9 When and why does entropy increase?

MARTIN BARRETT AND ELLIOTT SOBER

1 Introduction

Like the Sirens singing to Ulysses, the concept of entropy has tempted many a thinker to abandon the straight and narrow course. The concept is well-defined for chambers of gases. However, the temptation to extend the concept has been all but irresistible. As a result, entropy is used as a metaphor for uncertainty and disorder. We have nothing against such extensions of usage and, in fact, will indulge in it ourselves. However, we do believe that the price of metaphor is eternal vigilance.\(^1\)

One example of the temptation can be found in R. A. Fisher's *Genetical Theory of Natural Selection*.\(^2\) In the second chapter of that book, Fisher states a result that he dubs the *fundamental theorem of natural selection*. The theorem states that the average fitness of the organisms in a population increases under selection, and does so at a rate given by the additive genetic variance in fitness. Fisher then proposes the following analogy:

> It will be noticed that the fundamental theorem ... bears some remarkable resemblances to the second law of thermodynamics. Both are properties of populations, or aggregates, true irrespective of the nature of the units which compose them; both are statistical laws; each requires the constant increase of a measurable quantity, in the one case the entropy of a physical system and in the other the fitness ... of a biological population [p. 39].

Fisher then quotes Eddington's famous remark that 'the law that entropy always increases - the second law of thermodynamics - holds, I think, the supreme position among the laws of nature.' Fisher comments that 'it is not a little instructive that so similar a law should hold the supreme position among the biological sciences.'

\(^1\) We borrow this phrase from R. C. Lewontin.

\(^2\) Fisher (1930).
9 When and why does entropy increase?

Fisher's analogy between the second law and the fundamental theorem does not lead him to deny that they differ. Among the differences he notes is the idea that 'entropy changes lead to a progressive disorganization of the physical world, at least from the human standpoint of the utilization of energy, while evolutionary changes are generally recognized as producing progressively higher organization in the organic world' (p. 40). Here we see the familiar idea that entropy is not simply a technical concept applicable to chambers of gases, but has a larger meaning as a measure of organization and order.

If order declines in thermodynamic processes, but increases in the process that Fisher described, why is this so? One standard suggestion is that the second law of thermodynamics applies to closed systems, but populations are able to evolve new adaptations because they extract energy from their surroundings. Although it is true that the second law applies to closed systems and that organisms need energy to survive and reproduce, we will see in what follows that this does not answer the question we have posed. For the fact of the matter is that entropy can decline in closed systems and it can increase in open ones. The distinction between closed and open systems is an important one, but it is not the key to the problem we wish to explore.

In order to fix ideas, we begin with an informal and we hope intuitive description of what entropy is, one that we will refine in what follows. Entropy is a property of probability distributions. If there are \( n \) kinds of events, each with its own probability \( p_1, p_2, \ldots, p_n \), then the entropy of the distribution is given by \(- \sum p_i \log p_i\). For \( n \) states, the highest entropy distribution occurs when the \( n \) possibilities are equiprobable; the lowest entropy distribution occurs when one of the states has probability 1 and the others have probability 0. Additional properties of entropy are given in appendix A.

To the extent that both gases and evolving populations can be characterized by probability distributions, each can also be described by the entropy concept. A population of organisms may be characterized by the \( n \) genes that can occur at a given locus. The frequency of a gene in the population defines what we mean by the probability that a randomly chosen organism will contain a copy of that gene. A population with the \( n \) genes in equal frequencies will have a higher entropy than a population in which all but one of those \( n \) genes has disappeared.

---

3 The entropy concept in physics is a property of the particular probability distributions employed in statistical mechanics.
Consider, by analogy, a chamber divided down the middle by a wall. In the left hand side, there are some molecules of oxygen; in the right hand side, there are some molecules of helium. When the dividing wall is removed, the molecules will reach a mixed homogeneous state. If we associate probability distributions with the gases before and after the wall is removed, we can compare the entropy of the system at those two times.

When in the process just described is entropy well-defined? If each isolated subsystem is at equilibrium before the wall is removed, a probability distribution can be provided and the entropies of each may be calculated. However, immediately after the wall is removed, the system is not at equilibrium. At this point, the theory of equilibrium thermodynamics says nothing about the system's probability distribution, and so its entropy is not defined. However, if we wait a while, the coupled system will move to a new equilibrium, and entropy once again makes sense. Within equilibrium thermodynamics, one cannot talk about the change in entropy that occurs when a system moves from an out-of-equilibrium state to its equilibrium configuration.\(^4\) What, then, does the second law of thermodynamics mean when it says that entropy must increase?

What we need to focus on is the coupling process.\(^5\) Two or more isolated systems are each at equilibrium. Then they are coupled and the new conjoint system is allowed to reach its new equilibrium. We then look at the equilibrium entropy before coupling and the equilibrium entropy after. The second law does not say that entropy increases on the way to equilibrium; it says that equilibrium entropy increases under coupling.

This picture of what the thermodynamic concept means should guide how we frame the question about entropy in nonthermodynamic contexts. The gene frequency distribution in an evolving population changes in every generation until it reaches an equilibrium point. An equilibrium occurs when the population stops changing.\(^6\) If we look for biological parallels to the thermodynamic idea, we must consider a coupling process and ask what is true of the probability distributions that characterize populations at equilibrium. So the appropriate analog is not a single population moving from an out-of-equilibrium state to

---

\(^4\) We leave to others the investigation of how entropy should be understood in non-equilibrium situations.

\(^5\) This is not the term customarily employed in thermodynamics, but it well describes the process, e.g., of bringing a system into thermal contact with a heat bath.

\(^6\) The definition of the equilibrium concept in population biology would have to be refined to take account of random fluctuations around equilibrium values due to sampling error. This refinement is not worth pursuing here.
an equilibrium configuration. Rather, we should consider two or more populations, each at equilibrium, which then are coupled. The new conjoint population is then allowed to reach its equilibrium. We then compare the pre-coupling equilibrium entropy with the equilibrium entropy that obtains after coupling. Under what circumstances will the second have a higher value than the first?

In section 2, we describe with more care what entropy increase means in thermodynamics. Then, in section 3, we discuss several biological examples in which populations are coupled and their pre- and post-coupling equilibrium entropies are compared. In section 4, we address the question that is the title of this chapter.

2 Entropy increase in statistical mechanics

The textbook explanation\(^7\) for the increase of entropy in statistical mechanics is appealingly simple. We start with two systems of particles (\(S_1\) and \(S_2\)). These could be chambers of gases or blocks of metal or any other kind of homogeneous physical system. Each system has, in this idealization, a definite energy \(E_i\); the total energy of the two systems is \(E = E_1 + E_2\). The energy \(E_i\) is realized by summing the contributions from each of the \(N_i\) particles in \(S_i\). The pair of numbers \((N_i, E_i)\) fixes the macrostate of system \(S_i\), since it describes macroscopic quantities which are (at least approximately) observable. There are many ways to assign individual energies to the \(N_i\) particles compatible with the system energy \(E_i\). Each such way is a microstate of \(S_i\), accessible to the macrostate \((N_i, E_i)\).

A simple example illustrates these ideas. Let each \(S_i\) consist of a row of coins. Each coin can be in one of two states: \(H\) (energy 1) and \(T\) (energy 0). Each microstate of the system is just a particular configuration of heads and tails. The energy of the system in that microstate is just the number of heads in the configuration.\(^8\) Table 1 shows \(S_1\) and \(S_2\) in microstates accessible to the macrostates \((N_1 = 2, E_1 = 1)\) and \((N_2 = 3, E_2 = 2)\) respectively. The total energy of the combined systems is \(1 + 2 = 3\).

The Fundamental Postulate of Statistical Mechanics provides a way of assigning probabilities to the microstates. It states that in a closed system (one with no changing external influences) at equilibrium, all microstates accessible to the

---


\(^8\) There are real physical systems, magnetic spin systems, whose statistical mechanics is virtually isomorphic to our example. We prefer the example for its vividness.
Table 1. Accessible microstates of systems 1 and 2

<table>
<thead>
<tr>
<th>System 1 (energy 1)</th>
<th>System 2 (energy 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H T</td>
<td>H H T</td>
</tr>
<tr>
<td>T H</td>
<td>H T H</td>
</tr>
<tr>
<td></td>
<td>T H H</td>
</tr>
</tbody>
</table>

Macrostate of the system are equally probable. In $S_1$, there are two accessible microstates in the macrostate $(2,1)$, since there are two ways of assigning one head to two coins. Each microstate thus has probability $1/2$. The system $S_2$ has three accessible microstates in the macrostate $(3,2)$, since there are three ways of assigning two heads to three coins; each microstate has probability $1/3$.

Underlying this assignment is the so-called ergodic assumption that if we could somehow observe a closed system at equilibrium repeatedly (without disturbing the equilibrium) to pin down its microstate, the accessible microstates would appear equally often.

The probability space specifying accessible microstates of equal energy and equal probability is called in statistical mechanics the microcanonical ensemble (or the microcanonical distribution). The concept of entropy provides a way to measure the width of this probability distribution. If there are $n$ accessible microstates with probabilities $1/n$, the entropy is given by $\log n$ (see appendix A). This is the quantity that the textbooks take as the definition of entropy, denoted $\sigma$, of a closed physical system. If we write $\#(N,E)$ for the number of accessible microstates in macrostate $(N,E)$, then $\sigma = \log \#(N,E)$.

Suppose that $S_1$ and $S_2$ are initially isolated from each other; this means that no energy or other physical quantity can be exchanged between them. Probabilistically this appears as the assumption that the two systems are independent. The macrostate of the two isolated systems taken together is given by $N = N_1 + N_2$ and $E = E_1 + E_2$; an accessible microstate is just a choice of an accessible microstate of $S_1$ together with an accessible microstate of $S_2$. This means that there are $\#(N_1,E_1) \times \#(N_2,E_2)$ accessible microstates of the combined system. All $2 \times 3 = 6$ accessible microstates for the combined example systems are shown in table 2. Since the logarithm of a product is the sum of the logarithms, we have for the entropy of the combined system $\sigma = \sigma_1 + \sigma_2$. 

234
9. When and why does entropy increase?

Table 2. Accessible microstates of combined system before coupling

<table>
<thead>
<tr>
<th>Microstates of uncoupled system (energy 3)</th>
<th>System 1</th>
<th>System 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>H T</td>
<td>H H T</td>
<td></td>
</tr>
<tr>
<td>T H</td>
<td>H H T</td>
<td></td>
</tr>
<tr>
<td>H T</td>
<td>H T H</td>
<td></td>
</tr>
<tr>
<td>T H</td>
<td>H T H</td>
<td></td>
</tr>
<tr>
<td>H T</td>
<td>T H H</td>
<td></td>
</tr>
<tr>
<td>T H</td>
<td>T H H</td>
<td></td>
</tr>
</tbody>
</table>

Entropy is additive for independent systems. In the example, \( \sigma_1 = \log 2 \), \( \sigma_2 = \log 3 \), and \( \sigma = \sigma_1 + \sigma_2 = \log 6 \).

Now suppose that \( S_1 \) and \( S_2 \) are coupled by placing them in thermal contact. In our example, remove the wall between the two systems so that energy may flow between them. The total energy of the combined system remains unchanged, but when equilibrium is reestablished, any configuration of the \( N = N_1 + N_2 \) heads and tails with that energy is an accessible microstate. In our example, the combined energy is 3, and there are \( 5! / 3! 2! = 10 \) ways of assigning three heads to five coins. These are shown in Table 3. According to the Fundamental Postulate, all 10 microstates are equally probable, and the new entropy is \( \sigma = \log 10 \). Coupling has increased the entropy from \( \log 6 \)

Table 3. Accessible microstates of combined system after coupling

<table>
<thead>
<tr>
<th>Microstates of coupled system (energy 3)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H H H T T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H H H T T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H H T T H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H T H H T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H T H T H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H T T H H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T H H H H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T H H H H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T T H H H</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
to log 10. To summarize the steps that lead to the increase of entropy after coupling:

1. Prior to coupling the system is composed of two isolated subsystems in a definite macrostate.
2. The accessible microstates are constrained by the requirements that the energy of $S_1$ be $E_1$, that the energy of $S_2$ be $E_2$, and that the energy of the whole system be $E = E_1 + E_2$.
3. Coupling the two subsystems results in the removal of the first two of these constraints.
4. As a result, the space of accessible microstates is enlarged; formerly accessible microstates remain accessible, while new ones (possibly) are added.
5. Hence, the entropy, which is a measure of the number of accessible microstates, increases.

Call this the classical explanation of entropy increase. We believe that the simplicity of this explanation conceals two defects, one intrinsic to physics and the other extrinsic.

The first involves the justification of the Fundamental Postulate, on which the explanation intimately depends. The assumption that the systems in question, after being perturbed from equilibrium by coupling, eventually settle down to an equilibrium of equally likely microstates is far from obvious. What is obvious is that coupling must be implemented by some sort of physical operation, such as the insertion of a thermal (conducting) membrane between the systems where before there was an adiabatic (insulating) wall, or the removal altogether of the wall between two chambers of gases. But why this removal of a barrier in the physical sense should amount to a removal of constraints in the mathematical sense of (3) above is unclear.

The proper setting for the formulation of this difficulty is quantum mechanics, and we do not propose to detour into this theory here. We mention in this connection only that some have found the solution to the difficulty to lie in the notion of coarse-graining, which involves defining probabilities not over individual microstates, but over groups of microstates lumped together. We discuss coarse-graining in the biological examples in section 3.

We can give a sense of the nonobviousness of the Fundamental Postulate with an extension of our simple example. Although at equilibrium there is assumed to be a probability distribution over the space of microstates, the system at any
When and why does entropy increase?

<table>
<thead>
<tr>
<th>Table 4. Time evolution before coupling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$t + 0$</td>
</tr>
<tr>
<td>$t + 1$</td>
</tr>
<tr>
<td>$t + 2$</td>
</tr>
<tr>
<td>$t + 3$</td>
</tr>
<tr>
<td>$t + 4$</td>
</tr>
<tr>
<td>$t + 5$</td>
</tr>
<tr>
<td>$t + 6$</td>
</tr>
</tbody>
</table>

particular instant is in fact in one or another particular microstate. From there it evolves (in time) according to a set of deterministic physical laws. Let us specify a (fictitious) dynamics for our system. At time $t$ the system is in a particular microstate; at time $t + \Delta t$ each unit of energy (i.e., each occurrence of heads) moves one coin to the right, unless it reaches a wall, in which case it comes back in on the left. (In other words, the row of coins is really a circle, and the heads march around it.) The first six time-steps of the uncoupled systems' evolution are shown in table 4; after six steps the pattern repeats.

Under the ergodic assumption, we may derive the equilibrium probabilities from the long-term frequencies of the microstates; the result is that each of the two microstates of $S_1$ are equally probable (occur equally often), each of the three microstates of $S_2$ are equally probable, and the two subsystems are probabilistically independent. This agrees with the probability assignments demanded by the Fundamental Postulate. After coupling, however, the situation is as shown in table 5. The time evolution repeats after only five steps,

<table>
<thead>
<tr>
<th>Table 5. Time evolution after coupling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$t + 1$</td>
</tr>
<tr>
<td>$t + 2$</td>
</tr>
<tr>
<td>$t + 3$</td>
</tr>
<tr>
<td>$t + 4$</td>
</tr>
<tr>
<td>$t + 5$</td>
</tr>
</tbody>
</table>
and this is true no matter what the initial configuration is at the time of coupling. The result using the ergodic assumption is that the probability is spread evenly over only five accessible microstates, not ten.\(^{11}\)

Of course, the dynamics we have specified is not real physics (neither, for that matter, are our coins). But the problem of convincingly justifying the Fundamental Postulate for systems of real particles obeying real physical laws is well illustrated by this made-up example. To the extent to which this crucial assumption of the explanatory story remains ungrounded, the explanation itself is unsatisfying.

The extrinsic difficulty with the classical explanation of entropy increase concerns its applicability to problems outside of statistical mechanics. We wish to examine a variety of systems governed by probability distributions and which consequently allow entropy to be defined. We ask when and why entropy increases after coupling such systems. The formulation of a particular problem may reveal an analogy to statistical mechanics, leading us to hope that we can apply the classical explanation. But the classical explanation requires that the set of accessible states after coupling be an enlargement of the set before; accessible states before coupling remain accessible afterward. Unfortunately, this condition is not met in all real examples; a specific case is discussed in the next section.

Our conclusion, then, is that the classical explanation of entropy increase may be true (as far as it goes), but its assumptions are opaque and it offers little guidance as to whether we may properly apply the second law in applications outside of physics.

So to better illuminate the problem of entropy outside physics, we seek an alternative explanation of entropy increase within physics itself. Such an explanation, in its essentials, is provided by Khinchin.\(^{12}\) His exposition applies strictly to systems with a continuous range of admissible energy values, but it can be adapted to systems with discrete admissible energies.

In order to understand the key points of Khinchin's discussion, we first reconsider our simple coin example when the number of coins is very large. We use the term *orbital* to refer to the states which a single particle may attain; this term comes from quantum mechanics. In the coin example, there are two orbitals: \(H\) and \(T\). Under the microcanonical ensemble, the marginal distribution over the orbitals is the same for every particle. Consider table 3,

---

11 Notice that with the stipulated dynamics, coupling leads the entropy to decline.

12 Khinchin (1949).
at the time of
the probability is

neither, for that
pointing the Funda-
mental physical laws is
which this crucial
the explanation

tropy increase
mechanics. We
dency distributions
whether and why
of a particular
us to hope that
ation requires
of the set
ward. Unfortu-
case is discussed

tropy increase
ue and it offers
law in applica-
ics, we seek an
itself. Such an
position applies
energy values, but it
ussion, we first
as is very large.
e particle may
example, there
le, the marginal
clude table 3.

tropy increase for a homogeneous system. His development assumes that the
particles are distinguishable.\(^\text{13}\)

9 When and why does entropy increase?

in which there are five particles (i.e., coins) and 10 equally probable microstates. Coin 1 has a marginal distribution of \(P(H) = 0.6\) and \(P(T) = 0.4\). A perusal of the table confirms that the other coins have the same marginal distribution.

Now compute the probability that coin 1 is \(H\) and coin 2 is \(H\), which we write \(P(\text{HH})\). From the table we see that this probability is 0.30, and we have

\[
P(\text{HH}) = 0.30 \approx 0.36 = (0.6)(0.6) = P(H)P(H).
\]

Similar approximate equalities hold for \(P(HT), P(TH),\) and \(P(TT)\). This looks suspiciously like probabilistic independence. In fact, as the number of particles gets very large, the approximate equality in this equation approaches equality, and the marginal distributions of the particles become independent of one another. A derivation is given in appendix B. (In applications outside of physics, it is often appropriate, or at least not incorrect, to stipulate this independence at the outset.)

The probability distribution in which the marginal distributions of the particles are strictly independent is called in physics the canonical ensemble, or canonical distribution. It is of practical importance that even in a system distributed according to the microcanonical ensemble (a closed system), if the system is a large one, small chunks of it will be distributed approximately according to the canonical ensemble. This is discussed in appendix B. To put it another way, in the microcanonical ensemble there are correlations among the particles; but if the system is large, the correlations are weak, and small chunks of the system behave as though the particles were truly independent.

Now we can summarize the main postulates of Khinchin's demonstration of entropy increase for a homogeneous system. His development assumes that the particles are distinguishable.\(^\text{13}\)

1. The system consists of \(N\) identical particles whose (discrete) states are called orbitals. A microstate for the system is given by specifying the orbital that each particle occupies.

2. At equilibrium, there is a probability distribution over the microstates, and consequently marginal distributions for each of the particles. Each particle has the same marginal distribution, and (in the limit of large \(N\)) the marginal distributions are independent.

3. Associated with each orbital is an energy. Energy is additive; the energy of the system in any microstate is the sum of the energies of the particles in their

\(^{13}\) However, Khinchin's treatment can be adapted to apply, for example, to Bose and Fermi gases, in which the particles are not distinguishable.
orbitals. (This implies that interactions between the particles are weak or absent.)

(4) When two isolated systems are coupled, total energy is conserved. (If the probability distribution is spread over a range of energies, the mean total energy is conserved.)

(5) The probability distribution over the microstates is governed by one or more underlying parameters, such as the number of particles and the total energy. Another such parameter is the temperature, denoted \( \tau \), which in many systems is proportional to the average energy per particle. The parameters determine the distribution uniformly; i.e., they apply in the same way both before and after coupling.

(6) The probability that a given particle will be found in a specific orbital of energy \( \epsilon \) when the system is at temperature \( \tau \) is proportional to an exponential: \( P \propto \exp(-\epsilon/\tau) = \left\{ \exp(-1/\tau) \right\}^{\epsilon} \). (See the remarks following.)

(7) The value of \( \tau \) for the system is determined uniquely by the requirement that the mean energy for each particle, multiplied by \( N \), must give the (known) observable energy for the whole system. (According to assumption (5), this must hold before and after coupling.)

(8) When the conditions above are fulfilled, and in particular when \( \tau \) is constrained to give the correct value for the total energy, then Khinchin's theorem states that the entropy of the coupled system must be at least as great as that of the system before coupling.

With reference to condition (6): when \( \tau \) is positive, as it is for most systems in nature (including gases), the quantity \( \exp(-1/\tau) \) is a positive constant \( c < 1 \). (The smaller \( \tau \) is, the smaller \( c \) is.) Suppose for simplicity that the energy values \( \epsilon_0, \epsilon_1, \epsilon_2, \ldots \) are equally spaced integers 0, 1, 2, \ldots Then the probabilities are proportional to \( c^{\epsilon_0}, c^{\epsilon_1}, c^{\epsilon_2}, \ldots \). This means that the probability of a state of a particular energy is always a constant factor times the probability of a state of the next lower energy. This gives us a picture of gradually declining probability as the energy of the orbital gets higher. In fact, whether the temperature is positive or negative, (6) means that the probabilities decrease or increase monotonically with increasing energy.

When \( N \) (the number of particles) is large, the system is governed by the microcanonical distribution, and the temperature is not disturbed by small changes in the total energy, it can be proved that conditions (6) and (7) hold (see appendix B). We list them as independent postulates because in examples outside of physics (such as those in the next section), the microcanonical distribution may not obtain, or energy and temperature may not be definable in such a way as to have the listed properties.
9 When and why does entropy increase?

In this development, much of the simplicity of the classical explanation has been lost. In particular, it is not obvious from the abbreviated discussion just given why Khinchin's Theorem (item (8)) is true. (See Khinchin for the mathematics.) What is important, however, is that the conditions preceding item (8) collectively constitute a true and applicable sufficient condition for entropy increase. When statistical mechanical principles are applied outside of physics, all or some of the conditions above are often met. When some of them fail to be met, it is possible for entropy to decline. We will see from the biological examples how this can occur.

3 Some biological examples

Natural selection is a process that is set in motion by the existence of variation in fitness. The result of the process is that variation in fitness is eliminated. Variation in fitness is like the oxygen that fuels a fire; it is a precondition for the fire to exist, but the fire uses up oxygen and so will bring itself to a halt.

Although it is not standard to think of natural selection as initiated by the coupling of two populations, it does no violence to the biological concepts to impose this formulation. We will think of two populations that are each at equilibrium and then are coupled; after selection has run its course, acting on the variation created by the coupling process, we can compute the entropies before coupling and the entropy after.

The simplest case involves two populations and two traits, A and B. Each population has some equilibrium frequency of the two traits. Then the populations are coupled and the joint population moves to an equilibrium frequency. What will happen to the entropy in this case?

Let's imagine that A and B are the two alleles that can occur at a given diploid locus. There are then three genotypes AA, AB, and BB; their frequency independent fitnesses are \( \omega_{AA}, \omega_{AB}, \) and \( \omega_{BB} \). Table 6 describes what happens when two genetically different populations are coupled. Each is at equilibrium before coupling; after coupling, the fitness ordering determines the new equilibrium that the conjoint population attains.

When a population is monomorphic, its entropy is zero. So, in the first and last processes described, the summed entropy of the uncoupled populations is zero, and the entropy of the coupled population is zero as well.
Table 6. The effect of coupling two populations

<table>
<thead>
<tr>
<th>Fitness ordering</th>
<th>Before coupling</th>
<th>After coupling</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_{AA} &gt; \omega_{AB} &gt; \omega_{BB}$</td>
<td>100% $A$</td>
<td>100% $A$</td>
</tr>
<tr>
<td>$\omega_{AB} &gt; \omega_{AA} &gt; \omega_{BB}$</td>
<td>100% $A$</td>
<td>100% $B$</td>
</tr>
<tr>
<td>$\omega_{AA} &gt; \omega_{BB} &gt; \omega_{AB}$</td>
<td>100% $A$</td>
<td>100% $B$</td>
</tr>
</tbody>
</table>

The two middle cases, which involve heterozygote superiority, are more interesting. When each of the populations before coupling is monomorphic, coupling makes the entropy increase. And when one of the populations before coupling is polymorphic, the entropies before and after coupling are identical and nonzero.

So in the one-locus setting we have considered, entropy never declines under coupling. However, it is not hard to find selection processes in which this pattern is violated. Let us simply add a third allele $C$ to our consideration. The fitnesses of the genotypes are given in table 7.

Now consider two populations. In the first, $A$ and $B$ are the only alleles present; they are stably maintained because there is heterozygote superiority. In the second population, $C$ is monomorphic. If these two populations are coupled, $C$ will sweep to fixation. In this case, the sum of the entropies of the pre-coupled populations is nonzero, but the entropy of the coupled population after it reaches equilibrium is zero. Entropy has declined.

The situations described thus far bear some analogy to the statistical mechanical setup described in the previous section. The populations are homogeneous systems, whose macrostates are defined by the number of organisms and the gene frequencies involved. Microstates corresponding to

Table 7. Hypothetical fitnesses for genotypes with three alleles

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$B$</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td></td>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>
the macrostates are assignments of allowable genotypes to the individuals in the populations. We may define probabilities over the microstates by treating as equally probable all (and only those) microstates which have individual genotypes in the proportions prescribed by the macrostate. There are one or more underlying parameters which determine the macrostate, and hence the probability distribution over the microstates. Here the parameters are the trait fitnesses and frequencies, and the recipe for determining from them the equilibrium macrostate is that the gene frequencies adjust until the overall fitness of the population attains a maximum (Fisher’s Fundamental Theorem). This recipe applies uniformly to the populations both before and after coupling.

There are important differences as well. There is no clear analog in a selection process to the energy, a quantity which is additive and which is conserved, when two systems are coupled. (Fitness could be regarded as additive – take the fitness of each organism as its expected number of offspring; and the fitness of the population as the sum of the fitnesses of the individuals – but it is certainly not conserved when two systems are coupled.) Also, if we order the genotypes AA, AB, BB, the genotype frequencies will not necessarily be monotonically increasing or decreasing (both before and after coupling), as condition (6) in section 2 would require.

The result of the failure of these examples to correspond to important parts of the statistical mechanical picture is that entropy may increase or decrease, depending on the specific circumstances. An artifact of the selection process is that something or other will happen to entropy. But there is nothing intrinsic to the process of natural selection that allows us to say anything general about the direction of entropy change.

Matters are quite different when we turn away from selection and consider another example in population genetics. Wahlund’s Principle describes the consequences of coupling populations that were previously isolated. Suppose that an allele A occurs in each of k populations with frequencies \( p_1, p_2, \ldots, p_k \) and that the alternative allele a occurs with frequencies \( q_1, q_2, \ldots, q_k \), where \( p_i + q_i = 1 \). Within each population there is random mating. Equilibrium in each population is reached after at most one generation, with the genotypes AA, Aa, and aa in the Hardy–Weinberg proportions of \( p_i^2, 2p_iq_i \), and \( q_i^2 \). If the populations sizes are \( n_1, n_2, \ldots, n_k \), the average frequency of the A allele

---

Thermodynamics and time's arrow

across these separate populations is

\[ \frac{n_1 p_1 + n_2 p_2 + \cdots + n_k p_k}{n_1 + n_2 + \cdots + n_k} = \bar{p}. \]

Here we may view \( p \) as a random variable with possible values \( p_1, \ldots, p_k \) defined on the space \( \{1, \ldots, k\} \) with weights \( n_1 / \sum n_i, \ldots, n_k / \sum n_i \); so \( \bar{p} \) is simply the mean of \( p \). Similarly, the average frequency of the AA homozygote across the separate populations is

\[ \frac{n_1 p_1^2 + n_2 p_2^2 + \cdots + n_k p_k^2}{n_1 + n_2 + \cdots + n_k} = \bar{p}^2. \]

If these populations are combined, the allele frequency of \( A \) in the new pooled population remains \( \bar{p} \), so the frequency of \( AA \) homozygosity after random mating is \( \bar{p}^2 \).

The variance in the random variable \( p \) is by definition \( V_p = \bar{p}^2 - \bar{p}^2 \). Since variance is always nonnegative, this means that there is less (or at least, no more) \( AA \) homozygosity in the pooled population than there was, on average, before the populations were pooled. The same holds, \textit{mutatis mutandis}, for \( aa \) homozygosity. Or to put the point in terms of heterozygosity (i.e., the frequency of the \( Aa \) genotype), pooling increases heterozygosity.

Heterozygosity reflects the degree of mixing in the population of the alleles \( A \) and \( a \). In thermodynamical situations, mixing of disparate species of particles is accompanied by an increase in entropy; so we were naturally led to inquire what happens to the entropy of populations of organisms coupled in this manner.

When we look at this problem carefully, a close correspondence emerges between the biology and statistical mechanics. The organisms are 'particles'. The possible genotypes of the organisms are 'orbitals'. The assignment of a genotype to each organism is a 'microstate'. The number of organisms and the total number of \( A \) (or \( a \) alleles in the population together comprise the macrostate. We initially employ the artifice that the genotypes \( Aa \) and \( aa \) are distinct; we will correct this later. We may assign 'energies' to the orbitals as follows: the allele \( A \) has energy 1 and the allele \( a \) has energy 0. Then the orbitals have energy 0, 1, or 2; this is just the number of \( A \) alleles in the genotype. Energy is additive and sums to give the number of \( A \) alleles in the whole population. At Hardy–Weinberg equilibrium, the orbitals have probabilities as given in table 8.
9 When and why does entropy increase?

The last column rewrites the probabilities to show how they are in geometric proportion with the energy as the exponent, as in clause (6) of Khinchin's development. For random mating we may validly employ the beanbag model, in which each organism determines its genotype by 'drawing' randomly-with-replacement twice from the gamete pool; this entails that the orbital distributions of the 'particles' are independent. This means that the entropy of the i-th population is just $n_i$ times the orbital entropy:

$$\sigma_i = n_i (p_i^2 \log p_i^2 + 2p_iq_i \log p_iq_i + q_i^2 \log q_i^2).$$

The isolation of the populations means that the population distributions are independent of each other, so the entropy of the populations taken all together is just the sum of the population entropies. (See appendix A.)

When the populations are coupled (pooled), the total 'energy' (i.e., the number of A alleles) is conserved. The pooled population reaches an equilibrium in which each organism has the same (new) orbital distribution. Thus all the conditions of Khinchin's theorem are fulfilled, and we may conclude that entropy in the combined population increases.

We now must attend to the detail referred to above: in population genetics the genotype Aa is not in general distinguishable from aa. So there are really three (not four) orbitals, corresponding to AA, Aa, and aa. The probabilities and entropy have to be adjusted to take this into account. It turns out (we omit the proof) that this small change in the setup makes no difference to the conclusion; entropy increases anyway.

Notice that in this example we could not derive the conclusion that entropy increases by employing the model of equally probable microstates in which

---

15 This means that temperature may be defined so that $\exp(-1/\tau) = q_i^2(p_i/q_i)$, and then clause (6) is satisfied by this example.
Thermodynamics and time's arrow

coupling entails removal of constraints on the admissible microstates. This model requires that microstates of the (combined) system which are admissible before coupling remain admissible after coupling. But this condition is not fulfilled here. Take for example two populations of equal size: one with 80% A and the other with 20% A. In the first population the genotype frequencies for AA, Aa, and aa are 64%, 32%, and 4%; in the second population the genotype frequencies are 4%, 32%, and 64%. In the first population, take as equally probable microstates all genotype-to-organism assignments which have exactly the correct proportions of AA, Aa, and aa; do the same in the second population. In the combined (but uncoupled) population, all admissible microstates have frequency of AA equal to \((64\% + 4\%)/2 = 34\%\). After coupling, however, all admissible microstates have frequency of AA equal to \((50\%)^2 = 25\%\); so none of the previously admissible microstates are admissible. The classical explanation of entropy increase cannot be applied here; but Khinchin's alternative viewpoint, as we have seen, proves fruitful.

When we try to extend the result to more general situations in population biology, we begin to see ways in which the statistical mechanical analogy must be weakened. The most obvious extension is to consider loci possessing more than two alleles, with random mating as before. It is no longer possible to assign in a consistent way a single energy to the genotypes, so that the total energy is conserved when the populations are coupled and random mating produces a new equilibrium. However, the gene frequency of each of the alleles \(A_1, \ldots, A_m\) is preserved after random mating, so we could imagine \(m\) different species of energy, each representing the total number of alleles \(A_1, \ldots, A_m\) in the population, and each separately conserved upon coupling of the populations. With no single quantity to play the role of energy, we must abandon assumption (6) but the other assumptions remain in force. In this situation, it turns out that entropy increases.

The question then arises as to whether the list of conditions would suffice for entropy increase if assumption (6) were deleted. To address this question, consider a situation in which this condition is violated. Suppose that we have a single locus with two or more alleles, and a set of phenotypes which supervene on the genotypes at this locus. The simplest example is furnished by a recessive trait and two alleles \(A\) and \(a\). Suppose \(a\) is recessive. The organism exhibits the

---

\(A\) set of properties \(P\) is said to 

\(\text{supervene}\) on a set of properties \(Q\) precisely when the \(Q\) properties of an object determine what its \(P\) properties will be. See Kim (1979) for discussion.
microstates. This condition is not one with 80% type frequencies in a population that the population, take assignments which do the same in a population, all % + 4%) / 2 = 34%. frequency of AA possible microstates cannot be have seen, provesions in population mechanical analogy under loci possessing no longer possible types, so that the upled and random frequency of each of, so we could the total number of A, B, and O, and four phenotypes which we denote a, b, ab, and o. The phenotypic supervenience is as follows: phenotype a corresponds to genotypes AA and AO, b to BB and BO, ab to AB, and o to OO. We know from the results already described that the entropy of the genotypic distribution (over the six genotypes) must increase after coupling with random mating. However, it is by no means obvious what happens to the entropy of the phenotypic distribution. We employed a computer program to calculate the phenotypic entropies before and after coupling with various values for the gene frequencies. It turns out that for most values of the gene frequencies, the entropy after coupling increases, but that for a small range of gene frequencies, entropy decreases. For example, a population of 10% A, 87% B, and 3% O will produce an entropy decrease when coupled to an equal sized population of 10% A, 60% B, and 30% O. It is an empirical accident that the values which produce entropy decrease happen to be far
removed from the actual values of the gene frequencies in naturally occurring human populations.17

4 Conclusion

In the biological examples we have surveyed, the entropy of a population is determined by the frequency distribution of traits of organisms. It is important to realize that the entropy of a population, defined in this way, has nothing to do with the orderliness of the organisms in the population. Under Wahlund coupling, entropy goes up. Yet, the organisms in the post-coupling population may be no less ordered than the organisms were when the populations were isolated. Although heterozygosity increases under coupling, there is no clear sense in which heterozygotes are less ordered than homozygotes. What has changed here is the entropy of the population, not the orderliness of the organisms in the population.18

Although the entropy of the genotype distribution increases under Wahlund coupling, the entropy of the phenotype distribution need not do so. It is of some interest that supervening distributions can move in one direction while the distributions on which they supervene go in another. This point applies to chambers of gases no less than it does to Mendelian populations. Coupling two closed systems to produce a third can reduce entropy. The second law of thermodynamics says what will happen to some probability distributions, not to all of them. It does not take an energy input to get entropy to go down. Merely reconceptualizing the traits on which distributions are defined can do the trick.

17 Because our discussion of thermodynamic entropy was restricted to equilibrium situations, we have discussed entropy in population genetics also from the point of view of pre- and post-coupling equilibria. However, it is quite easy to think about entropy in population genetics for nonequilibrium situations. An evolving population on its way to equilibrium will move through a sequence of gene frequencies; since probability distributions are thereby well-defined in every generation, so too are the entropies. What then happens to the entropy of evolving populations? For many evolutionary processes, entropy declines. A novel variant X that sweeps to fixation by selection will lead entropy to rise and then fall. Imagine that it is introduced into a population of 100% Y individuals. The frequency of X increases and the entropy reaches a maximum when the frequencies are 50% X and 50% Y. When X reaches 100%, the population's entropy is again zero. Drift also tends to destroy variation and so reduce entropy. Mutation and recombination, however, tend to have the opposite effect, since they increase the range of variation.

For a sampling of opinion concerning how the entropy concept might be applied to evolutionary processes, see Brooks & Wiley (1986), Collier (1985), Morowitz (1986), Wicken (1987), and Layzer (1969).

When and why does entropy increase?

Table 9. Properties of the examples

<table>
<thead>
<tr>
<th></th>
<th>Statistical mechanics</th>
<th>Selection</th>
<th>Genotypic Wahlund</th>
<th>Supervening Wahlund</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle distributions</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>identical and (almost)</td>
<td>independent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same process generates</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>probabilities, before and</td>
<td>and after coupling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additive energy analog</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>conserved by coupling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exponential probabilities</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Must entropy increase?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

In the classical explanation of entropy increase in thermodynamics, entropy goes up because the removal of a physical barrier is taken to imply that constraints on admissible microstates are removed. The post-coupling system thus can exhibit all the microstates associated with the pre-coupling systems, and some more besides. The Khinchin explanation of entropy increase is less transparent than this simple account, which is why physicists rarely advert to it. Yet, a more general treatment of the problem of entropy increase suggests that the classical account may be limited in important ways. The deep affinity of Wahlund coupling to thermodynamic processes becomes clear when we use the Khinchin account, a parallelism that the classical account fails to describe.

The Khinchin account has a further virtue. It provides a template for analyzing cases in which entropy fails to increase. The Khinchin account lists a set of jointly sufficient conditions; entropy decline must violate at least one of them, as table 9 shows.

Discussions of entropy have tended to vacillate between two extremes. On the one hand there are unstructured and metaphorical explorations of the issues, which allow one to see analogs to thermodynamic processes everywhere. On the other hand, there is the tendency to restrict the concept to its strict physical meaning; though this reduces the scope (and, dare we say, the glamour) of the issue, at least it can be said that the problem of entropy increase remains in such treatments a precise and intelligible one. We have tried to steer a middle course. We have tried to isolate abstract features of the physical process in which isolated equilibrium systems are coupled and

249
allowed to find their new equilibria. The question we have explored is what it is
about this abstract formulation that forces entropy to increase. Our results are
summarized in table 9.

In order to give the problem some structure, we have assumed that two popu-
lations of objects attain an equilibrium frequency distribution by some definite
process. Then the populations are coupled, and the same process at work before
coupling then takes the coupled system to a new equilibrium. As the example of
selection illustrates, this minimum way of framing the problem does not
guarantee an increase in entropy. The reason entropy need not increase
under selection is that there is no additive energy analog, which is conserved
under coupling. In the Wahlund process as well as in thermodynamics, there
is a measure on each population that is the sum of the measures of all the
individuals. To be sure, fitness is a measure of this sort – the productivity of
a population is just the sum of the productivities of its member organisms.
However, the point is that the fitnesses at equilibrium of the coupled popula-
tion need not be the same as the sum of the fitnesses before coupling. There
is no principle of fitness conservation. This is because the coupling operation
triggers a process whose dynamics is controlled by the principle that fitter
traits increase in frequency while less fit traits decline.

However, even if we add the constraint that the process must possess an
additive energy analog that is conserved under coupling, this still is not
enough to guarantee an increase in entropy. The genotypic Wahlund process
generates an entropy increase, but the phenotypic Wahlund process does not.
The difference here is due to the fact that genotypes in the Wahlund coupling
process can be represented as obeying an exponential probability distribution
with energy as a parameter, whereas the phenotypes (in our examples of the
recessive trait and the ABO blood groups) cannot. The Khinchin proof does
not apply to entropies defined on arbitrary probability distributions.

One of the most interesting features of the second law is its description of a
process that is temporally asymmetric. It is therefore worth considering how
this asymmetry is represented in our account of the second law as a proposi-
tion concerning the coupling processes. If entropy increases under coupling,
what happens under decoupling? That is, if a population is subdivided, what
will happen to its entropy? Here we must recognize that the actual energy
(or actual gene frequency, etc.) of populations at equilibrium fluctuates about
the mean energy (or mean gene frequency, etc.) derived from the governing
distribution. If the population size is extremely large (on the order of $10^{23}$),

250
these fluctuations almost never depart measurably from the mean value. When such populations are subdivided and separated, we may therefore expect that the temperatures (or the temperature analogs) in the two parts are equal and equal to that of the system before decoupling; entropy therefore will be unchanged. Coupling and decoupling have qualitatively different effects.

In the case of populations that may be large but not ‘virtually infinite’, we must modify the contrast a bit. Fluctuations about the mean will be more significant, and there is a small chance that subdividing a population will create subpopulations that differ from each other appreciably. But, in expectation, the subdivided populations will differ from each other only a little. Decoupling leads to very minor entropy declines. It is a contingent fact about the large populations found in the world that they often differ from each other appreciably. It follows that coupling two such populations often leads to entropy increases that are much larger than the entropy decline that would result from subdividing any one of them. This fact, we emphasize, depends on both the ‘laws of motion’ of the coupling process and on the initial conditions that happen to obtain.

For the sake of completeness, we must consider the difference between coupling and decoupling for populations that are small. Here fluctuations around the equilibrium values can be large – indeed, large enough to obliterate the asymmetries noted for the previous two cases. For small populations, everything will depend on contingent facts concerning the actual sizes and frequencies at the moment of decoupling. One can imagine a universe in which populations are small enough that subdividing makes entropy go down more than coupling makes entropy go up. For example, consider a universe in which there are 10 populations of nearly identical small size and temperature. Coupling any two of them will leave entropy virtually unchanged; but, since the populations are small, subdividing and separating any one of them is likely to reduce entropy more.

In summary, the asymmetry described by the second law owes something to the ‘laws of motion’ governing the processes of coupling and subdividing, but also depends on contingent facts about the populations found in nature. The world we live in – whether we are talking about thermodynamic systems or Mendelian populations – consists of populations that are large enough and different enough that coupling and subdividing yield qualitatively different results.
Our strategy in exploring the general question of entropy increase has been to forget the physical meaning of concepts like 'orbital' and 'energy' and to focus on their abstract mathematical representation. In Wahlund coupling, the number of alleles of each type is conserved; it is quite irrelevant that the physical energy of a population (whatever that might mean) is unrelated to the number of alleles the population contains. Although entropy increases in thermodynamic processes because of the physical processes that take place, we have tried to understand why entropy increases as a mathematical problem. This strategy conflicts with one of the contrasts that Fisher\textsuperscript{2} draws between entropy and fitness:

Fitness, although measured by a uniform method, is qualitatively different for every different organism, whereas entropy, like temperature, is taken to have the same meaning for all physical systems [pp. 39–40].

Fisher's point about fitness is that it is multiply realizable. The fruitfulness of this abstract representation of the organic world is beyond dispute; it is astonishing that a single framework can encompass the evolution of life forms that are so physically diverse. In a sense, we have attempted in this paper to follow Fisher's example - to describe entropy as a mathematical property that also is multiply realizable. It remains to be seen to what extent this abstract treatment will bear fruit.

Appendix A: Some simple properties of entropy

The entropy of a finite partition \( \{X_i\}_{i=1}^n \) of a probability space is defined by the formula \( \sigma = -\sum_{i=1}^n P(X_i) \log P(X_i) \). (We take \( 0 \log 0 = 0 \).) Entropy is a measure of the spread of the distribution: roughly, the more possibilities over which the probability is spread, the greater the entropy.

1. \( \sigma \geq 0 \). Proof: If \( 0 < x < 1 \) then \( \log x < 0 \).
2. If the distribution consists of \( n \) equiprobable alternatives, that is, \( P(X_i) = 1/n \) for \( i = 1, \ldots, n \), then \( \sigma = \log n \). Proof:

\[
\sigma = -\sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} = -n(\frac{1}{n} \log \frac{1}{n}) = \log n.
\]

3. If the distribution is spread over \( n \) alternatives, the maximum \( \sigma \) is attained when the alternatives are all equiprobable. Proof: omitted.

Recall that partitions \( \{X_i\}_{i=1}^n \) and \( \{Y_j\}_{j=1}^n \) of a probability space are said to be independent if \( P(X_i \land Y_j) = P(X_i)P(Y_j) \) for all \( i = 1, \ldots, n \) and
9. When and why does entropy increase?

\[ j = 1, \ldots, m. \] When this holds, we also say that the probability space (partition) \( \{X_i \land Y_j\}_{i=1,\ldots,n, j=1,\ldots,m} \) is the product of \( \{X_i\}_{i=1,\ldots,n} \) and \( \{Y_j\}_{j=1,\ldots,m} \).

(4) If the partitions \( \{X_i\}_{i=1,\ldots,n} \) and \( \{Y_j\}_{j=1,\ldots,m} \) are independent, then the entropy of the product partition is the sum of the entropies of \( \{X_i\}_{i=1,\ldots,n} \) and \( \{Y_j\}_{j=1,\ldots,m} \). Proof:

\[
\sigma = \sum_{i,j} P(X_i \land Y_j) \log P(X_i \land Y_j) = \sum_{i,j} P(X_i)P(Y_j)\log P(X_i) + \log P(Y_j) = \sum_{i,j} P(Y_j)P(X_i) \log P(X_i) - \sum_{i,j} P(X_i)P(Y_j) \log P(Y_j) = \sum_{i,j} P(X_i) \log P(X_i) - \sum_{j} P(Y_j) \log P(Y_j).
\]

(5) Corollary: if the partition \( \{X_i\} \) with entropy \( \sigma \) is the product of \( N \) independent copies of a partition \( \{Y_j\} \) with entropy \( \sigma' \), then \( \sigma = N \sigma' \).

Appendix B: Some statistical mechanics

In this appendix we derive the exponential form of the marginal distributions for a single particle from the Fundamental Postulate of Statistical Mechanics, and show that single-particle distributions are identical and approximately independent. We consider a system \( S \) with \( N \) distinguishable particles, where \( N \) is very large. Each of the particles may occupy one of \( m \) possible states (orbitals) of varying energy. (Different orbitals may have the same energy.) We also assume that these orbital energies, \( e_1, \ldots, e_m \), remain fixed throughout. A microstate of \( S \) is given by specifying an orbital for each particle. Paramagnetic spin systems (with a fixed external magnetic field) are one example of a system satisfying these conditions with \( m = 2 \).

We suppose that \( S \) is a closed system with (unchanging) total internal energy \( U \). Then according to the Fundamental Postulate of Statistical Mechanics, each of the accessible microstates of energy \( U \) is equally probable. We denote the number of such states by \( \#(N, U) \) to indicate the dependence on \( N \) and \( U \). The entropy of a closed system with \( n \) particles and energy \( u \) is \( \sigma(n, u) = \log \#(n, u) \) (see appendix A). We write \( \sigma_S = \sigma(N, U) \) for the entropy of \( S \). The temperature \( \tau \) of \( S \) is defined in terms of \( N \) and \( U \) by

\[
\frac{1}{\tau} = \left[ \frac{\partial \sigma(n, u)}{\partial u} \right]_{n=N, u=U}.
\]
Now let us focus on a small chunk of $S$ with $k$ particles; $k \ll N$. The remainder of $S$ we can call the reservoir and denote it $R$. The reservoir is so large and contains so much internal energy that we can assume that small amounts of energy added to it or taken from it or small numbers of particles removed from it make no difference to its temperature. In other words, the derivative $\partial \sigma(N - k, u)/\partial u$ is constant for values of $u$ near $U$ and small $k$. This means that near $U$, the graph of $\sigma$ versus $u$ is a straight line with slope $[\partial \sigma(n, u)/\partial u]_{n=N-k,u=U} = 1/\tau$, and we can write

$$\sigma_R(U - \Delta u) = \sigma(N - k, U - \Delta u) = \sigma(N - k, U) - \frac{1}{\tau} \Delta u. \quad (1)$$

Now a state $s$ of the chunk is simply a $k$-tuple of orbitals: $s = (i_1, \ldots, i_k)$, with energy $U_s = \epsilon_{i_1} + \cdots + \epsilon_{i_k}$. (This assumes, as we usually do in statistical mechanics, that there is no interaction energy between the particles.) We seek the probability $P_s$ that the chunk is in state $s$. Since all microstates of $S$ are equiprobable, $P_s$ is proportional to the number of states of $S$ in which the chunk is in state $s$. Since each such state is simply a state in which the reservoir $R$ has energy $U - U_s$, $P_s$ is proportional to the number of states of $R$ with energy $U - U_s$, i.e., $\#(N - k, U - U_s)$. Using (1) we compute the logarithm of this number:

$$\log \#(N - k, U - U_s) = \sigma(N - k, U - U_s) = \sigma(N - k, U) - \frac{1}{\tau} U_s.$$

So we can write $P_s \propto \exp[\sigma(N - k, U) - U_s/\tau]$, or $P_s = c[\exp(-U_s/\tau)]$, where we have absorbed the factor $\exp[\sigma(N - k, U)]$ into the constant of proportionality $c$. Probabilities must be normalized to sum to 1, so we have

$$1 = \sum_s P_s = \sum_s c e^{-U_s/\tau} = c \sum_s e^{-U_s/\tau}.$$

The quantity $Z_k = Z_k(\tau) = \sum_s \exp(-U_s/\tau)$ is called the partition function (for the $k$-particle chunk) and is ubiquitous in statistical mechanics. We obtain, finally, $c = 1/Z_k$ and

$$P_s = \frac{e^{-U_s/\tau}}{Z_k}. \quad (2)$$

When $k = 1$, the chunk is composed of a single particle, and the energies $U_s$ are simply the orbital energies $\epsilon_1, \ldots, \epsilon_m$. The result (2) then becomes:

$$P_1 = e^{-\epsilon_1/\tau}/Z_1, \quad (3)$$

254
9 When and why does entropy increase?

where \( Z_1 = \sum_{i=1}^{n} \exp(-\epsilon_i/\tau) \) is the single-particle partition function, and \( P_i \) is the probability of being in orbital \( i \). Equation (3) holds no matter which particle we consider, so we have shown that the single-particle distributions are identical and that the single-particle probabilities have exponential form.

We next want to show that within a small \( k \)-particle chunk, the single particle distributions are independent. (This is how we formulate the notion of approximate independence.) First we relate the partition functions \( Z_k \) and \( Z_1 \). We have

\[
Z_k = \sum_{\{i_1, \ldots, i_k\}} \exp\left(-\frac{U_{i_1} + \cdots + U_{i_k}}{\tau}\right) = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_k} \exp\left(-\frac{U_{i_1}}{\tau}\right) \sum_{i_2} \cdots \sum_{i_k} \exp\left(-\frac{U_{i_2}}{\tau}\right) = Z_{i_1} \cdots Z_{i_k} = Z_{1}^k.
\]

Using this result we can write

\[
P(i_1, \ldots, i_k) = P_i = \frac{e^{-U_{i}/\tau}}{Z_k} = \frac{e^{-(\epsilon_{i_1} + \cdots + \epsilon_{i_k})/\tau}}{Z_1^k} = \left(\frac{e^{-\epsilon_{i_1}/\tau}}{Z_1}\right) \cdots \left(\frac{e^{-\epsilon_{i_k}/\tau}}{Z_1}\right) = P_{i_1} \cdots P_{i_k}
\]

and the independence property is established.