NOTE TO FILE:
Garvin H Boyle
Dated: R5:180214

Entropy and Units of Measure in a Histogram

Frontispiece

Based on Shannon:

$${}_S S = {}_S C \times \left[-\sum_{i=1}^{k}(p_i \times \ln[p_i])\right]$$

Based on Boltzmann:

$${}_B S = {}_B C \times [k_B \ln(\Omega)]$$

Boltzmann in terms of Shannon:

$${}_B S \approx \left[\frac{{}_B C}{{}_S C}\right] \times A \times {}_S S$$

	bit/Sh	nat/nit/nepit	dit/decit/ban/Hart
bit/Sh	1	0.69315	0.30103
nat/nit/nepit	1.44270	1	0.43429
dit/decit/ban/Hart	3.32193	2.30259	1

There are three measurement systems, based on $Log_2()$, $Ln_e()$ and $Log_{10}()$.

K =	5					A =		
a_i =	74	218	212	79	43	626		Relative
Probabilities $p_i = a_i / A$ =	0.11821	0.34824	0.33866	0.12620	0.06869	1.00000	= Sum(p_i) = 1	Sizes
Surprisals in Bits	0.36416	0.52997	0.52902	0.37686	0.26540	2.06540	= S In Bits	1.00000
Surprisals in Nats	0.25241	0.36735	0.36669	0.26122	0.18396	1.43163	= S In Nats	0.69315
Surprisals in Dits	0.10962	0.15954	0.15925	0.11345	0.07989	0.62175	= S In Dits	0.30103
$\ln(a_i!)$ =	247.573	959.512	927.275	269.291	121.533	1.41225	= S in HNats	0.68377
Base for Bits =	2							
Base for Nats =	2.71828							
Base for Dits =	10							

The equation for conversion from nats to hnats is dependent upon A and is:

$$S_{hnats} \approx (1.0041 * (A * S_{nats}) + 8.5)/A$$

The equation for conversion from hnats to nats is:

$$S_{nats} \approx \frac{A \times S_{hnats} - 8.5}{A * 1.0041}$$

Table of Contents

1 - References

A. A. Drăgulescu and V.M. Yakovenko (2000) "Statistical mechanics of money", Eur. Phys. J. B 17, 723-729.
B. 140218 Email from Yakovenko R1.pdf
C. 140409 NTF Discussion With Dr Yakovenko R1.pdf
D. 150527 PPR - Definition of EI R17.pdf
E. 180118 NTF Custom Functions in Excel R3.pdf
F. 180210 NTF Shannon Vs Boltzmann R3.pdf
G. 180123 Arrhenius equation – Wikipedia.pdf
H. 180123 Fenske equation – Wikipedia.pdf
I. 180123 Nernst equation – Wikipedia.pdf
J. 160214 Table of Dimensional Formulas.pdf
K. 150312 Dimensional Analysis.pdf
L. www.phys.ufl.edu/courses/phy2049/sum13/Dimensional-2049-Ramond.pdf
M. 180121 Shannon (unit) – Wikipedia.pdf
N. 180121 Nat (unit) – Wikipedia.pdf
O. 180121 Hartley (unit) – Wikipedia.pdf
P. 180212 XLS Units of informational entropy R4.xlsm
Q. Cohen and Stewart (1994) "The Collapse of Chaos: Discovering the Simplicity in a Complex World", Penguin Books.

2 - Background

This series of diary notes is a rework of a set of notes partially completed in 2013-2014. In 2000 Dr Victor Yakovenko and his student (Drăgulescu) published a set of eight capital exchange models which have come to be known as the BDY model (for Benatti-Drăgulescu-Yakovenko). Later, on his website, Dr Yakovenko had produced a demonstration of rising entropy in his BDY models, and I decided to do the same in my own agent-based models (ABMs). This was all with the goal of understanding the role of the Maximum Entropy Principle (MEP) and Maximum Entropy Production Principle (MEPP) in ABMs such as ModEco or PSoup. That study and the associated diary notes were set aside for a while as I studied Odum's MPP. Now, in 2018, I want to review, update, and complete this set of diary notes.

Ref A is the paper in which I first saw the BDY models described, and Ref B is a series of email in which I discuss entropy with a friend, with comments from Dr Yakovenko. I had opportunity in April 2014 to meet with Dr Yakovenko in Toronto, and Ref C captures my discussion with him, and a set of thoughts that occurred to me over the four hours as I drove home again.

Ref D is a draft paper in which I develop a measure of entropy as exhibited in operating ABMs similar to the BDY models. This series of NTFs is being (re-)written in support of that document.

Ref E is a technical note describing how to write a custom function in MS Excel. This skill is needed to pursue a study of entropy in ABMs.

Ref F is a note in which I use the combinatorial multinomial coefficient and the basic form of "Stirling's Approximation" of ln(A!) to derive Shannon's equation for entropy [1] from Boltzmann's equation [2].

Based on Shannon:

$$ {}_{S}S = {}_{S}C \times \left[-\sum_{i=1}^{k} (p_i \times \ln[p_i]) \right] \tag{1} $$

Based on Boltzmann:

$$ {}_{B}S = {}_{B}C \times [k_B \ln(\Omega)] \tag{2} $$

Boltzmann in terms of Shannon:

$$ {}_{B}S \approx \left[\frac{{}_{B}C}{{}_{S}C} \right] \times A \times {}_{S}S \tag{3} $$

Where ${}_{S}C$ and ${}_{B}C$ are dimensionless scaling factors each associated with what I am calling the "Shannon regime" of equations, and the "Boltzmann regime". In the Boltzmann regime, ${}_{B}C$ would be Boltzmann's constant. I am not sure what ${}_{S}C$ would be in the Shannon regime. In Ref F I showed that these two expressions for entropy are related by equation [3].

The approximation (symbolized as ≈) arises from the use of Stirling's approximation to convert Boltzmann's formula into Shannon's formula. One confusing aspect of the Ref F note is the units of measure for the two formulae for entropy.

3 - Purpose

The purpose of this diary note is to investigate issues around units of measure for entropy, so that I can better interpret the meaning of my entropy calculations.

4 - Discussion

4.1 - Why Explore Units of Measure of Entropy

After many years of consideration, I have come to believe that entropy is always a dimensionless number. However, it was not immediately clear to me that this was the case. When preparing the Ref F note the intrusion of the factor "A" into equation [3] was a surprise. It raises a question about possible conflict of dimensionality. And, certainly, the text books and published papers that I have read do not say that entropy is dimensionless, and, to the contrary, imply otherwise. So I will explore and record the thoughts that have led me to that conclusion.

4.2 - What is "Dimensionality"?

Dimensionality is an abstract concept that underlies all empirical measurements and all theoretical mathematical formulae in the hard sciences. Every number or variable used in a formula can have as many as three different parts: a numeric part, an indication of error, and a unit of measure. For example, the area of my patio is about 5.5 square meters, more or less, and we might write this as 5.5 [± 0.5] m^2. 5.5 is the numeric part. [±.05] is the estimated error. And m^2 is the units of measure. The dimensionality of this tripartite number is abstracted from the part called the "units of measure".

Dimensionality is a concept best understood by physicists and mathematicians. In mathematics, you cannot add numbers with different units of measure. For example, 1 centimeter plus 2 meters cannot be added until you choose a common unit of measure. The problem is solved by converting one unit of measure (centimeters) to the other (meters) and using a common unit. The answer is either 2.01 meter, or 201 centimeters. But dimensionality is more fundamental than that. Suppose the numbers are 1 centimeter and 2 square meters. Now we have a measurement of distance and one of area. These measurements are associated with two fundamentally different physical quantities – distance and area. Not only can they not be added as is, there is no conversion from distance to area that would make them compatible with the concept of addition. They do not just suffer from different units of measure – they suffer from different dimensionality. They have incommensurable units of measure.

But, here's a curious thing to note. When you multiply two quantities, the resulting units of measure have a different dimensionality from the originals – so 1 meter times 1 meter has a different dimension from 1 meter. We can start with a few types of physical quantities (types of dimensionality) and produce 100s of types of physical quantities, simply by multiplying or dividing by other types. It is therefore possible to produce 100s (maybe millions – though the pragmatic use of them would be questionable) of derived types of physical quantities by such means. Physicist who work with a large set of types of physical quantities distinguish between those that are derived (such as area) and those that are fundamental (such as distance).

Dimensional analysis is a process taught to physics students by which any equation can be validated as dimensionally valid or invalid.

- The dimensionality of any terms added together must be the same.
- The dimensionality of the two sides of an equation must be the same.

To complete a dimensional analysis, dimensionality of any term in the formula must be convertible to an expression in terms of fundamental physical quantities such as distance, mass and time.

There is some controversy amongst physicists about the number of fundamental physical quantities available to science. Most physics text books have long lists of "fundamental physical quantities" with associated agreed-upon units of measure (similar to those seen in the Ref J article).

Figure 01 - Excerpt from Ref J

Table of Units,dimensional Formulas of physical quantities.

Fundamental Physical Quantities:

S.No	Fundamental Physical Quantity	Formula	Dimensional Formula	S.I Unit of physical quantity
1.	Mass	Amount of matter in the object	M	kg
2.	Length		L	meter
3.	Time		T	sec
4.	Electric current		I or A	ampere
5.	Amount of substance		N	mole(mol)
6.	Luminous intensity		J	candela(cd)
7.	Temperature		K or θ	Kelvin

Derived Physical Quantities:

S.No	Derived Physical Quantity	Formula	Dimensional Formula	S.I Unit of physical quantity
1.	Area	$l \times b$	$[M^0 L^2 T^o]$	m^2

Under the schema given in Figure 01 we have an answer to the question of the dimensionality of entropy. When Clausius first discovered the strange behaviour of the ratio Q/T (quantity of heat divided by temperature), the obvious dimensionality for such a quantity was Joules per Kelvin. This translates into a dimensionality formula of $M^1L^2T^{-2}K^{-1}$.

But a small number of people argue that these list of physical quantities can and should all be reduced to a list of only three fundamental physical quantities. The Ref K document can be downloaded from the Ref L URL. It comes from the physics department of the University of Florida, and I presume was prepared by someone who falls into this renegade group of extreme reductionists. The three fundamental types of dimensionality are mass (represented by M), distance (represented by L), and time (represented by T) – all others being derivative.

This was a controversy of sorts that existed in the late 1960s when I studied physics, and it appears to be unresolved today.

Some numbers are dimensionless. For example, the number π is dimensionless. Consider the formula for the area of a circle.

$$A = \pi r^2 \quad \text{or} \quad \pi = A/r^2 \qquad [4]$$

Under dimensional analysis this becomes $L^2 = \pi L^2$ or $\pi = L^2 / L^2$. In the rules of dimensional analysis, the common dimensions cancel out of the ratio, and the dimensionality of π is 1, usually referred to as "dimensionless". A dimensionless number cannot be added to or subtracted from a dimensional number, but when multiplying another dimensional number, does not change the dimensionality. Dimensionless numbers seem to have two different sources.

Some, like π, are ratios, in which the dimensions of the numerator and denominator cancel. The other source is more subtle – they are simply counts of things that are discrete and not measurable as physical quantities. Like 6 cycles, or 6 dimensions, or 6 atoms. Purists might argue that "amount of substance" measured in moles would fall into this category, being considered dimensionless because it is a count. They might question this list of fundamental physical quantities.

Personally, I can see value in both sides of that argument. It is very similar to the controversy in computer science over casting of numbers. I can "cast" a number 6 as a long integer, as a floating point number, as a short integer, etc. Each cast determines the way the number is stored within assigned bytes in computer memory, and determines how the bits are interpreted when retrieved. Some computer languages have very rigid casting rules determined entirely by the hardware implementation of types of numbers. Other languages allow user-specified casts. So I could create a cast called "moles", and then any number of moles can only be added to another number of moles. This is called strong type casting. It prevents me (the programmer) from accidentally adding two numbers that should not be added. Just so, the approach implied by the list of fundamental physical quantities in Figure 01is similar to strong type casting. But, it hides creates barriers that hide the truth I seek. I get more into that argument below.

I would like to argue that all formulae involving entropy present entropy as a dimensionless physical quantity.

4.3 - Dimensionality and e^x or ln(x) – In General

In my experience, all scientific expressions in which there is an exponent or a logarithm are such that the argument of exponentiation or of the logarithm is dimensionless.

4.3.1 - Universal Law of Gravitation

An obvious example would be the universal law of gravitation:

$$F = G\frac{m_1 m_2}{r^2} \qquad [5]$$

The exponent 2 on r is a count of dimensions in space-time, and has no other physical dimensionality in terms of units of measure. I.e. in r^2, 2 is a "dimensionless number". I note in this example that the concept of "dimension" has two closely-related meanings, and I must be careful not to confuse them. I try to distinguish them by using "dimension" versus "dimensionality".

- **Dimensions represented in a number** – Space-time has four dimensions (3 of space and 1 of time). "r^2" has two space dimensions. "r^3" has three dimensions in space. This is the common everyday meaning when used in expressions like "3-dimensional movie".
- **Dimensionality of a number** – On the other hand the physical quantity associated with any number is referred to as the dimensionality of the number. Any number representing r^2 will have dimensionality consisting of some unit of measure of space (e.g. meter, or cm) squared – $meters^2$ or cm^2 or something like that. When talking about area, these two concepts coincide, but when talking about something like velocity, they do not.

For example:

- Area is calculated as $A = \pi r^2$.
- It has units of measure m^2 or cm^2, and so represents two dimensional space.

- It has two dimensions of distance, and therefore has dimensionality L^2.
- Velocity is calculated as V = d/t.
- It has units of measure of kilometers per hour, or centimeters per second.
- It has dimensionality of L/T, or more often written as L^1T^{-1}.
- Acceleration has dimensionality of L/T^2, more often written as L^1T^{-2}.

The distinction is, perhaps, subtle when talking about distance measures, but less so in other circumstances. For example, the frequency of a tuning fork is expressed in cycles per second – i.e. (a count of cycles)/(a unit of measure of time). All counts are without dimensionality. So frequency has dimensionality of T^{-1}. Equation [5] clearly mixes distance (L) and mass (M), but the inclusion of time is not evident.

But, back to my point, the exponent 2 on r^2 is a count, and so is dimensionless, not being a measure of mass, time or distance.

4.3.2 - Arrhenius' Equation

A more instructive example would be Arrhenius' Equation. According to the Ref G article, Arrhenius' equation gives the dependence of the rate constant of a chemical reaction on the absolute temperature, a pre-exponential factor and other constants of the reaction.

$$k = A \times e^{(E_a/RT)} \qquad [6]$$

Where:

- k is the rate constant,
- A is the pre-exponential factor, a scaling constant for each chemical reaction (according to collision theory, A is the frequency of collisions in the correct orientation),
- E_a is the activation energy for the reaction (in the same units as R*T),
- R is the universal gas constant, and
- T is the absolute temperature (in kelvins).

The expression E_a/RT has dimensionality energy / energy, and is therefore a dimensionless number. It seems to always be the case that all exponents in scientific formulae are dimensionless ratios. This makes some sense. We all know what the dimensionality of energy is ($M^1L^2T^{-2}$). But what is the dimensionality of e^{energy}?

4.3.3 - Tsiolkovsky rocket equation

Similarly, in my experience, all scientific expressions in which there is a logarithm are such that the argument of the logarithm is dimensionless.

As an example, the Tsiolkovsky rocket equation, or ideal rocket equation, describes the motion of vehicles that follow the basic principle of a rocket: a device that can apply acceleration to itself using thrust by expelling part of its mass with high velocity and thereby move due to the conservation of momentum. The equation relates the delta-v (the maximum change of velocity of the rocket if no other external forces act) with the effective exhaust velocity and the initial and final mass of a rocket, or other reaction engine.

For any such maneuver (or journey involving a sequence of such maneuvers):

$$\Delta v = v_e \ln\left(\frac{m_0}{m_f}\right) \qquad [7]$$

Where:

- Δv – the maximum change of velocity of the vehicle (with no external forces acting).
- m_0 is the initial total mass, including propellant.
- m_f is the final total mass without propellant, also known as dry mass.
- v_e is the effective exhaust velocity.
- ln() is the natural logarithm function.

The argument of the natural logarithm function is a dimensionless ratio.

4.3.4 - Other Examples

The Wikipedia articles at Refs H and I present the Fenske and Nernst equations. In all cases the argument of the logarithm is dimensionless.

Fenske Equation [8]

$$\boxed{N = \frac{\log\left[\left(\frac{LK_d}{HK_d}\right)\left(\frac{HK_b}{LK_b}\right)\right]}{\log \alpha_{avg}}}$$

Nernst Equation [9]

$$\boxed{E = E^0 + \frac{0.05916\text{V}}{z}\log_{10}\frac{a_{\text{Ox}}}{a_{\text{Red}}}}$$

Again this makes sense. For example, we know the dimensionality of energy ($M^1L^2T^{-2}$), but what is the dimensionality of ln(energy)?

In my long experience, I cannot remember seeing an exponential function or logarithmic function for which the argument was other than dimensionless. In this brief search on the internet, I have had the same result.

At the same time, I do not remember ever hearing a discussion of this phenomenon. Perhaps most people find it too obvious to remark upon.

4.4 - Dimensionality of Entropy In Thermodynamics

In the later study of thermodynamics, under Boltzmann, the equation for entropy became:

$$_BS= k_B \ln(\Omega) \qquad [10]$$

The term $\ln(\Omega)$ is dimensionless for two reasons. First, and most obvious, Ω is a count of the number of microstates that are associated with one macrostate. All counts are dimensionless, by definition. But, in my application of the formula (see equation [14]) I replace Ω with the combinatorial multinomial coefficient $\Omega = A!/\prod_{i=1}^{k} a_i$. Is this dimensionless? One could argue that the dimension is (count)/(count^k). But, since counts are always dimensionless, this reduces to a dimensionless count.

k_B is the Boltzmann constant and has a measured value of $1.38064852 \times 10^{-23}$ m^2 kg s^{-2} K^{-1}. This is a VERY small number, and is used to scale the entropy to a size compatible with the conservation of energy regime when discussing atoms and molecules. The units of measure for

k_B is read as [meters squared kilograms per second squared, per degree Kelvin], or $\frac{m^2 kg}{s^2 K}$. The units of energy can be pulled from Newton's laws, as in Work = Force x Distance = Mass x Acceleration x Distance = kg x m/s^2 x m =$\frac{m^2 kg}{s^2}$. So, it seems the units of measure of k_B are energy per degree Kelvin. Energy is definitely NOT a dimensionless quantity, and most physics text books consider degrees Kelvin to be a fundamental physical dimension in its own right, similar to mass, time and space. (For example, see Ref K.) If that is true, then the ratio (energy/temperature) would not be dimensionless, but would, rather, have a very complex dimensionality [$M^1L^2T^{-2}K^{-1}$].

However, with some research, I discovered two interesting things:

- First, I note when consulting the Ref K document that temperature is not just passed over as a fundamental physical quantity, it is not even included in the short list of useful derived physical quantities. But, since the first line of the document declares that ALL (in bold) units come from three fundamental physical quantities, I can presume that temperature was not left out of the list of fundamental physical quantities by accident. So, then, the question that cannot be avoided is, what the dimensionality of temperature, expressed in terms of those fundamental or derived units found on the Ref K list?
- Second, that absolute temperature and energy are related as follows. By the "equipartition of energy" law, the energy in a thermodynamic system is automatically spread over (partitioned over) all possible modes of storage. Each mode of storage is called a "degree of freedom" for the partition. Temperature is generally proportional to the energy per degree of freedom. So, for a given quantity of energy, the temperature is reduced as you add degrees of freedom.

But, this means that entropy with units of measure Joules/Kelvin is actually Joules / (Joule / degrees of freedom) which resolves to a count of degrees of freedom. All counts are dimensionless numbers. And voila! The units of measure for thermodynamic entropy resolve to a dimensionless number.

This is consistent with the equations for conservation of energy in which the term S x T is a form of energy. [degrees of freedom] x [energy per degree of freedom] resolves to [energy], as required in the equation.

4.5 - Dimensionality of Entropy In Information Theory

The articles at Refs M, N and O are about units of measure in information theory. Information theory was born about 1948 and has gone through a very rapid and tumultuous period of exploration and expansion, with, literally, thousands of papers being produced. In my experience, I have found a few papers that were helpful, but many were too sophisticated for me to follow, and a great many were just simply confusing, or wrong.

A turning point in my personal search for understanding of entropy came when I read the Ref Q book by Cohen and Stewart. It pointed out the absurdity of many of the arguments based on concepts of randomness, and the distinction between order and disorder.

Then I heard this story on a CBC program (citation lost) in which Boltzmann's own views were revealed. In those difficult days when he was developing the concept of entropy, there was great controversy about the very existence of atoms, about the nature of which he had to make assumptions. He had to defend his views against harsh public criticism. When releasing his first serious description of entropy, he added a section on order and disorder. In the vicious and very personal attacks he suffered from his opponents, a misrepresentation of this section was used as fodder for sarcasm, and became the focus of public debate. After a while he re-issued a more carefully written document, leaving out the controversial section on order and disorder. He died by his own hand, presumably depressed by the endless verbal assaults he had to endure in public debates. The first incontrovertible evidence of the existence of atoms was produced only a year later (1906), and his theories were vindicated. For over 100 years, all university texts presented entropy in terms of the arguments about order and disorder, since that is what the public controversy was about. But, that was a straw man issue. Boltzmann never meant his theory to be presented the way the public and academia had seen it to be. It is only in recent years (post 2000) that the confusion about order and disorder is being written out of university texts when discussing thermodynamic entropy. Unfortunately, writings about information entropy tend to be rife with such confusing terminology.

So, based on my reading of Cohen and Stewart, augmented by my understanding from the CBC documentary, I learned to ignore any papers or books that talked about randomness or disorder and to focus on those that showed mathematical developments using arguments I could follow. It turned out, that was a very small set of papers and books – especially when I read into information theory. So, it is almost certain that there are profound insights to be had from information theory that were/are simply over my head. I would be brash to say that ALL talk of order, disorder and randomness is wrong, but Cohen and Stewart very ably argue that MOST of that kind of argument is nonsense.

Nevertheless, I think that a look at units of measure as used in information theory provides some perspective on the tumultuous development of this field of study, and some insights into my own calculations of entropy.

In the Ref P spreadsheet (see Figure 02) I have collected together some notes about units of measure, as used for entropy by theorists of "information theory". It comes down to this. Just as there are different units of measure for mass (grams, pounds, kilograms) used in different physical measurement systems (CGS, British, MKS, SI), there are three different systems for units of measure for entropy.

Figure 02 – Drawn from the Ref P spreadsheet.

	bit/Sh	nat/nit/nepit	dit/decit/ban/Hart
bit/Sh	1	0.69315	0.30103
nat/nit/nepit	1.44270	1	0.43429
dit/decit/ban/Hart	3.32193	2.30259	1

There are three measurement systems, based on $Log_2()$, $Ln_e()$ and $Log_{10}()$.

And, there are conversion factors that let you convert from one system to the other. However, different theorists have given these three units many different names, so each unit of measure has several different names:

- **Bit** – a bit is the smallest unit of measure for entropy. It is associated with a measurement system using $\log_2()$. A bit also goes by the name "a Shannon" with symbol Sh;
- **Nat** – a nat has a more abstract definition, being associated with a measurement system based on natural logarithms, i.e. based on $\log_e()$. A nit is also known as a nit or a nepit.

- **Dit** – a dit is associated with a measurement system based on $\log_{10}()$. A dit is also known as a decit, a ban, or a Hart.

The conversion factors for these three units of measure are shown in Figure 02, and in the Ref P spreadsheet. These conversion factors seem to play the same role as Boltzmann's constant (k_B) plays in thermodynamics. That is to say that they are scaling factors that enable measurements made in one measurement system to be convertible for use in another measurement system. Scaling factors are often dimensionless constants, similar to Boltzmann's constant k_B.

However, exactly what they measure is difficult to determine. The "bit" is used in two distinctly different measurement systems. I believe that the nat and dit are only used in one. The two systems for the "bit" unit are:

- **Information** – First, the information technology industry, bits are counts of units of information as found in a computer memory bank or in a digital data transmission. It seems that "bit" in this context is a dimensionless unit – a count of minimal quanta of information. All counts are considered dimensionless. 8 bits make a byte, and both storage capacity and transmission size is measured in bits or bytes. Neither nats nor dits are used in such an information measurement system, to my knowledge.
- **Entropy** – Second, in information theory, bits are measures of entropy of a transmission or stored message. It is related to the "number of ways you can re-arrange the bits and deliver the same information". Then it would be closer in meaning to entropy in thermodynamics, or in my to-be-proposed ABM/histogram system. In any case, regardless of the exact words used to describe what it measures, it seems pretty clear that a bit in this context is a dimensionless number. Proponents of information theory seem to claim that this has the same meaning as in the first instance. The difference, as I see it, is that in the first instance a person can actually count the bits in a chip to arrive at an integral value, but in the second instance it is calculated for some ephemeral bit-stream and a positive real number is returned.

I think it is important for me to understand how the unit of measure "bits" functions in each of the two measurement systems in which it is used. (See the table in Figure 03, also drawn from the Ref P spreadsheet.)

4.5.1 - Entropy of a Bit-Stream – All Ones

I am going to use this same table several times, so it requires some explanation:

- The yellow cells are intended for varied input and contain the variables a_i. In this case there is a bit-stream of five bits, all of them turned on. In later cases, there will be a stream of numbers. The five yellow cells are the only cells into which I will be inserting new values as this exercise proceeds. All others are computed automatically.
- K=5 indicates the number of bins in the histogram.
- The blue cell to the right of the yellow cells is the sum of the values of a_i, i.e. the sum of the yellow cells. In my notation $\sum a_i = A$.
- The bases of the logarithm for the three systems of measurement are shown for bits, nats and dits at the bottom left.
- The green area contains temporary calculations for four systems of measurement: bits, nats, dits and hnats (i.e. my own developing system for use with general histograms, and for use in ABMs).

- Finally, the orange area to the right is where the four values of entropy for this histogram (this bit stream) is calculated. Note that each system of measurements produces a different value.

Figure 03 – Single Bits as Input – Various Units as Output

K =	5					A =		
a_i =	1	1	1	1	1	5		Relative
Probabilities $p_i = a_i / A$ =	0.20000	0.20000	0.20000	0.20000	0.20000	1.00000	= Sum(p_i) = 1	Sizes
Surprisals in Bits	0.46439	0.46439	0.46439	0.46439	0.46439	2.32193	= S In Bits	1.00000
Surprisals in Nats	0.32189	0.32189	0.32189	0.32189	0.32189	1.60944	= S In Nats	0.69315
Surprisals in Dits	0.13979	0.13979	0.13979	0.13979	0.13979	0.69897	= S In Dits	0.30103
$\ln(a_i!)$ =	0.00000	0.00000	0.00000	0.00000	0.00000	0.95750	= S in HNats	0.41237
Base for Bits =	2							
Base for Nats =	2.71828							
Base for Dits =	10							

In this version of the table, the bits are all on (i.e. = 1). Note that the relative sizes of the values for entropy (when divided by the entropy S in bits) exactly matches the conversion factors as shown in Figure 02. This tells me my table is set up in a valid fashion.

The relevant equations for the four orange cells are:

- S in bits: $S = \sum_{i=1}^{5}\left[\frac{a_i}{A} \times \log_2\left(\frac{A}{a_i}\right)\right]$ [11]
- S in nats: $S = \sum_{i=1}^{5}\left[\frac{a_i}{A} \times ln\left(\frac{A}{a_i}\right)\right]$ [12]
- S in dits: $S = \sum_{i=1}^{5}\left[\frac{a_i}{A} \times \log_{10}\left(\frac{A}{a_i}\right)\right]$ [13]
- S in hnats: $S = \left[\ln[A!] - \sum_{i=1}^{5}[a_i!]\right]/A$
  using the custom function shown in Figure 04. [14]

Figure 04 – Custom function used in calculation of hnats.

```
'///////////////////////////////////////////////////////////////////////////////////
'// Orrery Software, orrery@rogers.com, Garvin H Boyle.
'//////////////////////////////////////////////////////////////////////////////////
'// This function computes ln(x!) using Stirling's approximation, but only when
'//   needed.
'// Stirling's approximation is: ln(x!) ~ ( x * ln(x) ) - x
'// Stirling's more accurate version is: ln(x!) ~ ( ( x + .5 ) * ln(x) ) + 1.0 - x
'// MS Excel is not able to calculate x! if x > 170.
'// So, when it can, this function does the exact calculation (i.e. when x <= 170).
'// And when it cannot do that, it uses Stirling's approximation.
'///////////////////////////////////////////////////////////////////////////////////
Function LawnOfXFactorial(x)
   If x < 170 Then
     LawnOfXFactorial = Application.Ln(Application.Fact(x))
   Else
     LawnOfXFactorial = ((x + 0.5) * Application.Ln(x)) + (1 - x)
   End If
End Function
```

The unit of measure "hnats" is one I am inventing here. I do not want to call what I measure "nats" because (a) I am using a different formula from the one used for nats in information theory, and (b) I find it difficult to make a direct and clear link between what I am measuring (entropy of a histogram) and nats. Nevertheless, there is a close connection between what I am measuring and "information entropy", so I retain the "nats" designation, and prepend the 'H' to signify application to a histogram, and not a bit stream or message in the normal sense considered in information theory. At some point I might decide the distinction is not-needed, but for now it reminds me that there are unresolved issues.

Here are some observations from this table:

- K is the number of bits in the bit stream as the word "bits" is used in information technology language. In my schema, K is the number of bins in the histogram.
- The number in blue (5) is the count of the turned-on bits. This corresponds to the number of agents (A) in my system.
- The top number in orange is the measure of the information entropy and is designated the number of "bits" of information entropy.

I have read that "information entropy" is often interpreted as "information". So, we might say that this 5-bit hardware configuration contains only 2.32 bits of information. That does not really make sense to me, and that is one reason I find information theory confusing. But, I don't think my failure to understand this particularly arcane aspect of information theory makes my ultimate conclusions below wrong.

4.5.1 - Entropy of a Bit-Stream – Single One

Figure 05 is the same table as shown in Figure 03, except I have changed the contents of the five yellow cells. That is, I have changed the values of the a_i. In particular, I have put all of the value into a single bin, and made the others empty. The results are the same whether I put a '1' into the single non-empty bin, or 1,000 or any other number. All formulae produce S = 0.00000.

Figure 05 – Bit Stream with a Single 1

K =	5					A =		
a_i =	0	0	1	0	0	1		Relative
Probabilities $p_i = a_i$ / A =	0.00000	0.00000	1.00000	0.00000	0.00000	1.00000	= Sum(p_i) = 1	Sizes
Surprisals in Bits	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	= S In Bits	0.00000
Surprisals in Nats	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	= S In Nats	0.00000
Surprisals in Dits	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	= S In Dits	0.00000
ln(a_i!) =	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	= S in HNats	0.00000
Base for Bits =	2							
Base for Nats =	2.71828							
Base for Dits =	10							

I have tried a variety of sets of 0/1 bit values in the yellow cells. Two bins must contain 1s in order for the entropy to rise above zero. Obviously, the role of 0s and 1s is not symmetric. A bit-stream of all 1s should have the same amount of information as a bit-stream of all 0s, in my way of thinking about it, but that is clearly not the case.

4.6 - Entropy of a Histogram

Now, I will try the same table, but with inputs other than 0s and 1s. In other words, it will be the entropy of a 5-bin histogram.

Figure 06 – Calculating Entropy of a Histogram – Four Formulae

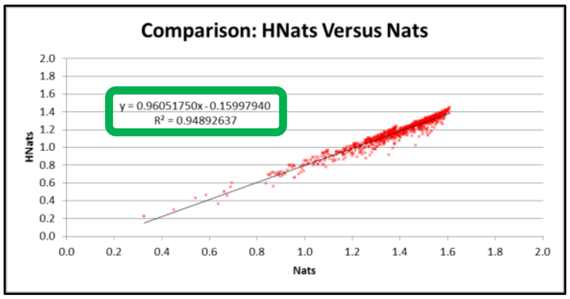
K =	5					A =		
a_i =	74	218	212	79	43	626		Relative
Probabilities $p_i = a_i$ / A =	0.11821	0.34824	0.33866	0.12620	0.06869	1.00000	= Sum(p_i) = 1	Sizes
Surprisals in Bits	0.36416	0.52997	0.52902	0.37686	0.26540	2.06540	= S In Bits	1.00000
Surprisals in Nats	0.25241	0.36735	0.36669	0.26122	0.18396	1.43163	= S In Nats	0.69315
Surprisals in Dits	0.10962	0.15954	0.15925	0.11345	0.07989	0.62175	= S In Dits	0.30103
ln(a_i!) =	247.573	959.512	927.275	269.291	121.533	1.41225	= S in HNats	0.68377
Base for Bits =	2							
Base for Nats =	2.71828							
Base for Dits =	10							

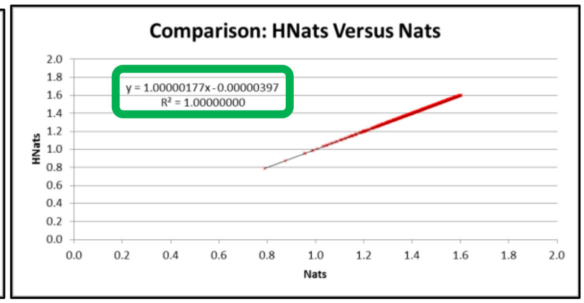
I note that the relative size of my formula for histogram entropy does not have a consistent size with respect to the bits, nats and dits. For example, as I change the values in the yellow cells, the derived conversion factors (relative sizes at the right) are rock solid for three of the four measures, but the conversion factor for hnats varies. There is no one constant conversion factor. This is NOT a surprise, due to the implicit use of Stirling's approximation in three out of four of the formulae. If the numbers in the yellow cells are close to zero, there is wide variation between the value of entropy calculated in nats and in hnats. But, if the numbers are large, then the relative size of nats and hnats are almost the same.

Figure 07 – Comparison – Small numbers vs Big numbers

K =	5					A =		
a_i =	3	9	5	7	2	26		Relative
Probabilities $p_i = a_i / A$ =	0.11538	0.34615	0.19231	0.26923	0.07692	1.00000	= Sum(p_i) = 1	Sizes
Surprisals in Bits	0.35948	0.52979	0.45741	0.50968	0.28465	2.14100	= S In Bits	1.00000
Surprisals in Nats	0.24917	0.36722	0.31705	0.35328	0.19730	1.48403	= S In Nats	0.69315
Surprisals in Dits	0.10821	0.15948	0.13769	0.15343	0.08569	0.64451	= S In Dits	0.30103
ln(a_i!) =	1.792	12.802	4.787	8.525	0.693	1.25624	= S in HNats	0.58675
Base for Bits =	2							
Base for Nats =	2.71828							
Base for Dits =	10							

K =	5					A =		
a_i =	3000000	9000000	5000000	7000000	2000000	26000000		Relative
Probabilities $p_i = a_i / A$ =	0.11538	0.34615	0.19231	0.26923	0.07692	1.00000	= Sum(p_i) = 1	Sizes
Surprisals in Bits	0.35948	0.52979	0.45741	0.50968	0.28465	2.14100	= S In Bits	1.00000
Surprisals in Nats	0.24917	0.36722	0.31705	0.35328	0.19730	1.48403	= S In Nats	0.69315
Surprisals in Dits	0.10821	0.15948	0.13769	0.15343	0.08569	0.64451	= S In Dits	0.30103
ln(a_i!) =	41742377	135114625	72124751	103329954	27017324	1.48403	= S in HNats	0.69315
Base for Bits =	2							
Base for Nats =	2.718282							
Base for Dits =	10							





4.7 - In The Proposed Histogram/ABM system

So, the purpose for this series of diary notes (NTFs) is to formulate a system for the measurement of the entropy of histograms that can be used in agent-based models (ABMs). The key question for this NTF is proposing a unit of measure for such a measurement of entropy. I hereby decide that the definition of hnats as described in the section 4.5 (especially at Equation [14]) will be my standard unit of measure for entropy in the histogram system of entropy measurement. When dealing with very large numbers in the histogram, this converges to nats as used in information theory. But for the numbers typically found in ABMs, an hnat is slightly different from an nat.

The formula for calculating entropy can then be written as:

$$_HS = \frac{_HC}{A} \times \ln\left(\frac{A!}{\prod_{i=1}^{k}(a_i!)}\right) = \frac{_HC}{A}\left[\ln(A!) - \sum_{i=1}^{k}(a_i!)\right] \qquad [15]$$

$_H$S is entropy, and it is a dimensionless number. The unit of measure (hnat) is also dimensionless. And, associated with that arbitrary decision, I also assume that the scaling factor $_H$C in equation [15] is also a dimensionless quantity. This assumption seems to be consistent with all I have read about entropy, and has not lead to any logical anomalies so far.

How would I interpret a calculated value of $_H$S for some histogram? I suppose I could say that it is a count (therefore dimensionless) of bit-like morsels of information, but that would imply that I understand what a bit, a nat, or a dit actually is, within the context of information theory. And I cannot really claim that. I think the better understanding of what an hnat is comes from its application, as in the Ref D draft paper. But when I consider the definition of an "entropic index" the units of measure cancel out, and it becomes a moot question.

The study that I did previously on the connection between Boltzmann's formula for entropy and Shannon's formula (at Ref F) provides the connection between hnats, as defined here, and nats, as used in information theory and as produced using Shannon's definition of entropy. (See equations [21] and [22] of that NTF.)

The equation for conversion from nats to hnats is dependent upon A and is:

$$S_{hnats} \approx (1.0041 * (A * S_{nats}) + 8.5)/A \quad [16]$$

The equation for conversion from hnats to nats is:

$$S_{nats} \approx \frac{A \times S_{hnats} - 8.5}{A * 1.0041} \quad [17]$$

4.8 - Concern

This does not look at tidy as I would like it to be. This conversion from hnats to nats and back is dependent upon A.

Perhaps that concern will go away when I address the issue of S_{max} and S_{index} in my next note.