Shannon Entropy and Productivity: Why Big Organizations Can Seem Stupid

Dr. Rich Janow

February 28, 2004

Department of Physics
New Jersey Institute of Technology
Newark, NJ 07102-1982
janow@njit.edu

and

The Analytic Solutions Group, LLC
514 North Wyoming Avenue, South Orange NJ 07079
973-762-4987
janow@att.net

Abstract:

There is a fairly common perception that large organizations tend to behave much less “intelligently” than their size suggests. They often lose the decisiveness seen in small groups and may seem "stupid" to people within them who work with ideas and knowledge. This paper offers the first quantitative basis for these observations. It adapts some elements of Shannon's information theory to organizations by defining "organizational entropy" and related notions of decision complexity and productivity. Organizational entropy measures extra decision information needed when partitioning functions onto a structure.

For an idealized model organization of \( n \) decision nodes the \( \text{per capita} \) limit on the sustainable decision rate is proportional to \( 1/\log_2(n) \). Additionally, a distinct saturation effect with a higher threshold switches on when individual's decision throughput limits are reached.

Quantitative tools for managing entropy and productivity in business firms and in command and control situations are feasible as practical outgrowths of this work. They involve conscious restructuring designed to limit entropy effects.

Keywords:

Management, information theory, organizational entropy, decision networks, social networks, business reengineering, technology management, decision complexity, command and control, C^3I, entropy-based warfare, knowledge management, complex systems, control theory applications.
Abstract: ............................................................................................................................................ 1
Keywords: ......................................................................................................................................... 1
1. Introduction .................................................................................................................................... 3
2. A model for decision processes ........................................................................................................ 4
   2.1. Symbol transmission in information theory ................................................................................... 4
   2.2. Management decisions and "dits" .................................................................................................... 4
   2.3. The upper limit M(n) on the maximum management decision flow rate: ...................................... 4
   2.4. The maximum "dit" capacity R(n) .................................................................................................. 5
   2.5. Organizational entropy and the decision complexity A(n) .............................................................. 5
3. Approximations ................................................................................................................................... 6
   3.1. A form factor for organization structure .......................................................................................... 6
   3.2. An exactly solvable model: the fully connected organization ...................................................... 7
   3.3. Sparsifying topology, limiting choice (partitioning into sub-units) ............................................... 8
4. Applications ....................................................................................................................................... 8
References: ......................................................................................................................................... 9
Appendix A: The choice of a definition for M(n) ................................................................................. 10
Appendix B: The smallest fully connected organization .................................................................... 10
About the author: ............................................................................................................................... 11
1. Introduction

Knowledge managers in large organizations often come to feel a limit on their efficiency and effectiveness that they attribute to the organization's size. For example, an individual may ask his firm to assimilate his work and make decisions based on it, but he/she may have to wait an inordinately long time for a response - even when the issues to be decided seem straightforward. People may become frustrated by these perceived impediments and assign them to bureaucracy, resistance to fresh initiatives, turf protection, complacency, etc.

These phenomena may actually, though, be the result of individuals' own decision processing over-running the organizations' collaboration channels' ability to keep up with them. The sluggish response may become more pronounced with growth. Small groups of people sometimes self-organize and bypass the structure to gain efficiency and avoid frustration. The knowledge managers and customers who find that an organization cannot keep up with them may conclude simplistically that it (the organization) is "stupid". When a small firm merges into a large one, culture shock often results.

Decision-intensive tasks are often integral to an organization's product as well as providing control of the organization itself. These are all lumped together and called "management decisions" in the discussions that follow.

Organizations are a type of control system in which the decision network (perhaps several of them) has humans as nodes [1]. The individual decision-makers use the control network they are in as a communication channel: they insert management decisions into it and expect to eventually receive decision information back from it. The network has some maximum decision flow rate it can support; if knowledge managers try to exceed it their work may be dropped altogether and responses may be delayed or inappropriate.

Conventional communication channels have similar capacity limits and this prompts the speculation that an analogy may be useful. The key events for communication systems are the transmission of symbols (such as letters of the alphabet) rather than decisions. Claude Shannon [2] developed information theory for communication systems and perhaps most notably recognized and quantified entropy as information.

Organizational entropy is defined and introduced below in analogy with the Shannon entropy; it contributes to the average collaborative decision complexity (choice) just as entropy for a communication channel measures the average information content (bits per symbol) of symbols transmitted over it. Organizational entropy defined this way increases as a decision network grows, if the complexity of the tasks themselves remains unchanged. As organizations grow, often having to take on increasingly complex tasks, the decision structure adds nodes and partitions functions among more decision makers in order to "divide and conquer". That added structure increases the network entropy and adds to the overall decision complexity, unless the tasks themselves become simpler or more specialized.

Importantly, the entropy grows fast enough to more than offset growth in the total of individuals' capacities for making basic binary decisions. There appears to be a fundamental upper limit on the total management decision flow rate. That limit grows slower than linearly with the number of nodes in an organization, and so the maximum per capita management decision flow rate actually shrinks as the number of decision-makers in the network grows.

The intended inference is that the widespread, anecdotal perceptions mentioned earlier may be symptoms that a shrinking limit on per capita decision throughput is reached and perceived as impaired productivity. The limit is intrinsic to large control networks. It is emphatically not related to congestion on physical data networks that may be present: the effect has in principle been operating as long as humans have formed groups and divided up roles.

These suggestions clearly need experimental validation and experience with practical applications to become useful and accepted. They are potentially important for any organization dominated by complex cognitive tasks and teamwork. Businesses that primarily manage knowledge and create intellectual property may need to restructure to become more efficient and competitive in their markets. Defense applications include improving command and control efficiency for friendly assets or understanding how to impede it (entropy-based warfare) for opposing forces [3].
2. A model for decision processes

2.1. Symbol transmission in information theory

Claude Shannon [2] showed how to quantify the information contained in members of some symbol set (such as letters of the alphabet) that occur with particular frequencies. The entropy \( H \) of an information source or sink measures the average information content of each symbol transmitted and is expressed as the number of bits per symbol needed when using the optimal compression coding scheme. Information is intimately related to the element of choice, as is the entropy of physical systems.

Entropy pertaining to symbol transmission has the same characteristic functional form as physical entropy:

\[
H = - k \sum_{j=1}^{n} p_j \log_2(p_j)
\]

A communication channel's capacity \( C \) is the maximum rate for transmitting raw bits/second, determined by physical characteristics of the channel and modulation schemes. The Fundamental Theorem for a Noiseless Channel [2] showed that there is a maximum rate \( S \) for symbol transmission that is simply \( C \) divided by the entropy \( H \):

\[
S \equiv \frac{C}{H}
\]

A source that tries to go faster will over-run the channel, resulting in errors and lost information.

2.2. Management decisions and "dits"

Management decisions are composed of many independent choices - much as communication symbols are composed of many bits. In principle, management decisions can be reduced to binary decision elements [1]. The quanta of actionable decision information are analogous to the "bits" of classical information theory, and so it's natural to call them "dits" for short to emphasize the parallel.

Beer [1] noted that a dichotomous classifier can in principle generate the binary representations for decisions. A human (or a coding device) might parse statements to find actionable content, develop a set of (decidable) symbolic logic statements, map them onto a binary tree, and then code them as a string of "dits". In general, the task of building such binary representations would be tedious and subjective unless it can be automated.

There is no direct relationship between the way a decision is represented via a sequence of "dits" and representations of the decision as symbol and bit strings for transmission. Dits are "actionable", not just perceptual. For example, bitmap graphics and the like may require a huge number of bits but have little or no actionable content. A message authorizing a military attack might be just a few bits long but have a huge decision complexity (many "dits").

The complexity of management decisions (i.e., their average length in "dits") says nothing at all about their importance to the organization. It measures ultimately the amount of thought decisions require, which may or may not correlate with impact.

2.3. The upper limit \( M(n) \) on the maximum management decision flow rate:

The quantity analogous to Shannon's upper limit on symbol transmission is a function \( M(n) \) that represents an organization's maximum total management decision flow rate. An organization is regarded as a network of \( n \) decision nodes (knowledge workers) that create, consume, and communicate actionable information related to the organization's activities. \( M(n) \) is simply the quotient:

\[
M(n) = \frac{R(n)}{A(n)}
\]

The function \( R(n) \) measures the maximum dit flow rate (in dits/unit time) for the entire organization - analogous to the Shannon channel capacity \( C \).

The function \( A(n) \) represents the average collaborative decision complexity (with dimensions dits/decision). Its value depends on the inherent complexity of collaboration tasks themselves, and also on an entropy component due to the decision network structure. The entropy component fits an analogy with Shannon's entropy.

The per capita limit \( \mu(n) \) on the management decision rate is simply \( M(n) \) divided by \( n \), viz:

\[
\mu(n) = M(n) \frac{n}{n} = \mu(n) \quad (1b)
\]

---

1 This is analogous to compression of a symbol set in information theory.
2.4. The maximum "dit" capacity \( R(n) \)

Each of the \( n \) people in an organization can potentially have a decision path to any of the remaining \( n - 1 \) others. So \( R(n) \) scales as \( n(n-1) \), and it approximates \( n^2 \) for large \( n \). For an individual node \( i \), \( R_i(n) \) represents the maximum capacity involving as many as \( n-1 \) destination nodes. Each such linkage is represented by \( R_{ij} \rho_{ij} \) (with dimensions "dits/unit time).

- The factors \( R_{ij} \) are maximum (not expected value) dit rates flowing between pairs of decision-makers and might vary for each path. They depend on the people at either end of the paths, who are filling the role that codecs (coder-decoders) play in hardware communications.
- The coefficients \( \rho_{ij} \) toggle the flow on or off for each path and thus specify the topology of the organization: they range from 0 to 1. If for example all paths are fully open all the time, then \( \rho_{ij} = 1 \) for all values of the indices.

The maximum "dit" capacity is a sum over all the sources and destinations:

\[
R(n) = \sum_{i=1}^{n} R_i(n)
= \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \rho_{ij} R_{ij} \leq \sum_{i=1}^{n} D_i \tag{2}
\]

A detailed evaluation of this function for a large organization may require a complete social network map of who is connected to whom with values for the dit rates.

The extreme right hand side of Equation (2) portrays a saturation limit. Individual collaborators can become saturated if the maximum rates \( D_i \) (in dits/unit time) at which they can make or assimilate basic binary decisions are exceeded. Those rates depend on personal and resource characteristics. A binary dit rate \( D_i \) might be spent on a few complex or many simple management decisions.

The right hand inequality above becomes the upper limit on the "dit" rate when all of the nodes in an organization reach saturation (too many paths operating at full flow rate, for example). \( R(n) \) then scales proportionally to \( n \) rather than \( n^2 \).

If saturation is reached a dramatic collapse of productivity follows (discussed below). It may happen for example through down-sizing or if an organization grows very large but does not restructure to limit the paths and dit rates per knowledge worker. To avoid saturation the workload must be partitioned onto a larger staff (increasing \( n \)) accompanied by structural change.

2.5. Organizational entropy and the decision complexity \( A(n) \)

The connection between entropy, information, and choice has been known for a long time and it follows that decision processes should have an associated entropy. As early as 1894, physicist Ludwig Boltzmann [4] observed that the entropy of a physical system is related to "missing information" inasmuch as it counts the number of alternative ("degenerate") microscopic states of a physical system that might be chosen consistent with a single macroscopic (observable) state. The entropy grows with the size of the phase space (microscopic system state) volume that a system can occupy. When all phase space cells are equally probable the entropy is a maximum. The system is then highly disordered and it takes a lot of information to specify which of the microscopic states it is in. By contrast, the state of a highly ordered system (say, a solid at absolute zero temperature) has low entropy and takes comparatively little information to specify.

When Shannon quantified information for communications the notion of choice was a key ingredient. The freedom exercised when choosing a symbol from a symbol set led to a definition of information using a function identical to that of statistical entropy. Whenever a symbol sequence is highly predictable the choice is small and very little information is conveyed.

Choice applies also to destinations for actionable information in an organization. Choice grows as tasks grow in complexity and are partitioned to more nodes, increasing the amount of collaboration. Organizational entropy measures choice ("degeneracy") in the number of collaboration states. A subset of the possible states is being used at any one time on particular knowledge-intensive tasks.

When the range of decision network states is large so is the entropy; the organization is then also something of a general-purpose tool. Conversely, when entropy is small, the organization will probably do a prescribed set of specialized tasks efficiently and others not at all. Like physical entropy, organizational entropy is an extensive quantity: it grows with the size of the decision network.

The price of having the capability to execute a wide range of complicated, multi-person tasks may be a large organizational entropy. That broad capability may impair efficiency when doing simple tasks for which a multi-purpose structure is over-kill.
The overall decision complexity \( A(n) \) is the average number of "dits" per collaborative management decision task. It is a sum over contributions \( A_i(n) \) from each of the \( n \) decision nodes.

Each contribution is the product of the average complexity \( A_{0i} \) (pertaining to tasks that node \( i \) performs) and a factor \( H_i \) that measures the entropy of the decision network that node \( i \) sees. The entropy measures structural choice information (in bits). \( A_{0i} \) accounts for the content in dits per decision per bit of structure information.

The entropy is found by retracing the thread of Hamming's [5] discussion of Shannon entropy for a communication channel. The \( i \)'th person chooses one of the remaining \( n-1 \) people in the organization every time he issues a decision to the network. Let \( p_{i}(j) \) be the conditional probability that person \( i \) chooses person \( j \) as his target. If the destination is known in advance there is no surprise and no information is needed to know whom person \( i \) collaborates with. One of the \( p_{i}(j) \) is then equal to 1 and all the other probabilities are zero. If the probabilities are all equal (and small if \( n \) is large) then the "surprise" when one recipient is actually picked is at its maximum and that choice carries significant information. The information needed to pick a recipient is thus related to the inverse of the probability \( p_{i}(j) \) for making that choice.

When two decision recipients are chosen independently, the information associated with the joint event should simply be the sum of the information for each separately, viz:

\[
\text{Info } [(s_i, s_j) \text{ and } (s_i, s_k)] = \text{Info } (s_i, s_j) + \text{Info } (s_i, s_k)
\]

In the above, the notation \((s_i, s_j)\) represents the event in which source \( i \) chooses recipient \( j \).

A logarithmic function\(^2\) satisfies all these requirements, viz:

\[
\text{Info } (s_i, s_j) = \log_2 (1/p_{i}(j)) = -\log_2 (p_{i}(j))
\]

Zero information is involved when there is no surprise inasmuch as \( \log_2(1) = 0 \).

The expected value for the entropy \( h_{i}(j) \) (information in "dits") associated with the pair \((s_i, s_j)\) is just the conditional probability of choosing \( j \) multiplied by the information associated with the choice, i.e.,

\[
h_{i}(j) = -p_{i}(j)\log_2 (p_{i}(j))
\]

The entropy \( H_i \) for all the destinations that person \( i \) communicates with is just the sum of \( h_{i}(j) \) over \( j \).

\[
H_i = - \sum_{j=1, j\neq i}^{n} p_{i}(j)\log_2 (p_{i}(j))
\]

This is identical in form to physical entropy expressions and to Shannon's information expression [2]. After summing on source nodes, the result for the decision complexity is:

\[
A(n) = \sum_{i=1}^{n} A_i(n) = \sum_{i=1}^{n} A_{0i} H_i
\]

\[
= - \sum_{i=1}^{n} A_{0i} \sum_{j=1, j\neq i}^{n} p_{i}(j)\log_2 (p_{i}(j))
\]

The complexity coefficients \( A_{0i} \) factor out and can be replaced by a single average value \( A_0 \). The elementary dit capacities \( R_{0i,j} \) likewise factor and are replaced by an average \( R_0 \). Equations (1), (2), and (3) simplify when an organization can be approximated by a set of decision-makers having identical capabilities and collaborating on tasks of the same inherent complexity. The complexity coefficients \( A_{0i} \) factor out and can be replaced by a single average value \( A_0 \).

\[
M(n) = M_0 \frac{\sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} p_{i,j}}{\sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} p_{i}(j)\log_2 (p_{i}(j))}
\]

where \( M_0 = \frac{R_0}{A_0} \)

\[
\log_2 (x) = \log_e(x)/\log_e(2) = \log_e(x)/.6931
\]

---

\(^2\) With base 2 logarithms, the logarithm of a number is simply the number of binary integers (bits) needed to represent it. One can switch between base 2 and natural (base "e") logarithms without losing generality, apart from the constant 0.6931 which is the natural logarithm of 2., viz:
The ratio \( M_0 \) is the maximum management decision flow rate per collaboration channel (below saturation). To the right of \( M_0 \) is a dimensionless form factor that depends purely on the organization structure and entropy.

With the structure effects separated, equation (4) can be used for comparisons that reflect organization structure alone, and absolute measurements of the coefficients are not essential.

### 3.2. An exactly solvable model: the fully connected organization

A further approximation is the "fully connected model", which corresponds to setting \( p_{ij} = 1 \) for all \( i \) and \( j \) in Equation (4). It assumes that paths between all pairs of knowledge workers are open and have equal weighting, and that all decision-makers collaborate with equal probability. The maximum dit rate \( R(n) \) reduces to \( R_0, n(n-1) \), which approaches \( R_0, n^2 \) for large \( n \). All of the conditional probabilities \( p_{i}(j) \) are equal to \( 1/(n-1) \). The organizational entropy becomes \( H = n \log_2(n-1) \) while the decision complexity becomes \( A(n) = A_0 H \).

This model can be solved exactly as all terms in the summations are the same, so the sums become trivial. But it tends to overstate the entropy effects. In large organizations this model can most realistically be applied to individual functional units or processes that are then sparsely linked to each other.

The \( n \log_2(n) \) expression for the entropy can be arrived at another way: that expression is also the combinatorial complexity of the most efficient general method for sorting \( n \) objects [6]. The decision network in this model can alternatively be viewed as a sorting machine for management decisions each of which has complexity \( A_0 \). The entropy is thus proportional to the sorting time.

The maximum sustainable decision rate \( M(n) \) for an organization below the size \( n_S \) that triggers saturation is:

\[
M(n) = M_0 \cdot \frac{n-1}{\log_2(n-1)} \quad \text{for large } n \leq n_S
\]

This result grows sub-linearly; that is, more slowly than \( n \). It makes intuitive sense that people work fastest (although not necessarily most effectively) on a task when they can proceed autonomously on their own pieces of it. When there is a synergy or scale economy to be gained by increasing collaborative coupling, it will show up as simplification of individuals’ tasks: i.e., as a reduction of the decision complexity coefficients \( A_i \).

Parallel processor arrays obey a superficially similar rule called "Amdahl's Law", which sets an upper limit on the speedup ratio [7] achievable by adding processing nodes. When an algorithm is completely parallelizable the speedup can grow linearly as processors are added, since the nodes never need to wait while another is making decisions. Otherwise the speedup ratio is less than 1. Amdahl’s Law, though, does not apply entropy to a processor network, so the analogy is limited.

The per capita maximum decision rate \( \mu(n) = M(n)/n \) is an organization’s productivity limit for knowledge-intensive tasks. It declines as \( 1/\log_2(n) \) for a system below the saturation threshold.

Table (1) shows some relative values of \( M \) and \( \mu \) for a range of organization sizes. For example, a one million-person organization can utilize only about 50,000 times as much management decision capacity as a single individual, neglecting saturation and assuming it is fully connected (the worst case).

<table>
<thead>
<tr>
<th>N</th>
<th>Decision Rate N-1/log2(N-1)</th>
<th>Per Capita Factor 1/log2(N -1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.00</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>1.89</td>
<td>0.63</td>
</tr>
<tr>
<td>5</td>
<td>2.00</td>
<td>0.50</td>
</tr>
<tr>
<td>7</td>
<td>2.32</td>
<td>0.39</td>
</tr>
<tr>
<td>10</td>
<td>2.84</td>
<td>0.32</td>
</tr>
<tr>
<td>100</td>
<td>14.90</td>
<td>0.15</td>
</tr>
<tr>
<td>500</td>
<td>56</td>
<td>0.11</td>
</tr>
<tr>
<td>1,000</td>
<td>100</td>
<td>0.10</td>
</tr>
<tr>
<td>10,000</td>
<td>753.00</td>
<td>0.0753</td>
</tr>
<tr>
<td>100,000</td>
<td>6,020</td>
<td>0.0602</td>
</tr>
<tr>
<td>200,000</td>
<td>11,356</td>
<td>0.0568</td>
</tr>
<tr>
<td>500,000</td>
<td>24,409</td>
<td>0.0528</td>
</tr>
<tr>
<td>1,000,000</td>
<td>50,171</td>
<td>0.0502</td>
</tr>
</tbody>
</table>

\[
\log_2(x) = \log_2(x)/\log_2(2) = \log_{10}(x)/.6931
\]

When an organization is "saturated" the total dit rate is limited by \( nD_0 \), where \( D_0 \) is an average node's own maximum internal dit capacity. For fixed \( R_0 \) the number of nodes \( n_S \) that initiates saturation, along with organizational thrashing and productivity implosion, satisfies \( D_0 = R_0, n_S - 1 \). Above saturation the total decision rate \( M(n) \) actually falls as \( 1/\log_2(n) \) with further growth, due to the increased structural information that increases entropy, viz:
The per capita maximum decision rate $\mu(n)$ declines as $1/n \log_2(n)$ in the saturated regime - a factor of $n$ faster than before.

Figure (1) plots $\mu(n)$ for organizations with populations from 3 nodes up to 1 million nodes. Small organizations growing from a few individuals to about 1,000 experience a ten-fold fall-off in their per capita productivity that should be highly noticeable. Further growth (without saturation) from 1000 to 1 million knowledge managers would reduce productivity by only another factor of 2 - smaller but still with significant economic impact. A person from a small startup firm that is acquired by a large one would feel a sudden culture shock.

Productivity collapses markedly if saturation is reached, as plotted in figure (1) for a range of choices for $n_s$. In a real organization saturation would be reached less suddenly than shown.

3.3. Sparsifying topology, limiting choice (partitioning into sub-units)

The fully connected architecture maximizes entropy. "Sparsifying" the decision paths reduces it; i.e. it subdivides an organization into subunits and lets a small fraction of the people do most of the communication between them. The entropy denominator decreases faster than the "dit capacity" numerator in Equation (1a).

Managers of large organizations often seem to intuitively understand this principle. They try to control decision complexity by subdividing into weakly coupled business units or non-hierarchical process teams, sometimes declaring the intent to emulate "small firm environments". Only a small fraction of the knowledge managers interact across unit boundaries.

Sparsification may not gain as much efficiency as is hoped for if the people who handle external decision interfaces exceed their "saturation" limits $D_i$. This may happen easily. One estimate $[8]$ of a typical human's conventional communication bandwidth capacity is 50 bits/sec - about 250 English words per minute\(^3\). A human's maximum "dit" rate $D_0$ is likely to be much smaller.

\[ M(n) = M_0 \frac{n_s - 1}{\log_2(n - 1)} \]

(5b)

\[ \approx M_0 \frac{n_s}{\log_2(n)} \text { for large } n > n_s \]

4. Applications

One way of applying this work to re-engineering is to use the simple modeling results as heuristics in conjunction with a set of rules. Some strategies for improving knowledge managers' productivity that are consistent with these results have been used intuitively, but in an ad hoc way without benefit of a quantitative rationale:

- Reduce choice by creating specialized, dedicated organizations whenever they can be justified. Define and automate workflows so that they follow customer-centric processes and cross "silo" boundaries.
- Match organization size and structure to the complexity of the task.
- Use information hiding aggressively. Knowledge managers should work at a level of abstraction where many low level decisions are not seen.
- Use knowledge management systems to avoid "reinventing the wheel", especially in large firms.
- Alter the rewards system so that managers have an incentive to lower entropy. Decouple pay from the number of people supervised.

As experience provides knowledge of the model's coefficients, quantitative tools may emerge to make tuning organizations for high performance a much less chancy and ad-hoc process. For example, the metrics introduced here may be used as components of cost functions in linear programming methods.\(^4\)

In commercial markets efficiency in managing knowledge increasingly determines competitive advantage. Bigness can be a dis-economy of scale if large, functionally diverse, general-purpose organizations compete with focused, low entropy firms in markets where intellectual labor costs drive product economics. Where traditional scale economies (like manufacturing or distribution) are dominant, knowledge managers' productivity has little leverage on bottom line competitiveness and complex, high entropy organizations may nonetheless enjoy critical strategic advantages.

Defense applications include improving the speed and productivity of command and control, intelligence and sensor fusion, and real-time battle management. These functions all depend on rapid execution of many complex decisions by networks of skilled human decision-makers, who are a limited resource.

---

\(^3\) Shannon and others estimate about 2.6 bits/letter and about 4.5 letters/word for English.

\(^4\) U. S. patent pending
In "entropy-based warfare" [3] the goal is the opposite one of deteriorating all these functions for the enemy; extensions of this work may help to countermeasure the effectiveness of combat or terror organizations.

Network entropy probably applies to any control system, not only those consisting of human decision elements. Examples may include whole societies and neural networks.

References:

6. Aho, Alfred V., John E. Hopcroft, & Jeffrey D. Ullman: "The design and Analysis of Computer Algorithms", Addison Wesley, 1974, page.77. The sorting literature most often uses base 2 logarithms, which differ from natural logarithms by a factor of \(\log_2(2) \approx .6921\).
8. Lucky, Robert W. "Silicon Dreams: Information, Man and Machine", St. Martin's Press, NY, 1989, page 33. Here is the quote: "After Claude Shannon and others conceived the principles of information theory in the late 1940s, a number of studies were conducted to determine the channel capacity of a human being...It seems that a human being -- you and I lest there be any doubt -- is only good for about 50 bits per second of input or output. That is all the information that we are capable of taking in or putting out..."
Appendix A: The choice of a definition for $M(n)$

An organization is viewed abstractly as a decision channel whose throughput is limited by the maximum management decision flow rate $M(n)$, defined simply as the total maximum bit capacity $R(n)$ divided by the total decision complexity $A(n)$. $R(n)$ and $A(n)$ are both always positive and they grow as $n$ grows. $M(n)$ has dimensions of decisions/unit time.

The defining equation for $M(n)$ is written below. On the far right hand side the summation over source nodes highlights a point: each term $R_i(n)/A(n)$ in the sum is the maximum (average) decision rate seen by the $i$'th node with the total decision complexity $A(n)$ as its denominator:

$$M(n) \equiv \frac{R(n)}{A(n)} = \sum_{i=1}^{n} \left[ \frac{R_i(n)}{A(n)} \right]$$

The entropy of the entire organization limits the decision rate at node $i$, not just the contribution due to decisions at node $i$ alone which is smaller by roughly a factor of $n$. This makes intuitive sense inasmuch as the individual nodes are working on pieces of tasks that were distributed to the nodes.

An alternative definition that was rejected in favor of the above was:

$$X(n) \equiv \sum_{i=1}^{n} \left[ \frac{R_i(n)}{A_i(n)} \right]$$

In this summation, each term is a decision rate for the $i$'th node that includes only the node's own decision complexity $A_i(n)$ in the denominator. If this were the definition of $M(n)$, the upper limit on per capita decision rates would grow with organizational size and complexity rather than decrease; i.e., we would observe synergy rather than antisynergy as the result of growth.

Such results would be a priori absurd and contrary to experience: decision-making would become faster and simpler as the tasks and the organization grow more complex. The results would also be inconsistent with a related common-sense boundary condition called Amdahl's Law [7], which simply says that an array of $n$ parallel processors working on pieces of a problem cannot speed things up by more than a factor of $n$. Normally, the speed-up factor is much less than $n$ on account of inherently serial points in a problem's algorithms - often points where decisions must be made.

The largest ($n$-fold) gain is realized when pieces of a problem can be solved completely independently. For organizations working on tasks that have been partitioned onto decision-makers, the same limit and logic applies.

Appendix B: The smallest fully connected organization

Suppose there are just three people in an organization with all of the connections open and equally weighted. As a result, $p_{i,j} = 1$ for each of the $n(n-1) = 6$ terms in the double summation of Equation (2). The maximum dit rate is: $R(3) = 6R_0$ dits/unit time.

The decision complexity is calculated using Equation (3). Any one of the 3 people can select two others as recipients. If all have equal probability of being chosen, $p_{i,j} = \frac{1}{2}$ and $\log_2(p_{i,j}) = -1$ and thus $\bar{H}_i = 1$; there is just one dit (or bit) of information in this choice since there are 2 destinations. The total organizational entropy given by Equation (3) is $H(3) = 3$, since there are 3 information sources.

The decision complexity $A(3) = 3A_0$ dits/decision. The maximum decision rate is $M(3) = 2M_0$ decisions/unit time, using the results above for $R(3)$ and $A(3)$.

Three persons is the smallest organization that can be treated using Equation (1). If $n = 2$ there is no choice of topology for mapping problems onto nodes, $p_{i,j} = 1$, the entropy function becomes zero, and Equation (4) fails to be mathematically well behaved.

Three person organizations might be used as an experimental tool to evaluate the coefficient $M_0$: it is simply one half of the maximum decision rate.
About the author:

Rich Janow is the principal of The Analytic Solutions Group, LLC, consulting with government, industry, and academic organizations. He likes working on cross-disciplinary problems at the intersection of R&D, information systems, business and technology strategy, physical science, and disruptive technology. He is a faculty member in the Applied Physics Department at New Jersey Institute of Technology.

Rich's association with advanced technology markets and strategy, roadmaps, assessment, technological forecasting, and futures includes 18 years at Bell Laboratories, 4 years as an executive in a high technology company, and applied research in computer science, condensed matter and surface physics. He holds a Ph. D. from CUNY and an A. B. degree from Columbia College, both in Physics, and resides in South Orange, New Jersey.

He may be reached at janow@att.net, janow@njit.edu or by phone at (973) 762-4987.