NOTE TO FILE

G H Boyle

Dated: 130221

Revised: 141219

Revised: 141230

Revised: 150107

Entropy in an Agent-based Model (ABM)

# References:

1. A Dragulescu and V M Yakovenko, “*The Statistical Mechanics of Money*”, European Physical Journal B, 2000.
2. Cockshott, Cottrell, Michaelson, Wright and Yakovenko, “Classical Econophysics”, Routledge advances in Experimental and Computable Economics, Routledge, Taylor and Francis Group, London and New York, 2009.
3. V M Yakovenko, “Applications of statistical mechanics to economics: Entropic origin of the probability distributions of money, income, and energy consumption”, Chapter from an as yet unpublished book, Routledge ? (2012 ?).

 arXiv:1204.6483v1 [q-fin.ST] 29 Apr 2012

 http://arxiv.org/abs/1204.6483

1. 140124 PPR - Definition of EI with equations R12.docx
2. 141220 BinWidth vs Entropy R3.xlsm
3. 141230 Tiny BinWidth vs Entropy R1.xlsm
4. 140415 NTF Entropy Equations with Stirling's Approximation.docx
5. <http://en.wikipedia.org/wiki/Maxwell%E2%80%93Boltzmann_distribution>
6. <https://www.openabm.org/model/3613/version/1/view>
7. <http://modelingcommons.org/browse/one_model/4219#model_tabs_browse_info>

# Purpose:

To explore the causes of the distribution of wealth that is indicative of the production of entropy, and to plan the development of a concept of entropy to be used in ABMs.

# Discussion

## ModEco –In Which Effects of Entropy Production are Evident

In ModEco, in the PMM, a distribution of wealth appears that is shaped very similar to the Maxwell distribution of speeds in an idealized gas. (See Refs I, J and K.) (In Figure 01, the distribution of total wealth of 1,000,000 agents taken over many generations (blue) has been normalized and overlain with a normalized Maxwell distribution (red)).

Processes that produce thermodynamic entropy and cause it to rise are considered the causative source of this fascinating distribution when it arises in thermodynamic systems. But the PMM is a model economy that exists only as an insubstantial logical system having meaning only as an interpretation of bits and bytes in a computer. Thermodynamic entropy has little or nothing to do with a run of the PMM, or any other computer model, for that matter. Its only association is that electrical potential energy is degraded to heat as the computer runs. This is true of every computer model, even if it does financial accounting. So I am certain that this distribution of wealth in the PMM is not explained by thermodynamic entropy.

**Figure 01 – Comparison of Curves**



An argument that would explain it as the product of informational entropy is, perhaps, a little more difficult for me to discard. On the other hand, proponents of information theory argue that all is information, even mass, and so their definition is rather expansive and wholly reductionist. I think that the reductionist argument that would say that ‘this is just another example of informational entropy’ is a cop out, and is ignoring the deeper implications of the emergence of such distributions in so many places. Or, perhaps it is misunderstanding and miss-stating the deeper implications. Maybe their emphasis is just on the wrong idea. Instead of saying ‘all entropy is an example of informational entropy’ they might say ‘entropy can be exhibited in many places, and informational entropy is one, as is thermodynamic entropy’. The latter would be my view.

Before continuing with my general thoughts about entropy in ABMs in general, here’s a summary of the logical structures within the PMM and ModEco that gave rise to the distribution in Figure 01.

### Distribution of Wealth Between ModEco Agents

Based on the distribution of wealth in a ModEco-based steady-state economy (the Perpetual Motion Machine, or PMM), and its similarity to the Maxwell distribution of speeds in an ideal gas, and the connection of that distribution to the Maximum Entropy Principle (MEP), I believe that there is an entropy-like concept at work in the PMM.

### Distribution of Wealth Within ModEco Agents

In order to make that particular economic model (i.e. the PMM) sustainable it was necessary for me to insert ‘business factors’ into the model for each agent. At the moment, all agents use the same business factors, so the “Hold Recycled Factor” (or HRF) is the same for all farmers. Each agent has its wealth held in a number of asset classes, and the total of all assets adds up to the total net worth (WNet).

|  |  |
| --- | --- |
| $$W\_{Net}=\sum\_{Asset Classes}^{}W(Asset Class)$$ | Equ 01 |

Where W(Asset Class) is the monetary value or worth of the asset class. For farmers, the asset classes are cash, energy, recycled mass, inventory, supplies and waste mass. For workers the asset classes are cash, energy, supplies and waste mass.

An agent must have some assets in each class in order to participate in the economic transactions involving that asset. E.g. a farmer cannot hire workers to work the field if there is no recycled mass (analogous to manure) on the farmer’s field. The field is barren and unable to produce. A business factor is a fraction between zero and one, and they function best, it seems, if the factors for one agent add up to less than one. Each business factor is the target fraction of the net worth that the agent attempts to maintain in that asset class. So if the W(recycled mass) is less than HRF x WNet, then the farmer will attempt to purchase more recycled mass.

Here is a summary of the business factors, and their default values:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Agent Type | Factor Name | Acronym | Value | Purpose |
| Farmers | Hold Recycled Factor | HRF | 0.500 | Triggers the purchase of recycled mass |
| Hold Inventory Factor | HIF | 0.250 | Triggers harvesting into inventory |
| Hold Supplies Factor | HSF | 0.250 | Triggers the purchase of supplies |
| Workers | Hold Supplies Factor | HSF | 0.350 | Triggers the purchase of supplies |

Here is a summary of the asset classes:

|  |  |  |  |
| --- | --- | --- | --- |
| Agent Type | Asset Name | Flag | Type of asset |
| Workers | Cash | X | Money to be used to purchase assets |
| Energy | X | Metabolic exergy, to be used to do work |
| Supplies | HSF | Mass with associated energy for metabolic uses, for personal use |
| Waste | X | Mass without associated energy load |
| Farmers | Cash | X | Money to be used to purchase assets or hire workers. |
| Energy | X | Metabolic exergy, to be used to do work |
| Recycled mass | HRF | Mass without associated energy load |
| Inventory | HIF | Mass with associated energy for metabolic uses, to sell to consumers |
| Supplies | HSF | Mass with associated energy for metabolic uses, for personal use |
| Waste | X | Mass without associated energy load |

The logic for creating the limited number of business factors is as follows:

* For both farmers and workers:
* Cash is the medium of exchange that is used as available and as needed, and so is a special asset convertible into any other asset, and need not be controlled in this fashion.
* Waste and energy are produced by the biological imperative of eating the required amount of supplies each day, so these asset classes have no associated triggers.
* Workers always work (sell their energy) to obtain cash, and always sell waste to obtain cash. No trigger is required to control these actions.
* The only asset workers purchase with their cash is supplies, so they do not need to keep a store of cash to participate in any other activity. You would think they should purchase supplies at every opportunity when they can afford it, in order to be ready for future lack of work. However, if wealthy workers are allowed to convert their entire net worth into supplies, there is nothing left for the poor workers, and the economy collapses. So, the purpose of the HSF for workers is to spread the supplies between agents.
* Workers purchase supplies when W(supplies) falls below HSF x WNet.
* Farmers, on the other hand, must use their cash to purchase recycled mass, hire workers, and buy supplies for their personal consumption. If they spend all of their cash on one asset, they cannot participate in the other necessary business opportunities that come along. So they must be restrained:
* They purchase recycled mass when W(recycled mass) falls below HRF x WNet.
* They hire workers when W(inventory) falls below HIF x WNet.
* They purchase supplies when W(supplies) falls below HSF x WNet.

This interaction of asset classes and business factors causes a curious phenomenon which I have called the “ratio-based distribution of wealth” among asset classes. (See figure 02.) This is, curiously, vaguely analogous to the equidistribution of energy among types of energy (translational kinetic, rotational kinetic, vibrational, etc.) for molecules in a gas. I suppose, if that was my goal, I could try to build in sufficient factors to achieve an ‘equidistribution’ of wealth.

**Figure 02 – Ratio-Based Distribution of Wealth**



But, my goal is sustainability, so that would be a diversion.

In the NetLogo version of ModEco (Ref K) I have used an intermediate version of HRF, HIF and HSF held in the variables HRL, HIL and HSL. In these acronyms the ‘L’ stands for “Limit” and is the amount of supplies, in mu, meu and meu respectively, below which purchase is triggered.

Finally, as part of the biophysical subsystem, the metabolic functions of each agent are controlled by certain “life function” parameters. Those of interest to this topic are the ones that cause death and removal from the economy. The relevant parameters are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Agent Type | Function Name | Acronym | Value | Purpose |
| Workers | Death Energy Threshold | DET | 4 | Causes death by emaciation |
| Energy Per MoveEnergy Per Tick | EPM orEPT | 4 | Causes death by starvation |
| Farmers | Death Energy Threshold | DET | 4 | Causes death by emaciation |
| Energy Per MoveEnergy Per Tick | EPM orEPT | 16 | Causes death by starvation |

Acronyms and default values may differ between the C++ version and the NetLogo version.

These are biophysical thresholds, but they have implications for wealth. If the Net Worth of an agent falls below W(DET), they die of emaciation. If W(supplies) falls below W(MPT) they starve. The implications are that the poorest end of the spectrum of possible wealth values is partially shaved by the removal of that fraction of poor agents that temporarily abrogate these lower limits on certain assets.

Since two out of three modes of death (old age not included) are associated with levels of wealth within an asset class, the associated life function parameters will have an effect on the shape of the distribution of wealth within the population. Certain combinations of low wealth will not exist.

## Argument For Existence of ‘Economic Wealth Entropy’

I hypothesize that there is a type of entropy that is associated specifically with economic activity in which wealth can be accounted and measured.

Yakovenko and Dragulescu show in their paper, at Ref A, that the distribution of wealth produced by a simple agent-based model is a reasonable approximation of distributions of energy in gasses.

In “Classical Econophysics”, at Ref B, one of the authors presents an essay in which he discusses the equivalency of Boltzmann entropy and Shannon entropy, that is the entropy as described by Clausius and Boltzmann in thermodynamics, and that described by Shannon in information theory. I can follow the argument. The argument seems sound enough. But I lack sufficient understanding of entropy to say whether I agree or disagree with the concept of the equivalence of the two. However, it is VERY clear to me that there is a common behind-the-scenes phenomenon that makes these two instances of entropic production to be, if not equivalent, then at least reasonably accurate maps one of the other, with energy and information playing the analogous roles.

Taking that idea one step further, Yakovenko, at Ref C, seems to point out that the definition of entropy is able to stand without reference to physical units, and that the physical units are supplied via the scaling factor that is included in the formula for entropy (my interpretation). That idea is further bolstered by dimensional analysis of the units of thermodynamic entropy, which are { Joules} per { degree Kelvin }. The dimensions of a { degree Kelvin }, in turn, are { Joules } per { degree of freedom of motion of the atoms }. But, { degree of freedom } is a dimensionless count, like { cycles } is { cycles } per { second } which is usually written as s-1. So, the units of measure of thermodynamic entropy are Joules per Joule, making it a dimensionless real number.

Putting these three ideas together, I wonder if it is possible to define, in addition to thermodynamic and informational entropy, a third kind of entropy called, say, economic entropy, or wealth entropy, which has a similar mathematical formulation, and which plays a role in both the internal and external distribution of wealth, within people’s assets classes, and across society.

I need to be careful how I define wealth, now, I think. I mean real wealth having a biophysical basis, and not the quasi-wealth that comes from holding cash. Cash (money, currency) is a really difficult thing to handle. A dollar is an outstanding IOU, a claim on future biophysical production, and can be exchanged for biophysical wealth at some time. Within ModEco cash is NOT generated via the trick of simultaneous creation two offsetting accounts of debt and credit, as is common in reserve banking systems (except, of course, for the MMgr when it goes into debt), so the cash should/could be included as wealth. For the purposes of this discussion I will call wealth that arises from ownership of biophysical assets as ‘real’ wealth, and the wealth that arises from cash as ‘claimable’ wealth, total wealth being the sum of the two. The problem is that total net worth is now a combination of current real wealth and future claimable wealth. I suspect that is not a purely “biophysical economics” approach to defining wealth. It becomes a combination of real and imaginary wealth.

## Can we Define ABM Wealth Entropy?

Going one step further, since this kind of entropy-producing activity seems to be happening, not only in real-world economies, but also in very-highly-abstracted agent-based models of economies, I wonder if it is possible to define some kind of ABM wealth entropy.

## A Hypothetical Vision of Many Types of Entropy

To summarize the above development, there is a progression of types of entropy that I see:

* Thermodynamic entropy, à la Clausius and Boltzmann
* Informational entropy, à la Shannon
* Wealth entropy, à la Dragulescu and Yakovenko (this is where I first saw it implicitly demonstrated)
* ABM entropy, à la Yakovenko’s website (this is where I saw it explicitly demonstrated)

This implies that (I hypothesize that) entropy is a fundamental mathematical property of measurements of conserved qualities of collections of objects, whether physical (molecules in a gas), logical (information encoded in bits in a signal), economic (wealth values for people), or imaginary (numbers fabricated in Models). I hypothesize that entropy of many kinds can be calculated for a single system, if many qualities are simultaneously conserved. I hypothesize that entropy is important because, for such collections of objects found in conservative self-organizing systems, entropy of all kinds tends to rise as the system self-organizes.

**Figure 03 – Realms of Application of Entropy Formulae**



Now I may be trying to whittle a sharp point on a soft stick, and wasting time doing it, but I need to think through the stuff in Figure 02.

In this diagram I am comparing the formulae for entropy (middle section) with the formula for calculating the arithmetic mean (right section). I suppose I am doing that to convince myself that an entropy calculation is no different from any other platonic mathematical concept that finds application in the real world.

### Platonic Realm

In the Platonic realm, any old set of numbers can be turned into a histogram, and then the entropy of that histogram calculated. But, this is not a fully-defined process. Certain decisions need to be made as this is done. I use the word ‘bin’ to mean the same as ‘non-overlapping interval’ in a tiling of the domain. A bin might also be called a class, or a macro-state. The starting point for the first bin, and the ending point for the last bin need to be decided, allowing room for an integral number of contiguous non-overlapping bins, and the bin width needs to be determined. These three numbers are arbitrary, and will determine exactly what the histogram will look like. But, choose any two and the third number is determined, so two arbitrary decisions are need. Once a standard procedure is decided upon to create the histogram, then you can consider that, given the same set of numbers, and the same procedure for making the histogram, the entropy formula will always produce the same value for entropy of the set of numbers. So, we have a rule (how to construct a histogram) and a formula (how to compute the entropy of a histogram), and that enables us to consistently compute the entropy of any set of measurements.

The formula for the mean of the numbers does not need the setup of construction of a standard histogram first. It can be applied directly to the numbers. On the other hand, if the data is already binned into a histogram, there is an accepted rule (or procedure) and formula for estimating the mean of such binned data. The mean so calculated will be close to the mean of un-binned data, but not necessarily the same. Some precision is lost in the process of binning the data.

Hmmm! Does this mean to imply that entropy computations on binned data necessarily has a loss of precision? That’s a tricky question. I cannot conceive of how entropy could be computed on unbinned data. All real numbers are, by their very nature, binned by the process of rounding them to some standard precision prior to use. So, precision is lost. In fact, the same is true whether you are computing a mean or entropy. So, precision is always lost by binning, unless you work with exact symbols such as π. The difference is entirely an issue of bin width.

### Insubstantial Logical Realm

This is the ‘realm’ in which ABMs exist. Agents in ABMs are not things. They are representations or simulations or conceptualizations of things, and those conceptualizations exist only in the bits of computers and the imaginings of our minds when we watch the computers.

This is also the realm in which much writing about informational entropy exists. It is often disembodied concepts that have been abstracted away from real biophysical systems. That is fine, but it is also confusing at times.

I have broken this realm into two parts: the static not-necessarily-conservative part, and the dynamic conservative part. I suppose I could actually break it into three (adding dynamic in one step and conservative in another), but I don’t see the point.

#### Logical Realm

The static part is there mostly as an instantiation of the “Platonic Realm”, which, by definition, I think, is static in nature. In this realm any collection of concepts (e.g. agents) that have a measureable characteristic, (such as height, age, or wealth) can be used as the source of a set of numbers. A histogram of those measurements can be built, and the entropy can be calculated, or, the mean can be calculated from the pure measurements or from the binned measurements.

#### Conservative dynamic logical Realm

For the conservative dynamic logical part we are talking about a system of conceptualized things that change with time. I add the dimension of time. As time moves forward, the characteristics of each object change, the shape of the histogram changes, and the entropy of the measurements changes, as does the mean of the measurements. If the agents’ characteristic being measured is not extensive, or not conserved, in the system (e.g. the age of the agents is neither extensive nor conserved, and the sum of the ages of all agents will vary from tick to tick) then I hypothesize that the associated entropy may rise or fall over time. If, on the other hand, the agents’ characteristic being measured is both extensive and conserved in the system (e.g. the total cash in the system is conserved from tick to tick) then I hypothesize that the associated entropy will rise asymptotic to a maximal value.

### Conservative Dynamic Realm

This realm is in two parts, again.

The top part, already discussed above, includes all insubstantial logical systems that are dynamic and conservative with respect to an extensive characteristic.

#### The Conservative Dynamic Biophysical Realm

The bottom part includes all biophysical systems that have real biophysical existence, real instantiations. These include data transmission systems (in which informational entropy was discovered as a biophysical phenomenon), thermodynamic systems (in which thermodynamic entropy was discovered as a biophysical phenomenon), economic systems (in which I am hypothesizing economic entropy will be shown to be an important phenomenon).

## Logical Spaces and Fields

Behind all of this, I see a fundamental characteristic relationship between the entropy of a system and the probability field of the state space of the agents in the system. Here I must distinguish between the state of an agent, the state space of an agent, the state of the system, and the state space of the system. The state space of an agent can be populated by many agents, each represented by a point in the state space – all agents being represented by a cloud of points. On the other hand, the state space of a system is populated by a single system, represented by a single point having a trajectory, through time, in the state space. An ABM has entropy based on measurements of a characteristic of the state of each of its contained agents.

* State changes for agents are driven by the states of nearby agents, and are dependent on them.
* State changes in the state space of the system, however, are not dependent on things outside of the system (the model).

So, again, within the state space of the agents, each agent is represented by a dot (an n-tuple) and the population of all agents is a cloud of dots that form a shape, having varying density of placement. Conserved quantities are each represented by one dimension in this relatively small state space. The dot density, when projected on this dimension, becomes the histogram, the probability distribution for that conserved characteristic. When the system as a whole self-organizes, the agents move in their small state space, the dots move, and form a characteristic configuration such as the Maxwell distribution. For each agent, the set of available next states would be relatively easy to define, but the information about probability of moving into one of those next states is not found within the agent itself. From the point of view of the agent, this probability is unknown and unknowable without reference to externalities.

But, the system as a whole must be viewed differently. The state variables of each agent each form dimensions of the system’s state space. So, suppose there are X state variables that are not state variable of the agents, and Y state variables for each agent, and A(t) agents, a number that changes over time. Then the total number of state variables (e.g. in ModEco) for the system is N = X + ( A(t) \* Y ). As agents are born or die A(t) rises and falls, and the associated dimensions in that all-encompassing system state space appear or disappear. As a generation of agents are born, live out their lives, and then die, all of the dimensions that are not in the X group are removed and replaced with (A(t) \* Y) new dimensions. At each step along the way, the entropy of the system can be calculated. The system is represented by a single point in this shimmering ever-changing space with volatile multi-dimensionality. The number of next possible states has to be immense, but it is knowable without reference to externalities. Additionally, the probability of moving into each of those next possible states must be knowable, and calculable, if minute.

Define a ‘transition pair’ of states, for either the agent state space or the system state space, as two states (two n-tuples) one of which can be transformed to another during one tick of the ABM. So, in the agents’ state space, if an agent exhibits state P and it can move to state Q within one tick, then [P,Q] is a transition pair. Similarly, it the system, the model, as a whole, is in state S in the system state space, and it can move to state T in one tick, then [S,T] is a transition pair.

I hypothesize that:

* the entropy associated with any conserved extensive quantities rises asymptotically to a maximal value and hovers close to that value.
* that the maximal value of entropy is associated with:
	+ a signature distribution (e.g. Maxwell distribution) of points in that dimension of the agents’ state space, and
	+ some sort of ephemeral limit cycle and stationary state in the state space of the ABM itself.
* the difference in entropy rise between the two states of a transition pair in the state space of the ABM will be correlated in some way with the probability of transition from state to state within the transition pair. This concept is explored in some detail in the NTF in which I define the concept of entropic index for ABMs at Ref D.
* the trajectory of the system through this shimmering volatile state space is *most probably* going to follow the trajectory of *most probable transitions*, and will end up in the ephemeral limit cycle of the stationary state.

## Calculation For ABM Entropy

So far, I have not mentioned the formula to be used. The standard formula for entropy would look like this, using Stirling’s approximation:

|  |  |
| --- | --- |
| $$S\_{ABM}=F ln\left(ω\right)=F\sum\_{All States}^{}P(Macrostate)×ln⁡(1/P\left(Macorstate\right))$$ | Equ 02 |

where:

* SABM = ABM Entropy
* F = a scaling factor with suitable units of measure, possibly always dimensionless.
* State refers to the state of an agent in an ABM with reference to a conserved quantity. It is a macro-level measurement of that conserved quantity, such as cash, mass, energy or wealth in a ModEco economy. For example, if an agent’s net worth (real plus claimable wealth) can vary from $0 to $100,000 you might partition that into ‘macro-states’ or levels of wealth, each macro-state being the wealth rounded to the nearest $1,000.
* The sum is taken over all possible such macro-states (e.g. of agent wealth), from zero to maximum.
* P(macro-state) is the probability that a randomly selected agent will be in that macro-state, calculated as { the number of agents in that macro-state } / { the total number of agents }.
* and ln(1/P(macro-state)) is the natural logarithm of the reciprocal of that probability.

The factor F is of no particular interest to me at the moment. My inclination is to say it is usually a 1, and can be left out, but I really don’t know, so I leave it in as a place holder.

I can make this more specific by indicating what the conserved quantity is. I can also make it more compact with some notation. Let A be the total number of agents in the model. Let the conserved quantity of interest be wealth, the sum of real and claimable wealth. Let the macrostates of wealth be enumerated from smallest to largest by the number i, a positive natural number. Let Ai be the number of agents in macro-state i. Then:

|  |  |
| --- | --- |
| $$S\_{ABM}^{Wealth}=F\sum\_{State=i}^{}\left(\frac{A\_{i}}{A}\right)×ln⁡\left(\frac{A}{A\_{i}}\right)$$ | Equ 03 |

What are the limits on the size of the interval used to define a macro state? Using wealth as an example, an agent cannot have wealth less than zero. In the PMM, a closed system, total intrinsic value is conserved, so the maximum possible wealth of any one agent cannot exceed the total intrinsic value (TIV) of the model economy as a whole. But, in pragmatic terms, you can never have a single farmer or worker, because each sector of the economy is entirely dependent on the other sector remaining vibrant. So, in fact, the maximum wealth of a farmer would be a very small fraction of TIV, in practice. It could be determined empirically, but for the sake of figuring out a reasonable size for a macro-state determination, that would not help. Having never actually seen a worked example of a calculation of thermodynamic entropy for a gas, I am not sure how the range of possible values is determined there. I suppose it’s based on the Kelvin temperature scale, and a whole degree separates different macro-states.

I suppose the easy answer is to determine the agent with the maximum wealth, and divide that wealth level by {some fraction of the number of agents }. For example, if the maximum wealth held by an agent is M, if I want there to be an average of n agents per state, and if there are a total of A agents in the model, then a formula for calculating a suitable interval V for the macro-state would be:

|  |  |
| --- | --- |
| $$V=\frac{M}{^{A}/\_{n}}=\frac{nM}{A}$$ | Equ 04 |

But A varies with time, M varies with time, and an average value for A and M only makes sense in a steady-state (or, at least, stationary state). At best, it seems, I must establish the best V empirically for the PMM.

This raises a few questions:

* Does SABM vary a lot as V is varied? Or is it constant? Do I need to define SABM(V) as a function of V?
* Given that the PMM has two different types of mortal agents (workers and farmers), with radically different average wealth:
	+ Do I need a different value of V for each type of agent?
	+ Can I calculate S for each type and add them, or must I calculate the entropy all at once.
* As A varies, is the computation of ABM entropy comparable from tick to tick, or does the change in A make the values non-comparable?
* Is Stirling’s approximation the best formula to use in this case? Or should I be using the GammaLn() formula, a somewhat more difficult formula, or maybe, the original combinatorial formula?

An approach to resolving these questions might be:

1. Build an Excel spreadsheet that generates random wealth data for two populations, farmers and workers.
2. For each population:
	1. See if varying the size of V makes a difference.
	2. Calculate the entropy for a fraction of the population;
	3. Calculate the entropy for the whole population;
	4. See whether the entropy is comparable for small A and larger A.
3. Develop three equations for SABM using Stirling’s approximation, the GammaLn() function, and the combinatorial formula.
4. Determine whether any formulae indicate circumstances for being additive, or comparable.
5. For both populations, calculate entropy and see under what circumstances the entropy is additive, or can be combined.
6. ??

Hmmm?! Some of this has been addressed already. I just need to clean it up a bit with the goal of answering the questions more pointedly.

Material added 141230

Step 1 of the above plan was completed and is the spreadsheet at Ref E.

Step 2.1 was done as part of the same spread sheet, and the results are below. It was tricky to set up. In the first sheet I randomly generate the populations of agents. In the second sheet I select and fix one such set of populations (wrkrs and frmrs), create automated histograms using pivot tables, and compute entropy for each histogram. I then run a macro that alters the V (bin width), refreshes the pivot tables and histograms, recomputes entropy, and stores the result.

When I constructed the histograms, I used two different routines for determining the start point of the first bin: fixed and float. For ‘fixed’ start points, each interval starts on a multiple of the bin width V. So, if V is 5, the first bin starts on some multiple of 5. If V is 2.144553, then the first bin starts on some multiple of that weird number. For ‘float’ the process is a little more complicated. I compute the extent of the range as R1 = Xmax – Xmin. I then compute how many bins are needed to completely cover this range as K = ceiling( R1 / V ). I then compute an inclusive cover for the range R2 = K x V and a mid-point of the range as M = (Xmax + Xmin) / 2. Finally, the start point of bin 1 is S1 = M – (R2 / 2). I thought that placing the range R1 precisely in the middle of the interval covered by the histogram would be a good idea, but, apparently it introduced some bias away from end bins, some edge effect for the histogram. The ‘fixed’ approach seems to avoid such accidental edge-effect bias.

**Figure 04 – Wrkrs**



To be clear, there is a fixed population of wrkrs having fixed wealth. I make two wealth histograms for the wrkrs, one for each method of finding S1. I compute the entropic index of the histogram using the formula derived using Stirling’s approximation for ln(N!) = N\*ln(N)-N. (See Ref D.)

I set it up so R1 for wrkrs is about 75 and for frmrs is about 150. Bin width (V) rises in increments of 0.25, and entropy is calculated for each increment, for both methods of finding the start point of the cover, for both wrkrs and frmrs. The graph for wrkr data is above in Figure 04.

I have also included the graph line of the reciprocal of the number of bins required to cover the range of values.

Because all agents have wealth of approximately the same amount, I expected an entropic index of about 0.8 or 0.9. The calculated entropic index for bin widths from 0.25 to 10.0 seems to be relatively stable, with a peak around V = 3.0. It seems neither method is trustworthy if V > R1/3

**Figure 05 – Wrkrs – detail**



Figure 05 shows some detail from Figure 04. The peak and most stable values for the entropic index are found, roughly speaking, in the range V ∈ [3.0, 6.0] with some regular variation starting to appear at about V=6.0.

Following, in Figures 06 and 07, are the same types of graph but for the population of frmrs.

Based on these charts, it would seem that the ‘floating’ method gives stable but bad results for large V whereas the ‘fixed’ method gives unstable but bad results in that range. For both methods, too small a V results in a slightly depressed value, but not markedly so. Both techniques seem to give a stable peak value for entropic index when V ∈ [R1 / 25, R1 / 12 ]. Or, if K is the number of bins, K ∈ [ 12, 25 ].

**Figure 06 – Frmrs**



**Figure 07 – Frmrs - detail**



Here’s an observation, before moving on. Why does it dip on the left? Shouldn’t it rise? If V is very small, most bins will be empty, contributing nothing to the entropic index, and all non-empty bins will have 1 count each. The entropic index should be 1.0, should it not? Maybe I can do a run for teeny tiny V and see what happens. [ Done and done! See figure 08. ]

**Figure 08 – Wrkrs and Frmrs – Detail – V ∈ [0,3.5] in increments of 0.01**





Now, that is a surprising result! I did this by making a modification to the while loop of the macro in Ref E and saved it as Ref F, and making it terminate at zero vice R1. I also set the increment from 0.25 to -0.01, and initialized V to 3.5.

Returning to my ‘plan’. I have not done steps 2.2 – 2.4 yet.

Step 3 was done at Ref G back in April.

Step 4 I know I have done somewhere, but can’t find it at the moment. Maybe I didn’t diarize it. Oops!

(ZZZ find it, or redo it.)