Review: Statistical physics of Self-Replication by Jeremy England

This work argues for thermodynamic constraints placed on any physical unit which (unaided[1]) makes copies of itself. The fundamental result is that the heat released during such copying depends not merely on the forward process of replication but also on the hypothetical process of "inverse replication" (or less fantastically the degradation of one of the replicated units). The author applies this work to establish constraints on the heat released during the self-replication of an RNA enzyme and to estimate the minimum energy released during bacterial cell division.

- **What questions does this work answer?** This work seeks to establish thermodynamic limits on the net growth rate or maximum fitness associated with self-replication as determined by the durability of the replicator, the heat released, and the entropy change during a single round of replication.
- **Why are these questions important?** Self-replication exists as the foundational motif of asexual reproduction but it also extends to any biological context when a macromolecule manages to make a copy of itself through the function of a single template. Understanding the thermodynamic limits of this process would represent a solid step in developing basic physical theories of living systems.
- **How does the work answer these questions?** The work uses a non-equilibrium statistical physics equation relating forward and backward transition rates to the heat produced during a forward transition. The end result is a form of the second law of thermodynamics which is corrected by an "entropic contribution due to irreversibility".
- **What are applications to systems/problems?** The author applies his results to self-replicating RNA molecules and *E. coli* reproduction. The author uses the former application to consider why RNA (and not DNA) was the genetic material of prebiotic life, and he uses the latter to consider the efficiency of bacterial metabolism.

Originality Disclaimer: These notes represent the transcriber's most faithful attempt at interpreting many of the results in [1] and thus do not at all represent any original work on the part of the transcriber. All credit for this work goes to the original author.

1 Introduction

The process of organismal self-replication is statistically irreversible in that it is much more likely for a bacterium, for example, to split itself into two copies than for two copies to combine back into a single bacterium. From a physical standpoint this irreversibility contains a hint for how replication processes can be constrained by physical laws. In statistical physics notions of irreversibility are always concerned with entropic considerations. However, because living systems operate far from thermodynamic equilibrium and hence cannot be modeled by a Boltzmann distribution describing energy configurations, it it difficult to speak generally of *the* theory of the statistical physics of living systems.

Still, to motivate a small aspect of a theory of self-replication, we begin by establishing a consequence of detailed balance which relates entropy changes to irreversibility. Consider a system with a fixed particle number N and volume V in contact with a heat bath of inverse temperature β. If we label the microstates of

[1] Transcriber Note: The fact that this application is unaided makes DNA replication with DNA polymerase inapplicable in this scenario.

this system as $i,\ j,\ k, \ldots$ with associated energies $E_i,\ E_j,\ E_k, \ldots$, then at thermal equilibrium the following relation holds:

$$T(i \to j; \tau)\frac{e^{-\beta E_i}}{Z(\beta)} = T(j \to i; -\tau)\frac{e^{-\beta E_j}}{Z(\beta)} \tag{1}$$

where $Z(\beta)$ is the canonical partition function of the system. The parameter τ is a time scale defining the time between occupying the two states i and j. The quantity $T(i \to j;\ \tau)$ is the conditional probability that the system will be found to be in a microstate j at time $t = \tau > 0$ given that it started in microstate i at time $t = 0$. Eq.(**??**) asserts that when the system has relaxed to thermal equilibrium and achieved a Boltzmann distribution over microstates (i.e., $\lim_{t\to\infty} P_1(i,t) = e^{-\beta E_i}/Z(\beta)$) then the probability currents connecting i to j in the forward and reverse direction are equal. We must emphasize that contrary to the claims of the paper, Eq.(1) does not follow from detailed balance, but is instead a mathematical property of stationary Markov Processes. If Eq.(1) was symmetric in time and T referred to a conditional probability rate, then it would be equivalent to detailed balance. See Appendix Eq.(**??**).

The heat expelled into the bath during the transition from i to j is $\Delta Q_{i\to j} = -\Delta E_{i\to j} = -(E_j - E_i)$. By the relationship between (dimensionless) entropy and heat, $\beta\Delta Q_{i\to j} = \Delta S_{\text{bath}}^{i\to j}$, the amount by which the entropy changes over the transition is similarly constrained. We therefore have

$$\frac{T(j \to i; \tau)}{T(i \to j; -\tau)} = e^{-\Delta S_{\text{bath}}^{i\to j}}. \tag{2}$$

Eq.(2) relates the irreversibility of a transition to the amount of entropy which is introduced into the surroundings arising from the forward transition. The paper asserts that although this result was motivated/derived from the detailed balance condition which holds primarily for systems in thermal equilibrium, it is in fact (according to [2]) much more general and is valid for any transition between two microstates. Eq.(2) also exists as a starting point for achieving the larger goal of the work which is to show that the microscopic and quantitative relationship between irreversibility and entropy production has important physical consequences and establishes thermal production limits on far from equilibrium macroscopic biological processes.

2 England's Inequality

The paper's goal is to obtain thermodynamic constraints on the macroscopic[2] transition between two arbitrary course grained states. We begin by characterizing the state of our system by a variable $x(t)$. The starting and ending points of the system are given by $x(0) = i$ and $x(\tau) = j$. Starting from the equation

$$\frac{T[x(\tau - t)]}{T[x(t)]} = \exp\left(-\beta\Delta Q[x(t)]\right) \tag{3}$$

(taken from a definition in [2]), we average the exponential of forward heat production over all paths connecting i to j to obtain

$$\frac{T(j \to i;\ \tau)}{T(i \to j;\ -\tau)} = \left\langle \exp\left[-\beta\Delta Q_{i\to j}^{\tau}\right]\right\rangle_{i\to j}. \tag{4}$$

We next move toward talking about our system in macroscopic terms. Suppose there is a course grained condition **I** for the system. For example, **I** could define the system as containing one healthy, exponential growth phase bacterium at the start of a round of cell division. If the criteria defining **I** are satisfied, we can can then associate this macroscopic state with a space of microstates $\Omega(\mathbf{I})$ and a probability $P_1(i)$ for $i \in \Omega(\mathbf{I})$ to be in a microstate i given that the system was prepared under controlled conditions and observed to be

[2] Transcriber Note: The distinction in this section between microscopic and macroscopic seems reminiscent of the distinction between two *scales* of physics, and the idea that we average a quantity over all paths connecting two points is reminiscent of the *path integral*. So it seems these ideas can be made somewhat more concrete, or at least more physically rigorous if they are couched in the language of the renormalization group and probability functionals. See the last section for more details

in a macrostate **I**.

We now let a time interval τ pass, while keeping the system in contact with a heat bath of inverse temperature β. After this time τ, we define another coarse grained observable whose criteria we label as **II** and the associated space of microstates as $\Omega(\mathbf{II})$. For example, **II** could define the system as containing precisely two healthy bacteria cells at the start of cellular division. If criteria **II** are satisfied, then we can define $P_1(j)$ for $j \in \Omega(\mathbf{II})$ as being the likelihood of the system being in a microstate j given that it was found to be in macrostate **II**.

Now to give a macroscopic definition of irreversibility, we define[3] To give a macroscopic definition of irreversibility, we must introduce some formalism. First we establish the probability density definitions: $P(i, t_1; j, t_2)$ is the probability density to be at i at t_1 and j at t_2; $P(j, t_1; i, t_2)$ is the probability to be at i at t_2 and j at t_1. If we let $\Omega(\mathbf{I})$ and $\Omega(\mathbf{II})$ define the space of microstates associated with the macrostates **I** and **II** respectively, then

$$\Pi(\mathbf{I} \to \mathbf{II}; t_2 - t_1) \equiv \int_{\Omega(\mathbf{I})} di \int_{\Omega(\mathbf{II})} dj\, P_2(i, t_1; j, t_2) \qquad (\text{for } t_2 - t_1 > 0) \tag{5}$$

is the total probability to be in macrostate **I** at t_1 and macrostate **II** at t_2, while

$$\Pi(\mathbf{II} \to \mathbf{I}; t_2 - t_1) \equiv \int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, P_2(j, t_1; i, t_2) \qquad (\text{for } t_2 - t_1 > 0) \tag{6}$$

is the total probability to be in macrostate **II** at t_1 and macrostate **I** at t_2. We take both definitions to be only valid for $t_2 > t_1$, so that we are only considering forward propagation in time, and we thus cannot move from Eq.(5) to Eq.(6) by merely switching time endpoints. By the formal definition of conditional probabilities, and (taking this process to be stationary) we can then write

$$\begin{aligned} \Pi(\mathbf{I} \to \mathbf{II}; t_2 - t_1) &\equiv \int_{\Omega(\mathbf{I})} di \int_{\Omega(\mathbf{II})} dj\, P_2(i, t_1; j, t_2) \\ &= \int_{\Omega(\mathbf{I})} di \int_{\Omega(\mathbf{II})} dj\, T(i \to j; t_2 - t_1) P_1(i) \qquad (7) \\ \Pi(\mathbf{II} \to \mathbf{I}; t_2 - t_1) &\equiv \int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, P(j, t_1; i, t_2) \\ &= \int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, T(j \to i; t_2 - t_1) P_1(j) \qquad (8) \end{aligned}$$

The quantity $\Pi(\mathbf{I} \to \mathbf{II})$ is the likelihood that a system prepared in **I** is observed to satisfy **II** after some time $t_2 - t_1$, and $\Pi(\mathbf{II} \to \mathbf{I})$ is the likelihood that after another time $t_2 - t_1$ the same system would be observed again to satisfy **I**. We depict this process in Fig. 1.

Taking the ratio of these two quantities we find

$$\begin{aligned} \frac{\Pi(\mathbf{II} \to \mathbf{I}; t_2 - t_1)}{\Pi(\mathbf{I} \to \mathbf{II}; t_2 - t_1)} &= \frac{\displaystyle\int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, T(j \to i; t_2 - t_1) P_1(j)}{\displaystyle\int_{\Omega(\mathbf{I})} di \int_{\Omega(\mathbf{II})} dj\, T(i \to j; t_2 - t_1) P_1(i)} \\ &= \frac{\displaystyle\int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, \frac{T(j \to i; t_2 - t_1)}{T(i \to j; t_2 - t_1)} \frac{P_1(j)}{P_1(i)} T(i \to j; t_2 - t_1) P_1(i)}{\displaystyle\int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, T(i \to j; t_2 - t_1) P_1(i)} \end{aligned}$$

[3] Transcriber Note: The main result of the paper hinges mostly on the form of these definitions. I have not found any references which affirm them.

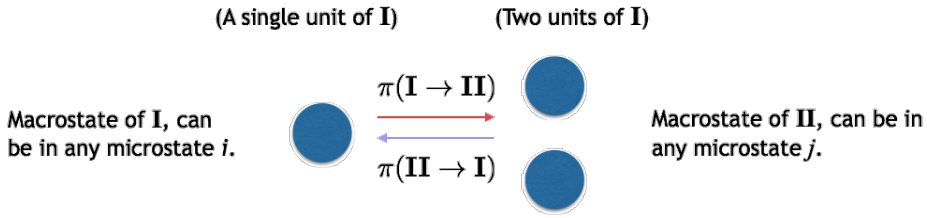

Figure 1: Self-Replication Process: A single unit in the system defines **I** and is converted into two units, which define **II**, at a probability rate $\Pi(\mathbf{I} \to \mathbf{II})$. The reverse process (from **II** to **I**) also occurs and does so at a probability rate $\Pi(\mathbf{I} \to \mathbf{I})$.

$$= \frac{\int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, e^{-\beta \Delta Q_{t_2-t_1}(i\to j)}\, e^{-\ln P_1(i)/P_1(j)}\, T(i \to j; t_2 - t_1) P_1(i)}{\int_{\Omega(\mathbf{II})} dj \int_{\Omega(\mathbf{I})} di\, T(i \to j; t_2 - t_1) P_1(i)} \tag{9}$$

and so

$$\frac{\Pi(\mathbf{II} \to \mathbf{II}; t_2 - t_1)}{\Pi(\mathbf{I} \to \mathbf{II}; t_2 - t_1)} = \left\langle e^{-\beta \Delta Q_{t_2-t_1}(i\to j)} e^{-\ln P_1(i)/P_1(j)} \right\rangle_{\mathbf{I}\to\mathbf{II}}, \tag{10}$$

where $\langle \cdots \rangle_{\mathbf{I}\to\mathbf{II}}$ denotes an average over all paths from some i in the initial ensemble to some j in the final ensemble **II** with each weighted by its likelihood. We may rearrange Eq.(10) to write

$$1 = \left\langle e^{-\ln\left[\frac{\Pi(\mathbf{II} \to \mathbf{II}; t_2 - t_1)}{\Pi(\mathbf{I} \to \mathbf{II}; t_2 - t_1)}\right] + \ln\left[\frac{P_1(j)}{P_1(i)}\right]} e^{-\beta \Delta Q_{t_2-t_1}(i \to j)} \right\rangle_{\mathbf{I}\to\mathbf{II}}. \tag{11}$$

Given that $e^x \geq 1 + x$ for all (even negative) x, we can expand all of the exponentials to first order and find

$$1 \geq 1 - \ln\left[\frac{\Pi(\mathbf{II} \to \mathbf{II}; t_2 - t_1)}{\Pi(\mathbf{I} \to \mathbf{II}; t_2 - t_1)}\right] + \left\langle \ln\left[\frac{P_1(j)}{P_1(i)}\right] \right\rangle_{\mathbf{I}\to\mathbf{II}} - \beta \left\langle \Delta Q_{t_2-t_1}(i \to j) \right\rangle_{\mathbf{I}\to\mathbf{II}}, \tag{12}$$

which leads to England's inequality

$$\boxed{\beta \langle \Delta Q \rangle_{\text{bath},\, \mathbf{I}\to\mathbf{II}} + \ln\left[\frac{\Pi(\mathbf{II} \to \mathbf{I})}{\Pi(\mathbf{I} \to \mathbf{II})}\right] + \Delta S_{\text{sys},\, \mathbf{I}\to\mathbf{II}} \geq 0,} \tag{13}$$

where we suppressed the time dependence and we defined the change in the system's entropy[4] to be

$$\Delta S_{\text{sys},\, \mathbf{I}\to\mathbf{II}} \equiv - \left\langle \ln\left[\frac{P_1(j)}{P_1(i)}\right] \right\rangle_{\mathbf{I}\to\mathbf{II}}, \tag{14}$$

and the heat released into the bath as

$$\beta \langle \Delta Q \rangle_{\text{bath},\, \mathbf{I}\to\mathbf{II}} \equiv \beta \left\langle \Delta Q_{t_2-t_1}(i \to j) \right\rangle_{\mathbf{I}\to\mathbf{II}}. \tag{15}$$

In Eq.(13), if **I** and **II** are taken to represent identical macrostate criteria with the same space of microstates

[4]We note that this entropy is not 1, because j and i correspond to distinct spaces of the microstates.

$\Omega(\mathbf{I}) = \Omega(\mathbf{II})$ then we have $\Pi(\mathbf{I} \to \mathbf{II}) = \Pi(\mathbf{I} \to \mathbf{I})$ and the result reduces to

$$\beta \langle \Delta Q \rangle_{\text{bath, } \mathbf{I} \to \mathbf{II}} + \Delta S_{\text{sys, } \mathbf{I} \to \mathbf{II}} \geq 0, \tag{16}$$

which is the second law of thermodynamics. Thus the general relation Eq.(13) says something specific: the more irreversible a macroscopic process is (i.e., the greater $\Pi(\mathbf{I} \to \mathbf{II})$ is in comparison to $\Pi(\mathbf{I} \to \mathbf{I})$) the greater the minimum entropy released for the forward direction of the process. The paper notes that Eq.(13) applies to a wide range of far from equilibrium transitions between coarse grained starting and ending states.

3 Fitness, Durability, Heat, and Entropy

The generalized 2nd law of thermodynamics Eq.(13) applies in many far from equilibrium (non-replicatory) systems, but is particularly useful as a means by which to understand self-replication. The exponential growth which typifies self-replication is highly irreversible but Eq.(13) allows us to be precise about the properties of that irreversibility.

Suppose there is a simple self replicator living at inverse temperature β with a population $n \gg 1$ governed by the master equation

$$\dot{p}_n(t) = gn \left[p_{n-1}(t) - p_n(t) \right] - \delta n \left[p_n(t) - p_{n+1}(t) \right], \tag{17}$$

where $p_n(t)$ is the probability of having a population of n at time t and $g > \delta > 0$. When one new replicator forms (at a rate g) the internal entropy of the system is taken to change by Δs_{int} and an amount of heat Δq is released into the surroundings. A decay event defined by δ amounts to a reversion of the replicator back into the reactants which created it.

The probability that a single replicator will give birth to two replicators in a short time dt is $\Pi(\mathbf{I} \to \mathbf{II}) = gdt$. Similarly, the probability that two replicators will decay into one is $\Pi(\mathbf{II} \to \mathbf{I}) = \delta dt$. We can thus write Eq.(13), in its infinitesimal time form as

$$\beta \Delta q + \Delta s_{\text{int}} \geq \ln \left[\frac{g}{\delta} \right]. \tag{18}$$

The paper argues that a factor of 2 need not be included for the conversion of two replicators into a single replicator. The reasons are claimed to extend from two directions both associated with classical distinguishability: one, if we think of the process as bounding the entropy of self-replication and the mixing of particles then the possible mixing entropy term cancels the conjectured factor of 2; two, we can define the coarse grained state to only consider the probability current for the reversion of the replicator which was just created.

Eq.(18), by the requirement of positive total entropy production implies $g > \delta$ which is the basic requirement for population growth[5]. Assuming we have fixed δ, Δs_{int}, and Δq we can find the maximum net growth rate:

$$g_{\text{max}} - \delta = \delta \left(\exp[\beta \Delta q + \Delta s_{\text{int}}] - 1 \right). \tag{19}$$

Eq.(19) also defines the maximum fitness of our replicator. We see that this fitness is fixed by three properties of the self-replicator: internal entropy production, Δs_{int}; durability, $1/\delta$; and heat dissipated into the bath during self-replication, Δq . This result implies that (for fixed durability and entropy change) replicators which release more heat during self-replication have a higher fitness. Thus Eq.(19) provides a physically motivated and quantitative link between reproductive fitness and metabolic efficiency: a self-replicator's maximum fitness is defined by how efficiently[6] it uses the energy resources in its own environment.

[5] Transcriber Note:This assertion in the paper seems incorrect. For example, does it imply that if a population is dying that total entropy production is decreasing? This would seem false seeing as a system of dead/degraded biological units would surely have more entropy than their formerly alive biological antecedents.

[6] Transcriber Note: Can this efficiency be defined more concretely if we considered the replicator to be a heat engine?

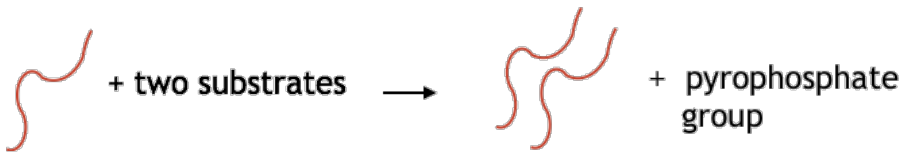

Figure 2: Self Replicating RNA: In [3] the catalysis by an RNA enzyme of a reaction between two substrate leads to a replication of the RNA enzyme and the release of a pyrophosphate group.

A subtler result is that the less organized a replicator (i.e., the higher Δs_{int}) and the less durable (i.e., the higher δ) the less metabolic energy must be harvested at minimum to achieve a certain growth rate. Thus it is to the advantage of certain replicators to be simpler and more prone to spontaneous degradation.

4 Self-Replicating Polynucleotides

The paper applies the formalism of Eq.(18) to self-replicating polynucleotides. Namely, [3] describes a reaction where two substrates A and B combine to form an RNA enzyme under a reaction which is catalyzed by a copy of the final RNA enzyme itself. The reaction displayed in Fig. 2 is

$$\text{RNA-ase} + \text{A} + \text{B} \quad \longrightarrow \quad 2\ \text{RNA-ase} + \text{P}_2\text{O}_7^{4-} \tag{20}$$

with $\text{P}_2\text{O}_7^{4-}$ a pyrophosphate group released upon catalysis. In [3], it is found that this reaction takes about an hour to occur and in [4] another RNA enzyme is argued to have a half-life on the order of 4 years. Thus $g \simeq \ln 2/1$ hr and $\delta \simeq \ln 2/\ 4$ yrs.

Since Eq.(20) begins with three units and ends with three units (we assume the RNAs are not identical), the entropy change for this reaction is presumed to be negligible. Thus Eq.(18) establishes the total heat bound (per mole) to be

$$\langle \Delta Q \rangle_{\text{RNA}\to 2\ \text{RNA}} \gtrsim RT \ln \left[\frac{4\ \text{years}}{1\ \text{hour}} \right] \simeq 6\ \text{kcal/mol} \tag{21}$$

where we took the reaction to occur under room temperature[7] $T = 298$ K conditions. Eq.(21) indicates that the heat released in this reaction needs to be at least 6 kcal/mol. Experimental data indicates that the actual the heat released for the ligation reaction of nucleotides is on the order of 10 kcal/mol [5], and thus this RNA enzyme operates close to its thermodynamic efficiency.

The paper compares this result to that obtained from a similar (but entirely hypothetical) process that could occur with single stranded DNA. Given that DNA is much more stable than RNA (the former having a half life of 3×10^7 years), the minimum enthalpy of the reaction is much higher. We find then that

$$\langle \Delta Q \rangle_{\text{single DNA}\to\ \text{double DNA}} \gtrsim RT \ln \left[\frac{3 \times 10^7\ \text{years}}{1\ \text{hour}} \right] \simeq 16\ \text{kcal/mol}, \tag{22}$$

which suggests that the heat released for this reaction is higher than the 10 kcal/mol associated with a ligation reaction. Thus this self replication process for DNA is forbidden. The paper interprets this result as indicating that the higher durability of DNA requires a much greater entropic-heat expenditure associated with self-replication of DNA if this self-replication is to occur at the same rate as that of RNA.

[7] Transcriber's note: We note that the value cited in the paper is 7 kcal/mol. This appears to be because the chosen temperature is different.

The paper thus interprets (according to Eq.(19)) RNA's greater fragility (i.e., larger degradation rate δ) as a fitness advantage which allows it to grow at a given rate given a smaller necessary energy release Δq than other replicators which are more durable. The paper suggests this fact is interesting given the current understanding that RNA (and not DNA) functioned as the replicating nucleic-acid molecule for pre-biotic life. The implicit argument is that for life forms in which the complex machinery of polymerases was not present, it was thermodynamically advantageous to have a genetic molecule which was sufficiently unstable as to be thermodynamically allowed to undergo self-replication without additional an molecular apparatus.

5 Bacterial Cell Division

For this section the paper considers preparing a large system initially containing a single *E. coli* bacterium in a nutrient rich medium which is itself in contact with a heat bath held at the cell's optimal growth temperature ($1/\beta \sim k_B T \sim 4.3 \times 10^{-21}$ J). We take the cell to be in its exponential growth phase and that the volume and mass of the entire system (that is, bacterium and nutrient medium) is held fixed, but the system is well aerated and open to the earth's atmosphere. We summarize these aforementioned experimental conditions as **I**.

The paper makes three points extending from the fact that the system consists not only of the bacterium but also of the surrounding nutrient medium and the oxygen the bacteria respirates.

1. **Defining $\mathbf{\Delta S}_{\text{sys, }\mathbf{I}\to\mathbf{II}}$:** The entropy changes associated with the bacterium's metabolic processes during growth and division are parts of $\Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}}$.

2. **Shannon Entropy and Statistical Mechanics:** Even far from equilibrium, the Shannon entropy remains a measure of statistical disorder, and we can write $\Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}} = \ln[\Omega_{\mathbf{II}}/\Omega_{\mathbf{I}}]$. That is, the Shannon entropy effects a log-scale comparison of phase-space volumes before and after the transition. The main implication is that changes in internal entropy involve, for example, changes in the partial volumes (or pressures) of surrounding gasses.

3. **Transition Probability and Entropy:** Over any interval of time τ, the transition probability $T(i \to j; \tau)$ obeys the equilibrium result Eq.(2). The eventual steady state of the system reduces to one described by the canonical ensemble and in which detailed balance holds. Therefore, what makes this scenario a far-from-equilibrium one is that the initial conditions $P_1(i)$ for $i \in \Omega(\mathbf{I})$ (i.e., microstate i in the space of microstates corresponding to macrostate **I**) correspond to a *highly* non-Boltzmann distribution[8] over microstates.

The natural definition of the final state **II** for this self-replication is a system with two bacteria in the media with the surrounding atoms rearranged into new molecular combinations.

Our goal is to estimate a lower bound for $\beta\langle\Delta Q\rangle_{\text{bath, }\mathbf{I}\to\mathbf{II}}$ in this self-replication process. By Eq.(13), we have

$$\beta\langle\Delta Q\rangle_{\text{bath, }\mathbf{I}\to\mathbf{II}} \geq -\ln\left[\frac{\Pi(\mathbf{I}\to\mathbf{I})}{\Pi(\mathbf{I}\to\mathbf{II})}\right] - \Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}}. \tag{23}$$

If we denote $p_{\text{div}}(t)$ as the probability that a cell divides in time t, we have simply

$$\Pi(\mathbf{I}\to\mathbf{II}) = p_{\text{div}}(t). \tag{24}$$

Now to turn our attention to $\Pi(\mathbf{I}\to\mathbf{I})$. First, the scenario presumably described by the transition $\mathbf{II}\to\mathbf{I}$ (i.e., the recombination of two daughter cells into a single mother cell) is so unlikely as to bring to mind the the quantum calculations of classically forbidden events such as the chances your coffee mug will "tunnel"

[8] Transcriber's note: We should note the claim of a "highly non-Boltzmann" distribution over microstates $P_1(i)$ (for $i \in \Omega(\mathbf{I})$) seems to contradict the authors point 2. above where he presumably takes the probability to be of a constant microcanonical ensemble form.

through your desk to the floor below. Intuition would suggest such events are zero, but quantum physics tells us they are just absurdly small.

For our purposes, we can similarly argue that the spontaneous recombination of two daughter cells is certainly unlikely but has a non-zero probability, and more importantly is less likely than other more easily calculable events. Specifically, although the probability (which we cannot calculate) that two daughter cells recombine into a mother cell is very low, the probability (which we can calculate) that all the peptide bonds within a daughter cell spontaneously hydrolyze/break is also low and acts as an upper bound of the former.

The paper computes the probability for complete spontaneous hydrolysis of a cell using the parameters n_{pep} (the number of peptide bonds in the cell), τ_{div} (the division time of the cell), and τ_{hyd} (the half-life of the peptide bond). With the first two parameters the rate of peptide growth r is given by

$$r = n_{\text{pep}}/\tau_{\text{div}} \tag{25}$$

The author considers each daughter cell in **II** in turn but first computes the probability that all the peptide bonds in a single cell hydrolize.

The probability that all the peptide bonds in a single cell hydrolize has two parts. Starting our origin of time at $t = 0$, the probability that a single peptide bond is hydrolized at time t is taken to be

$$p_{\text{hyd},1}(t) = 1 - e^{-t/\tau_{\text{hyd}}} \simeq t/\tau_{\text{hyd}}, \tag{26}$$

where we make the assumption that for the time scales of interest $t \ll \tau_{\text{hyd}}$.

We can take the probability that $n < n_{\text{pep}}$ peptide bonds are hydrolized to be modeled as a decay process akin to that in Chapter 5 of [6]. Namely, this probability[9] is

$$p_{\text{hyd},n}(t) = \binom{n_{\text{pep}}}{n} \left(1 - e^{-t/\tau_{\text{hyd}}}\right)^n e^{-(n_{\text{pep}}-n)t/\tau_{\text{hyd}}} \simeq \binom{n_{\text{pep}}}{n} \left(t/\tau_{\text{hyd}}\right)^n \left(1 - t/\tau_{\text{hyd}}\right)^{n-n_{\text{pep}}} \tag{27}$$

Thus the probability that all n_{pep} peptide bonds are hydrolized is $p_{\text{hyd},n_{\text{pep}}} \simeq (t/\tau_{\text{hyd}})^{n_{\text{pep}}}$. However, during this time from 0 to t, the paper assumes the cell is continually making peptide bonds, so that at time t, n_{pep} is no longer the total number of peptide bonds. Instead there is an additional number of peptide bonds $n_{\text{new}} = rt = n_{\text{pep}}t/\tau_{\text{hyd}}$. To find the probability that the all the bonds (both old and new) in the cell hydrolyze, we need to compute the probability that these new bonds hydrolize and multiply our $p_{\text{hyd},n_{\text{pep}}}$ by this new probability. Using a similar argument, the probability for the complete hydrolysis of the new bonds is

$$p_{\text{hyd, new}}(t) = \left(1 - e^{-t/\tau_{\text{hyd}}}\right)^{n_{\text{new}}} \simeq (t/\tau_{\text{hyd}})^{rt}. \tag{28}$$

Thus the probability for the complete hydrolysis of a cell is given by

$$p_{\text{hyd, complete}}(t) = p_{\text{hyd},n_{\text{pep}}}(t) p_{\text{hyd, new}}(t) \simeq (t/\tau_{\text{hyd}})^{n_{\text{pep}}+rt}. \tag{29}$$

The author states it would be pertinent to compute the maximum of Eq.(29) to obtain an upper bound for this probability. We expect such an upper bound because for $t \simeq 0$, we see that this probability goes to 0 and also for large t (still with $t \ll \tau_{\text{hyd}}$)this probability again goes to zero because the rt term is much larger than 1. Presuming it does not remain constant between those limits, Eq.(29) should have a local maximum. The time where this local maximum exists can be found by differentiating $\ln p_{\text{hyd, complete}}(t)$. Doing so we find the time t_{max} for maximum probability is given by the condition

$$1 + t_{\text{max}}/\tau_{\text{div}} + t_{\text{max}}/\tau_{\text{div}} \ln[t_{\text{max}}/\tau_{\text{hyd}}] = 0 \tag{30}$$

or

$$\tau_{\text{div}}/t_{\text{max}} + 1 + \ln[t_{\text{max}}/\tau_{\text{hyd}}] = 0 \tag{31}$$

[9] Transcriber's note: We do not model this as a poisson process because the number of events is bounded above by n_{pep}.

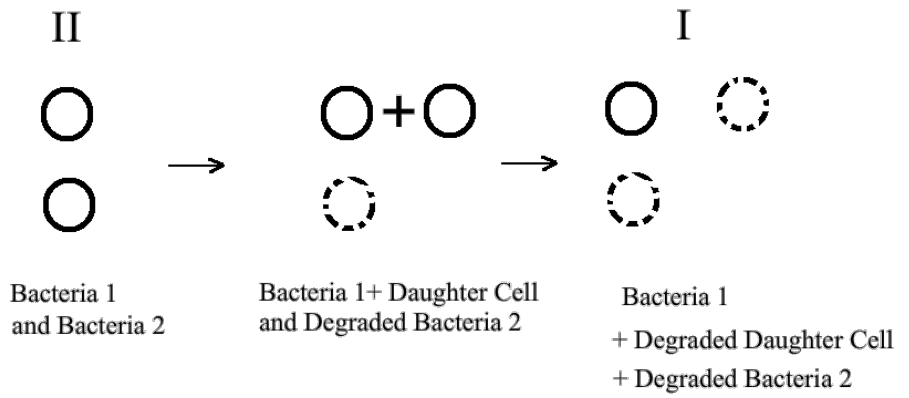

Figure 3: Transcriber's interpretation of transition from **II** to **I**: In moving from **II** to **I**, one of the cells degrades and the other divides into two cells one of which degrades.

which implies (given $t_{\text{max}} \ll \tau_{\text{hyd}}$) that $t_{\text{max}} \ll \tau_{\text{div}}$. We can then write the inequality

$$p_{\text{hyd, complete}}(t) \leq (t_{\text{max}}/\tau_{\text{hyd}})^{n_{\text{pep}}+rt_{\text{max}}} < (t_{\text{max}}/\tau_{\text{hyd}})^{n_{\text{pep}}}. \tag{32}$$

And thus we have

$$\ln p_{\text{hyd, complete}}(t) < n_{\text{pep}} \ln[t_{\text{max}}/\tau_{\text{hyd}}] = -n_{\text{pep}} \left(\tau_{\text{div}}/t_{\text{max}} + 1\right), \tag{33}$$

where we used Eq.(31) in the last equality. From, here the paper now considers other cell in **II** which has remained alive. In the words of the author:

> We must, however, consider an alternative mechanism for the most likely $\mathbf{II} \to \mathbf{I}$ transition: it is possible that a cell could grow and divide in a time slightly less than τ_{div}. If, subsequent to such an event the daughter cell of the recent division were to spontaneously disintegrate back into its constituent nutrients (with log-probability at the most on order of $\ln p_{\text{hyd, complete}}(t_{\text{max}})$), we would complete the interval with one recently divided continually growing bacterium in our system, that is, we would have returned to the **I** ensemble.

The transcriber interprets this to mean that we begin with a state of two cells defining **II**. One of these cells die by disintegration with a probability as was computed in the previous paragraphs (thus leaving us with one cell). The other cell divides in a time less than τ_{div} into two daughter cells (thus giving us two cells). Then one of these second generation daughter cells also disintegrates, thus returning us to our initial state **I** of a growing single cell. This situation is depicted in Fig. 3. The probability of this entire process (the two disintegrations and a single division)[10] gives the upper bound to the probability of the $\Pi(\mathbf{I} \to \mathbf{I})$ transition:

$$\Pi(\mathbf{I} \to \mathbf{I}) \leq p_{\text{hyd, complete}}(t)^2 p_{\text{div}}(t). \tag{34}$$

And thus, we find

$$\ln\left[\frac{\Pi(\mathbf{I} \to \mathbf{I})}{\Pi(\mathbf{I} \to \mathbf{II})}\right] \leq 2 \ln p_{\text{hyd, complete}}(t) < -2n_{\text{pep}} \left(\tau_{\text{div}}/t_{\text{max}} + 1\right), \tag{35}$$

where we used Eq.(24) and Eq.(33). Given that $a < -b$ implies $-a > b$ and returning to Eq.(23), we obtain

$$\beta\langle \Delta Q\rangle_{\text{bath, }\mathbf{I}\to\mathbf{II}} \geq -\ln\left[\frac{\Pi(\mathbf{I} \to \mathbf{I})}{\Pi(\mathbf{I} \to \mathbf{II})}\right] - \Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}} > 2n_{\text{pep}} \left(\tau_{\text{div}}/t_{\text{max}} + 1\right) - \Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}}, \tag{36}$$

[10]The convoluted nature of this process deserves more discussion than what is provided here or in the original paper

or simply

$$\beta\langle\Delta Q\rangle_{\text{bath, }\mathbf{I}\to\mathbf{II}} > 2n_{\text{pep}}\,(\tau_{\text{div}}/t_{\text{max}}+1) - \Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}}. \tag{37}$$

Eq.(37) establishes the predicted heat bound for a replicating bacterium. It indicates that the minimum heat for self-replication is determined by the number of amino acids in the cell (n_{pep}), the entropy change for the system ($\Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}}$), how rapidly the bacteria copies itself (τ_{div}), and how long it takes the bacterium to disintegrate (t_{max}).

We now compare the theoretical bound predicted by Eq.(37) to what is known of the experimental bound. Taking the dry mass[11] of the single bacterium to be 0.3 picograms, we have $n_{\text{pep}} = 1.6\times10^9$.

The author assumes the entropy change in going from $\mathbf{I}\to\mathbf{II}$ comes in two parts: The conversion of oxygen into carbon dioxide and the compacting of free amino acids into their protein shape. The first part results in an increase in entropy due to the larger volume (and smaller partial pressure) associated with CO_2. Taking each polypeptide bond to take part in such a reaction, we find the entropy change is

$$\begin{aligned}\left(\Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}}\right)_{O_2\to CO_2} &\sim n_{\text{pep}}\Big[\ln V_{CO_2} - \ln V_{O_2}\Big]\\ &= n_{\text{pep}}\ln\left(\frac{p_{O_2}}{p_{CO_2}}\right)\\ &\simeq n_{\text{pep}}\ln\left(\frac{2\times10^3}{30}\right)\sim 6\,n_{\text{pep}}.\end{aligned} \tag{38}$$

The second part results in a decrease in entropy due to the smaller volume of the compacted amino-acid. Taking the volume the amino acid can explore within the compacted protein to be $V_f = 0.001$ nm^3 and the initial free space volume to be $V_i = 100$ nm^3, we find an entropy change of all amino acids to be

$$\begin{aligned}\left(\Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}}\right)_{\text{free a.a.}\to\text{bound a.a.}} &\sim n_{\text{pep}}\ln\left(\frac{V_f}{V_i}\right)\\ &= n_{\text{pep}}\ln\left(10^{-5}\right)\sim 12\,n_{\text{pep}}.\end{aligned} \tag{39}$$

Adding these two entropic contributions gives us

$$\Delta S_{\text{sys, }\mathbf{I}\to\mathbf{II}} \sim -6\,n_{\text{pep}}. \tag{40}$$

Evaluating the first term in Eq.(37) is more straightforward. Assuming a cell division time of $\tau_{\text{div}}\sim 20$ min and a spontaneous hydrolysis lifetime of $\tau_{\text{hyd}}\sim 600$ yrs, we find by Eq.(31) $t_{\text{max}}\sim 1$ min. And so

$$2n_{\text{pep}}\,(\tau_{\text{div}}/t_{\text{max}}+1)\sim 42\,n_{\text{pep}} \tag{41}$$

Assembling these results gives us

$$\beta\langle\Delta Q\rangle_{\text{bath, }\mathbf{I}\to\mathbf{II}} \gtrsim 48\,n_{\text{pep}}.. \tag{42}$$

Experimentally, the total amount of heat produced in a single division cycle for an *E. coli* bacterium is $\beta\langle Q\rangle_{\text{exp.}} = 220\,n_{\text{pep}}$. Two comments are thus in order.

- **Dominance of Irreversibility:** The relative contributions of Eq.(41) and Eq.(40) reveal that contrary to intuition, the dominant factor in determining the heat bound for self-replication is not the entropic cost of assembling the components of the cell in the appropriate way, but is rather the sheer irreversibility of the self-replication reaction as it creates cells that are not easily degradable.
- **Metabolic Efficiency:** The comparison between the theoretical lower bound for bacterial self replication ($48\,n_{\text{pep}}$) and the experimental result ($220\,n_{\text{pep}}$) would seem to suggest that *E. coli* operates far from its metabolic efficiency threshold and is thus rather inefficient. The paper argues that given that the

[11] I'm assuming the author calculated this by dividing 0.3 picograms by the average weight of two amino acids

bacterium is highly adaptable to its environmental conditions, it is not a surprise that it is not perfectly optimized for the one presented in this problem.

6 Concluding Comments

The crux of this work rests on the general idea that any non-equilibrium process (whether biological or not) in contact with a heat bath has certain constraints on heat exchanges as can be derived from a generalized second law of thermodynamics. This generalized second law includes the typical heat changes for the system and the surrounding bath, but includes an additional term which defines the entropy change due to the irreversibility of the process.

The paper asserts that the applicability of these ideas to self-replication rests on the ability to precisely define a "self" in a system according to a coarse graining of microstates. Such an explicit definition of a "self" then allows one to track probabilistically the dynamics of that self through time using the methods of non-equilibrium statistical physics.

The paper concludes by hoping that the work can lead to future developments which explore the connection between natural selection and non-equilibrium systems.

7 Stationary Markov Process and Eq.(1)

In this section we review the formalism behind Eq.(1). This discussion is mostly transcribed from Chapter 3 of [6].

For a stationary Markov Process, the conditional probability density for beginning in a state x_1 at t_1 and then later being found in a state x_2 at t_2 is a time dependent function of only the *difference* in times between the two states. Namely, we can write[12]

$$P_{1|1}(x_2, t_2|x_1, t_1) = T_\tau(x_2|x_1), \tag{43}$$

or, using the notation in the paper,

$$P_{1|1}(x_2, t_2|x_1, t_1) = T(x_1 \to x_2;\, \tau), \tag{44}$$

where T is a conditional probability density. By Eq.(44), the Chapman-Kolmogorov equation becomes

$$T(x_1 \to x_3;\, \tau + \tau') = \int_{\Omega_x} dx_2\, T(x_1 \to x_2;\, \tau)\, T(x_2 \to x_3;\, \tau') \tag{45}$$

which can be written as a matrix product as $T_{\tau+\tau'} = T_\tau T_{\tau'}$.

For this paper, the more important property for a stationary Markov process is the way it connects reversed transition processes. First, by basic probability theory, for two events A and B we have

$$P(A \cap B) = P(B \cap A). \tag{46}$$

Thus if we have two events x_1 at t_1 and x_2 at t_2, then we have

$$P_2(x_1, t_1; x_2, t_2) = P_2(x_2, t_2; x_1, t_1) \tag{47}$$

Eq.(47) states (arguably obvious fact) that the probability density for finding a particle at x_1 at time t_1 and at x_2 at time t_2 is independent of whatever ordering we choose for writing the events. Then by Bayes' Rule

[12]The notation $P_{k|\ell}$ denotes a conditional probability with ℓ events in the initial state and k events in the final state. The notation P_n denotes a conditional probability which is a function of n events.

and Eq.(44), we have

$$P_2(x_1, t_1; x_2, t_2) = T(x_1 \to x_2; t_2 - t_1)P_1(x_1) \tag{48}$$

and

$$P_2(x_2, t_3; x_1, t_1) = T(x_2 \to x_1; t_1 - t_2)P_1(x_2). \tag{49}$$

Thus we have by Eq.(47) the mathematical identity

$$\boxed{T(x_1 \to x_2; t_2 - t_1)P_1(x_1) = T(x_2 \to x_1; t_1 - t_2)P_1(x_2).} \tag{50}$$

We should emphasize that this is *not* a physically derived result and as such is distinct from the equation for detailed balance. The analogous result for detailed balance, would have the form

$$W(x_1 \to x_2;\, \tau)P_1(x_1) = W(x_2 \to x_1;\, \tau)P_1(x_2), \tag{51}$$

where τ is some infinitesimal time scale. Eq.(51) is distinguished from Eq.(50) by the fact that across the equal sign it is symmetric in time, and W refers to a conditional probability per unity time (i.e., a transition rate) while T refers to a conditional probability.

In contradiction to the claim in the paper, the starting point for the derivation of England's Identity is Eq.(50) and not Eq.(51)

8 Transcriber Comments

Transcriber's Note: The goal of these notes is to take the arguments and derivations of the paper and rewrite them in an as explicit form as possible. In this section, we interrogate and try to understand some of the more important arguments in the paper.

- **Experimental Probing on Eq.(4):** Common sense notions of heat changes might lead us to believe that the heat released in a thermal process should only be a function of the forward reaction rate. Eq.(4) contradicts such notions and if it is indeed true there should be a way to devise an experiment testing it. That is is there a way that one could fix the forward reaction rate for a process, modulate the reverse reaction rate, and then observe variations in the heat released?

- **Heat and Irreversibility:** The connection between heat and irreversibility posited in Eq.(3), stems from the paper . In the paper this connection seems to be posited more as a definition/assertion rather than a result derived from the foundations of non-equilibrium statistical mechanics. This is not to say that the result is wrong; only that it perhaps needs to be placed on a firmer footing.

- **Definition of Macroscopic Transition Amplitude:** The main theoretical work of the paper Eq.(13) rests on the definition of the macroscopic transition amplitude provided in Eq.(**??**). Conceptually, it consists of averaging over the probability of a transition from one initial microstate to a final microstate weighted by the probability to be in the initial microstate. The averaging occurs over the space of states. I have never seen such a definition of a macroscopic transition amplitude and it would help make these results more rigorous if such a definition existed.

- **Renormalization Group:** The connection between microscopic and macroscopic scales seems akin to the basic ideas of the renormalization group, namely whether the second law of thermodynamics for non-equilibrium statistical physics takes on a different (transition amplitude dependent form) when we move from one length scale (i.e., the microscopic system) to a larger length scale (i.e., the macroscopic system).

- **Non-Traditional Definition of Path Average:** The paper references a path average in its definition Eq.(4), but never explicitly writes one out. Such a path average is then extended to macroscopic states

in the definitions which make up Eq.(13). Intuitively, if I were to average a quantity over all paths connecting some initial state to some final state I would end up writing something which looks like a path integral. Would such a mathematical definition make this work more concrete? What would such a path integral look like?

- **Case of $\delta > g$:** If we consider Eq.(18), in the case where $\delta < g$ (a population which dies faster than it grows) we find that the total entropy change of the universe (that is both the bath represented by $\beta\Delta q$ and the system represented by Δs_{int}) is less than zero which contradicts the second law of thermodynamics. Whether such a result proves fatal to this formalism is contingent on whether Eq.(18) can really be applied in the case of an exponentially dying population (Although I do not see why it can?t).

- **Bacterial Cell Division:** In Section 5, I outlined the calculation for the reverse transition amplitude in bacterial cell division as faithful to the text as possible. However, the process which is claimed to define the **II** to **I** transition seems convoluted and unnatural. I think a more careful analysis of the factors relevant to cell division would need to be undertaken to in order to properly apply this formalism to the example provided in the paper.

References

[1] J. L. England, "Statistical physics of self-replication," *The Journal of chemical physics*, vol. 139, no. 12, p. 121923, 2013.

[2] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Physical Review E*, vol. 60, no. 3, p. 2721, 1999.

[3] T. A. Lincoln and G. F. Joyce, "Self-sustained replication of an rna enzyme," *Science*, vol. 323, no. 5918, pp. 1229–1232, 2009.

[4] J. E. Thompson, T. G. Kutateladze, M. C. Schuster, F. D. Venegas, J. M. Messmore, and R. T. Raines, "Limits to catalysis by ribonuclease a," *Bioorganic chemistry*, vol. 23, no. 4, p. 471, 1995.

[5] C. A. Minetti, D. P. Remeta, H. Miller, C. A. Gelfand, G. E. Plum, A. P. Grollman, and K. J. Breslauer, "The thermodynamics of template-directed dna synthesis: base insertion and extension enthalpies," *Proceedings of the National Academy of Sciences*, vol. 100, no. 25, pp. 14719–14724, 2003.

[6] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, vol. 1. Elsevier, 1992.