2-2009

Ross Ashby's Information Theory: A Bit of History, Some Solutions to Problems, and What We Face Today

Klaus Krippendorff
University of Pennsylvania, kkrippendorff@asc.upenn.edu

Recommended Citation

Krippendorff, K. (2009). Ross Ashby's Information Theory: A Bit of History, Some Solutions to Problems, and What We Face Today. *International Journal of General Systems, 38* (2), 189-212. https://doi.org/10.1080/03081070802621846

The correction is included as an additional file. Its information is as follows: *International Journal of General Systems, 38*(6), 667-668.

Ross Ashby's Information Theory: A Bit of History, Some Solutions to Problems, and What We Face Today

Abstract

This paper presents a personal history of one strand of W. Ross Ashby's many ideas: using information theory to analyse complex systems empirically. It starts with where I entered the evolution of the idea as one of his students, points out a problem that emerged as a consequence of generalising information measures from simple to complex systems, i.e. systems with many variables, shows how this problem was eventually solved, and ends with how his idea of decomposing complex systems into smaller interactions reappears in one of the most complex technologies of our time: cyberspace. While nobody could anticipate the complexities that developed since, Ashby's idea of understanding complex systems in terms of manageable interactions, which I call electronic artefacts, is actually practised today and cyberspace is again worth analysing in information theoretical terms.

Keywords

Information Theory, Cybernetica, Complexity, Interaction, Cyberspace, Computational Artifacts

Disciplines

Communication | Other Social and Behavioral Sciences

Comments

The correction is included as an additional file. Its information is as follows: *International Journal of General Systems, 38*(6), 667-668.

International Journal of General Systems
Vol. 38, No. 2, February 2009, 189–212

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, USA

(*Received 13 February 2008*; *final version received 24 August 2008*)

This paper presents a personal history of one strand of W. Ross Ashby's many ideas: using information theory to analyse complex systems empirically. It starts with where I entered the evolution of the idea as one of his students, points out a problem that emerged as a consequence of generalising information measures from simple to complex systems, i.e. systems with many variables, shows how this problem was eventually solved, and ends with how his idea of decomposing complex systems into smaller interactions reappears in one of the most complex technologies of our time: cyberspace. While nobody could anticipate the complexities that developed since, Ashby's idea of understanding complex systems in terms of manageable interactions, which I call electronic artefacts, is actually practised today and cyberspace is again worth analysing in information theoretical terms.

Keywords: communication theory; complexity; cybernetics; cyberspace; decomposition; electronic artefacts

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 was 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, Ashby's (1956a) *An Introduction to Cybernetics* and Ludwig Wittgenstein's (1922) *Tractatus Logicus Philosophicus*. I cannot 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 place to study. I met Heinz von Foerster at the 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 my becoming a student at the University of Illinois. I enrolled in Ashby's two-semester course in 1962–1963. He became an important member on my dissertation committee. The dissertation reconceptualised content analysis as a research method in the social sciences 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*

* Email: kkrippendorff@asc.upenn.edu

ISSN 0308-1079 print/ISSN 1563-5104 online

DOI: 10.1080/03081070802621846
http://www.informaworld.com

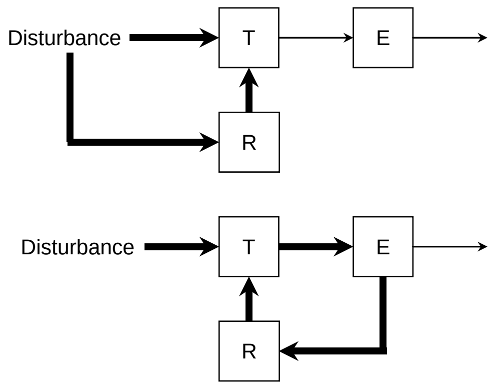


Figure 1. Ashby's regulators R and two versions of the *Law of Requisite Variety*.

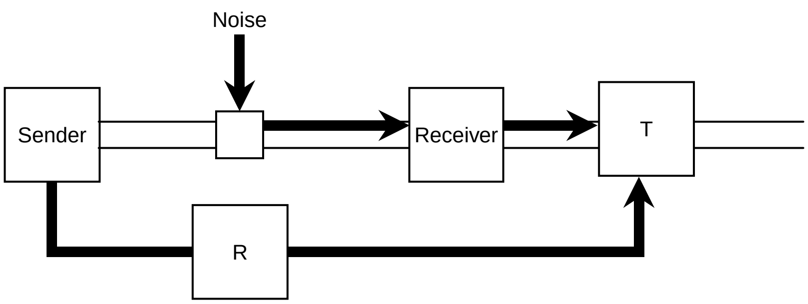


Figure 2. Noisy communication channel and correction channel R.

Variety (Ashby 1956, pp. 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 that 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.

Today, Ashby's *Law of Requsite Variety* is considered a generalization of Shannon's 10th theorem which 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 Ashby's *Introduction to Cybernetics* went.

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 ability to understand systems is limited by the variety available to the observer'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.

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 had no need to distinguish systems from their environment, or to generalise from living systems what makes them viable. Ashby always insisted that anything can afford multiple descriptions and what we know of a system always is 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 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 cohort preceding mine to figure out its input-output relationships. This must have been a most frustrating assignment because every hypothesis advanced to explain the box seemed to 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 analysing complex systems became an important part of Ashby's work. It is now recognised that the ability to determine the nature of a system by observation is limited to trivial machines (Foerster 1984).

Before fully embracing information theory, Ashby (1964b) 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*. 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, adapted from Roger Conant's (1981a) account of constraint analysis, demonstrates the two operations involved. Here, the result of a constraint (of three out of eight possible cells not occurring), the relation R(123), is projected onto each plane in two variables R(12), R(13), and R(23), and on each individual variable R(1), R(2), and R(3). Then, the inverse of projections is used in an attempt to reconstruct the original relation from some of its projections. Among the four examples shown here, the first does not account for any constraint, the second and third shows some constraint but not enough to reconstruct the original relation. The fourth demonstrates that the original relationship R(123) can be reconstructed from relations R(12) and R(13) and is hence simplifiable into these without loss: R(123) = R(12:13). 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 separately.

Ashby was attracted to information theory, not only because of his *Law of Requisite Variety*, but also because it promised to generalise his constraint analysis to probabilistic systems and finding an elegant algebra of relations. Shannon's theory had distinguished signals from noise or 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.

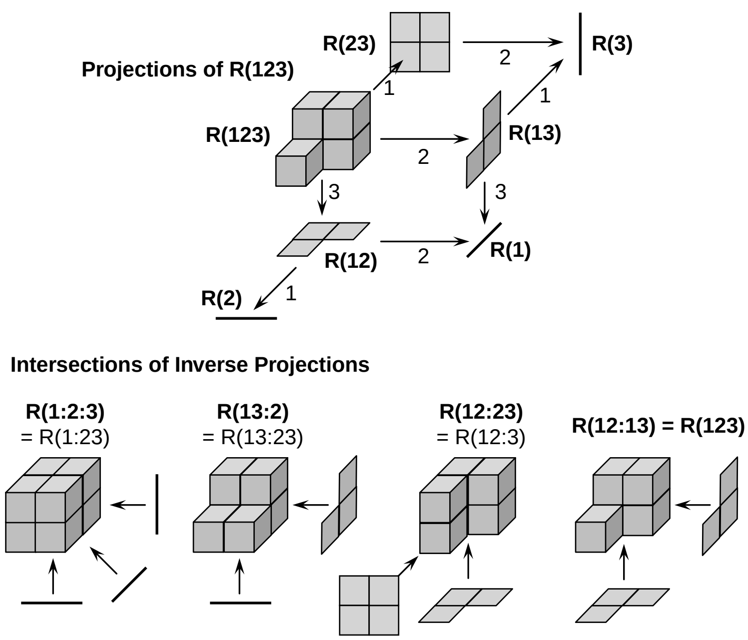


Figure 3. Geometrical interpretation of the constraint analysis of a three-variable relation.

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

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

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

Noise: $H_{\mathrm{A}}(\mathrm{B}) = \sum_{a\in A} p_a \left[-\sum_{b\in B}(p_{ab}/p_a)\log_2(p_{ab}/p_a)\right]$

Transmission: $T(\mathrm{A}{:}\mathrm{B}) = H(\mathrm{A}) + H(\mathrm{B}) - H(\mathrm{AB}) = H(\mathrm{A}) - H_{\mathrm{A}}(B)$

McGill (1954) and Garner's (1962) uncertainty analysis extended Shannon's measures to three variables for analysing psychological data. Entropies:

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

$$H_{\mathrm{A}}(\mathrm{BC}) = H(\mathrm{ABC}) - H(\mathrm{A}),$$

$$H_{\mathrm{AB}}(\mathrm{C}) = H(\mathrm{ABC}) - H(\mathrm{AB}).$$

Transmissions:

$$T(\mathrm{A}{:}\mathrm{B}) = H(\mathrm{A}) + H(\mathrm{B}) - H(\mathrm{AB}),$$

$$T_{\mathrm{C}}(\mathrm{A}{:}\mathrm{B}) = H_{\mathrm{C}}(\mathrm{A}) + H_{\mathrm{C}}(\mathrm{B}) - H_{\mathrm{C}}(\mathrm{AB}),$$

$$T(\mathrm{A}{:}\mathrm{B}{:}\mathrm{C}) = H(\mathrm{A}) + H(\mathrm{B}) + H(\mathrm{C}) - H(\mathrm{ABC}),$$

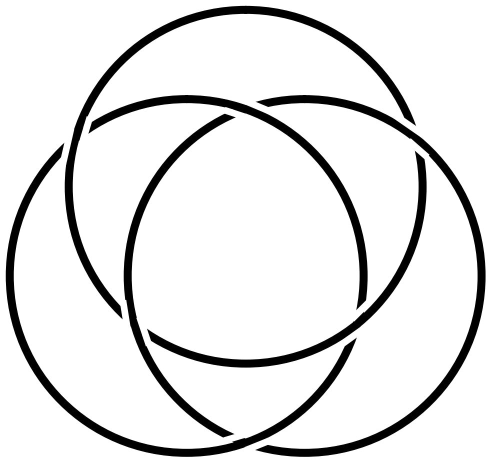


Figure 4. Ashby's chain necklace.

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

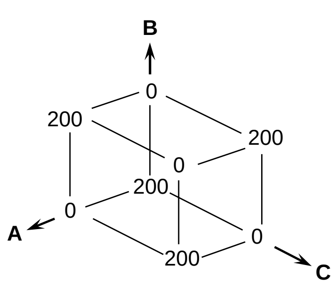
$$Q(\mathrm{ABC}) = T_{\mathrm{C}}(\mathrm{A{:}B}) - T(\mathrm{A{:}B}) = T_{\mathrm{B}}(\mathrm{A{:}C}) - T(\mathrm{A{:}C}) = T_{\mathrm{A}}(\mathrm{B{:}C}) - T(\mathrm{B{:}C}).$$

Guided by the idea of his constraint analysis, Ashby saw the possibility of decomposing unanalysed multi-variable systems into its constituent relationships among fewer than all variables, manifest in non-zero quantities of the information calculus. McGill provided this accounting equation:

$$T(\mathrm{A{:}B{:}C}) = T(\mathrm{A{:}B}) + T(\mathrm{A{:}C}) + T(\mathrm{B{:}C}) + Q(\mathrm{ABC}),$$

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 subsets. 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.

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



$Q(ABC)$	=	1.00
$T(B{:}C)$	=	0.00
$T(A{:}C)$	=	0.00
$T(A{:}B)$	=	0.00
$T(A{:}B{:}C)$	=	1.00

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

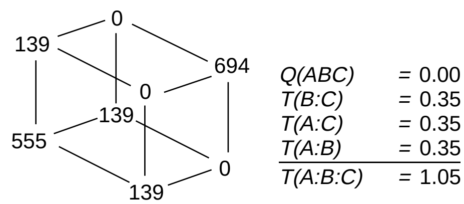


Figure 6. Frequency distribution with zero Q(ABC).

binary transmission measures may be seen as justified as the projections of the three-dimensional distribution onto its three two-dimensional tables are uniform and exhibit no structure, and the three-dimensional distribution could not possibly have been predicted from them.

We uncritically accepted that zero values for the Q-measures, as exemplified in Figure 6, signalled the absence of interactions.

As students we computed many of these accounts by hand, using an $n \log_2 n$ table, which was tedious without electronic computers, and we followed Ashby's lead to generalise information theory to any number of variables, which was easier than computing their numerical quantities. This effort culminated in the publication of two lists of some 50 accounting Equations (Ashby 1969), amounting to the beginning of an elaborate information algebra. The Q-measures were Ashby's prime candidates. By extending McGill and Garner's Q-terms to fewer and to more than three variables

$$Q(\mathrm{A}) = -H(\mathrm{A}),$$

$$Q_{\mathrm{B}}(\mathrm{A}) = -H_{\mathrm{B}}(\mathrm{A}),$$

$$Q_{\mathrm{BC}}(\mathrm{A}) = -H_{\mathrm{BC}}(\mathrm{A}),$$

$$Q(\mathrm{AB}) = Q_{\mathrm{B}}(\mathrm{A}) - Q(\mathrm{A}) = Q_{\mathrm{A}}(\mathrm{B}) - Q(\mathrm{B}) = T(\mathrm{A}:\mathrm{B}),$$

$$Q_{\mathrm{C}}(\mathrm{AB}) = Q_{\mathrm{BC}}(\mathrm{A}) - Q_{\mathrm{C}}(\mathrm{A}) = Q_{\mathrm{AC}}(\mathrm{B}) - Q_{\mathrm{C}}(\mathrm{B}) = T_{\mathrm{C}}(\mathrm{A}:\mathrm{B}),$$

$$Q(\mathrm{ABC}) = Q_{\mathrm{C}}(\mathrm{AB}) - Q(\mathrm{AB}) = Q_{\mathrm{B}}(\mathrm{AC}) - Q(\mathrm{AC}) = Q_{\mathrm{A}}(\mathrm{BC}) - Q(\mathrm{BC})$$

$$\begin{aligned} Q(\mathrm{ABCD}) &= Q_{\mathrm{D}}(\mathrm{ABC}) - Q(\mathrm{ABC}) \\ &= \text{other expressions by permutation of these variables}, \end{aligned}$$

$$\begin{aligned} Q(\mathrm{ABCDE}) &= Q_{\mathrm{E}}(\mathrm{ABCD}) - Q(ABCD) \\ &= \text{other expressions by permutation of these variables, etc.}, \end{aligned}$$

a general accounting equation emerged (Ashby 1969, p. 6). In its terms, we assumed able to quantitatively decompose the total amount of information transmission T in a system of any number of variables into its unique interaction quantities Q:

$$T(\mathrm{A}:\mathrm{B}) = Q(\mathrm{AB}),$$

by definition

$$T(\mathrm{A:B:C}) = Q(\mathrm{AB}) + Q(\mathrm{AC}) + Q(\mathrm{BC}) + Q(\mathrm{ABC}),$$

$$T(\mathrm{A:B:C:D}) = Q(\mathrm{AB}) + Q(\mathrm{AC}) + Q(\mathrm{AD}) + Q(\mathrm{BC}) + Q(\mathrm{BD}) + Q(\mathrm{CD}) + Q(\mathrm{ABC})$$
$$+ Q(\mathrm{ABD}) + Q(\mathrm{ACD}) + Q(\mathrm{BCD}) + Q(\mathrm{ABCD})\,\text{etc.}$$

Stated generally:

$$T(\mathrm{S}) = \sum\nolimits_{\alpha \subset S} Q(\alpha),$$

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 (Broekstra 1976, 1977, 1979, 1981, Conant 1976, 1980). I too developed equations and algorithms for simplifying complex systems in these terms (1974), and aimed at a spectral analysis of multi-valued relations (1976, 1978, 1981). 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 Figure 5 and absent in Figure 6, there remained something odd: Q could be negative, as may be seen in Figures 7 and 8.

McGill (1954) had acknowledged this possibility and considered any deviation from zero a signal that interaction existed in the data. 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 two-dimensional faces of the cube are sufficient to reconstruct or uniquely determine its

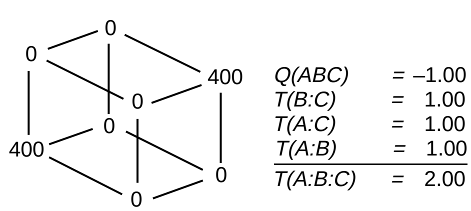


Figure 7. Sparse frequency distribution with negative Q(ABC).

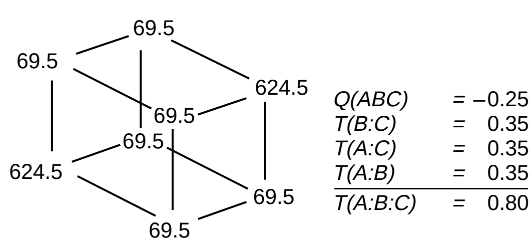


Figure 8. Frequency distribution with negative Q(ABC).

three-dimensional distribution, incidentally much like the example in Figure 3. The third projection is redundant, implied and not needed to obtain that distribution. If $T(\mathrm{A{:}B{:}C}) = T(\mathrm{A{:}B}) + T(\mathrm{A{:}C})$, a third-order interaction should be absent by this conception, yet $Q(\mathrm{ABC})$ has a value other than zero. In fact, it seemingly corrects for redundant measures, here of $T(\mathrm{B{:}C})$. 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 $T(\mathrm{A{:}B}) = T(\mathrm{A{:}C}) = T(\mathrm{B{:}C}) = 0.35$. But the distribution in Figure 6 is most unlike chance or maximally entropic, satisfying the three two-dimensional distributions and therefore suggesting the presence of an interaction, stronger than in Figure 8, but not measured by Q. We all followed a faulty logic.

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 Systems Research, Klir (1978) presented a paper reporting 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 1964a) 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

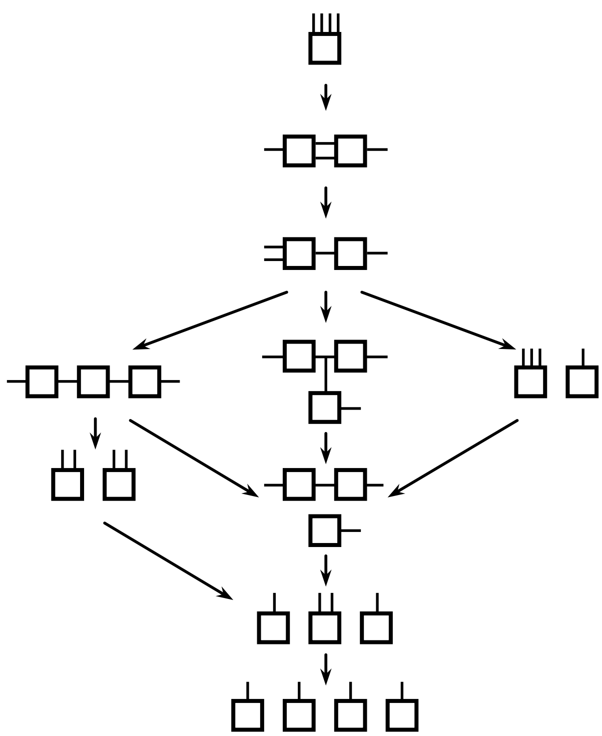


Figure 9. Lattice of simplifications of models of systems in four variables without loops.

connecting them. This simple gestalt switch allowed me to visualise interactions inside Klir's boxes, without 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 variables.

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 components that defined interactions, one by one. Although Klir was not concerned with information theory, his lattice visualised 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:

$$T(\mathrm{A:B:C:D}) = T(\mathrm{C:D}) + T(\mathrm{A:D}) + T(\mathrm{B:C}) + T_{\mathrm{C}}(\mathrm{B:D}) + T_{\mathrm{D}}(\mathrm{A:C}) + T_{\mathrm{CD}}(\mathrm{A:B})$$

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 within 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.

With such lattices, it became easy to reconceptualise the information quantities of interest, not in terms of *Q*-measures, but in terms of the differences between the maximum entropies within 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 original and unanalysed whole system, m_{ind} is the model of the 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.

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:

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

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 circular relations among its constituent variables.

4. Information in circular systems

Shannon called his theory *A Mathematical Theory of Communication* and attended to processes that proceeded in one direction only. Accordingly, a message received could have no effect on the message sent. Noise that entered a channel could only degrade what was communicated. A prior choice necessarily limited subsequent choices. It could not have

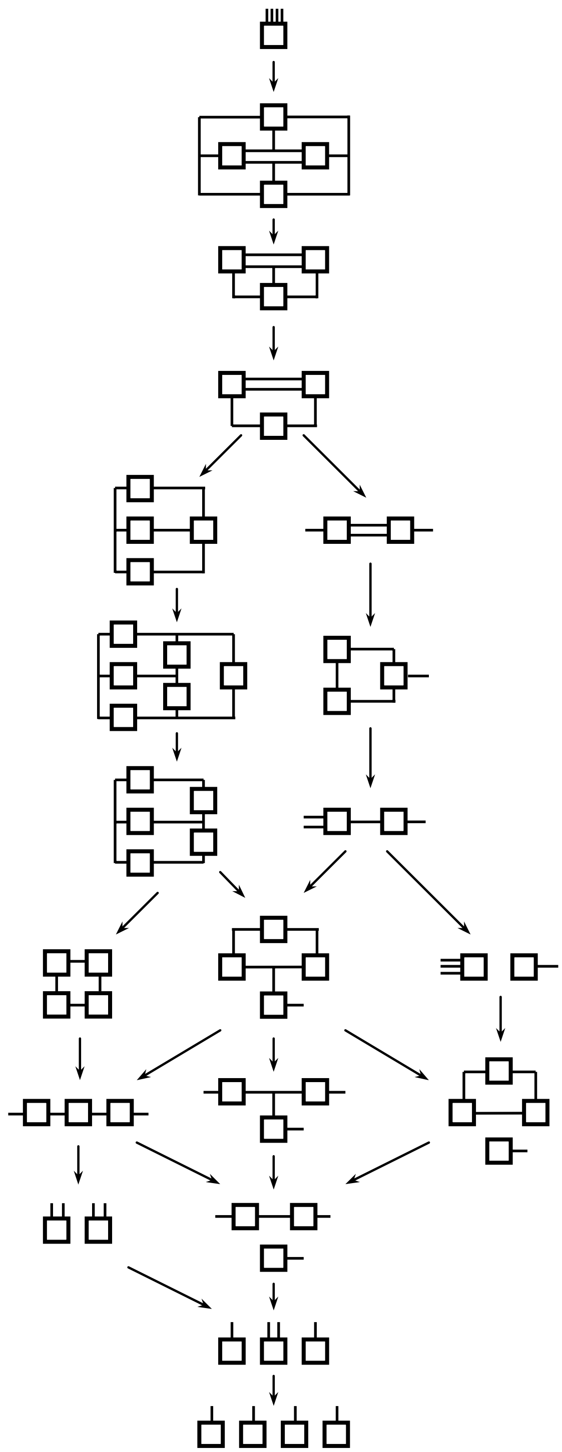


Figure 10. Lattice of all possible of models of systems in four variables.

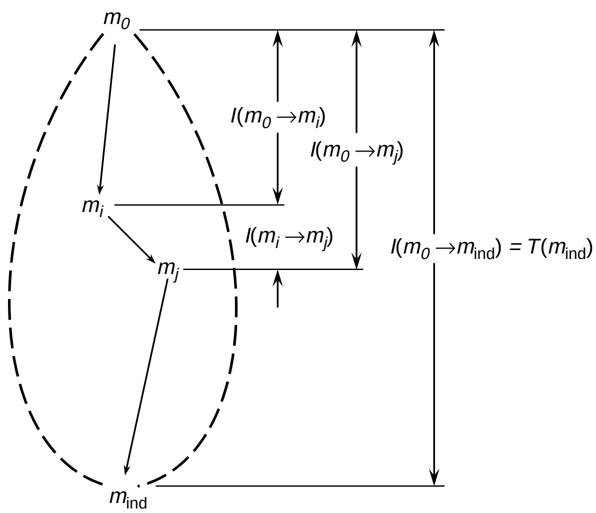


Figure 11. Generalised lattice of all possible models of a system m_0 and information measures of their differences (with interactions successively removed).

an effect on itself. This linearity may not have been entirely intentional 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 his theory was grounded in a far more basic conceptual commitment: *probability theory*. Probability theory axiomatically requires that the probabilities in any one set sum to 1, and expected probabilities of joint events are obtained by multiplication of the probabilities of their components.

Shannon's second theorem (Shannon and Weaver 1949, p. 19) relied on the logarithm function which converts products into sums, and established that the entropy function was the only function that afforded the intuition of information being an additive quantity. This additivity is fundamental to all the entropy and information quantities defined above. However, as it turns out, the additively that created the Q-measures violate the axiom of probability theory. This may be seen when expressing Q in terms of probabilities:

$$Q(\mathrm{A}) = \sum\nolimits_{a\in A} p_a \log_2 p_a = -H(\mathrm{A}),$$

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

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

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

All numerators of these expressions are proper probabilities: $\sum_{a\in A} p_a = 1$, $\sum_{a\in A}\sum_{b\in B}\ p_{ab} = 1$, etc. and so is the denominator in Q(AB): $\sum_{a\in A}\sum_{b\in B} p_a p_b = 1$.

But the denominators, starting with Q(ABC), no longer are: $\sum_{a\in A}\sum_{b\in B}\sum_{c\in C}\frac{p_{ab}p_{ac}p_{bc}}{p_ap_bp_c}$ $\neq 1$, $\sum_{a\in A}\sum_{b\in B}\sum_{c\in C}\sum_{d\in D}\frac{\frac{p_{abc}p_{abd}p_{acd}p_{bcd}}{p_{ab}p_{ac}p_{ad}p_{bc}p_{bd}p_{cd}}}{p_ap_bp_cp_d}\neq 1$, etc. Thus, for three or more variables, the denominators of Q are not probabilities, and Q-measures are incompatible with the 2nd theorem of information theory, which presumed probability theory to be able to obtain expected or maximum entropies.

This incompatibility with probability theory stems from the fact that removing unique interactions from systems with three or more variables, which Ashby wanted to distinguish and into which he wanted to decompose complex systems, created *circular relationships* among the remaining components. However, obtaining maximum entropy probability distributions of systems by multiplying their component probabilities assumes *linear relationships* among them. Thus, circularities in systems defy the possibility of obtaining maximum entropy probability distributions by multiplication. None of us who applied information theoretical measures to complex systems at this time realised this mathematical limit of probability theory. In retrospect not seeing this right away is all the more surprising as circularity is fundamental to cybernetics.

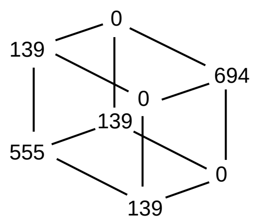
However, the idea of additive quantities that measure the unique contributions of higher-order interactions in systems (leaving circularities behind) can be retained by calculating the maximum entropy probability distribution, subject to the constraints of the probabilities of its components – not by 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 1982a, 1982b). 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: $\ldots$:K_e: $\ldots$: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 eth 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 $\omega^{(0)}_{abc\ldots} = 1/N_{\mathrm{ABC}\ldots}$, where $N_{\mathrm{ABC}\ldots}$ 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: $\omega^{(rt+e)}_{abc\ldots} = p_{k_e}\left(\omega^{(rt+e-1)}_{abc\ldots}/\sum_{\bar{k}_e}\omega^{(rt+e-1)}_{abc\ldots}\right)$

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: $\ldots$:K_e: $\ldots$: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 but excluded from m_i becomes:



$I(ABC \rightarrow AB{:}AC{:}BC)$	$= 0.70 \neq Q(ABC)$
$I(AB{:}AC{:}BC \rightarrow AB{:}AC)$	$= 0.10 \neq T(B{:}C)$
$I(AB{:}AC \rightarrow AB{:}C)$	$= 0.35 = T(A{:}C)$
$I(AB{:}C \rightarrow A{:}B{:}C)$	$= 0.35 = T(A{:}B)$
$I(ABC \rightarrow A{:}B{:}C)$	$= 1.50 = T(A{:}B{:}C)$

Figure 12. Correct account of interactions in Figure 6.

$$I(m_0 \rightarrow m_i) = \sum\nolimits_{abc\ldots \in ABC\ldots} p_{abc\ldots} \log_2 \frac{p_{abc\ldots}}{\omega_{abc\ldots}}.$$

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:

$$I(m_i \rightarrow m_j) = I(m_0 \rightarrow m_j) - I(m_0 \rightarrow m_i) = \sum\nolimits_{abc\ldots \in \mathrm{ABC}\ldots} p_{abc\ldots} \log_2 \frac{\omega_{abc\ldots(m_i)}}{\omega_{abc\ldots(m_j)}},$$

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, unlike in Figure 5, which manifests such interactions. 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.

I presented these developments at a 1980 conference on cybernetics in Vienna (Krippendorff 1982a), 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 (1976, 1978, 1979a, 1979b), and many more – excluding Conant, who never trusted the Q-measure – 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 conceptualised, is measurable not by Q but by $I(m_i \rightarrow m_{i+1})$, the entropy present in m_i and absent in its successor m_{i+1}. 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 old program and several recent developments on his website http://www.pdx.edu/sysc/research_dmm.html (last accessed 20 May 2008). Martin also reminded me that log-linear modelling has fully embraced the iterative computation of probabilities for interactions with loops.

5. Material and informational numbers

One can say that 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 decipherability of encoded messages without knowledge of the key, and in Ashby's terms, on the ability to regulate a system that faces disturbances. Crucially is that it takes for granted the existence of differences. Gregory Bateson (1972, p. 381) defined information as 'any difference which makes a difference in some later event'. But differences do not exist in nature. They result from someone drawing distinctions and noticing their effects. Therefore, substituting 'recognizable change' for 'difference' leads one to something that can be recognised and observed. 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 meet Hans-Joachim Bremermann at the second conference on self-organising systems. Bremermann (1962) recognised 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

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

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	$\cong 10^{-10}$ s
	One year	$\cong \pi \times 10^7$ s
	Time since the earth solidified	$\cong 10^{20}$ s
Mass	Mass of the Earth	$\cong 6 \times 10^{27}$ g
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	$2 \times 10^{47} \times 6 \times 10^{27} \cong 10^{75}$ bits/s
	or	$10^{75} \times \pi \times 10^7 \cong \pi \times 10^{82}$ bits/year
	Since the earth solidified, that ideal computer could have computed no more than	$10^{20} \times 10^{75} \cong 10^{95}$ bits

From which Ashby (1968) concluded:

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 built or found in nature*, he was led to enumerate possibilities rather than actual observations and the numbers that emerged may be called combinatorial or informational. For example,

Possibilities:	Number of configurations displayable by an array of $20 \times 20 = 400$ 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	$2^{13,297} \cong 3 \times 10^{4002} \gg 10^{100}$
	Number of images presentable on a HDTV screen with 1920×1080 pixels and 32 bits for colour	$2^{2,073,632} \cong 10^{624,000} \ggg 10^{100}$
	Number of distinctions between good and bad images on that screen	$\cong 2^{1 \text{ followed by } 624,000 \text{ zeros}} \ggg 10^{100}$

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

$$\begin{aligned}\textit{Cyberneticians have to cope with numbers } &\gg 10^{100}\\ \textit{with material resources for computation } &\ll 10^{100}.\end{aligned}$$

Eight years before I concluded my part in the development of information theory, in 1972, I attended a conference on cybernetics in Oxford, England where we learned from the cybernetician, Gray Walter, that Ashby was mortally ill with a brain tumour. Another former Ashby student from Switzerland, by the name of Burckhardt (regrettably, I am not recalling his first name), and I took a train to visit him. His wife told us to be brief and not to mention to him the terminal nature 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. When asked what he intended to do once he got better, he told us of planning a book that would start with Bremermann's limit. Subsequent inquiries did not turn up notes of how he would have proceeded. Roger Conant (1981c) edited Ashby's writings. 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 accordingly – without rational prescriptions or pre-established determinisms. Bremermann's finding implicates human agency by stating not what exists but what we *CAN* or *CANNOT* do 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 favour the latter and have distinguished four epistemologies regarding understanding systems (Krippendorff 2008).

- Systems whose behaviour is deducible from a finite history of recorded observations are *observationally determinable*. This reflects the epistemological stance of detached observers who seek to discover a system's properties by testing all possible hypotheses about that system's 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 know the structure of a system having determined its makeup.
- Systems that can be utilised by skilfully interacting with them are *hermeneutically determinable*, contemporary computers, for example.
- Systems that can be understood by participating in them are *constitutively determinable*. The latter applies to social systems, constitutively involving knowledgeable human participants. This includes what second-order cybeneticians do.

Theorising the experiences with the above-mentioned 'Ashby box', Heinz von Foerster (1984) has shown that observational determinability is limited to trivial machines – systems with few states and simple structures. Non-trivial machines, involving internal memories, defy observational determinability but can be understood by building them or taking them apart and reassembling them. Computers, for example, are non-trivial by this definition. They can be built but hardly understood by merely observing what they do. The designer of non-trivial systems faces informational limits as well, however, these limits concern the number of components available to them. The history of computing started with programming small procedures, assembling them to larger and larger procedures. The elements of current computer languages are far removed from the changes in zeros and ones they ultimately control – but always within the limit of what designers can handle. Most competent computer users have no clue and do not need to care about how their machines are built and function 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. Computer users deal with information quantities other than computer designers. These quantities have to do with the details that users can distinguish among the pixels on their computer screens and how fast they can change them by their actions.

I suggest that information processes of the kind we are facing today can no longer be understood by discovering and identifying interactions in observed systems. Reconstructability analysis, for example, 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 technology requires an understanding of how possibilities are created and realities are constructed within them. Bremermann's limit merely defines the space within which human agency is physically possible.

7. Cyberspace

Considering the above, *space* is not a metaphor or a mathematical artefact. *Space* is created and recognised by human actors in the process of realising (making real) their artefacts. It is a way to understand human abilities, is manifest in the auxiliary verb 'can' and becomes evident in material artefacts that could not emerge in unattended nature and be explained causally or entropically. Space is constituted in the possibilities that human actors perceive in their world. Here are five propositions concerning that 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) Choices among possible actions have consequences, often social ones, i.e. pertaining to other actors. For example, dialling 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 neighbours might build theirs.
(iii) Choices among technologies almost always trade 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 human interfaces with them. For example, the human use of cyberspace 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 on the internet within seconds for something that would take humans a lifetime. The openness experienced by internet users makes it difficult if not impossible to formulate a single elegant theory of cyberspace.

History of cyberspace

Cyberspace consists of technologically supported possibilities for human actions. To me, cyberspace originated when early humans found sticks, stones, and fire to be separable from where they could be found and moved to where they might accomplish something previously thought impossible. As such sticks, stones, and fire may have been the first human artefacts. 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 in information, as current writers on information society insist, than to an increase in our collective ability to draw more and finer distinctions, to recognise more and finer differences, to handle, assemble, use, and communicate what we distinguished more efficiently than before, and to construct worlds that enhance our collective ability to realise ourselves. The great Cheops pyramid, built 5000 years ago during a 20 year period, amounted to moving 2.3 Billion stones into a descriptively very simple arrangement. The mass production of same-size bricks enabled the building of a great many and descriptively far more complex 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 US Library of Congress is estimated to contain 10-terra bytes (Lyman and Varian 2003), including the characters its collection contains, about 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 and generated numerous 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 artefacts, 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.

The current size of cyberspace (Krippendorff 2009, pp. 299–321)

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 200×10^9 bytes, or $200 \times 8 \times 10^9 = 1.6 \times 10^{12}$ 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 $10^9 \times 1.6 \times 10^{12}$ binary choices or store 1.6×10^{21} bits of data.
- Considering the speed of computation, say $1\,\text{GHz} = 10^9/\text{s}$, during one year of continuous processing – 1 year $\cong \pi \times 10^7$ s – we could collectively compute about $1.6 \times 10^{21} \times 10^9 \times \pi \times 10^7 \cong 5 \times 10^{37}$ 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 $\pi \times 10^{82}$, the capacity of a computer of the mass of the earth running for a year 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 $\pi \times 10^{82}$ is practically unachievable, Moore's law is soon doomed. Cyberspace may then become more user-friendly and integrated in everyday life but no longer grow as fast as it is now.

There is of course much happening outside computer technology, not reflected in these numbers. People grow and eat food, drive cars to work, construct buildings and cities, publish, read, and communicate 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 are used to accomplish a great many things, but their production drawings, sales documents, repair records and service instructions are transmitted among the manufacturer's computers. Through registration numbers, insurance and repair records cars occupy cyberspace. Books, newspapers and theatrical performances increasingly are available online. Web pages are read and inform decisions outside cyberspace but their results reenter cyberspace variously. Everything in cyberspace is connected to what I call externalities via their users. These externalities are essential to keeping cyberspace meaningful and alive but they do not add significantly to its estimated size.

Artefacts in cyberspace

Unlike traditional machines, which are designed to serve particular functions, the utility of cyberspace depends on the artefacts with which it is furnished. Electronic artefacts consist of documents, software, and networks that define dependencies among finite numbers of binary variables.

As a matter of definition, *artefacts in cyberspace*

(1) *Occupy space* (in bits of cyberspace) by relating individual bits, for example, the neighbourhood 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 artefacts are manifest are precisely what Ashby had conceptualised as higher-order interactions and hoped to discover and quantify with the ill-fated Q-terms. Artefacts in cyberspace do occupy space, but identifying their structure by observation (observational determinability) is virtually impossible while their structure is easily established by design (synthetic determinability).
(2) *Selectively interact with one another,* form clusters of cooperative ecologies due to interface protocols, common programming languages, or storage in proximity of each other.
(3) *Are preserved* under a variety of recursive transformations (Foerster 1981), for example, during their transmission. Artefacts cannot be experienced at their location but where they are reproduced, in the process of their communication, or while doing actual work. While relatively stable the location of artefacts in cyberspace remains mostly uncertain.
(4) *Can be controlled*, installed, composed, removed, activated, monitored, and terminated by their users (not necessarily by everyone alike).
(5) *Are meaningful in their users' lives* in the sense of being understood and usable (hermeneutic determinability) and relate to the cultural externalities of cyberspace.

Let me mention a few kinds of artefacts by their properties:

The artefacts that determine the size of cyberspace are *physical memories*, hard drives, storage devices, media of communication, and networks. These artefacts do not physically move. The rates of their production less 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 the user *interfaces* with computers, usually part of a computer but doing no work other than providing users access to cyberspace by bridging user cultures with the operation of computers.

Data, textual, visual and sound records, files, and web pages, usually kept as whole documents, are the most common, most space consuming, and least intelligent artefacts. They largely inform individual users about externalities. Lyman and Varian (2003) estimated that most computers hold no more than 1% original data, the remainder are duplicates, representing redundancies in cyberspace. Duplicates replicate traditional mass media products and compete with libraries. Specialised software is first of all data, until it is instructed to cooperate with the operating system and compute data other than themselves, combining them, or connecting them with each other on own or other computers.

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 consumes a considerable amount of cyberspace.

The need for privacy, allocating privileges, and protecting data bases has created *security systems* that organize cyberspace around communities of users with the effect of limiting access to it.

Another increasingly important category of artefacts is *intelligent assistants* or agents that either learn to serve user needs as a function of their habits or can be instructed to assume chores that the user prefers not to undertake or cannot undertake as speedily, reliably, or efficiently as an assistant.

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

One can argue over the categorization of such artefacts but not that they are designed, programmed, or captured to aid users' practices. Except for the self-replicators, which are a nuisance precisely because they cannot easily be controlled, electronic artefacts provide access to possibilities generally not available otherwise. Without a diverse population of artefacts, cyberspace would be an empty shell.

Despite 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 installing software, we must be wary 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 opens to their users and that the other artefacts exhaust by enabling their users to do computational work, communicate with each other, and most importantly, to move among and explore the artefacts in cyberspace.

Human interface capacities

The size of cyberspace exceeds by far several individual capacities, a fact that limits human interaction with computers and how we can operate in cyberspace. Whereas Bremermann's limit concerns physical responsiveness to differences, here I am concerned with the implications of human responsiveness to 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 and are unavailable for perception. Experiments have suggested that human comprehension is about two bits per second or $2 \times \pi \times 10^7$ 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 $2 \times \pi \times 10^7 \times 10^9 \times 10^{-1} \cong 10^{16}$ bits of cyberspace annually.
- Comprehension does not mean responding to every letter, pixel, or option available on *computer screens*. Perception is selective and holistic and what appears on an individual's computer screen necessarily is richer than can be perceived and be responded to. A computer screen with 1280×1024 pixels, 32 bits for colours, 75 Hz refresh rate, observed 10% of a year by one billion computer users would take up not more than $1280 \times 1024 \times 32 \times 75 \times 10^{-1} \times \pi \times 10^7 \times 10^9 \cong 10^{22}$ 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 article), 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 $5.5 \times 5 \times \pi \times 10^7 \times 10^{-1} \times 10^9 \cong 10^{17}$ 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 much as can be comprehended. This may well be the difference between understanding whole words as opposed to individual letters. *Second*, the amount of information that can be displayed must always be 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 colours. We see and think in chunks, not pixels. *Third*, although I do not dare to estimate the cyberspace occupied by all of its artefacts, (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 artefacts enter cyberspace not by individual construction but by being captured by powerful systems, digital cameras, medical imaging, video recorders, and surveillance systems that operate with minimal human involvement. The volume they fill far exceeds human comprehension and ability to enter them bit-by-bit. (c) The majority of artefacts 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 download, copy or transmit an artefact may require very few human actions. The amount actually looked at and individually comprehended is a miniscule fraction of what occupies cyberspace. *Fourth*, artefacts in cyberspace are packages of bits and organised 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 the details of what it does, nor the data it analyses. The volumes searched on the internet remain largely hidden from the user's view.

While cyberspace must be larger than the artefacts it houses, it is perfectly sensible to conclude that the space they collectively occupy far exceeds human comprehension and the human ability of designing them and that their growth expands cyberspace as well. 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, doing things not programmable by any computer language, travelling paths nobody paved for them, and constructing new artefacts to support one's own practices of living and share their use with others to live and co-create that space. As long as these artefacts do not consume the whole cyberspace or prevent access to most willing users, the possibility of human agency is preserved.

8. Conclusion

It should be clear that what we now call cyberspace cannot remotely approximate Bremermann's limit. Much of the earth consists of hot or dull matter 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 – generalising from many cases – nor deductive – deriving knowledge from known theory – but evolutionary – rooted in the idea of the recursion of mutation and selection (Bateson 1972). Ashby (1956, p. 2) defined cybernetics as the study of all possible systems that is informed (constrained) by what cannot be built or found in nature. 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 artefacts in cyberspace – occupying finite spaces but being difficult to localise and no longer identifiable by algebraic accounting equations. They can no longer be identified by observation, but by construction on the part of experts, and by handling them on the part of users. Equating them as packages of bits, created by software companies, programmers, and users, manipulable and useable seems natural to us now. This article has shown, I hope, that creating artefacts in cyberspace goes far beyond the computational resources available to discover their complexities from their outside. The paradigm shift from methods of discovery to methods of design has overcome the computational limits on analysis and reconstructability and outdates the approach taken by earlier systems theories.

Notes on contributor

Klaus Krippendorff, PhD in Communication, University of Illinois Urbana/Champaign, where he studied with W. Ross Ashby; Design Grad., Ulm School of Design, now the Gregory Bateson Term Professor for Cybernetics, Language, and Culture at the University of Pennsylvania's Annenberg School for Communication. He is a Fellow of the American Association for the Advancement of Science, the International Communication Association, the East-West Centre in Hawaii, and Netherlands Institute for Advanced Studies. He is a Past President of the International Communication Association, Chair of the Council of the International Federation of Communication Associations, active in the American Society for Cybernetics, and in other professional organizations. He is on the editorial board of several communication journals. He published widely on cybernetics and systems theory, on methodology in the social sciences, and on human communication theory. Among his books are *The Analysis of Communication Content* (1969) (Co-ed.); *Communication and Control in Society* (1979) (Ed.); *Content Analysis* (1980) (translated into Italian, Japanese, Spanish and Hungarian), 2nd Edition (2004); *Information Theory* (1986); *A Dictionary of Cybernetics* (1986); *The Semantic Turn, a New Foundation for Design* (2006); *On Communicating, Otherness, Meaning, and Information* (2009); and *The Content Analysis Reader* (2009) (Co-ed.). He initiated '*Product Semantics*,' and is the author of a reliability statistics, Krippendorff's alpha. Recent publications have focused on second-order cybernetics, constructivist epistemology, theory of conversation and discourse, critical (emancipatory) theory, and the role of language in the social construction of realities.

References

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.

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., 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., 1969. Two tables of identities governing information flows within large systems. *The American society for cybernetics communications*, 1 (2), 3–8.

Bateson, G., 1972. *Steps to an ecology of mind*. New York: Ballentine Books.

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., 1976. Constraint analysis and structure identification. *Annals of systems research*, 5, 67–80.

Broekstra, G., 1977. Constraint analysis and structure identification II. *Annals of systems research*, 6, 1–20.

Broekstra, G., 1979. Probabilistic constraint analysis of structure identification; An overview and some social science applications. *In*: B. Zeigler, *et al*. eds. *Methodology of systems modeling and simulation*. Amsterdam: North Holland, 305–334.

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.

Conant, R.C., 1976. Laws of information which govern systems. *IEEE transactions on systems, man, and cybernetics*, 6 (4), 240–255.

Conant, R.C., 1980. Structural modeling using a simple information measure. *International journal of systems science*, 11 (6), 721–730.

Conant, R.C., 1981a. Set-theoretic structure modeling. *International journal of general systems*, 7 (1), 93–107.

Conant, R.C., 1981b. How to ignore Bremermann's limit. *In*: *Proceedings of the southeastern regional meeting of the society for general systems research*. Louisville, KY, April 21–23, 1981, 398–404.

Conant, R.C., ed., 1981c. *Mechanisms of intelligence: Ross Ashby's writings on cybernetics*. Seaside, CA: Intersystems Publications.

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, 1979. Cybernetics of cybernetics. *In*: K. Krippendorff, ed. *Communication and control in society*. New York: Gordon and Breach, 5–8.

Foerster, H. von, 1981. Objects; tokens for (eigen)behaviors. *In his: observing systems*. Seaside, CA: Intersystems, 274–285.

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, *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. *In*: *Proceedings of the 22nd annual SGSR meeting*. Washington, DC: February 13–15, 151–155.

Krippendorff, K., 1967. *An examination of content analysis; a proposal for a general framework and an information calculus for message analytic situations*. Dissertation (PhD). Urbana, IL: University of Illinois.

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

Krippendorff, K., 1976. A spectral analysis of relations. Philadelphia, PA: The Annenberg School of Communication, University of Pennsylvania, Mimeo. *Presented to the International Congress of Communication Sciences*, Berlin, May 31, 1977.
Krippendorff, K., 1978. A spectral analysis of relations; further developments. *Paper presented to the Fourth European Meeting on Cybernetic and Systems Research*, Linz, Austria, March 1978.
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* 1979. Delft, The Netherlands, 15 November, 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., 1981. An algorithm for identifying structural models of multi-variate data. *International journal of general systems*, 7 (1), 63–79.
Krippendorff, K., 1982a. Q; an interpretation of the information theoretical Q-measures. *In*: R. Trappl, G.J. Klir, and F. Pichler, eds. *Progress in cybernetics and systems research*. Vol. VIII. New York: Hemisphere, 63–67.
Krippendorff, K., 1982b. On the identification of latent functions in multi-variate data. *In*: R. Trappl, G.J. Klir, and F. Pichler, eds. *Progress in cybernetics and systems research*. Vol. VIII. New York: Hemisphere, 31–42.
Krippendorff, K., 1986. *Information theory; structural models for qualitative data*. Newbury Park, CA: Sage.
Krippendorff, K., 2006. *The semantic turn; a new foundation for design*. Boca Raton, FL: Taylor & Francis/CRC Press.
Krippendorff, K., 2009. *On communicating, otherness, meaning, and information*. *In*: F. Bermejo, ed. New York: Taylor & Francis/Routledge.
Krippendorff, K., 2008. Four (in)determinabilities, not one. *In*: J.V. Ciprut, ed. *Indeterminacy; the mapped, the navigable, and the uncharted*. Cambridge, MA: MIT Press.
Lyman, P. and Varian, H.R., 2003. *How much Information? 2003* [online]. Berkeley, CA: School of Information Management and Systems. Available from: http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm#summary [Accessed 20 May 2008].
McGill, W.J., 1954. Multivariate information transmission. *Psychometrica*, 19, 97–116.
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.