communication between a sender and a receiver, measured at different points in time. The entropy H in the sender A, the noise received by the receiver B gave rise to the amount of information transmitted T (stated in Ashby's terms):

Entropy in sender A: $H(A) = -\sum_{a\in A} p_a \log_2 p_a$

Entropy in receiver B: $H(B) = -\sum_{b\in B} p_b \log_2 p_b$

Joint entropy in the channel AB: $H(AB) = -\sum_{a\in A}\sum_{b\in B} p_{ab} \log_2 p_{ab}$

Noise: $H_A(B) = H(AB) - H(A) = \sum_{a\in A} p_a \left[-\sum_{b\in B} \frac{p_{ab}}{p_a} \log_2 \frac{p_{ab}}{p_a} \right]$

Transmission: $T(A:B) = H(A) - H_A(B) = H(A) + H(B) - H(AB)$

McGill (1954) and Garner's (1962) Uncertainty Analysis extended Shannon's measures to three variables for analyzing psychological data. Entropies:

$$H(ABC) = -\sum_{a\in A}\sum_{b\in B}\sum_{c\in C} p_{abc}\, log_2\, p_{abc}$$
$$H_A(BC) = H(ABC) - H(A)$$
$$H_{AB}(C) = H(ABC) - H(AB)$$

Transmissions:

$$T_C(A:B) = H_C(A) + H_C(B) - H_C(AB)$$
$$T(A:B:C) = H(A) + H(B) + H(C) - H(ABC)$$

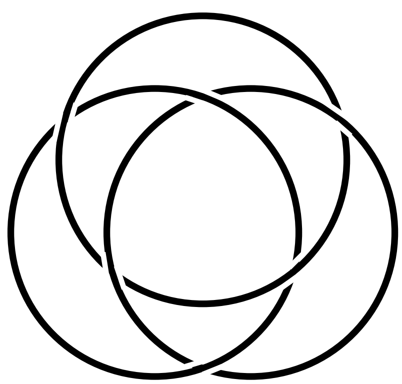
and, last but not least, the amount of interaction involving three variables:

$$Q(ABC) = T_C(A:B) - T(A:B) = T_B(A:C) - T(A:C) = T_A(B:C) - T(B:C).$$

Guided by the idea of his constraint analysis, Ashby saw the possibility of decomposing complex unanalyzed systems into its constituent relationships among variables, manifest in non-zero quantities of the information calculus. McGill provided him this accounting equation:

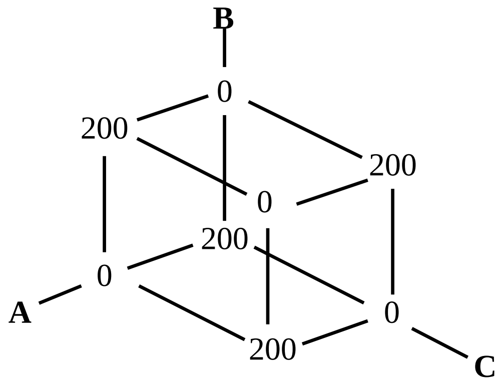
$$T(A:B:C) = T(A:B) + T(A:C) + T(B:C) + Q(ABC),$$

showing the total amount of transmission within three variables as the algebraic sum of the three transmissions between pairs of variables plus the amount of interaction unique to all three. Ashby explained the *Q*-measure as the amount due to the unique combination of a number of variables, not reducible to any of its subset. To illustrate, he had made and wore a necklace consisting of three interlinked chains, schematically shown in Figure 4, which had the property of falling into separate chains once any one of them was cut.



Ashby's Chain Necklace
Figure 4

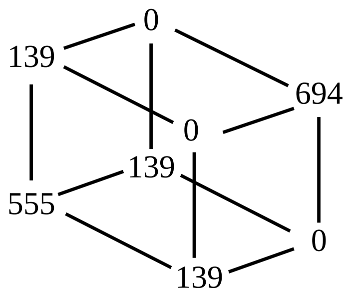
Figure 5 shows a three-dimensional table of frequencies whose distribution is typical of a non-decomposable interaction between all three variables, which can be seen in the corresponding breakdown of the overall transmission measures. The zero values of the binary transmission measures may be seen as justified as the projections of the three-dimensional distribution onto its three faces are uniform and exhibit no structure, and the three-dimensional distribution could not possibly have been predicted from them.



$$\begin{aligned} Q(ABC) &= 1.00 \\ T(B:C) &= .00 \\ T(A:C) &= .00 \\ \underline{T(A:B)} &= \underline{.00} \\ T(A:B:C) &= 1.00 \end{aligned}$$

Frequency Distribution with Q(ABC) fully accounting for T(A:B:C)
Figure 5

The absence of interactions, we uncritically accepted, was indicated by zero values for the *Q*-measures, as exemplified in Figure 6.



$$\begin{aligned} Q(ABC) &= .00 \\ T(B:C) &= .35 \\ T(A:C) &= .35 \\ \underline{T(A:B)} &= \underline{.35} \\ T(A:B:C) &= 1.05 \end{aligned}$$

Frequency Distribution with zero *Q(ABC)*
Figure 6

As students we computed many of these accounts by hand, using an $nlog_2n$ table, which was tedious without electronic computers, and we followed Ashby lead to generalize information theory to any number of variables, which was easier. This effort culminated in the publication of two lists of some 50 accounting equations (Ashby, 1969), the beginning of an information algebra. The *Q*-measures were Ashby's prime candidates. By extending McGill and Garner's *Q*-terms to fewer and more than three variables

$$Q(A) = -H(A)$$
$$Q_B(A) = -H_B(A)$$
$$Q_{BC}(A) = -H_{BC}(A)$$
$$Q(AB) = Q_B(A) - Q(A) = Q_A(B) - Q(B) = T(A:B)$$
$$Q_C(AB) = Q_{BC}(A) - Q_C(A) = Q_{AC}(B) - Q_C(B) = T_C(A:B)$$
$$Q(ABC) = Q_C(AB) - Q(AB) = Q_B(AC) - Q(AC) = Q_A(BC) - Q(BC)$$
$$Q(ABCD) = Q_D(ABC) - Q(ABC) = \text{three other expressions by permutation of these variables}$$
$$Q(ABCDE) = Q_E(ABCD) - Q(ABCD) = \text{four other expressions by permutation of these variables}$$

etc.

a general accounting equation emerged (Ashby, 1969:6) by which we assumed able to decompose the total amount of transmission in a system of any number of variables into its unique interaction quantities *Q*:

$$T(A:B) = Q(AB) \quad \text{by definition}$$

$$T(A:B:C) = Q(AB) + Q(AC) + Q(BC) + Q(ABC)$$

$$T(A:B:C:D) = \\ = Q(AB) + Q(AC) + Q(AD) + Q(BC) + Q(BD) + Q(CD) + Q(ABC) + Q(ABD) + Q(ACD) + Q(BCD) + Q(ABCD)$$

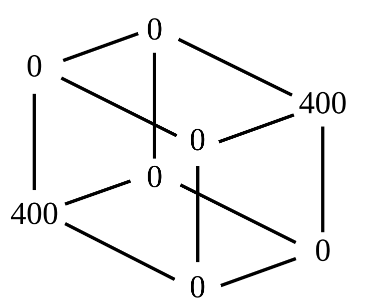
etc.

Stated generally:

$$T(S) = \sum_{\alpha \subset S} Q(\alpha)$$

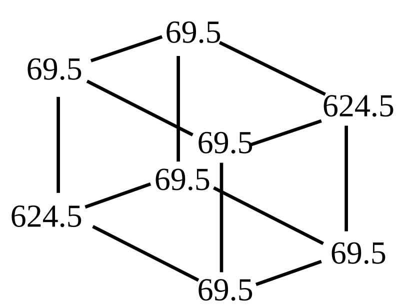
where S is the set of variables of a chosen system and α is a subset of S of two or more variables.

Accounting for the complexity of a system in terms of additive quantities was appealing to many researchers (Ashby, 1965, 1970; Broekstra, 1976, 1977, 1979, 1981; Conant, 1973, 1976, 1980). I too developed equations and algorithms for simplifying complex systems in these terms (Krippendorff, 1974), and aimed at a spectral analysis of multi-valued relations (Krippendorff, 1976; 1978). Nevertheless, despite the compelling logic and obvious simplicity of these accounting equations, suggesting that *Q*-measures would quantify higher-order constraints, for example, present in the data of Figure 5 and absent in those of Figure 6, there remained something odd: *Q* could be negative, as may be seen in Figures 7 and 8.



$$\begin{aligned} Q(ABC) &= -1.00 \\ T(B:C) &= 1.00 \\ T(A:C) &= 1.00 \\ \underline{T(A:B)} &\underline{= 1.00} \\ T(A:B:C) &= 2.00 \end{aligned}$$

Frequency Distribution with negative *Q(ABC)*
Figure 7



$$\begin{aligned} Q(ABC) &= -.25 \\ T(B:C) &= .35 \\ T(A:C) &= .35 \\ \underline{T(A:B)} &\underline{= .35} \\ T(A:B:C) &= .80 \end{aligned}$$

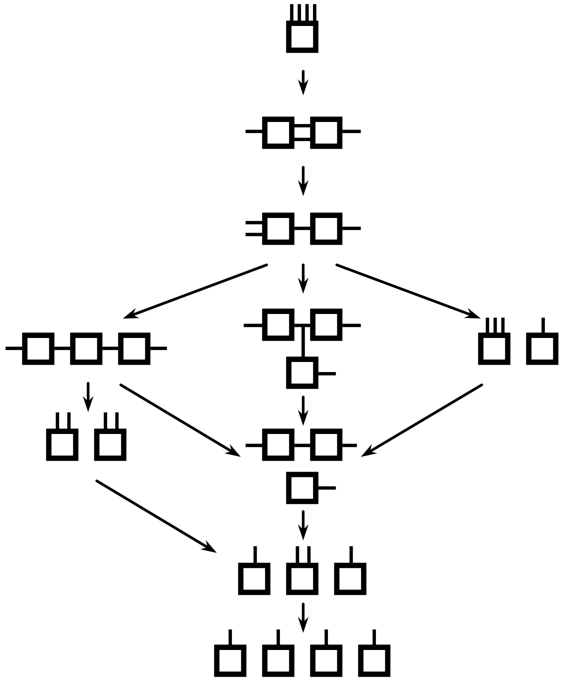
Frequency Distribution with negative *Q(ABC)*
Figure 8

McGill (1954) had acknowledged this possibility and interpreted any deviation from zero a signal that interaction in the data existed. Ashby deferred to his interpretation, and we all continued developing this calculus. The promises of an algebraic account of complexity were too appealing to be wrong.

However, observe in Figure 7 that any two of the three projections onto the faces of the cube are sufficient to reconstruct or uniquely determine its three-dimensional distribution. The third projection is redundant, implied and not needed to obtain that distribution. If $T(A:B:C) = T(A:B)+T(A:C)$, a third-order interaction should be absent by this conception, yet $Q(ABC)$ has a value unlike zero. In fact, it seemingly corrects for redundant measures, here of $T(B:C)$. This suggested to me that the Q-measures did not only respond to higher-order interactions but also compensated for the over-determination by redundant lower-order interactions. If true, this finding would cast serious doubt on the ability of Q-measures to indicate the presence or absence of higher-order interactions. For example, the projections of the distributions in Figures 6 and 8 onto its faces are the same, as evident in $T(A:B)=T(A:C)=T(B:C)=0.35$. But the distribution in Figure 6 is most unlike chance or maximally entropic, satisfying the three two-dimensional distributions and therefore suggests the presence of an interaction, stronger than in Figure 8, but not measured by Q. We all seemed to follow a convenient logic with questionable interpretations.

3. A gestalt switch

Meanwhile, George Klir (1976) had picked up on Ashby's constraint analysis (Ashby, 1964). At the 1978 conference of the Society of General System's Research Klir (1978) presented a paper of his explorations. Two seemingly unimportant things struck me. First, whereas Ashby diagrammed systems in terms of his set theoretically motivated "diagram of immediate effects" (Ashby,1967) between variables, the variables being represented by boxes and the effects by lines, as in Figures 1 and 2, Klir had inversed that convention, putting the effects among variables into boxes and showing variables as lines connecting them. This simple gestalt switch allowed me to visualize interactions inside Klir's boxes, not in terms of lines connecting variables. Second, whereas Ashby dealt with interactions algebraically, as an unordered list of quantities that summed to a total, Klir presented an algorithm for generating a lattice of simplifications of the models of a system from the most to the least complex one, covering the same set of variables in each case, shown in Figure 9 involving four.



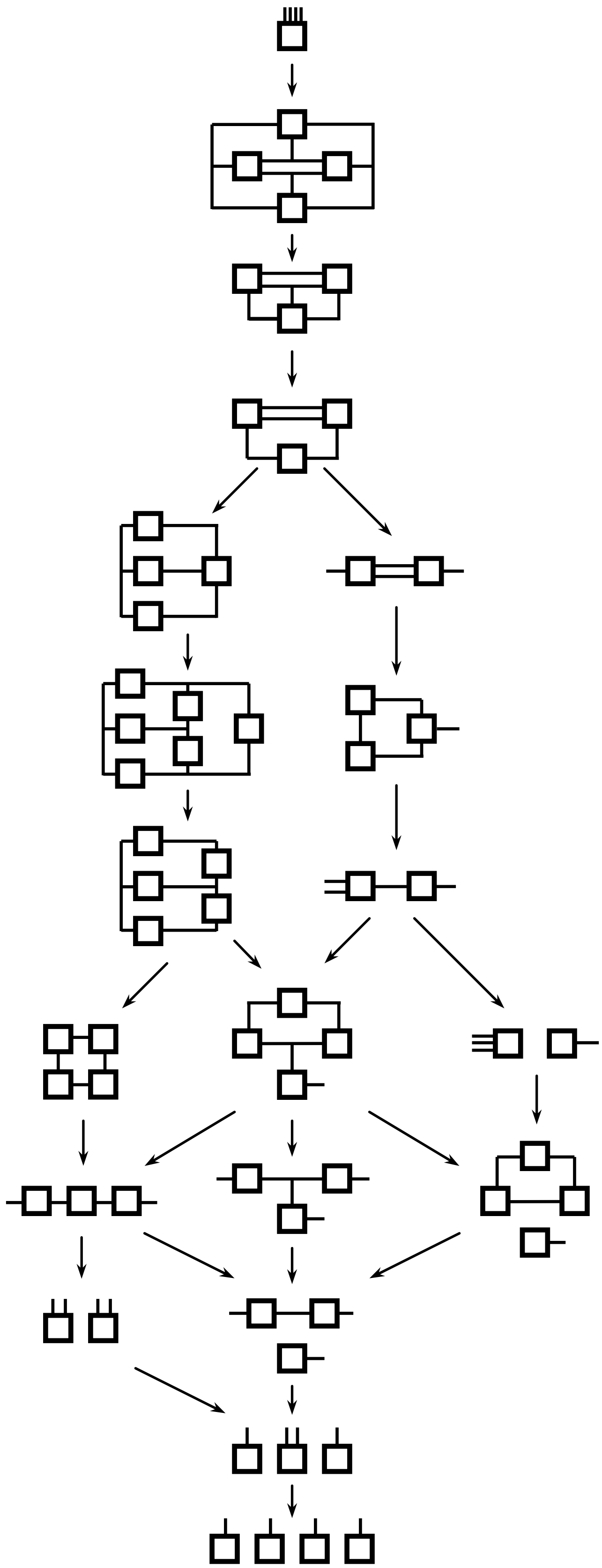
Lattice of Simplifications of Models of Systems in Four Variables Without Loops
Figure 9

In effect, each of Klir's models consisted of several components which (a) were shown as linked through the variables they shared, (b) contained all subordinate interactions, and (c) could be "degraded" into simpler ones by removing the component defining interactions, one by one. Although Klir was not concerned with information theory, his lattice visualized the relationships between the components of a system and implied an ordering of the interactions to be removed. This suggested to me that each simplification could be linked to a specific information quantity. Indeed, with variables named A, B, C, and D, the leftmost path of six steps up the lattice in Figure 9 amounts to this accounting equation:

$$T(A:B:C:D)=T(C:D)+T(A:D)+T(B:C)+T_C(B:D)+T_D(A:C)+T_{CD}(A:B)$$

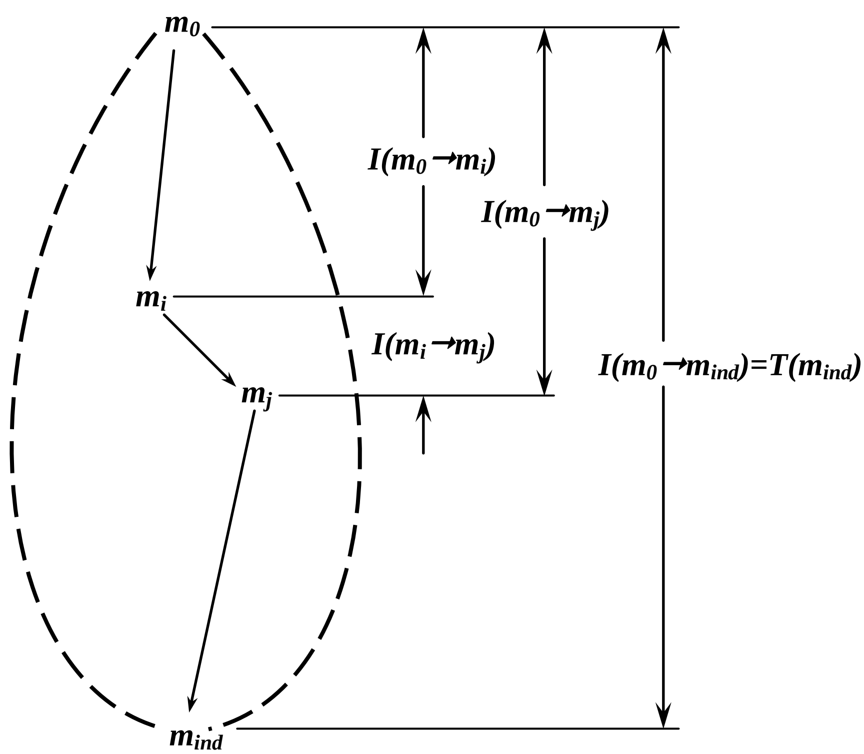
Another path through this lattice would have produced the same six terms save for their order and permutations of the variable names.

But as a cybernetician, I could not help noticing the conspicuous absence of circular relations among Klir's models. An examination of these models revealed that whenever an interaction among three or more variables was absent or analytically removed – Ashby's idea – all lower order interactions formed models with loops. The accounting equations in terms of *Q*-measures hid these facts. Figure 10 shows the lattice of all possible models involving four variables, half of which happen to be models with loops.



Lattice of All Possible of Models of Systems in Four Variables
Figure 10

With such lattices, it became easy to reconceptualize the information quantities of interest, not in terms of *Q*-measures, but in terms of the differences between the quantities of information that can be handled by any two models, one being a descendent of the other. Figure 11 shows a schematic lattice and the measures of interest where m_0 is the unanalyzed whole system, m_{ind} is the model of that system with all of its variables regarded as independent, m_i is a simplification of m_0 and m_j is simplification of m_i regardless of the number of steps involved.



Generalized Lattice of all Possible Models of a System m_0 and
Information Measures of their Differences (with Interactions Successively Removed)
Figure 11

Figure 11 also shows how the total amount of information transmitted within a system can be algebraically decomposed into quantities along a path of simplifications of models of m_0, within a lattice of possible models:

$$T(m_{ind}) = I(m_0 \to m_{ind}) = \sum_{i=0}^{i=ind-1} I(m_i \to m_{i+1})$$

This gestalt switch was conceptual and enormously convenient notationally. But the information quantities could be applied only to Klir's models without loops. The biggest nut to crack was how to cope with systems that did contain loops.

4. Information in circular (causal) systems

Shannon called his theory *A mathematical theory of communication* and attended to processes that proceeded in one direction only. A message received had no effect on the message sent. Noise that entered a channel could only degrade what was sent. A prior choice limited subsequent choices but could not have an effect on itself. This linearity may not have been intended as Shannon constantly struggled with notions of feedback, how a corrupted message could be restored, which implied an observer who could refer back to the original. But it was grounded in a far more basic adoption: *probability theory*. That theory axiomatically requires that the probabilities in any one set sum to 1, and it combines probabilities from independent sets by multiplication.

Shannon's 2nd theorem (Shannon & Weaver, 1949:19) relied on the logarithm function to convert products into sums, and established that the entropy function was the only function that afforded the intuition of information being an additive quantity. As it turns out, the algebraically obtained *Q*-measures violate the axiom of probability theory. This may be seen when expressing *Q* in probabilistic terms:

$$Q(A) = \sum_{a \in A} p_a \log_2 p_a = -H(A)$$

$$Q(AB) = \sum_{a \in A} \sum_{b \in B} p_{ab} \log_2 \frac{p_{ab}}{p_a p_b} = T(A:B)$$

$$Q(ABC) = \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} p_{abc} \log_2 \frac{p_{abc}}{\dfrac{p_{ab} p_{ac} p_{bc}}{p_a p_b p_c}}$$

$$Q(ABCD) = \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} \sum_{d \in D} p_{abcd} \log_2 \frac{p_{abcd}}{\dfrac{p_{abc} p_{abd} p_{acd} p_{bcd}}{\dfrac{p_{ab} p_{ac} p_{ad} p_{bc} p_{bd} p_{cd}}{p_a p_b p_c p_d}}}$$

etc.

All numerators of these expressions and the denominator in *Q(AB)=T(A:B)* are probabilities proper: $\sum_{a \in A} \sum_{b \in B} p_{ab} = 1$ and $\sum_{a \in A} \sum_{b \in B} p_a p_b = 1$. But the denominators, starting with *Q(ABC)*, no longer qualify as such: $\sum_{a \in A} \sum_{b \in B} \sum_{c \in C} \frac{p_{ab} p_{ac} p_{bc}}{p_a p_b p_c} \neq 1$. Thus, for three or more variables, *Q*'s denominators are incompatible with an axiom of probability theory which underlies the 2nd theorem of information theory.

This incompatibility stems from the fact that separating the interactions that Ashby conceptualized as being unique to a system creates circularities that defy obtaining their probability distributions by multiplication. None of us who applied information theoretical measures to systems at this time realized this contradiction, which is all the more surprising as circularity is fundamental to much of cybernetics.

However, the idea of additive quantities that measure the unique contributions of higher-order interactions in systems (leaving circularities behind) can be retained by calculating the maximum entropy probability distribution, subject to the constraints of the probabilities of its components – not by multiplication – but by following the circularity iteratively, going around and around the circle, through each component, in either direction, until that joint probability is found. Solomon Kullback (personal communication) directed my attention to an algorithm developed by Darroch and Ratcliff (1972), which I could adapt for this purpose (Krippendorff, 1980a, 1980b). Omitting here a generalization of this algorithm to fixed and zero probabilities, which are considered elsewhere (Krippendorff, 1986), this algorithm is defined as follows:

- Let $p_{abc\ldots}$ be the joint probabilities of variables $A, B, C, \ldots$ of a system m_0 chosen for analysis
- Given a model m_i of m_0 consisting of r components K_1: K_2: …:K_e: … :K_r (Klir's boxes), each defined by a subset of the system's variables, jointly covering all.
- Let p_{k_e} be the probabilities within the e^{th} component K_e of m_i obtained by summing over all values $\bar{k}_e \in \bar{K}_e$ of variables not in K_e: $p_{k_e} = \sum_{\bar{k}_e \in \bar{K}_e} p_{abc\ldots}$
- Set all cells $abc\ldots \in ABC\ldots$ to $\omega^{(0)}_{abc\ldots} = \frac{1}{N_{ABC\ldots}}$, where $N_{ABC\ldots}$ is the number of cells in $ABC\ldots$

Iterate $t = 0, 1, 2, \ldots$ until $\omega^{(rt+e)}_{abc\ldots} = \omega^{(rt+e-1)}_{abc\ldots}$ for all components K_e.

For all components: K_e, $e = 1, 2, 3, \ldots, r$

For all cells $abc\ldots \in ABC\ldots$, compute: $\omega^{(rt+e)}_{abc\ldots} = p_{k_e} \frac{\omega^{(rt+e-1)}_{abc\ldots}}{\sum_{\bar{k}_e} \omega^{(rt+e-1)}_{abc\ldots}}$

It yields the maximum entropy distribution of probabilities $\omega_{abc\ldots}$ (expected by chance) in the variables of a system m_0, satisfying the constraints of components K_1: K_2: …:K_e: … :K_r of the model m_i of m_0. In terms of these maximum entropy probabilities the amount of information in the original system m_0 excluded from m_i becomes:

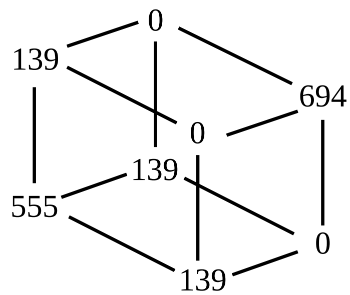
$$I(m_0 \to m_i) = \sum_{abc\ldots \in ABC\ldots} p_{abc\ldots} \log_2 \frac{p_{abc\ldots}}{\omega_{abc\ldots}}$$

The difference between any two models m_i and m_j of the system m_0, m_j being a descendant of m_i, then becomes:

$$I(m_i \to m_j) = I(m_0 \to m_j) - I(m_0 \to m_i) = H(m_j) - H(m_i) =$$
$$= \sum_{abc\ldots \in ABC\ldots} \omega_{abc\ldots(m_i)} \log_2 \omega_{abc\ldots(m_i)} - \sum_{abc\ldots \in ABC\ldots} \omega_{abc\ldots(m_j)} \log_2 \omega_{abc\ldots(m_j)}$$

which associates quantities of information with the expressions in Figure 11.

Unlike what Q was thought to measure, Figure 8 exemplifies a system without ternary interactions, not Figure 5. With the new quantities in place, the correct account of the data in Figure 5 are shown in Figure 12. Here, it may be observed that the information in the two bivariate components *AB* and *AC* add to the information in *AB:AC*, but the third *BC* (in this case any third binary component) adds less to a model consisting of all three bivariate components *AB:AC:BC.* The unique interaction (deviating from the distribution of frequencies in Figure 8) has a positive value.



$$\begin{array}{lll}
I(ABC \rightarrow AB{:}AC{:}BC) & = 0.25 \text{ bits} & \neq\ Q(ABC) = 0.00 \\
I(AB{:}AC{:}BC \rightarrow AB{:}AC) & = 0.10 \text{ bits} & \neq\ T(B{:}C) = 0.35 \\
I(AB{:}AC \rightarrow AB{:}C) & = 0.35 \text{ bits} & =\ T(A{:}C) \\
\underline{I(AB{:}C \rightarrow A{:}B{:}C)} & \underline{= 0.35 \text{ bits}} & =\ T(A{:}B) \\
I(ABC \rightarrow A{:}B{:}C) & = 1.05 \text{ bits} & =\ T(A{:}B{:}C)
\end{array}$$

Correct Accounts of the Interaction Information in Data in Figure 6

Figure 12

I presented these developments at the 1980 conference on cybernetics in Vienna (Krippendorff, 1980a), once and for all disposing of Q as a viable measure in information theory, showing that we all (McGill, 1954; Garner, 1962; Ashby, 1969; Broekstra, 1976, 1977, 1979, 1981; Krippendorff, 1974, 1976, 1978, 1979a, 1979b, and others, but not Roger Conant) were wrong in assuming we could account for interactions in systems with loops *algebraically*, when we should have followed the circularity in these system *iteratively*. Thus, when m_i is decomposed into m_{i+1} by removing just one interaction, the unique contribution of that interaction, which Ashby had conceptualized, is measurable not by Q but by $I(m_i \rightarrow m_{i+1})$, the entropy in m_{i+1} less in its predecessor m_i.

While Q-quantities turned out not to measure interaction information in complex systems as assumed by the above mentioned information theorists, the resolution of this negative finding came with the discovery that $Q(ABC)$ was the difference between the correct amount of interaction information $I(ABC \rightarrow AB{:}AC{:}BC)$ and a measure of over-determination or redundancy $R(AB{:}AC{:}BC)$:

$$Q(ABC) = I(ABC \rightarrow AB{:}AC{:}BC) - R(AB{:}AC{:}BC)$$

It shows Q not to be a stand-alone measure of either entropy or information but of the extent to which I exceeds R, explaining Q's odd behavior, and solving the problem of interpreting Q's negative values. This relationship gives rise to a more general measure of redundancy:

$$R(m_1) = I(m_0 \rightarrow m_1) - Q(m_0)$$

For data in Figures 6 and 12:

$$R(AB{:}AC{:}BC) = I(ABC \rightarrow AB{:}AC{:}BC) - Q(ABC) = 0.25 - 0 = 0.25 \text{ bits.}$$

One may recognize this amount of redundancy in the difference between $T(B{:}C) = 0.35$ bits and the information lost when BC is removed from the model of the data, $I(AB{:}AC{:}BC \rightarrow AB{:}AC)=$ 0.10 bits. T quantifies the interaction in BC without reference to its context, I quantifies the same interaction but in the context of the interactions in AB and AC that together form a loop and partly imply BC.

The three-dimensional frequency distribution in Figure 7 can be reconstructed from any two faces of the data cube, a ternary interaction being absent and one binary interaction being redundant and ignorable without loss. This is reflected in the amount of redundancy, $R(AB{:}AC{:}BC) = 0 - (-1)$ $= 1$ bit, which equals the amount of information in any one redundant binary interaction, $T(A{:}B)$, $T(A{:}C)$, or $T(B{:}C) = 1$ bit each.

For data in Figure 5, redundancy measures $R(AB{:}AC{:}BC) = 1 - 1 = 0$ bits. Indeed, the ternary interaction in this data cube is unique. The three faces of the data cube, AB, AC, and BC tell the analyst nothing about the frequency distribution in ABC. It is noteworthy that the absence of redundancy is the only condition under which $Q(m_0)$ correctly measures the information of an interaction, $I(m_0 \rightarrow m_1)$.

Whenever circularities exist in multi-variate data, Q-measures are confounded by the redundancy of their algebraic calculations. It should be noted that $R(m_1)$ can be negative, a condition that pertains when algebraic accounts of the interactions in a system underdetermine these interactions. I am grateful to Loet Leydesdorff (personal communication, 2009.4.17) for providing an example of this condition. I am sure Ashby would have been pleased to see this development, especially since it proved us all mistaken.

The above algorithm added a new chapter to Shannon's theory: the possibility of measuring the information flows in systems with loops, which had heretofore defied adequate accounts and it added a meaningful measure of the complexity of systems. Martin Zwick has put my original program calculating $I(m_i \rightarrow m_j)$ for up to ten variables and ten states each and several recent developments on his website http://www.pdx.edu/sysc/research-discrete-multivariate-modeling (last accessed 2009.4.24). It runs on UNIX and MacOS and probably on other systems as well.

5. Material and informational numbers

Shannon's information theory foremost is a theory of limits. It states limits on how much information can be transmitted through a noisy channel of communication, on the enciphering of encoded messages without knowledge of the key, and in Ashby's terms on the ability to regulate a system faced with disturbances. However, it assumes and takes for granted the existence of differences. Gregory Bateson (1972:381) defined information as "any difference which makes a difference in some later event." But differences do not occur in nature. They result from acts of drawing distinctions, need to be recognizable as such, and only then can they enter explanations of their effects. Thus, substituting "difference" by "recognizable change" leads one to something that responds to that change or to someone who can observes it. Ashby was interested in whether there was a limit to that recognition, a limit that cannot be overcome, even with all conceivable technological advances.

It was fortuitous for Ashby to met Hans-Joachim Bremermann at the second conference on self-organizing systems. Bremermann (1962) recognized that information transmission or information processing systems need to respond to differences, which cannot be arbitrarily small, thus entailing a limit, not part of information theory. In terms of Einstein's mass-energy equivalence and Heisenberg's Uncertainty Principle, he argued that the transmission or processing capacity of any circumscribable system, artificial or living, cannot exceed

$$mc^2/n \quad \textit{bits per second},$$

where m = the mass of the system (including its power source)
c = the velocity of light
n = Plank's constant

By inserting the two constants, Bremermann concluded that

No material system can exceed a processing capacity of
approximately 2×10^{47} bits per second per gram of its mass.

To get a sense of this limit, Ashby presented us with several humbling numbers:

Times:	A distinguishable atomic event takes	$\cong 10^{-10}$ sec
	One year	$\cong \pi\times10^{7}$ sec
	Time since the earth solidified	$\cong 10^{20}$ sec
Mass:	Mass of the Earth	$\cong 6\cdot10^{27}$ grams
Counts:	Number of atomic events since the earth solidified	$\cong 10^{30}$
	Number of atoms in the visible universe	$\cong 10^{73}$
	A computer the size of the entire Earth, operating at Bremermann's limit could perform no more than	$2\cdot10^{47}\cdot6\cdot10^{27} \cong 10^{75}$ bits/sec
	or	$10^{75}\cdot\pi\cdot10^{7} \cong \pi\cdot10^{82}$ bits/year
	Since the earth solidified, that ideal computer could have computed no more than	$10^{20}\cdot10^{75} \cong 10^{95}$ bits

From which he concluded (Ashby, 1968):

Everything material stops at 10^{100}.

This is a pretty solid limit. But cyberneticians, he argued, are concerned mainly with another kind of number. True to his conception of cybernetics as *the study of all possible systems that is informed (constrained) by what cannot be build or found in nature*, he was led to enumerate possibilities rather than actual observations and the numbers that emerged may be called combinatorial, informational, or transcomputational. For example,

Possibilities:	Number of configurations displayable by an array of 20×20=400 light bulbs, which are either on or off	$= 2^{400} \cong 10^{120} > 10^{100}$
	Number of non-trivial machines (Foerster, 1984) with only 3 binary inputs and 4 internal states	$2^{13,297} \cong 3\cdot10^{4002} >> 10^{100}$
	Number of images of a HDTV screen with 1920×1080 pixels and 32 bits for color	$2^{2,073,632} \cong 10^{624,000} >>> 10^{100}$
	Number of distinctions between good and bad images on that screen	$\cong 2^{1 \text{ followed by } 624,000 \text{ zeros}} >>>> 10^{100}$

The enormity of these numbers and the fact that they often appear as exponents of 2 is one reason for expressing them in $\log_2$ or "bits" rather than in actual counts. Ashby (1968) concluded that

Cyberneticians have to cope with numbers >> 10^{100}
with material resources for computation << 10^{100}.

——— o ———

In 1972, I attended a conference on cybernetics in Oxford, England where we learned from Gray Walter that Ashby was mortally ill with a brain tumor. Another former Ashby student from Switzerland by the name of Burckhardt (regrettably I am not recalling his full name) and I took a train to visit him. His wife told us to be brief and not to mention the seriousness of his situation. I gave him a copy of my conference paper (Krippendorff, 1974) drawing on his work. He was pleased and promised to read it when he felt better. We saw the working space he had set up after retiring in 1970 from Urbana to a beautiful old school house with a lovely garden. Asked what he planned to do, he told us of writing a book that would start with Bremermann's limit. Subsequent inquiries did not turn up notes of how he would have proceeded. But I kept his idea in mind.

6. A paradigm shift

Meanwhile, computational technology made enormous leaps. Cybernetics became more self-reflective to the point of suggesting its evolution from first-order to second-order cybernetics. I pursued interests far removed from Bremermann's limit, the design of human interfaces with technology (Krippendorff, 2006). Such interfaces cannot be understood without the participation of *human agency,* the ability to draw distinctions, decide among the alternatives thus distinguished, and act creatively, i.e., without rational prescriptions or pre-established determinisms. Bremermann's finding implicates human agency by stating not what exists but by distinguishing what *CAN* or *CANNOT* be done within the laws of physics.

Given that we can cope with numbers beyond available computational resources, Ashby's conclusion can signal two things. Either numbers $> 10^{100}$ are meaningless or our dominant epistemology has not kept up with the technology we are facing today. I favor the latter and have distinguished elsewhere four epistemologies regarding understanding systems (Krippendorff. 2008).

- Systems whose behavior is deducible from a finite history of recorded observations are *observationally determinable.* This reflects the epistemological stance of detached observers who seek to discover systems properties by testing all possible hypotheses about that systems structure against the data it produces
- Systems that can be built and set in motion are *synthetically determinable*. This reflects the epistemological stance of designers who have access to the structure of a system having determined its makeup
- Systems that can be lived with or utilized by interacting or communicating with them competently are *hermeneutically determinable,* for example computers as well as people
- Systems that can be understood by participating in them are *constitutively determinable*. The latter especially applies to social systems, constitutively involving knowledgeable human participants. They also include what second-order cybeneticians do.

Heinz von Foerster (1984) showed that observational determinability is limited to trivial machines – systems with few states and simple structures. Non-trivial machines, involving internal memories, including the above mentioned "Ashby box", defy observational determinability but can be understood by building them (or dissembling them into trivial parts and reassemble them). Computers, for example, are non-trivial by this definition. They have internal loops and memories, can be built, but their operating structure is incomprehensible by merely observing what they do. Regarding computers, most competent computer users have no clue and do not need to care about how their machines actually function, their synthetic determinability, yet have no problems learning how to use them. Indeed, computers are designed for hermeneutic determinability. It is when users install software and reconfigure their interfaces that they approach being designers – at least of the contours of what is going on inside them.

I suggest that understanding information processes of the kind we are facing today can no longer be accomplished by discovering and identifying interactions in observed systems. In fact, reconstructability analysis quickly runs into transcomputational numbers. In a little known paper Conant (1981b) found a way to bypass Bremermann's limit by not selecting a solution from all possible alternatives but *constructing* a solution based on a simpler representation of the problem. In effect, he moved beyond the limit of observational determinability by designing a solution. Technology is not discovered, it is designed. To understand complex technological systems requires an understanding of how they are designed, how their realities are constructed within the possibilities created by their designers. Bremermann's limit defines the space within which human agency is physically possible.

7. Cyberspace

7.1 Space

Considering the above, to me space is not a metaphor or a mathematical artifact. Space is created and recognized by human actors in the process of realizing (making real) their artifacts. It is a construct that is necessary to understand humans as having abilities, manifest in linguistic constructions with the auxiliary verb "can" and becomes evident in material artifacts that cannot emerge from unattended nature or be explained causally or entropically. Space is constituted in the possibilities that human actors perceive in their world. Here are six obvious propositions concerning that concept of space. (i) Actions consume possibilities. For example, writing a document occupies a certain amount of space – on paper or in computer memory – thereafter not available for expressing other things. (ii) Choosing among possible actions has consequences, often social ones, i.e., pertaining to other actors. For example, dialing a telephone number establishes a connection with someone at the expense of connecting with someone else, or building a house not only changes a landscape, but where neighbors might build theirs. (iii) In using technologies one almost always trades constraints on less important possibilities for desirable possibilities that would not be available otherwise. For example, using the telephone limits communication to voice within a narrow bandwidth, but extends the ability to converse with people at distances far greater than could be reached acoustically. (iv) The human use of technology is limited to the possibilities it provides in the human interfaces with them. For example, use of the internet is limited to what computer interfaces enable their users to do. (v) Computers may amplify human intelligence (Ashby, 1956b) when the choices made by their users initiate processes that select among a far greater number of possibilities, for example, searching the world wide web within seconds for something that would take humans a lifetime to find. (vi) The great number of possibilities that information processing systems provide to their users makes theorizing such systems (predicting its states and structure) difficult if not impossible. The openness that internet users experience is their primary motivation to engage that net. There can be no single elegant theory of cyberspace without human agency.

7.2 History of Cyberspace.

To me, cyberspace consists of technologically supported possibilities for human actions. So conceived, cyberspace originated when early humans found sticks, stones, and fire to be separable from where they could be found, movable to and usable where they accomplish something previously thought impossible. Thus, sticks, stones, and fire may have been the first human artifacts. The path from that early beginning to where we are now took several millennia of technological development.

What has changed during this remarkable history, in my view, is due less to an increase of information, as current writers on information society insist, than to an increase in our collective ability to draw more and finer distinctions, to handle, assemble, use, and communicate what we distinguished more efficiently than before, and to construct worlds that enhance our collective ability to realize ourselves. The great Cheops pyramid, built 5000 years ago during a 20 year period, amounted to moving 2.3·Billion stones into a very simple arrangement. The mass production of same-size bricks enabled the building of a great many and rather different kinds of buildings. Writing, using combinations of letters from a small alphabet of characters added choices not available to painting naturalistically. The largest library of ancient times, the Royal Library of Alexandria, destroyed by fire about 2000 years ago, is estimated to have held between 40,000 and 700,000 books and scrolls among which users had about 10^6 binary choices. For comparison, the collection of printed matter of the U.S. Library of Congress is estimated to contain 10-terra bytes (Lyman & Varian, 2003), including the characters its publications contain, about $\cdot 10^{14}$ bits or 10,000 times the size of the library of Alexandria. The searchable World Wide Web contains about 136 times the number of bits in the Library of Congress. Already the library in Alexandria featured principles of mechanics and hydraulics that could be combined to generate numerous technological inventions. The 2000 years between the library of Alexandria and the World Wide Web witnessed numerous milestones. Gutenberg's invention of movable type, mass production of freely combinable technological artifacts, the printing press, Hollerith punch cards, radio tube computers, and digital communication. All afforded us options previously unavailable or time consuming. To me, the history of human technology is one of increasing the number of possibilities we can use to our advantage. Cyberspace began well before electronic possibilities emerged although the latter certainly have dwarfed all previous technologies in how much they offer.

7.3 The Current Size of Cyberspace.

Existing communication and computer technology operates far from Bremermann's limit. But one may appreciate the size of the space it collectively offers by estimating the unconstrained possibilities it currently provides.

- A byte is an atomic unit of data in a computer, increasingly used by computer manufacturers to quantify information processing and storage capacities. It consists of an 8-bit sting of 0s and 1s or eight binary variables and can keep 256 different characters. However, since I am interested in the choices human actors can collectively make rather than how data are stored inside a computer, I prefer to express possibilities in terms of the number of binary choices they enable. Accordingly one byte = 8 bits.
- A contemporary 200 gigabyte computer can store $200 \cdot 10^9$ bytes, or $200 \cdot 10^9 \cdot 8 = 1.6 \cdot 10^{12}$ bits.
- With an estimate of one billion (10^9) 200 gigabyte computers (personal and midrange servers) in use in 2008 worldwide (to err by exaggeration) one could collectively make $10^9 \cdot 1.6 \cdot 10^{12}$ binary choices or store $1.6 \cdot 10^{21}$ bits of data.
- Considering the speed of computation, say 1 GHz = 10^9/sec, during one year of continuous processing – 1 year $\cong \pi \cdot 10^7$ sec – we could collectively compute about $1.6 \cdot 10^{21} \cdot 10^9 \cdot \pi \cdot 10^7 \cong 5 \cdot 10^{37}$ bits, bringing the cyberspace that we can explore in 2008 to an upwardly rounded:

$$10^{38} \textit{ bits per year}$$

This growing number is large but far smaller than $\pi \cdot 10^{82}$, the capacity of a computer of the mass of the earth running at Bremermann's limit. According to Moore's law, which suggests that the capacity of computation doubles every two years, Bremermann's limit would be reached in about 150 years. Since $\pi \cdot 10^{82}$ is practically unachievable, Moore's law is soon doomed. Cyberspace may become more user-friendly and integrated in everyday life but no longer grow as fast as since the 1960s.

There is of course much happening outside the global computer technology, not reflected in these numbers. People grow and eat food, drive cars to work, construct buildings and cities, publish, read, and talk with one another. However, as observed by Lyman and Varian (2003), most of what is happening outside the electronic world migrates into it. Economic transactions may still take place at a cash register but are recorded electronically and tracked in this medium. Cars operate outside the electronic media, but their production drawings, sales documents, and service instructions are transmitted among the manufacturer's computers. Cars occupy cyberspace also in the from of registration numbers, insurance, repair, and tax records. Books, newspapers and theatrical performances increasingly are available online. Web pages are read and inform decisions outside cyberspace but their results reenter cyberspace in terms of email, blogs, and online purchases. Everything in cyberspace connects via its users to what I call externalities. These externalities are essential to keeping cyberspace meaningful and alive but they do not add significantly to its estimated size.

7.4 Artifacts in Cyberspace.

Unlike traditional machines, which are designed to occupy physical space and serve particular functions at their location, the utility of cyberspace depends on electronic artifacts with which it is furnished. Such artifacts consist of documents, software, networks, plus plus that define dependencies among finite numbers of binary variables. As a matter of definition, *artifacts in cyberspace*

(1) *Occupy space* (in bits of cyberspace) by relating individual bits, for example, the neighborhood relations among the pixels of images, the strings of characters comprising written documents, and the codes of computer software. The relations among otherwise free possibilities in which artifacts are manifest are precisely what Ashby had conceptualized as higher-order interactions and hoped to discover and quantify with the ill-fated *Q*-terms. Artifacts in cyberspace do occupy finite spaces, but identifying their structure by observation (observational determinability) is virtually impossible while their structure is easily established by design (synthetic determinability).

(2) *Are preserved* under a variety of recursive transformations (Foerster, 1981), for example, in the course of their transmission. Artifacts cannot be experienced at their location but where they are reproduced, seen on computer interfaces, printed, controlling something, or doing actual work. They are relatively stable while their location often remains uncertain.

(3) *Can be controlled*, installed, composed, removed, activated, monitored, and terminated by their users (not necessarily by everyone alike)

(4) *Interact with each other*. Software, one kind of artifact, is applied to data, images, or records, another kind of artifact, and its results modifies the data, images, or records or produces different kinds of artifacts. Users' actions intervene in what amounts to an ecology of electronic artifacts

(5) *Are meaningful in their users' lives* in the sense of being understood and usable (hermeneutic determinability) and can be related to the externalities of cyberspace.

Let me distinguish a few kinds of artifacts by their properties:

The artifacts that determine the size of cyberspace are *physical memories*, hard drives, storage devices, media of communication, and networks. These artifacts do not physically move. The rates of their production less of their retirement determine the growth of cyberspace.

A prerequisite of working computers are their *operating systems*. As each computer needs to be equipped with one, operating systems occupy a good deal of cyberspace. This also includes the software for running user *interfaces* with computers, usually part of a computer but doing no work other than providing users access to cyberspace by bridging user culture with the operation of computers.

Data, textual, visual, and sound records, files, folders, and web pages, usually kept as whole bit packages, are the most common, most space consuming, and least intelligent artifacts. They largely inform individual users about externalities. Lyman and Varian (2003) estimated that most computers have no more that 1% original data, the remainder are duplicates, represent redundancies in cyberspace. Duplicates simulate traditional mass media products and compete with the ideas of libraries. Specialized software too belongs in this category of artifacts until it is instructed to cooperate with the operating system and become part of the ecology of electronic artifacts.

Links among documents, web pages, and the organization of file systems occupy cyberspace as well, and so are *transmissions*, i.e., networks that temporarily coordinate computers for the purpose of reproducing data from one location to another. Traffic in contemporary cyberspace from email to data transmission consumes a considerable amount of cyberspace.

The need for privacy, allocating privileges, and protecting data bases has created *security systems* that organize and limit access to cyberspace around particular communities of users. Bank records, military communications, corporate accounts are thus protected.

Another increasingly important category of artifacts is *intelligent assistants* or agents that either learn to serve user needs as a function of their habits or can be instructed to assume responsibilities that the user prefers not to assume or cannot undertake as speedily, reliably, or efficiently as that assistant. Commercial use of assistants is already widespread, visible and invisible by users.

Finally, there are *self-replicators*, viruses, worms, and other artifacts substantially out of users' control. Often designed with malicious intents, they can make their way through cyberspace and create havoc to individual computers, hard drives, data bases, and networks. Self-replicators may be difficult to destroy, but because they occupy space, they often can be quarantined.

One can argue over the categorization of such artifacts but not that they are designed, programmed, or captured to aid users' practices – except for the self-replicators. They provide access to possibilities generally not available otherwise. Without a diverse population of artifacts, cyberspace would by an empty shell.

Despite valid claims that Shannon's quantities have little to say about everyday life, we experience these quantities everywhere. When buying a computer, we pay for the size of memory in bytes and speed. When considering the installation of software, we must be weary of how much valuable space it consumes. When attaching images of Kilobyte size to an email, we need to be concerned with how long it takes to send them and whether they can be received. Bits or bytes are measures of the space that the hardware of cyberspace makes available to computer users and that the artifacts they install exhaust by enabling them to do computational work, communicate with each other, and most importantly, to move among the artifacts in cyberspace and explore what they mean to them.

7.5 Human Interface Capacities.

The size of cyberspace exceeds by far several individual capacities, a fact that limits individual human interaction with computers and how we can operate in cyberspace. Whereas Bremermann's limit concerns physical responsiveness to differences, here I am addressing the implications of human responsiveness to bits in cyberspace.

- *Individual comprehension* – for Ashby – must be accomplished by the 10^{13} to 10^{15} synapses of the human brain, most of which are occupied with coordinating human bodily functions. Experiments have suggested that human comprehension is about two bits per second or $2 \cdot \pi \cdot 10^7$ bits per year. With one billion (10^9) computers in use, attended to 10% per day, the current population of cyberspace users could comprehend about $2 \cdot \pi \cdot 10^7 \cdot 10^9 \cdot 10^{-1} \cong 10^{16}$ bits of cyberspace annually.
- Comprehension does not mean responding to every letter, pixel, or option available on *computer screens*. Perception is selective and what appears on an individual's computer screen necessarily is richer than is seen and can be responded to. A computer screen with 1280×1024 pixels, 32 bits for colors, 75 Hz refresh rate, observed 10% of a year by one billion computer users would take up not more than $1280 \cdot 1024 \cdot 32 \cdot 75 \cdot 10^{-1} \cdot \pi \cdot 10^7 \cdot 10^9 \cong 10^{22}$ bits of cyberspace per year.
- *Typing* probably is the fastest way to direct the performance of a computer. If a very good typist can write about one word/second, a word contains on average 5.5 characters (as in this paper),

each character amounts to $\log_2 32=5$ bits, then one year of typing, 10% of each day, by 10^9 cyberspace users could determine $5.5\cdot5\cdot\pi\cdot10^7\cdot10^{-1}\cdot10^9 \cong 10^{17}$ bits of cyberspace annually – just ten times what one can comprehend.

The order of magnitude of these differences, rough as they may be, is not surprising. First, typing instructs a computer just ten times as can be comprehended. This may well be the difference between understanding whole words as opposed to individual letters. Second, the amount of information displayed on a monitor naturally is far greater than what can be comprehended. Perception is selective and each letter of an alphabet occupies more than 32 pixels plus 32 bits for colors. As George A. Miller (1956) taught us, we see and think in chunks, not pixels. Third, although I do not dare to estimate the cyberspace occupied by all of its artifacts, (a) computer languages, data bases, software and networks have histories of cumulative growth that exceed the lifespan and creativity of individual users, thus naturally exceeding the 10^{17} bits per year of typing. (b) Many artifacts enter cyberspace not by individual construction but by being captured by powerful systems, digital cameras, video recorders, medical imaging, and surveillance systems that operate with minimal human involvement and represent externalities of cyberspace. Their details far exceed human ability to read and to enter them bit by bit. (c) The majority of artifacts in cyberspace are copies. Lyman and Varian (2003) estimate as much as 99% on individual computers. Copies are easy to produce. Directing a device to copy or transmit an artifact may require very few human actions. The amount actually looked at and individually comprehended is a miniscule fraction of what occupies cyberspace. Fourth, artifacts in cyberspace are packages of bits and organized to be controllable by users with a minimum of choices. Getting to an image may need no more than a few clicks with a mouse. Applying a familiar statistical program on available data does not require the user to know their computer code, what exactly it does, nor the data it analyses. The volumes searched on the internet remain largely hidden from the user's view.

While by definition, cyberspace must be larger than the artifacts it houses, it is perfectly sensible to conclude that the space these artifacts jointly occupy far exceeds human ability to know them in all of their details, the actual bits, and the collective ability to design them bit by bit. In his information theory, Ashby talked about equivocation, the mathematical inverse of noise. Equivocation, chunking packages of bits into comprehensible distinctions, takes place on a massive scale at human computer interfaces. The increasing spaces that electronic artifacts occupy is at least one factor that drives the need of cyberspace to grow.

Far more important and unique to this technology is the unoccupied cyberspace. This is a measure of the openness for users to exercise their agency, make individual choices without rational justification, do things not programmable by any computer language, travel paths nobody has paved for them, customize their interfaces, and construct novel artifacts that support their evolving practices of living together. As long as these artifacts do not consume the whole cyberspace or prevent access to willing users, human agency is preserved if not enhanced.

8. Conclusion

It should be clear that what we now call cyberspace cannot remotely approximate Bremermann's limit. Much of the earth's matter is hot or dull and much of our biomass is concerned with itself. Although computation has become indispensable to contemporary society and everyday life, it can always only be a part of it. Estimating the size of cyberspace is an important step in acknowledging human agency as a non-naturalist explanation of the world we construct. It invokes a new paradigm. Ashby's method of first considering possibilities and then exploring which are empirically sustainable and which are not is neither inductive – generalizing from many cases –

nor deductive – deriving knowledge from known theory – but evolutionary – rooted in the idea of the recursion of mutation and selection. Ashby defined cybernetics as the study of all possible systems notwithstanding what has not yet been built or evolved in nature.[1] He modeled that definition after a theory of information that granted cyberneticians human agency. I suggest this ushered in a paradigm that enables us now to study the increasingly complex human use of information technologies which I describe as cyberspace.

Ross Ashby could not experience the technology we live with, which rapidly evolved from the mainframe computers he knew. His conception of a system did not exhibit the fluidity we are now facing. His notion of higher-order interactions in systems of many variables has morphed into the artifacts in cyberspace – occupying finite spaces but are difficult to localize and no longer identifiable by algebraic accounting equations. Conceptualizing them as packages of bits, created by software companies, programmers, and users, captured by powerful devices for recording externalities of cyberspace, and manipulable and useful to us seems natural now. This paper has shown that designing artifacts in cyberspace is not limited by the computational resources available to discover the complexities of the designed artifacts. The paradigm shift from computational methods of discovery to human methods of design has overcome Bremermann's limit and questions the approach taken by earlier systems theories.

References

Ashby, W.R., 1970. Information flows within co-ordinated systems. *In*: J. Rose, ed. *Progress of Cybernetics: Proceedings of the 1st International Congress of Cybernetics.* London: Gordon and Breach, 57-64.

Ashby, W.R., 1969. Two tables of identities governing information flows within large systems. *The American Society for Cybernetics Communications,* 1 (2), 3-8.

Ashby, W.R., 1968. Some consequences of Bremermann's limit for information processing systems. *In*: H. Oestreicher and D. Moore, eds. *Cybernetic Problems in Bionics*. New York: Gordon and Breach, 69-76.

Ashby, W.R., 1965. Measuring the internal information exchange in a system. *Cybernetica,* 8, 5-22.

Ashby, W.R., 1964a. The set theory of mechanisms and homeostasis. *General Systems Yearbook,* 9, 83-97.

Ashby, W.R., 1964b. Constraint analysis of many-dimensional relations. *General Systems Yearbook,* 9, 99-105.

Ashby, W.R., 1956a. *An Introduction to Cybernetics*. London: Chapman and Hall.

Ashby, W.R., 1956b. Design for an Intelligence-Amplifier. *In*: C.E. Shannon and J. McCarthy, eds. *Automata Studies*. Princeton, NJ: Princeton University Press, 215-234.

Bateson, G., 1972. *Steps to an Ecology of Mind*. New York: Ballentine Books.

[1] Ashby's (1956:2) original formulation reads: "(Cybernetics) takes as its subject matter the domain of 'all possible machines,' and is only secondarily interested if informed that some of them have not yet been made, either by Man or Nature." I took the liberty of rephrasing his formulation in contemporary terms. Norbert Wiener (1948) had added "in the animal and machine" to his definition in order to acknowledge the generality of "communication and control," like Ashby embracing naturally evolved and designed systems. Ashby's definition does not commit cybernetics to particular concepts, emphasizing instead the two epistemological stances of designing and discovering within what is possible.

Bertalanffy, L. von, 1968. *General Systems Theory. Foundations, Developments, Applications.* New York: George Braziller.

Bremermann, H.J., 1962. Optimization through evolution and recombination. *In*: M.C. Yovitz, G. Jacoby, and G. Goldstein, eds. *Self-Organizing Systems.* Washington, DC: Spartan Books, 93-106.

Broekstra, G., 1981. C-analysis of C-structures: Representation and evaluation or reconstruction hypotheses by information measures. *International Journal of General Systems,* 7 (1), 33-61.

Broekstra, G., 1979. Probabilistic constraint analysis of structure identification: An overview and some social science applications. *In*: B. Zeigler et al. (Ed.). *Methodology of Systems Modeling and Simulation*. Amsterdam: North Holland, 305-334.

Broekstra, G., 1977. Constraint analysis and structure identification II. *Annals of Systems Research,* 6, 1-20.

Broekstra, G., 1976. Constraint analysis and structure identification. *Annals of Systems Research,* 5, 67-80.

Conant, R.C., 1981a. Set-theoretic structure modeling. *International Journal of General Systems,* 7, 93-107.

Conant, R.C., 1981b. How to ignore Bremermann's limit. *Proceedings of the Southeastern Regional Meeting of the Society for General Systems Research*. Louisville, KY, April 21-23, 1981, pages 398-404.

Conant, R.C., 1980. Structural modeling using a simple information measure. *International Journal of Systems Science,* 11 (6), 721-730.

Conant, R.C., 1976. Laws of Information which govern systems. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC 6 (4), 240-255.

Conant, R.C., 1973. Measuring the strength of intervariable relations. *Kybernetes,* 2, 47-52.

Darroch J.N. and Ratcliff, D., 1972. Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics,* 43, 1470-1480.

Foerster, H. von, 1984. Principles of self-organization in a socio-managerial context. *In*: H. Ulrich and G.J.B. Probst, eds. *Self-Organization and Management of Social Systems.* New York: Springer Verlag, 2-24.

Foerster, H. von, 1981. Objects: tokens for (eigen)behaviors. *In his*: *Observing Systems*. Seaside, CA: Intersystems, 274-285.

Foerster, H. von, 1979. Cybernetics of cybernetics. *In*: K. Krippendorff, ed.. *Communication and Control in Society*. New York: Gordon and Breach, 5-8.

Foerster, H. von, et al., 1974. *Cybernetics of Cybernetics or the Control of Control and the Communication of Communication*. Urbana, IL: Biological Computer Laboratory at the University of Illinois.

Garner, W.R., 1962. *Uncertainty and Structure as Psychological Concepts*. New York : Wiley.

Klir, G.J., 1976. Identification of generative structures in empirical data. *International Journal of General Systems,* 3, 89-104.

Klir, G.J., 1978. Structural modeling of indigenous systems. *Proceedings of the 22nd Annual SGSR Meeting*. Washington, DC: February 13-15, pages 151-155.

Krippendorff, K., 2008 in press. Four (in)determinabilities, not one. *In*: J.V. Ciprut, ed. *Indeterminacy: The Mapped, the Navigable, and the Uncharted.* Cambridge, MA: MIT Press.

Krippendorff, K., 2006. *The Semantic Turn; A New Foundation for Design*. Boca Raton, FL: Taylor & Francis Group, CRC Press.

Krippendorff, K., 1986. *Information Theory; Structural Models for Qualitative Data.* Newbury Park, CA: Sage.

Krippendorff, K., 1981. An algorithm for identifying structural models of multi-variate data. *International Journal of General Systems,* 7, 63-79.

Krippendorff, K., 1980a. Q; An interpretation of the information theoretical Q-measures. *Proceedings of the Fifth European Meeting for Cybernetics and Systems Research*, Vienna, April 1980; *and in*: R. Trapple, G. Klir and F. Pichler, eds., 1982. *Progress in Cybernetics and Systems Research* Vol. VIII. New York: Hemisphere, 63-67.

Krippendorff, K., 1980b. On the identification of latent functions in multi-variate Data. *Fifth European Meeting for Cybernetics and Systems Research*, Vienna, April 1980 *and in*: R. Trapple, G. Klir and F. Pichler, eds., 1982. *Progress in Cybernetics and Systems Research*, Vol. VIII. New York: Hemisphere, 31-42.

Krippendorff, K., 1979a. On systems thinking. *In*: P. Broholm and N. van Dijk, eds. *Systems Thinking and Social Science,* Proceedings of a Symposium held at the Inter-Universitaire Interfaculteit Bedrijfskunde. Delft, The Netherlands, November 15, 1979, pages 13-21.

Krippendorff, K., 1979b. On the identification of structures in multi-variate data by the spectral analysis of relations. *In*: B.R. Gaines, ed. *General Systems Research: A Science, A Methodology, A Technology*. Louisville, KY: Society for General Systems Research, 82-91.

Krippendorff, K., 1978. A spectral analysis of relations: Further developments. *Fourth European Meeting on Cybernetic and Systems Research*, Linz, March 1978.

Krippendorff, K., 1976. A spectral analysis of relations. *International Congress of Communication Sciences*, Berlin, May 31, 1977. Philadelphia, PA: The Annenberg School of Communication, University of Pennsylvania, mimeo.

Krippendorff, K., 1974. An algorithm for simplifying the representation of complex systems. *In*: J. Rose, ed. *Advances in Cybernetics and Systems*. New York: Gorden & Breach, 1693-1702.

Krippendorff, K., 1967. *An Examination of Content Analysis; A Proposal for a General Framework and an Information Calculus for Message Analytic Situations*. Ph.D. Dissertation. Urbana: University of Illinois.

Kullback, Solomon, personal communication.

Lyman, P. and Varian, H.R., 2003. *How much Information? 2003*. http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm#summary, last accessed 2007.10.26.

McGill, W.J., 1954. Multivariate information transmission. *Psychometrica,* 19, 97-116.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review,* 63, 81-97.

Shannon. C.E. and Weaver, W., 1949. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Wiener, N., 1948. *Cybernetics or Control and Communication in the Animal and the Machine*. New York: John Wiley & Sons.

Wittgenstein, L., 1922. *Tractatus Logico-Philosophicus*. London: Kegan Paul, Trench, Trubner & Co.