Thorough Thoughts on Thermodynamics

by Grant Kokoszka
9/6/2003

Physical understanding may be regarded as the recognition that the multiplicity of observed phenomena may be resolved to a smaller number of basic observations. These basic observations provide the fundamental postulates on which physics is built, examples of which include Newton's postulates of mechanics, accounting for the motion of bodies, and Maxwell's equations, summarizing the properties of electromagnetism. Being resistant to any obvious context for deeper scrutiny, these underlying relations seem indivisible expressions of apparent fact. And since they perfectly complement the impressions and intuition man develops by experiencing the world, it follows that acceptance of these postulates be natural and their fundamental stature not be challenged. Yet, in the case of thermodynamics, its basic laws do not share this privilege of sanctity. Circumstances are readily contrived, such as those suggested by the kinetic theory of gases, that provoke suspicion as to the fundamental nature of these laws. This is because these circumstances, to which the laws seem essential, appear manageable from within the framework of other disciplines. In the interest of making its underlying tenets most basic, the question emerges as to whether the laws of thermodynamics serve an indispensable purpose, contributing a regard for some unique aspect of the universe that is otherwise unaccounted for. Further complicating the issue, the laws of thermodynamics do not reflect quantities that are readily imagined, but instead rely on foreign concepts for which common experience provides no direct preparation, such as entropy. From this troubling condition, the question persists: Is thermodynamics fundamental?

The intent of this paper is to address this question by purposing an abstract prospective built from Shannon's information theory that demonstrates how the central feature of thermodynamics, entropy, can be interpreted as an evolving character of a system otherwise escaping description. Although the dynamics that bring about this feature are consistent with the views provided by other disciplines (such as mechanics in the case of kinetic theory), it is suggested here that the long-term behavior exhibited by these systems is not completely described by these considerations, prompting the need for an additional assertion. This assertion is precisely what thermodynamics, through entropy, provides. By presenting an intuitively satisfying impression of thermodynamics developed through information theory, the hope is that the mystery and suspicion surrounding this subject will diminish, making it more meaningful and deeply appreciated. Thermodynamics makes a unique contribution to the tapestry of the universe, and in this paper that contribution is made clear.

1. Summary of Traditional Thermodynamics

First it may be worthwhile to examine the concepts of thermodynamics upon which this paper relies. This is a quick summary of thermodynamics.

Heat is a thermodynamic quantity concocted to provide a means of dealing with occurrences of temperature change. It is motivated by the apparent transmission that occurs between objects undergoing a mutually induced temperature change, and it is known to flow of its own accord only from hot to cold regions. Its influx, associated with positive temperature change, is defined in the positive sense. By considering cyclical processes, such as that of the Carnot cycle, it is clear that work can be done at the expense of this entity, suggesting that heat is in some way representative of energy. This conclusion may be reached by any cyclical process, approximated by a succession of infinitesimally slender Carnot cycles[1], and in addition any arbitrary process may be thought of as a part of a cyclical one. As a result, this conclusion may be applied generally. Heat is therefore a form of energy.

Internal energy is another contrivance serving a valuable purpose in thermodynamics. It accounts for the energy associated with the thermal agitation of a system, taking forms corresponding to any available degrees of freedom. An example is the energy of motion of particles in the case of an ideal gas. At this point it is useful to draw a distinction between two forms of energy, which for these purposes are designated as ‘organized’ and ‘disorganized.’ Organized energy is that which can be completely changed into work in a wholly recoverable manner. An example of this is the upward motion of the center of mass of an object, which can raise the mass in the presence of gravity at the expense of this motion. This occurs in a way that at its apex no trace of this energy remains anywhere in its original form, but the energy can be recovered by allowing the mass to fall. Disorganized energy, on the other hand, cannot undergo a wholly recoverable conversion of this type. An example of this energy is in the randomly directed motion about the center of mass of an object, such

---

[1] This technique is used in Appendix A, where the definition of differential entropy change is developed. The technique is explained there in detail.

as the above cited example of thermally induced motion of an ideal gas. This is the type to which internal energy is ascribed. No known process can completely convert this type of energy to work without leaving some other effect on the system, and it is this fact that predicates the Second Law of thermodynamics. But first things first.

The First Law of thermodynamics is a statement of the law of conservation of energy applied to quantities pertinent to thermodynamic consideration. It is given as

$$dU = dQ + dW \tag{1.1}$$

where dU is the differential change in internal energy,
dQ[2] is the heat into the system, and
dW[2] is the work done on the system.

Of primary importance, at least from a practical standpoint, is a way of expressing how much of a system's energy is available to do work. Entropy is a construct of thermodynamics entwined with this purpose. More specifically, entropy reflects the amount of energy unavailable in this regard. From a basic principle of this subject, established from a generalization of the Carnot cycle[3], entropy change is given by

$$dS \equiv \frac{dQ}{T} \tag{1.2}$$

where dS is the differential change in the entropy of a system,
dQ is the heat into the system, and
T is the system's temperature.

---

[2] A note on notation, at this point, is prudent. Heat is represented here in differential notation as dQ and work is represented as dW, though neither quantity is actually the differential of any function. Although there is a theoretical maximum amount of heat which can be evolved from a system and a theoretical maximum of work which can be performed (in each case being the system's total energy, a fact following from energy conservation), no definite function for heat or work exists. This is because these quantities are not unique to a system's state but instead are the consequences of the specific process through which the system is taken. This is juxtaposed against the differential of the state function for internal energy, dU, which is a perfect differential. A system in a given state has a definite internal energy, but the heat and work that contributed this energy are ambiguous since their contributions are merged by addition. This fact, rooted in the additive relationship of these terms within the First Law of thermodynamics, is a lucky feature of the universe, making engines (such as the cellular engines of life) possible. In this paper the differential notation scheme was chosen for conventional consistency and also to emphasize the 'differential magnitude' of the quantities.

[3] See Appendix A for thoughtfulness on entropy's underpinning as a state variable.

This leads to

$$\Delta S = \int_{rev} dS = \int_{rev} \frac{dQ}{T} = \frac{Q}{T} \tag{1.3}$$

for a sufficiently small, reversible process increment, over which the heat transferred does not appreciably affect the temperature. This allows the temperature, T, to be treated as constant, thereby validating the implicit assumption of equilibrium. An important consequence of this definition is its resulting additive properties. Since entropy change is given by a summation (integral), total entropy change over separate processes is then the algebraic sum over the individual entropy changes.

Repeated use is made of the fact that, since entropy is a variable of system state, it is therefore independent of the particular history of the system to bring about that state. Entropy change between two states is therefore independent of the particular process undertaken between them. This allows the entropy change over any arbitrary process to be determined by the change over a particular process that proceeds between the same initial and final states. The chosen process may be selected for matters of convenience, such as a relative ease in evaluation. This same argument can be made for any state variable, bringing about an important artifice in thermodynamics, the use of the quasi-static process. This is a process carrying a system between two states over which externally imposed changes are by design made to be arbitrarily minute at any moment in time, thus maintaining the system arbitrarily close to equilibrium. This is an important feature since thermodynamics makes extensive use of quantities such as temperature and volume, which rely on the assumption of equilibrium. The use of quasi-static processes lends the possibility of mathematical evaluation in determining exact changes in state variables over processes which themselves are nonnegotiable.

Of primary interest in this paper are the implications and interpretation of the Second Law of thermodynamics. As is the case with any physical law, this law is a summary of experimental observation. Specifically, it recounts the observation that systems having undergone certain processes cannot, by any possible faculty, restore themselves to their prior state without the influence of an external agent. Such processes are labeled *irreversible*, and in the course of a system's natural progression these are the ones preferred by the universe wherever possible. They are processes over which the entropy increases, a fact concisely stated as the inequality of the Second Law of thermodynamics,

$$\Delta S \geq 0. \tag{1.4}$$

The equality stands for the remaining possibility of the reversible process, where the system in its entirety can restore itself without external influence. But the irreversible processes are of more pressing interest here. Particularly illustrative examples of these processes include heat transference toward thermal equilibrium, damping, the expansion of a gas, and mixing action. Being particularly illustrative examples, special attention is devoted to each of these phenomena in the following sections.

2. Irreversibility of Heat Transference toward Thermal Equilibrium

A most basic occurrence motivating the subject of thermodynamics is heat transference toward thermal equilibrium, where a uniformity of temperature is established across a system initially without such uniformity. For a treatment of this phenomenon, consider two objects of differing temperatures placed in thermal contact with each other but otherwise isolated. Say the H side has a uniform temperature higher than the uniform temperature of the L side. Experience dictates that with time these objects approach a common, intermediate temperature, after which the temperature remains steady. This process is known to be irreversible, and it will be shown to be consistent with the Second Law, requiring that a process bringing this about be one of increasing total entropy. As given above in eqn. 1.3,

$$\Delta S = \frac{Q}{T}$$

for a sufficiently small increment of a process over which the temperature does not appreciably change. To satisfy the criterion of reversibility linked to this expression, the process may be viewed instead as an approximating quasi-static process[4]. Applying this to the situation being considered, we see that the entropy change of the H side over a brief instant is

$$\Delta S_H = \frac{Q_H}{T_H}, \tag{2.1}$$

and for the L side it is concurrently

$$\Delta S_L = \frac{Q_L}{T_L}. \tag{2.2}$$

Having recognized that heat represents a form of energy, the law of conservation of energy demands that

$$Q_H + Q_L = 0$$

---

[4] This process may be of the type encountered in Appendix A.

or

$$Q_H = -Q_L. \tag{2.3}$$

For convenience, a new variable representing the heat flow in a positive sense is introduced as

$$Q = Q_L, \tag{2.4}$$

chosen in this way since positive quantities of heat correspond to heat influx and heat naturally flows from the hotter H side into the colder L side. From this

$$\Delta S_H = \frac{-Q}{T_H} \tag{2.5}$$

and

$$\Delta S_L = \frac{Q}{T_L}. \tag{2.6}$$

Entropy is an algebraically additive quantity, so the total entropy change is the sum of these terms,

$$\Delta S = \Delta S_H + \Delta S_L = \frac{-Q}{T_H} + \frac{Q}{T_L}$$

which after factoring and a slight rearrangement yields

$$\Delta S = Q\left[\frac{1}{T_L} - \frac{1}{T_H}\right]. \tag{2.7}$$

The situation is prescribed such that

$$T_H > T_L, \tag{2.8}$$

from which it follows that

$$\frac{1}{T_L} - \frac{1}{T_H} > 0. \tag{2.9}$$

Also noting that Q by definition represents a positive quantity, then eqn. 2.7 implies

$$\Delta S > 0. \tag{2.10}$$

In the other possible case where the temperatures are initially equal, then eqn. 2.9 becomes

$$\frac{1}{T} - \frac{1}{T} = 0, \tag{2.11}$$

and the possibility

$$\Delta S = 0 \tag{2.12}$$

emerges. This exhausts the range of possibilities for initial conditions, having considering both matching and mismatching temperatures, and therefore leads to the broad conclusion

$$\Delta S \geq 0,$$

which is the Second Law of thermodynamics.

3. Irreversibility of Damping

To investigate damping phenomena, it is useful to reiterate the distinction between two forms of energy mentioned earlier. Organized energy is that which by some mechanism can be interchanged completely with work, such as in lifting a weight, while disorganized energy is that which cannot undergo such a complete conversion, such as internal energy.

Systems are known to progress in such a way that their energy of the organized kind dissipates with time, suggesting the conversion of organized to disorganized energy. In no case is it observed that disorganized energy, without other change, contributes to the development of organized energy, inhibiting the suggestion of the converse. By conservation of energy, the loss of organized energy in a closed system must contribute positively to internal energy of the system so as to keep the total energy constant. So this must occur in a form similar to the influx of energy in the case of heat transference toward thermal equilibrium. The implied transfer is accordingly one with a positive quantity of heat. Applying the formula for entropy change, given by eqn. 1.3 as

$$\Delta S = \frac{Q}{T}$$

with

$$Q \geq 0, \tag{3.1}$$

where the inequality denotes energy transfer as described above and the equality denotes no energy transfer. Also with

$$T > 0, \tag{3.2}$$

as is always the case with temperature, then we retrieve the Second Law once again,

$$\Delta S \geq 0.$$

4. Irreversibility of Gaseous Expansion

To treat gaseous expansion, consider a partitioned container with ideal gas confined only to one side, as shown in Figure 1. This side of the vessel has volume V_1. After removal of the partition, experience dictates that the gas will expand to fill the entire volume of the container, $V_1 + V_2$. To determine the associated entropy change, the expansion process is chosen to be an isothermal one, attractive for its simplicity in examination. The entropy change over this chosen process is applicable to any, since, as pointed out earlier, entropy is a function of the system's state alone and not of the history of the system in attaining that state. The definition for entropy change, given by eqn. 1.2, provides a good starting point

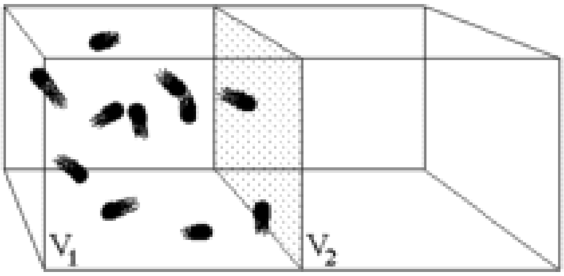

Figure 1: Partitioned container confining an ideal gas

$$dS \equiv \frac{dQ}{T}$$

or rewritten

$$dQ = T\,dS. \tag{4.1}$$

In the case of an ideal gas

$$dW = -P\,dV, \tag{4.2}$$

and also in the case of the gas undergoing an isothermal process where

$$\mathrm{dT} = 0, \tag{4.3}$$

then

$$\mathrm{dU} = \frac{3}{2}\,\mathrm{nR\;dT} \;=\; 0. \tag{4.4}$$

Inserting these into the First Law of thermodynamics, given by eqn. 1.1 as

$$\mathrm{dU} = \mathrm{dQ} + \mathrm{dW},$$

we find

$$0 = \mathrm{T\;dS} - \mathrm{P\;dV}$$

or following some algebraic manipulation

$$\mathrm{dS} = \frac{\mathrm{P}}{\mathrm{T}}\,\mathrm{dV}. \tag{4.5}$$

Next, to perform this integration, a substitution is made to express the right side of this equation as a function only of V using the ideal gas law,

$$\mathrm{PV} = \mathrm{nRT} \tag{4.6}$$

or

$$\frac{\mathrm{P}}{\mathrm{T}} = \frac{\mathrm{nR}}{\mathrm{V}}, \tag{4.7}$$

obtaining

$$\mathrm{dS} = \frac{\mathrm{nR}}{\mathrm{V}}\,\mathrm{dV}. \tag{4.8}$$

Integrating from the initial to the final state gives

$$\int_{\mathrm{i}}^{\mathrm{f}} \mathrm{dS} = \int_{\mathrm{i}}^{\mathrm{f}} \frac{\mathrm{nR}}{\mathrm{V}}\,\mathrm{dV} \tag{4.9}$$

$$\Delta\mathrm{S} = \mathrm{nR}\big[\ln|\mathrm{V}_1 + \mathrm{V}_2| - \ln|\mathrm{V}_1|\big]. \tag{4.10}$$

Since volume is never a negative quantity, the magnitude operators are neglected.

$$\Delta S = nR\left[\ln\left(V_1 + V_2\right) - \ln\left(V_1\right)\right]$$

$$\Delta S = nR\ln\left(\frac{V_1 + V_2}{V_1}\right). \tag{4.11}$$

Again making use of this restriction on sensible volume measurements, quantities for V_2 are limited to the range

$$V_2 \geq 0 \tag{4.12}$$

so

$$V_1 + V_2 \geq V_1$$

$$\frac{V_1 + V_2}{V_1} \geq 1 \tag{4.13}$$

and

$$\ln\left(\frac{V_1 + V_2}{V_1}\right) \geq 0. \tag{4.14}$$

With n and R necessarily non-negative, then

$$nR\ln\left(\frac{V_1 + V_2}{V_1}\right) \geq 0. \tag{4.15}$$

Comparing this to eqn. 4.11, the Second Law of thermodynamics has been manufactured yet again,

$$\Delta S \geq 0.$$

5. Irreversibility of Mixing Action

Consider this time a container with a barrier initially separating ideal gas of two distinct types.[5] This situation is depicted in Figure 2. Each type of gas has some distinguishing feature irrelevant to the mixing process. For this discussion, take that feature to be color, with the partition separating black and gray particles. The color of the particles has no bearing on their diffusion.[6] Additionally, the temperatures and pressures initially present in each of the sections are set to match.

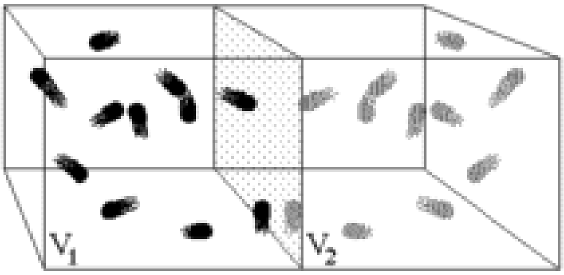

Figure 2: Partitioned container with two types of ideal gas

Entropy is additive in a way that the mixing process may be accomplished by the accumulation of simpler steps. Drawing from the preceding treatment of isothermal expansion of an ideal gas, the steps are chosen to be separate isothermal expansions of each of the initial collections of gases. That is, the particles starting in V_1 expand into $V_1 + V_2$, just as those starting in V_2 expand into $V_1 + V_2$. From eqn. 4.11

$$\Delta S_1 = n_1 R \ln\left(\frac{V_1 + V_2}{V_1}\right) \tag{5.1}$$

and

$$\Delta S_2 = n_2 R \ln\left(\frac{V_1 + V_2}{V_2}\right), \tag{5.2}$$

both of which are positive if not zero, as was concluded from the above treatment of expansion. The total entropy change is given by the sum of the entropy changes of these steps,

---

[5] 'Mixing' of gas of the same type is not covered here in detail, but in that case the entropy change is not of the same order as that involving different types. The obvious reason for the discrepancy is that the interchange of positions resulting from this kind of 'false mixing' produces little distinguishable effect. To some degree, however, the entropy is subject to increase because the number of particles contained within each volume of the initial compartments is no longer definite. This argument is essentially the topic of Gibbs' paradox.

[6] With removal of the partition in the dark, the mixing process is still expected to occur, showing that the mixing will take place even under conditions where the mixing process is necessarily oblivious to this feature.

$$\Delta S = \Delta S_1 + \Delta S_2$$

$$\Delta S = n_1 R \ln\left(\frac{V_1 + V_2}{V_1}\right) + n_2 R \ln\left(\frac{V_1 + V_2}{V_2}\right)$$

$$\Delta S = R\left[n_1 \ln\left(\frac{V_1 + V_2}{V_1}\right) + n_2 \ln\left(\frac{V_1 + V_2}{V_2}\right)\right]. \tag{5.3}$$

This represents the summation of two non-negative terms, as established above, and is therefore itself non-negative, so

$$\Delta S \geq 0,$$

showing yet another emergence of the Second Law.

6. Summary of Information Theory

In the midst of the explosive technological growth following the Second World War, Claude Shannon published his paper *A Mathematical Theory of Communication*[7], which provides the basis for the subject now known as information theory. This theory was motivated by engineering considerations spawned in the development of communications systems, but Shannon was keenly aware of the connection his work shared with thermodynamics. His theory is completely self-contained, formulated from its own abstract axioms, but it may be applied to science with complete consistency. From the theory, an alternative expression for entropy is developed, which is shown to be in agreement with that developed from a basis of thermodynamics.

The primary issue Shannon wished to address was that of communication, i.e. that which is necessary to reproduce the state of a local system on a remote one. Of all possible states available to the remote system, the state that it is to adopt is indicated in some way by the information presented to it from the local system. As a fundamental principle of information theory, this idea of information was quantified in such a way as to satisfy a basic defining characteristic: information is to increase as the number of possibilities from which a selection is made increases. This complements the intuition that more information be necessary to uniquely identify a state from among a larger number of possible states. The choice for mathematical expression for this quantity is

---

[7] *The Bell System Technical Journal,* Vol. 27, pp. 379–423, 623–656, July, October, 1948.

$$I = K \log(W) \tag{6.1}$$

where K is a constant, and

W is the number of equally likely possibilities from which one is identified by the information.

By the change of base formula, the base of this logarithm and the constant act in conjunction to describe the units of information. We may therefore say

$$I = K' \ln(W). \tag{6.2}$$

Further, we may introduce Boltzmann's constant in place of K', scaling the measure of information so that it is represented in units familiar to thermodynamics[8]

$$I_{\text{thermo}} = k \ln(W). \tag{6.3}$$

More may be said to lend credence to this particular choice for the mathematical expression of information. Although any monotonically increasing function of W would suffice to represent information according to its defining characteristic, the logarithm function is the most convenient choice. This is because the properties of the logarithm correspond most closely to the intuitive notion for information. (e.g. Two identical storage media should have twice the capacity for information than one, even though the total number of possible states is squared.) It is for reasons of convenience such as this that information is expressed in this way.

More generally, the quantity of information may be attributed to a set of states which do not share the same likelihood of occurrence. For this, the expression for the information associated with a particular state may be given with statistical reference as

$$I = -K \log(p) \tag{6.4}$$

where the argument of the logarithm, p, is the probability of the state's occurrence. This complements the intuition that more information be associated with the

---

[8] Boltzmann's constant is typically presented in units of energy over units of temperature, implying that information carries this same measure. This conclusion is misleading, however, and does not facilitate a proper regard for the role of information. Temperature is actually a measure of a form of energy, thermal energy, and can be expressed in such units. Therefore, by cancellation, information may also appear dimensionless. Clearly, this type of argument does not establish unequivocal units for information. Information is typically expressed in bits, units representing the most basic form of distinction. Each bit specifies one from among two possibilities with equal probability of occurrence.

specification of a state of lesser likelihood, an extension of the previously asserted presumption that more information be required to identify a particular state from among a larger number of possibilities.

In the case of equally likely states, this form produces the one established above. The probability of selection of each from among W of these states is

$$p = 1/W \tag{6.5}$$

which, by substitution into eqn. 6.4, gives

$$I = -K\log(1/W)$$

$$I = K\log(W), \tag{6.6}$$

in accordance with eqn. 6.1.

The average of this quantity of information over the range of possible selections is also of interest. In fact, it is this which Shannon refers to as entropy.[9] Denoted by the letter H, it is given per symbol by

$$H = \sum_i p_i I_i \tag{6.7}$$

$$H = -K\sum_i p_i \log(p_i) \tag{6.8}$$

or for consistency with the customary system of units in thermodynamics

$$S = -k\sum_i p_i \ln(p_i). \tag{6.9}$$

7. The Connection with Information

For an example of the physical connection to information theory, let's resume consideration of the example previously given for entropy of mixing. In the equilibrium state, the system is at a uniform temperature with total volume

---

[9] Further justification for this expression of entropy is the content of Appendix B. The argument there follows closely to and is clearly inspired by one appearing in Leon Brillouin's *Science and Information Theory*. (Brillouin, Leon, *Science and Information Theory*, Second Edition. New York: Academic Press Inc., 1962.)

$$V = V_1 + V_2 \tag{7.1}$$

and total population

$$n = n_{\text{black}} + n_{\text{gray}}. \tag{7.2}$$

From eqn. 5.3, the entropy change of the mixing process is

$$\Delta S = R\left[n_{\text{black}} \ln\left(\frac{V_1 + V_2}{V_1} \right) + n_{\text{gray}} \ln\left(\frac{V_1 + V_2}{V_2} \right) \right]$$

$$\Delta S = R\left[n_{\text{black}} \ln\left(\frac{V}{V_1} \right) + n_{\text{gray}} \ln\left(\frac{V}{V_2} \right) \right]. \tag{7.3}$$

Introducing a one in the form

$$n / n = 1 \tag{7.4}$$

and then distributing the denominator yields

$$\Delta S = nR\left[\frac{n_{\text{black}}}{n} \ln\left(\frac{V}{V_1} \right) + \frac{n_{\text{gray}}}{n} \ln\left(\frac{V}{V_2} \right) \right]. \tag{7.5}$$

Now introducing the ideal gas law in the form

$$V = \frac{nRT}{P} \tag{7.6}$$

we obtain

$$\Delta S = nR\left[\frac{n_{\text{black}}}{n} \ln\left(\frac{nRT/P}{n_{\text{black}} RT/P} \right) + \frac{n_{\text{gray}}}{n} \ln\left(\frac{nRT/P}{n_{\text{gray}} RT/P} \right) \right]$$

$$\Delta S = nR\left[\frac{n_{\text{black}}}{n} \ln\left(\frac{n}{n_{\text{black}}} \right) + \frac{n_{\text{gray}}}{n} \ln\left(\frac{n}{n_{\text{gray}}} \right) \right]$$

$$\Delta S = -nR\left[\frac{n_{\text{black}}}{n} \ln\left(\frac{n_{\text{black}}}{n} \right) + \frac{n_{\text{gray}}}{n} \ln\left(\frac{n_{\text{gray}}}{n} \right) \right]. \tag{7.7}$$

From the definition of probability as the ratio of the tally of a particular outcome to the total number of possible outcomes, with the particular outcome here being the differentiation between the black and gray particles, then

$$p_{\text{black}} = n_{\text{black}} / n \tag{7.8}$$

and

$$p_{\text{gray}} = n_{\text{gray}} / n. \tag{7.9}$$

Into eqn. 7.7, we get

$$\Delta S = -nR\left[p_{\text{black}} \ln\left(p_{\text{black}}\right) + p_{\text{gray}} \ln\left(p_{\text{gray}}\right)\right]$$

$$\Delta S = -Nk\left[p_{\text{black}} \ln\left(p_{\text{black}}\right) + p_{\text{gray}} \ln\left(p_{\text{gray}}\right)\right], \tag{7.10}$$

which corresponds exactly to a case of Shannon's formula for the entropy of a message, as given in eqn. 6.9, for all N particles. In this case the message is comprised of N symbols, each of which represents one of the two particle types, black and gray. Shannon's entropy is intended to describe the information content of a message, so the emergence of his formula suggests that a change in information be associated with the mixing process. This change may be regarded as the information lost as the original arrangement of particles is scrambled. And since it is this change in information that is equated to entropy change, entropy, therefore, appears in some way to be representative of the information specifying the arrangement of the system.

To further the connection between entropy and information, we return to the example of expansion of an ideal gas treated earlier. Expansion is a process over which the localization of the constituent particles is lost as they become uniformly distributed over a larger volume. Recall the vessel considered in this example is initially partitioned to create separate internal compartments with volumes V_1 and V_2, and prior to the expansion, the gas is confined to the compartment of volume V_1. The corresponding entropy change over the process is given by eqn. 4.11 to be

$$\Delta S = nR \ln\left(\frac{V_1 + V_2}{V_1}\right).$$

In approaching this change from the perspective provided by information theory, we begin by deducing the probability of a gas particle to be in V_1 prior to expansion. The gas particle is localized to region V_1, so the probability of finding it there is

$$p_A = \frac{V_1}{V_1} = 1 \qquad (7.11)$$

with associated information about the system state of

$$I_A = -k \ln\left(p_A\right) = -k \ln\left(1\right) = 0. \qquad (7.12)$$

Following the expansion, the system has reached equilibrium, suggesting a uniform density over the entire available volume $V_1 + V_2$. At this point, the probability of the particle's occurring in V_1 is

$$p_B = \frac{V_1}{V_1 + V_2}, \qquad (7.13)$$

with information in this case of

$$I_B = -k \ln\left(p_B\right) = -k \ln\left(\frac{V_1}{V_1 + V_2}\right) = k \ln\left(\frac{V_1 + V_2}{V_1}\right). \qquad (7.14)$$

In order to restore the system after such an expansion, information is needed in recognizing the localization of the particles. To know a particle is within the original boundary, the amount of information per particle is

$$I_{restore} = I_B - I_A = k \ln\left(\frac{V_1 + V_2}{V_1}\right) - 0 = k \ln\left(\frac{V_1 + V_2}{V_1}\right), \qquad (7.15)$$

and for N particles

$$I_{RESTORE} = N\, I_{restore} = Nk \ln\left(\frac{V_1 + V_2}{V_1}\right) = nR \ln\left(\frac{V_1 + V_2}{V_1}\right). \qquad (7.16)$$

This is exactly the change in entropy. From this and the result from mixing[10], an interpretation for entropy's role may be advanced.

Entropy is typically referred to as being a measure of disorder. This is not a description conducive to precise interpretation, but from the preceding examples the rational for this association becomes clear. An ordered condition is one about which information describing the particular arrangement of a system is held. Each of these examples begins with an ordered condition which over time evolves into one of lesser order. In the case of mixing, this information corresponds to the initial, ordered arrangement of the particles, sorted by color. In the case of expansion, this corresponds to the circumstance of the particles being contained within a certain volume, smaller than the obvious constraint of the vessel boundaries. In each of these examples, the entropy change reflects the information required to restore the systems to their initial, more ordered conditions following a process over which they have become disordered. So entropy change, then, may be viewed as representative of the information lost over the system's progression into a more disordered state. Therefore, it can be construed that entropy itself is a measure of the lack of information specifying a system's particular state. That is, entropy measures the uncertainty that exists about the specific, detailed structure of a system. In the typical considerations of thermodynamics, this may refer to the microscopic state of the system as understood from quantum theory. When applied more generally, however, this may also include larger scale detail, such as the identity of a particular playing card.

8. Bolstering Boltzmann

In the previous section, the practices of information theory are applied to selected physical phenomena to establish the connection between information and entropy. In this section we investigate the connection in a more general sense.

Imagine a system free to attain a state about which the observer has no foreknowledge. To that observer, each possible state has an equal probability of

---

[10] In each of these examples, the entropy change resulting from classical consideration has been shown to match a result derived from information theory. However, the aspects of information theory drawn upon in each case may seem different. In the case of mixing, the entropy change of thermodynamics is equated to Shannon's entropy of a message. In the case of expansion, the entropy change of thermodynamics is equated to the expression for information itself. The reason for this difference is that the nature of the information in each case is different. Mixing is a special case, where the numbers of particles of each type are fixed and the information is only associated with their arrangement. From the perspective of information theory, this means that the statistics of the message are fixed. With expansion, however, the statistics change. That is, the likelihood of a particle's being in each region changes as the populations of the regions change. As a result, the expression for information reverts to the more primitive one.

occurrence. Thus,

$$p_i = p \tag{8.1}$$

and the entropy is

$$S = -k \sum_i p \ln(p)$$

$$S = -k \ln(p) \sum_i p. \tag{8.2}$$

Since, by the nature of a summation over a complete set of probabilities,

$$\sum_i p = 1 \tag{8.3}$$

then the simplification can be achieved

$$S = -k \ln(p). \tag{8.4}$$

With W states available, the probability of occurrence of any single state is

$$p = 1/W \tag{8.5}$$

and therefore by substitution

$$S = -k \ln(1/W)$$

or

$$S = k \ln(W), \tag{8.6}$$

precisely the statistical formulation of entropy attributed to Boltzmann.

9. Subjectivity of Information and Entropy

One of the basic principles that underlies the philosophy of science is the belief of consistency in observations, irrespective of the personal condition of the observer. Quantities such as pressure, volume and temperature satisfy this principle in that they are determined through tests for which the experimenter's condition is irrelevant. Entropy, however, does not satisfy this principle. For evidence of this, consider an example of two people playing a card game, a newcomer and a card counter. The card counter, who has been paying close attention to the deck for some time, has developed a keen awareness of the order of the cards. The newcomer, who may exhibit equal proficiency at counting the cards, has no initial awareness of the deck's order. While the card counter can offer specific statements about the state of the deck, the newcomer can at best only offer probabilistic statements.

This inconsistency manifests itself in the entropy of the deck. Its entropy, according to Boltzmann, is

$$S = k \ln(W). \tag{9.1}$$

From the perspective of the newcomer, who possesses no familiarity with the state of the deck, the multiplicity, W, is

$$W = 52!. \tag{9.2}$$

This is arrived at from the fact that a card in a particular position in the deck may be any from among the 52 available possibilities, another card in a different position may be from among the remaining 51, etc.

The card counter, however, with his specific knowledge of the deck accepts less freedom for the possible order of the cards. With n cards whose positions are known to the card counter, the multiplicity is

$$W = (52 - n)! \tag{9.3}$$

where, by definition of the informed card counter, n is greater than zero. Clearly these differing results for multiplicity yield different results for entropy.

This conclusion is consistent, however, with the interpretation lent to entropy as a measure of the uncertainty about a system’s detail. Having more information, the uncertainty of the card counter should be different from (less than) that of someone with no information about the state of the deck. The characteristics of the system as it exists independently, though, must not be a function of the observers. Information and entropy jointly play a complementary role to assure this by maintaining a constancy in a system’s degree of detail (that is, how much about the system can be specified). This is a characteristic only of the system itself. To explore this in greater depth, let us further investigate the information and entropy associated with the deck of cards.

Generally, eqn. 9.3 stands for the multiplicity of the deck for the n cards known to the observer, regardless of what n is. The entropy for the observer is

$$S = k \ln\left((52 - n)!\right). \tag{9.4}$$

The corresponding information held by that observer is

$$I = k \ln\left(\frac{52!}{(52 - n)!}\right)$$

$$I = k \ln\left(52!\right) - k \ln\left((52 - n)!\right), \tag{9.5}$$

arrived at by recognizing that positions with unknown cards do not impose distinct possible arrangements of the deck, and therefore do not contribute to the number of possibilities from which the information makes a distinction.

The degree of detail is then the sum of the known information, I, and the lack of information, S.

$$I + S = k \ln\left(52!\right) - k \ln\left((52 - n)!\right) + k \ln\left((52 - n)!\right)$$

$$I + S = k \ln\left(52!\right) \tag{9.6}$$

which, as required, is independent of the characteristic of the observer, n.

By their nature, information and entropy account for the state of the observer. But although they individually play subjective roles, their confluence is irrespective of the condition of the observer, leading to the invariant truth demanded by underlying scientific principle. Information is obtained at the expense of entropy, and conversely entropy is obtained at the expense of information, further corroborating their underlying connection. But together their combined description of the system is independent of the observer. To put this in other words, information reflects the amount of detail known about a system, entropy reflects the amount of detail unknown, and their total reflects the total amount of detail that exists about the system.

10. The Meaning of Information

It has so far been established that information serves to specify the state occupied by a system in more detail. It has also been established that entropy serves to measure the uncertainty in the specific detail of a system. But the idea of knowing about a system may seem personified. It is worth addressing what is actually meant by having information about a system.

To borrow a phrase from Shannon[7], information is defined in mathematical form to quantify that which is necessary to "...[reproduce] at one point either exactly or approximately a message selected at another point." The information associated with a message conveys the selection of a particular state at the receiver as dictated by the transmitter. Working from this interpretation, this means that the existence of information between two independent systems implies a correlation in their states. The state of one system is in some way a function of the state of the other. This fact provides a good context for recognizing the purpose that information serves. It reflects this recognized correspondence between the states of independent systems. In other words, information results from a known relationship that exists between the states, allowing the state of one system to provide an indication as to the state of the other. An example of this is a pair of computers coupled by a network between which a message is sent. Upon receipt of the message, the contents of one computer's memory are made to reflect the contents of the other computer's memory. Such an occurrence is precisely what information theory is intended to depict. Another example returns to the card counter. His actions in the game are made in response to the order of the deck, suggesting that internally he is tracking the order of the cards and therefore has some mental representation of the deck. In his case, the information indicates the correspondence between the deck's order and the internal representation of this order within the card counter's mind.

Information is established through communication. In fact, historically it was the study of communication that provided the impetus for information theory's development. Just as information has a physical interpretation of being a correlation between states, communication, too, may be interpreted physically. Information is established by a coupling between the two systems so that the state of the observing system is made to be a function of the system being observed. It is this coupling that is meant by communication. In the example of the communication between computers, the coupling mechanism is the network hardware whose task it is to create a correspondence between the memories of the independent computer systems. In the case of the card counter, the coupling mechanism is the card counter's senses, as he watches the cards. By doing this, he learns their positions to establish his mental representation of the deck.

To summarize, a correlation between systems is made by communication, which is a coupling between the systems, so that their states are not free to be arbitrary with respect to one another. This results in the creation of information, which reflects the constraint on the freedom of one system in attaining a state relative to the other. With information, the state of one of the systems in some way indicates something about the other’s state. This is the physical interpretation of information.

11. Self-Depletion

In earlier sections, examples are highlighted of various types of systems exhibiting some initial imbalance which, over time, approach a condition of greater uniformity, i.e. a state of equilibrium. To explore a possible general explanation for this behavior, consider a system with some conserved characteristic labeled A. After a balanced division of the system into two regions, called 1 and 2, presume this quantity A is additive such that its value summed over the two regions, $A_1 + A_2$, equals the measure over the entire system, A. Examples of such quantities include mass, energy and population.

If the barrier to A separating the two regions is removed (or more generally any coupling mechanism is established between them), then, assuming that elemental units of A are agitated such that they are free to migrate between the regions, those elements in 1 diffuse by mixing into 2 and those in 2 similarly diffuse into 1.

Consider the details of this migration. Over some period, Δt, an elemental unit of A has some probability of crossing into the other region. The total amount crossing over this period is proportional to this probability, κ, but also proportional to the amount present on the original side. The amount transferred from V_1 into V_2 is

$$\Delta A_1 = -A_1 \kappa \, \Delta t \tag{11.1}$$

and from V_2 into V_1

$$\Delta A_2 = -A_2 \kappa \, \Delta t. \tag{11.2}$$

The difference existing between the two regions, with A_1 taken to be the more populous quantity, is

$$D = A_1 - A_2, \tag{11.3}$$

and the change over that period of time is

$$\Delta D = \Delta A_1 - \Delta A_2$$

$$\Delta D = -A_1 \kappa \, \Delta t + A_2 \kappa \, \Delta t$$

$$\Delta D = \left(A_2 - A_1\right) \kappa \, \Delta t$$

$$\Delta D = -D \, \kappa \, \Delta t. \tag{11.4}$$

In the limit of an infinitesimal slice of time

$$dD = -D \, \kappa \, dt. \tag{11.5}$$

Integrating reveals

$$\int \frac{1}{\mathrm{D}}\,\mathrm{dD} = \int -\kappa\,\mathrm{dt}$$

$$\ln|\mathrm{D}| + C = -\kappa \mathrm{t}. \tag{11.6}$$

An equation of exponentials is formed from this equation, giving

$$e^{\left(\ln|\mathrm{D}|+C\right)} = e^{-\kappa t}. \tag{11.7}$$

At this point, two potentially perplexing steps must take place. First, the absolute value operation is discarded since D is non-negative by its definition given in eqn. 11.3. Second, since C represents only an arbitrary constant, the result of an operation on it that yields a constant, such as e^C, is still represented as C. These steps reveal

$$\mathrm{D}C = e^{-\kappa t}. \tag{11.8}$$

Dividing to move the arbitrary constant to the other side, and for the last step, assigning it an appropriate meaning, the initial difference, (which suggests a non-zero value, in compliance with the division) gives

$$\mathrm{D} = \mathrm{D}_0 e^{-\kappa t}. \tag{11.9}$$

Therefore, with a system which can be divided into two regions over which a difference exists, and with the possibility of a symmetric exchange between these regions, the difference follows an exponential decay and therefore diminishes with time. This is exactly consistent with the evolution observed in the cited phenomena from preceding examples.

We take from this that a system exhibiting any form of imbalance tends to reduce its imbalance with time as a result of the ‘overpopulated’ side depleting itself faster than the ‘underpopulated’ side. This conclusion is based on the assumption that symmetric free exchange within the system is permitted, meaning that the exchange process favors no particular direction. This assertion is typically satisfied by the symmetry in the governing dynamics of the processes considered.[11] Equilibrium emerges as the circumstance when populations and, as a consequence, the flow between them become equalized.

Specifically applied to those examples to which special attention has been devoted, those elemental units exchanged between the regions are identified in the cases of heat transference and damping as units of energy. In the case of expansion, the measure is taken to be the population of particles, and finally in the case of mixing, the measure is taken as duel populations of particles of particular types. Thus, a general mechanism has been cited to explain the natural processes undertaken by these systems.

12. The Second Law of Thermodynamics as a Consequence

In his paper, Shannon noted that any reduction in the disparity among a message's statistics results in an increase in entropy for that message. In his own words[7], "Any change toward equalization of the probabilities p_1, p_2, …,p_n increases [entropy]. Thus if $p_1 < p_2$ and we increase p_1, decreasing p_2 an equal amount so that p_1 and p_2 are more nearly equal, then [entropy] increases." Making an observation of a system is in no way incongruous with receiving a message (in fact receiving a message may be thought of as making an observation of the transmitter). Therefore Shannon's insight may be applied here. As the probabilities of identifying distinct features of a system become increasingly proximate in value, the entropy attached to that system should consequently increase. The previously asserted self-depletion mechanism promotes just such a convergence of probabilities. The result is an increase in entropy that is reminiscent of the Second Law of thermodynamics. It is this line of reasoning that is the focus of this section. By again pursuing the fundamental example of expansion, we establish this result by demonstrating that the time derivative of the entropy function be necessarily positive if non-zero, thereby indicating that the entropy over a process, if changing, increases with time.

---

[11] This fact is considered in detail in another paper of mine. Lacking a formal title, it has been dubbed the paper on 'impossible machines.' The content of this paper is summarized well in its first two lines:

> In the course of trying to gain insight into the second law of thermodynamics, the consideration of machines whose operation could be imagined to violate this law has become a central focus. These machines rely on the appearance of a preferred direction in the coupling between thermal effects and an output of some kind with the potential to do work.

This quandary is discussed, leading to the conclusion:

> Any process, to the extent that it can be driven by its own thermal energy at equilibrium, is symmetric with respect to direction of energy flow because of the symmetry in coupling mechanisms, and therefore equilibrium is necessarily maintained. Thus, the second law is satisfied. And with this revealing explanation, I can be satisfied with the second law.

In the case of expansion the noteworthy feature of the constituent particles of a system is their position, whether they are found within their original confines or beyond them. From information theory, the entropy bound to an observation of each particle of the system is given in eqn. 6.9[12] as

$$S=-k\sum_{i} p_i \ln\left(p_i\right), \tag{12.1}$$

where in this case two exclusive possibilities exist. Those possibilities are of being found within or beyond the original confines. With a total population of N, the population within the original volume, V_1, is denoted N_1 with a probability of occurring for a given particle, p_1, and the population within the newly available volume, V_2, is denoted N_2 with a probability of occurring, p_2. To determine the time derivative of the entropy function, we start by finding its derivative with respect to the probability term p_1, and then, using the time derivative of the probability, we arrive at the sought after time derivative of the entropy function by the Chain Rule of calculus.

To get the derivative of the entropy function with respect to p_1, we must first express entropy as a function of p_1. This endeavor is begun by expanding the entropy summation for this case, revealing

$$S=-k\left[p_1 \ln\left(p_1\right)+p_2 \ln\left(p_2\right)\right]. \tag{12.2}$$

The events represented by p_1 and p_2 are individually exclusive but together all inclusive. Therefore the sum of these probabilities must be one. From this

$$p_2=1-p_1. \tag{12.3}$$

Making this replacement into eqn. 12.2 gives

$$S=-k\left[p_1 \ln\left(p_1\right)+\left(1-p_1\right)\ln\left(1-p_1\right)\right]. \tag{12.4}$$

To facilitate finding the derivative of this with respect with p_1, the multiplication of the

---

[12] In the earlier section where the connection between thermodynamics and information theory is drawn, a footnote comments that the entropy change of expansion is not represented by Shannon's entropy. Yet as the current section begins, Shannon's formula is proposed to serve this purpose. The reason given earlier for the failure of the formula is the changing nature of the statistics as the expansion occurs. Driven by the theme of that section, to establish a physical connection with information theory, these statistics are based on spatial details, implicitly relying on a condition of uniform density. The argument here presents an alternate set of statistics without implicit reliance on uniformity, thereby removing the difficulty previously encountered.

second term is carried out.

$$S=-k\left[p_1\ln\left(p_1\right)+\ln\left(1-p_1\right)-p_1\ln\left(1-p_1\right)\right] \tag{12.5}$$

The derivative is now found.

$$\frac{dS}{dp_1}=-k\left[\ln\left(p_1\right)+\frac{p_1}{p_1}+\frac{1}{1-p_1}\left(-1\right)-\frac{p_1}{1-p_1}\left(-1\right)-\ln\left(1-p_1\right)\right]$$

$$\frac{dS}{dp_1}=-k\left[\ln\left(p_1\right)+1+\frac{1}{1-p_1}\left(p_1-1\right)-\ln\left(1-p_1\right)\right]$$

$$\frac{dS}{dp_1}=-k\left[\ln\left(p_1\right)+1-1-\ln\left(1-p_1\right)\right]$$

$$\frac{dS}{dp_1}=-k\left[\ln\left(p_1\right)-\ln\left(1-p_1\right)\right] \tag{12.6}$$

Later, use will be made of this. Next, however, the derivative of p_1 with respect to time must be found. For this, p_1 must be found as a function of time. Drawing from the proceeding description of self-depletion, the difference, D, in this case is identified as the difference in populations of the two compartments. Mathematically, it is expressed as

$$D=N_1-N_2 \tag{12.7}$$

where N_1 is understood to remain greater than N_2 over the course of the expansion. Now recalling that

$$N=N_1+N_2$$

or

$$N_2=N-N_1, \tag{12.8}$$

the substitution is made into eqn. 12.7

$$D=N_1-\left(N-N_1\right)=2N_1-N. \tag{12.9}$$

The difference, D, was found in the proceeding section as eqn. 11.9 to be

$$D = D_0 e^{-\kappa t}$$

so equating these, we arrive at

$$2N_1 - N = D_0 e^{-\kappa t}$$

or

$$N_1 = \frac{N + D_0 e^{-\kappa t}}{2}. \tag{12.10}$$

From this, p_1 may be given as a function of time.

$$p_1 = \frac{N_1}{N} = \frac{N + D_0 e^{-\kappa t}}{2N} = \frac{1}{2} + \frac{D_0 e^{-\kappa t}}{2N} \tag{12.11}$$

and its time derivative is

$$\frac{dp_1}{dt} = \frac{d}{dt}\left[\frac{1}{2} + \frac{D_0 e^{-\kappa t}}{2N}\right] = -\frac{D_0 \kappa}{2N} e^{-\kappa t}. \tag{12.12}$$

By the Chain Rule, the time derivative for entropy may now be found

$$\frac{dS}{dt} = \frac{dS}{dp_1} \cdot \frac{dp_1}{dt} = -k\left[\ln(p_1) - \ln(1 - p_1)\right]\left(-\frac{D_0 \kappa}{2N} e^{-\kappa t}\right) \tag{12.13}$$

and with appropriate substitutions for p_1, this is expressed as a function of time

$$\frac{dS}{dt} = \frac{dS}{dp_1} \cdot \frac{dp_1}{dt}$$

$$\frac{dS}{dt} = -k\left[\ln\left(\frac{1}{2} + \frac{D_0 e^{-\kappa t}}{2N}\right) - \ln\left(1 - \left(\frac{1}{2} + \frac{D_0 e^{-\kappa t}}{2N}\right)\right)\right]\left(-\frac{D_0 \kappa}{2N} e^{-\kappa t}\right)$$

$$\frac{dS}{dt} = -k\left[\ln\left(\frac{1}{2} + \frac{D_0 e^{-\kappa t}}{2N}\right) - \ln\left(\frac{1}{2} - \frac{D_0 e^{-\kappa t}}{2N}\right)\right]\left(-\frac{D_0 \kappa}{2N} e^{-\kappa t}\right)$$

$$\frac{dS}{dt} = \frac{kD_0\kappa}{2N}\left[\ln\left(\frac{1}{2}+\frac{D_0 e^{-\kappa t}}{2N}\right) - \ln\left(\frac{1}{2}-\frac{D_0 e^{-\kappa t}}{2N}\right)\right]e^{-\kappa t}. \tag{12.14}$$

At this point we have a description of the change in entropy over time. Over the course of a natural process subject to the self-depletion mechanism previously established, the entropy of the system evolves according to this rule. It will now be demonstrated that the rule demands that the time derivative of entropy be positive or zero, in agreement with the Second Law of thermodynamics.

The expression may be broken into three distinct parts to discuss the signs of its factors. The first is the constant coefficient, $\frac{kD_0\kappa}{2N}$. This coefficient is positive since all of its factors are positive if non-zero. Boltzmann's constant, k, is known to be positive. The initial difference, D_0, is positive as a result of its use as the difference of the greater population minus the smaller one, or it is zero if no difference exists. The rate of exchange factor, κ, is positive following from the nature of its use in the eqns. 11.1 and 11.2 in the self-depletion section or zero if no coupling is allowed. In the denominator, the number two is positive, of course, and the total population, N, is a positive number for a realistic situation of interest to exist. Therefore, this coefficient is a positive number, if non-zero.

The next part of the expression is taken to be the bracketed quantity, $\left[\ln\left(\frac{1}{2}+\frac{D_0 e^{-\kappa t}}{2N}\right) - \ln\left(\frac{1}{2}-\frac{D_0 e^{-\kappa t}}{2N}\right)\right]$. This quantity is positive since it is a subtraction of a monotonically increasing function, the natural logarithm, with the argument of the positive term greater than that of the negative term. In other words, by drawing from the property of logarithms that the greater a number, the greater its logarithm, the positive term has the greater logarithm argument, so the positive term itself is greater, leading to a positive difference. Therefore this portion of the expression is positive.

The last part of the expression is that which remains, $e^{-\kappa t}$. By the nature of this function, it is always positive. Therefore, the product of these three factors, the time derivative of entropy, is positive if non-zero. Thus entropy is increasing over natural processes (or remaining constant in the case of zero), meaning the argument presented here is consistent with the Second Law of thermodynamics,

$$\Delta S \geq 0.$$

13. The Impression of the Second Law of Thermodynamics

In the arcane language of a physicist, the Second Law of thermodynamics states that the entropy of a closed system never decreases. A crucial step is to recognize what feature of the universe this law is meant to represent - that is, to recognize the essence of the law. Having granted a new prospective upon thermodynamics provoked by the statistical considerations of information theory, this becomes possible.

Information is a measure of how much coincidence there exists in the states of independent systems. Entropy is its counterpart, measuring the uncertainty of coincidence between them. Increasing entropy, as demanded by the Second Law, therefore implies a loss of information as the recognition of coincidence between the states is eroded. From the conceptual framework provided by information theory, the Second Law may be interpreted as stating that the amount of information held about a closed system must decrease, if it changes. In other words, we live in a forgetful universe. With the additional insight supplied by the meaning attributed to information, this means that independent systems, such as the closed system and its former observer[13], diverge as the state of the closed system, the state of the observer or both scramble. As this divergence progresses, the observer's information is exchanged for entropy, thereby causing the entropy to increase.

This theoretical justification may complement intuition, but it is also worthwhile to cultivate a more practical impression. Four examples are given earlier in this paper of natural processes that exhibit consistency with the Second Law, and they have provided a good stage for scrutiny to advance the preceding arguments. But they are contrived, idealized examples whose artificially simplified nature belies the complexity of real-world phenomena. So, consider more everyday examples.

Metal ores, crude oil and sand are among the raw materials processed and assembled to build many machines of the modern world. A clear parallel emerges between the construction of these machines and the organization of the initial state of sorted particles in the mixing example. For the materials fashioned into a machine, information is known about their state. Their organization is known to be within the bounds of the device into which they are formed. The car provides an example. It represents a narrow range of possible organizations from among those possible for its constituent atoms. Information is associated with this order. Furthermore, the role of the Second Law is in the car's journey to eventual failure, as 'wear and tear' scrambles the particular assemblage of atoms, diminishing the information known of the arrangement. The car goes from being a pristine duplicate from among seemingly identical new cars to a car about which less is known. It may have

---

[13] The observer must cease his spying activity. Otherwise the notion of a closed system is violated. This is because observation is a coupling between the states of the observer and the observed, meaning that the systems lack the isolation required of being closed.

developed a worn fan belt, a leaky radiator, a broken cam shaft, etc. as it diverges from the new car state. This is exactly the reason care is advised when purchasing a used car.

Another example is the operation of an engine. Every engine is required to engage in some form of consumption in order to sustain its operation. This fact follows from the Second Law of thermodynamics. According to this law, the operation of the engine results in an entropy increase and therefore a loss of information. As has been established, information represents a recognized correspondence between states. The information in this case reflects the correspondence that exists between the form of the fuel[14] and the structure of the engine. The structure of the engine makes reference to the particular nature of the fuel in order to recognize how to take advantage of it through consumption. A person does not derive sustenance from gasoline, nor does a car run on a sumptuous meal. In the same way, a solar cell and a windmill do not interchange. Any process with a coupling mechanism customized to take advantage of fuel in order to produce work serves as an engine. The operation of the engine, with its regard for the fuel's form, thereby consumes the fuel and diminishes this form. This takes place at the expense of information in exchange for entropy, occurring in a familiar manner for its resemblance to self-depletion. Earlier in this paper, entropy is described as being entwined with an expression of the energy unavailable for work. It is in this way that the account is made.

For a less concrete type of information, consider language. The grade-school telephone game, where a message is orally passed along a string of children, is an example. As each child couples with his neighbor to pass the message, the possibility of error accumulates and the uncertainty in the message increases. Increasing uncertainty implies increasing entropy, so this occurs in accordance with the Second Law. Such an accumulation of errors in oral communication is apparent as a natural driving mechanism in the evolution of language, for example, where dialects and accents are the result of perversion of their root language.

In an address to Congress after strife with President Truman, General Douglas MacArthur announced that, "...old soldiers never die; they just fade away." Though intended, no doubt, as a tribute to himself, this statement is also a testament to the Second Law of thermodynamics.

14. The Statistical Role in Thermodynamics

A troubling difficulty with thermodynamics is its apparent reliance on statistical results. Systems evolve with increasing entropy, as experience shows they approach

---

[14] The use of the word fuel, here, is intended in a more general sense than its typical function. Fuel may account for a chemical agent to consume, but also a temperature or density gradient, among other things.

a state of greater uniformity. This provides a strong temptation to accept a direct relationship between entropy and the degree to which a system is uniform. But the possibility exists that a system may happen to engage in a series of events purely by chance that have the effect of reducing uniformity. If entropy increase implies an approach to uniformity, it is troubling that, upon inspection of a system, a state of greater uniformity is not necessary found.

In the typical example of gaseous expansion, entropy increase is shown to coincide with the dispersion of gas particles, suggesting that entropy is in some way a measure of the degree of this dispersion. The expectation arises that a greater entropy be associated with a system state of greater expansion, so from the Second Law the observed property of expansion of a gas follows. Yet it is easy to imagine from the detailed mechanics of the particle model that, as a result of random fluctuation, spontaneous compression may occur naturally. Further, this compression may occur naturally on a large scale, should the random particle motion happen to arrange it. This is in clear violation of the expectation that the Second Law lead to the establishment of a uniform, equilibrium condition. Somewhat mitigating this unsettling possibility is the extreme improbability of such an event occurring to a highly significant degree. But the blemish on the credibility of thermodynamics stands. Entropy is a variable of system state, after all, and should therefore unambiguously indicate the condition of the system.

Key to unraveling this contradiction is the recognition that, although entropy is a state variable, it accounts for more than just the system in isolation. Entropy reflects the relationship between the system and its observer. It is this fact that removes the array of difficulties in conceptualizing thermodynamics in accordance with other branches of physics that operate from a detailed view. Although entropy increases with gaseous expansion, that increase does not follow from the physical separation of the particles. Entropy is not tied to the particular arrangement of particles, but instead in the uncertainty of precisely specifying that arrangement. The commonplace view that increased entropy suggests a condition of greater expansion, based on a blind association, is the cause of the difficulty by asserting that an observation of a less expanded condition is not possible, despite no such restriction from mechanics. The interpretation through information theory, however, presents a conceptual framework from which this difficulty does not arise. From this view, the entropy increase only notes the increase in uncertainty as the system evolves, regardless of whether that evolution be in a direction of increasing or decreasing uniformity. The natural favoritism toward uniformity is the result of that condition’s statistical advantage, where expanded states account for an overwhelmingly large proportion of all possibilities. But a more uniform condition is not directly endorsed by entropy increase. Expansion's ubiquitous role in describing entropy increase is inspired only by the fact that it represents an easily imagined case where an initial state of detail is recognized. But evolution from any particularly specified state is a case of increase in entropy.

One consequence of this outlook on entropy's role that may arouse suspicion is the decrease in entropy required of the system upon observation, as its state is learned by the observer and the uncertainty of the system's state accordingly decreases. This entropy decrease is not in contrast to the Second Law, however, since the prohibition of such decrease applies only to closed systems, of which a system under external observation is not. The entropy decrease of the observed system is offset by the entropy increase of the observer. (e.g. A candle must burn for one to read in a dark room). It is this fact that robs Maxwell's demon of his fantastic feature. And with this, suspicion is allayed.

Although the possibility of a natural spontaneous compression is unlikely, nothing in the mechanical point of view absolutely precludes this possibility. And it is now established that nothing in the thermodynamic view does either. Entropy is a function of probabilistic parameters of a system, not physical ones, and entropy change only recognizes probabilistic changes in identifying specific system states. So, entropy does not make any comment to indicate the detailed state of a system. To the contrary, entropy's purpose is to acknowledge the uncertainty in the occurrence of any specific detailed state.

15. Introduction of Temporal Asymmetry

The primary interest of this paper is to suggest an intuitively satisfying interpretation of thermodynamics. At this point, this in large part is accomplished,[15] with the application of information theory bestowing conceptual clarity to the nebulous notions of thermodynamics. Paramount among these notions is the idea of entropy, established to be a gauge of the uncertainty in precisely identifying the particular detailed state of a system. The Second Law of thermodynamics is then established from this to be a statement implying that this uncertainty increases, if changing at all, because of the scrambling action of natural mechanisms such as thermal fluctuation. These mechanisms continually disorder the system, thereby effacing any correspondence that may exist between the states of independent systems. With this theoretical framework in place, the final pursuit of the paper may be engaged: demonstrating the uniqueness of thermodynamics.

An illuminating contradiction appears from again examining the view of the kinetic theory of gases being contrast with thermodynamics. This contradiction sets the stage for demonstrating the unique role thermodynamics plays. It comes about from the familiar foundation which is typical for an argument in thermodynamics, the box of ideal gas. Again imagining this system as a collection of bouncing particles, the system appears as though it may be described in its entirety by an approach

---

[15] Satisfaction is at the discretion of the reader.

rooted in mechanics. This leads to a new troubling difficulty, reconciling the time asymmetry of thermodynamics with the time symmetry of mechanics. But the resolution to this quandary leads to an affirmation of the fundamental posture of thermodynamics.

The postulates of mechanics take exactly the same form whether operating forward or backward in time. Therefore mechanics is time symmetric. This is not the case, however, with thermodynamics, which is sensitive to temporal direction. Theoretically, a system governed by mechanics might be reversed, progressing backward through its past succession of states without further external influence. This is in clear contrast to the demands imposed on that system by thermodynamics. To resolve this contradiction, we turn to the new understanding afforded us by information theory and examine the suspicious underlying assumptions present in the point of view of mechanics (and also pervasive among many other branches of physics).

Mechanics tells us that a body's position is a function of time. Therefore, all information regarding a particle's position over time is completely contained in this function. This leads to the belief that, with information of this type in addition to the information of the precise initial conditions, the long-term behavior of systems may be tracked with infinite precision. This exacting condition, however, is unattainable in practice, because the measurements in ascertaining the initial conditions are always made only to within a limited precision[16]. Therefore the system is not completely resolved. Variability remains beyond the precision of the measurements. Further, with the procession of time, this variability builds on itself, accumulating a divergence of increasing significance. This occurs even in the presence of damping forces, since these forces do not produce energetic losses to the system, but only change the organized character of this energy. The cumulative effect of this divergence is shown through computer simulation in Figures 3a-c and 3d, where the paths of two particles are traced. These particles have identical initial conditions with the exception of a slight deviation in their initial velocities. The paths can be seen to overlap initially but diverge with time. In typical physics nomenclature, this divergence is described as 'error,' but the term seems misapplied since this effect is not the result of any avoidable deficiency but instead is a natural consequence of the act of measuring.

---

[16] This statement is built from classical mechanics, with no regard for the implications of quantum mechanics. This represents no threat to the argument, however, since often in practice measuring limitations occur on larger scales before quantum effects become significant, meaning that the cause of this contradiction is not rooted by necessity in quantum mechanics.

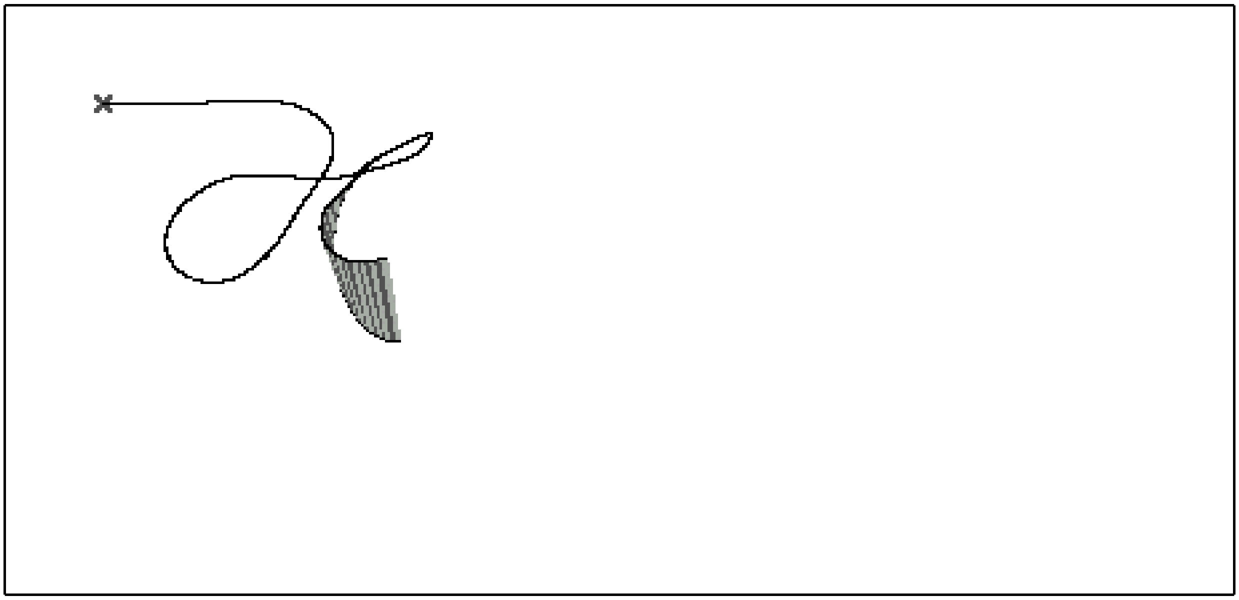

Figure 3a: Computer simulation indicating the difference between the paths of two identical particles with nearly identical initial conditions. Each path, marking the edge of the shaded region, is generated independently and represents the course taken by one free particle subject to $1/r^2$ forces supplied by six other free particles (not shown). The two paths differ in their particles' initial velocity by one one-thousandth. The starting points are indicated by an 'X.'

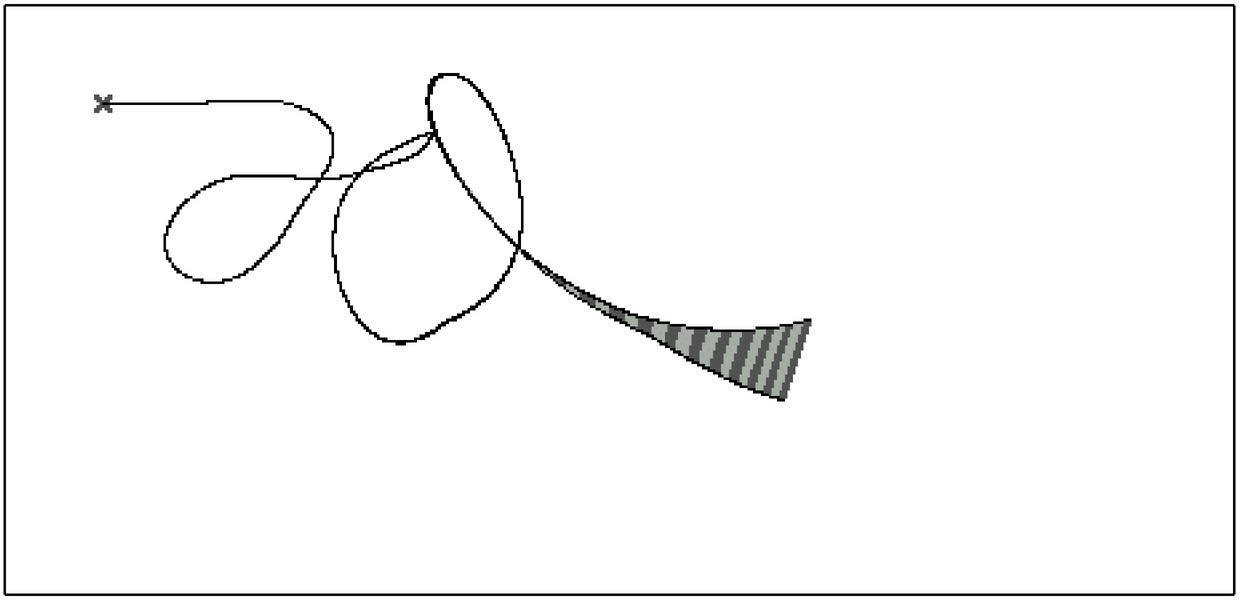

Figure 3b: The difference between two particle paths, again each representing the course of a free particle subject to $1/r^2$ forces from six other free particles (not shown). This time, though, the two paths differ in their particles' initial velocity by one one-millionth. And again, 'X' marks the spot from which their journeys began. Note the span over which the paths are coincident is longer with this smaller initial difference, but the paths show an eventual divergence.

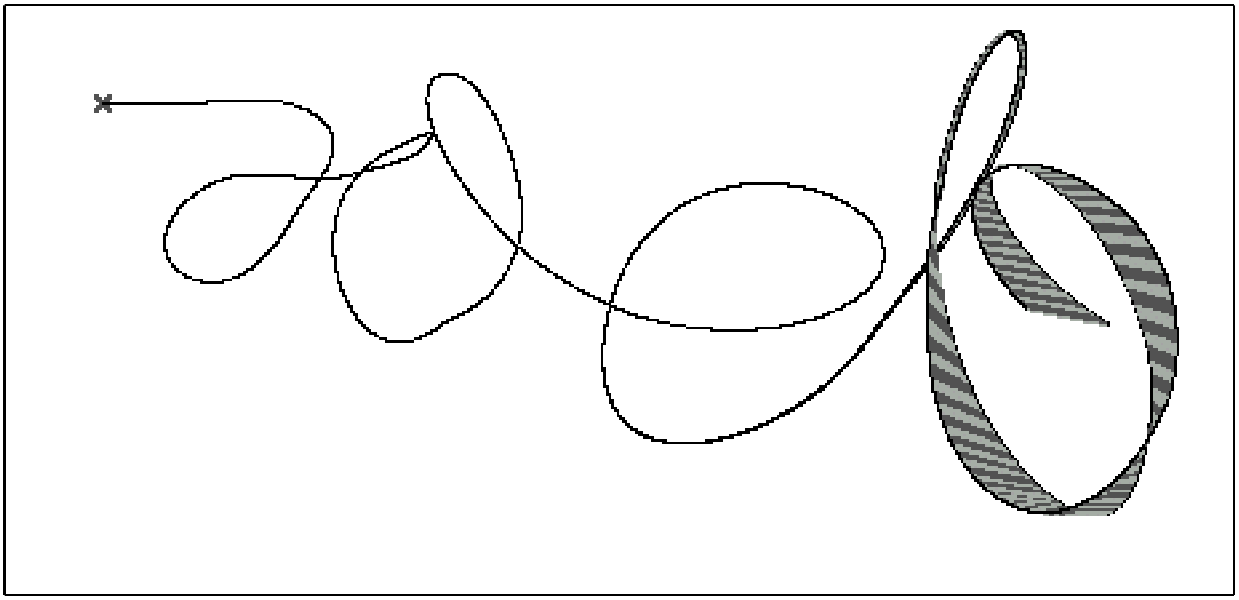

Figure 3c: Once again, the difference between two particle paths starting at the 'X.' This time, though, the two paths differ in their particles' initial velocity by one one-billionth. Note here the span over which the paths coincide is even longer, although they still eventually separate.

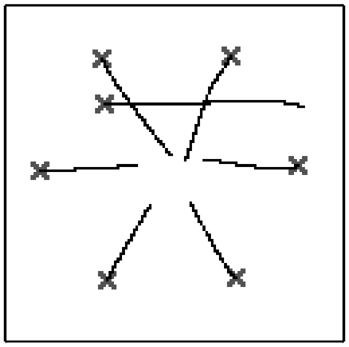

Figure 3d: A glimpse of the situation depicted in Figure 3a, this time with the paths of the six other particles shown.

Entropy, according to the development purposed in this paper, increases with increasing deviation between independent systems. Such a deviation is exactly what the preceding paragraph depicts as an unavoidable result of measuring limitations. Therefore entropy increase is the natural result of these limitations, as the variance they introduce accumulates as divergence between the systems. So from this we may conclude that entropy increase is compatible with, in addition to being independent of, mechanics and other physical descriptions. It serves to recognize the inherent impossibility permitted by these descriptions, that measurements exist to an infinite precision, thereby allowing for an unachievably complete description of a system's state.

Thermodynamics is frequently regarded as a mere statistical regurgitation of the detailed characterizations of large scale systems. This view is then accompanied by the associated comments about intractable numbers of coupled governing equations which, in theory, yield exact solutions. But doubt is now cast on this point of view. Physics is built on measurements, and its most basic purpose is to allow for successful predictions. But because of the unavoidable imprecision of the act of measuring, the uncertainty of predictions in specifying an outcome increases with time, making specific predictions decreasingly likely to be successful. This is what thermodynamics serves to recognize. And with this interpretation in place, the unique role of thermodynamics is established.

* * * * * * *

All deeper arguments aside, truth is nothing more than a belief that provides personal satisfaction. The purpose of this paper is to present an interpretation of thermodynamics which I currently find satisfying, making it truth enough for me. But, as is a convenient feature of science, this belief may be subject to revision or replacement in the future, for the sake of a more satisfying one. For now, though, this will do.

Appendix A

Expression for Entropy Change from Study of the Carnot Cycle

An obvious feature of any state variable is that its change between any two states be equal, regardless of the process undertaken to bring about the change of state. For such a quantity, we look to the Carnot cycle.

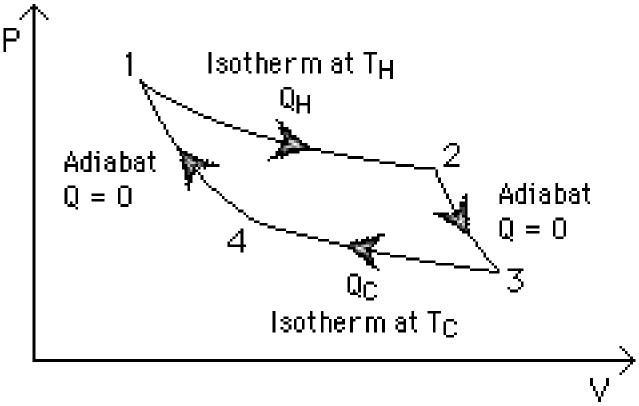

Figure A.1: The Carnot Cycle

The Carnot cycle is a reversible cyclical process imposed on an ideal gas. It is comprised of four subprocesses, as depicted in Figure A.1. The subprocesses extending from state 1 to state 2 and below that from state 3 to state 4 are isothermal. The constant temperature over these subprocesses is maintained by an exchange of heat with the surroundings. The subprocesses from state 2 to state 3 and then from state 4 back to state 1 are adiabatic. Over these, no heat is exchanged.

The amount of heat required in the procession of the isothermal subprocesses may be found in exact form. Since the system is comprised of an ideal gas whose temperature is kept constant, the internal energy of the system is also constant. This stems from the fact that the internal energy of an ideal gas is known to be a function of temperature by

$$U = \frac{3}{2} nRT, \tag{A.1}$$

so its change over temperature is

$$\frac{dU}{dT} = \frac{3}{2} nR \tag{A.2}$$

and

$$dU = \frac{3}{2} nR\, dT, \tag{A.3}$$

and so with

$$dT = 0 \tag{A.4}$$

then

$$dU = 0. \tag{A.5}$$

The First Law of thermodynamics, given in eqn. 1.1 as

$$dU = dQ + dW,$$

in this case simplifies to

$$dQ = -dW. \tag{A.6}$$

Work is done on the gas by compression, and the gas does work on its surroundings by expansion. Therefore,

$$dW = -P\ dV. \tag{A.7}$$

Applying this to eqn. A.6 and integrating for the isotherm occurring at the higher temperature gives

$$Q_H = -\left(-\int_1^2 P\ dV\right). \tag{A.8}$$

From the ideal gas law

$$PV = nRT$$

this produces

$$Q_H = -\left(-\int_1^2 \frac{nRT_H}{V}\ dV\right) = \int_1^2 \frac{nRT_H}{V}\ dV = nRT_H \int_1^2 \frac{1}{V}\ dV = nRT_H\ \ln\left|\frac{V_2}{V_1}\right|. \tag{A.9}$$

Volume measurements are restricted to non-negative numbers, so this simplifies to

$$Q_H = nRT_H\ \ln\left(\frac{V_2}{V_1}\right). \tag{A.10}$$

Similarly for the isotherm of lower temperature, we find

$$Q_C = -\left(-\int_3^4 \frac{nRT_C}{V}\, dV \right) = \int_3^4 \frac{nRT_C}{V}\, dV = nRT_C \int_3^4 \frac{1}{V}\, dV = nRT_C \ln\left| \frac{V_4}{V_3} \right|$$

$$Q_C = nRT_C \ln\left(\frac{V_4}{V_3} \right). \tag{A.11}$$

Later use will be made of these equations for heat over the isothermal subprocesses. But to relate them to each other, we must first find the relationship between the logarithms of the volume ratios they contain. This is done by focusing on the adiabatic subprocesses, where we note that

$$dQ = 0, \tag{A.12}$$

and the First Law becomes

$$dU = dW. \tag{A.13}$$

Again drawing on eqn. A.7, this may be expressed as

$$dU = -P\, dV, \tag{A.14}$$

and replacing from the ideal gas law as before, this becomes

$$dU = -\frac{nRT}{V}\, dV. \tag{A.15}$$

Again in the case of an ideal gas

$$dU = \frac{3}{2} nR\, dT.$$

Combining these equations leads to

$$\frac{3}{2} nR\, dT = -\frac{nRT}{V}\, dV$$

$$\left(\frac{3}{2} nR\, dT \right)\left(-\frac{1}{nRT} \right) = \frac{1}{V}\, dV$$

$$-\frac{3}{2} \frac{1}{T}\, dT = \frac{1}{V}\, dV. \tag{A.16}$$

Integrating this from general states A to B yields

$$-\frac{3}{2}\int_{B}^{A}\frac{1}{T}\,dT=\int_{B}^{A}\frac{1}{V}\,dV$$

$$-\frac{3}{2}\ln\left(\frac{T_B}{T_A}\right)=\ln\left(\frac{V_B}{V_A}\right)$$

$$-\ln\left(\frac{T_B{}^{3/2}}{T_A{}^{3/2}}\right)=\ln\left(\frac{V_B}{V_A}\right)$$

$$\ln\left(\frac{T_A{}^{3/2}}{T_B{}^{3/2}}\right)=\ln\left(\frac{V_B}{V_A}\right). \quad \text{(A.17)}$$

With this, the relation between the quantities of heat required for the two isotherms may be crafted. This is begun by first recognizing the relationship that exists between the logarithms of the volume ratios present in their equations, eqns. A.10 and A.11. From eqn. A.17 applied to each adiabatic process

$$\ln\left(\frac{V_2}{V_3}\right)=\ln\left(\frac{T_C{}^{3/2}}{T_H{}^{3/2}}\right) \quad \text{(A.18)}$$

from state 2 to state 3, and

$$\ln\left(\frac{V_4}{V_1}\right)=\ln\left(\frac{T_H{}^{3/2}}{T_C{}^{3/2}}\right)$$

$$\ln\left(\frac{V_1}{V_4}\right)=\ln\left(\frac{T_C{}^{3/2}}{T_H{}^{3/2}}\right) \quad \text{(A.19)}$$

from state 4 to state 1, and therefore

$$\ln\left(\frac{V_2}{V_3}\right)=\ln\left(\frac{V_1}{V_4}\right). \quad \text{(A.20)}$$

Raising both sides to the power V_3/V_1

$$\left[\ln\left(\frac{V_2}{V_3}\right)\right]^{V_3/V_1} = \left[\ln\left(\frac{V_1}{V_4}\right)\right]^{V_3/V_1}$$

$$\ln\left(\frac{V_2}{V_3}\frac{V_3}{V_1}\right) = \ln\left(\frac{V_1}{V_4}\frac{V_3}{V_1}\right)$$

$$\ln\left(\frac{V_2}{V_1}\right) = \ln\left(\frac{V_3}{V_4}\right)$$

$$\ln\left(\frac{V_2}{V_1}\right) = -\ln\left(\frac{V_4}{V_3}\right). \quad \text{(A.21)}$$

Making a replacement into eqn. A.11 with this, we get

$$Q_C = -nRT_C \ln\left(\frac{V_2}{V_1}\right). \quad \text{(A.22)}$$

Dividing eqn. A.10 by this equation gives

$$\frac{Q_H}{Q_C} = \frac{nRT_H \ln\left(\frac{V_2}{V_1}\right)}{-nRT_C \ln\left(\frac{V_2}{V_1}\right)}$$

$$\frac{Q_H}{Q_C} = -\frac{T_H}{T_C}$$

$$\frac{Q_H}{T_H} = -\frac{Q_C}{T_C}$$

$$\frac{Q_H}{T_H} + \frac{Q_C}{T_C} = 0. \quad \text{(A.23)}$$

This is an essential result for identifying a state variable. The cumulative effect on such a variable over a closed process (one which leads back to its starting state) must result in no change. This summation over the ratios of heat to temperature is a promising candidate since it adds to zero over the Carnot cycle, thereby representing a special case of this general property we expect of a state variable. (It should be noted that this conclusion presumes the variable be subject to the property of algebraic additivity. But, this presumption is verified by the fact that such additivity follows naturally from its development in evaluating a process synthesized through adiabatic and isothermal subprocesses.)

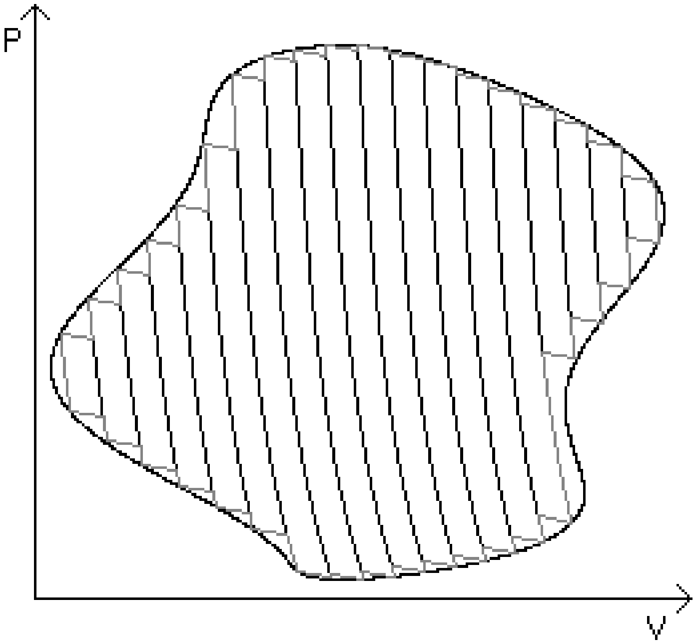

Figure A.2: An arbitrary process approximated by adiabatic and isothermal subprocesses.

We now endeavor to extend this result to an arbitrary process. As it stands, the quantity's essential feature, summing to zero, is only established over a Carnot cycle. Interestingly the adiabatic portions of this cycle do not contribute to this quantity. It accumulates exclusively from the isotherms, over which the temperature remains constant. This inspires a scheme for general application, a glimmer of which appears in the preceding paragraph. An arbitrary closed process may be approximated by a succession of adiabatic and isothermal subprocesses as depicted in Figure A.2. The summation of the heat to temperature ratios over the subprocesses, to which the adiabats do not contribute and over the isothermal remainder of which the temperature is constant, maintains the desired feature

$$\sum_{\text{closed, rev}} \frac{Q}{T} = 0. \tag{A.24}$$

An exact solution is achieved as the approximation approaches the actual process by its construction from infinitesimal subprocess segments. This turns the summation over the closed process into the integral

$$\oint_{\text{rev}} \frac{dQ}{T} = 0, \tag{A.25}$$

where it is noted that Q transforms to dQ only to reflect the fact that the quantity becomes infinitesimal in magnitude. This quantity sums to zero around a closed path in phase space for a reversible process. Further, the summation occurs in a non-trivial manner, not being zero over its entirety. It also appears to return a non-zero result in the absence of a closed process, with only one exception. Just as the state variable of temperature sustains no change over only the isothermal process, the state variable of pressure sustains no change over only the isobaric process and so on, no change occurs to this quantity over only one type of process, the adiabatic one. Therefore we have established the quantity’s viability to act as a state variable. In fact, this is the basis of the entropy of thermodynamics,

$$dS \equiv \frac{dQ}{T}\,. \tag{A.26}$$

Appendix B

Entropy: Abstract Axioms and Intuitive Development

Shannon's motivation for the quantity he refers to as entropy is to measure the average information contained in a message (that is, a particular arrangement of symbols, such as the text on this page) about which only statistical characteristics are known to the receiver. To quote him directly on this topic, "Can we find a measure of how much 'choice' is involved in the selection of the event or of how uncertain we are of the outcome?"[5] where the event he refers to is the recognition of a particular message. He developed this measure by imposing basic properties to which such a measure should adhere, and from this established the unique expression that solely satisfies these requirements. Briefly the requirements are:

1. The entropy, H, should be continuous over the probabilities of occurrence of the message's constituent symbols. The important implication of this requirement is that small changes in the statistics of the message (that is, the probabilities of occurrence of the symbols) should result in only small changes in the entropy of the message.

2. The entropy should be monotonically increasing with an increasing number of outcomes. This means that the uncertainty, which is what entropy measures, should be greater for a greater variability in the possible outcome.

3. The entropy should be additive in such a way that a single choice with multiple outcomes should yield the same result as a succession of choices leading to the same final outcomes. In other words, the total entropy should be the weighted sum of the entropies associated with all of the choices that lead to the final set of outcomes, a scheme which gives the same result regardless of how the choices are grouped.

Shannon provides an elegant proof built from these conditions that firmly establishes entropy be given as

$$H = -K \sum_{i} p_i \log\left(p_i\right) . \qquad (B.1)$$

An alternate approach, catering more closely to intuition, may be developed by straightforward application of information theory. To do this, consider the following two examples of messages assembled from collections of symbols.

First, for a message comprised of n distinct symbols, the number of different arrangements is

$$W = n! . \tag{B.2}$$

The mathematical expression assigned to information is

$$I = K \log(W) \tag{B.3}$$

with equal probability of occurrence of any of the W possible messages. The information content of a particular message composed of these symbols follows as

$$H = I \tag{B.4}$$

$$H = K \log(W). \tag{B.5}$$

Next, for a message composed of n symbols of m different types, where m is less than n, the number of distinct possible arrangements is

$$W = \frac{n!}{\prod_{i=1}^{m} n_i!} . \tag{B.6}$$

The denominator accounts for the fact that arrangements differing only among the positions of indistinguishable symbols of the same kind result in no difference in the outcome, so these arrangements are not recounted. Again with equal probability for the occurrence of any outcome, the information content in this case is

$$H = I \tag{B.7}$$

$$H = K \log\left(\frac{n!}{\prod_{i=1}^{m} n_i!}\right) \tag{B.8}$$

$$H = K\left[\log(n!) - \log\left(\prod_{i=1}^{m} n_i!\right)\right]$$

$$H = K\left[\log(n!) - \sum_{i=1}^{m} \log\left(n_i!\right)\right] \tag{B.9}$$

Substitution for the factorial may be achieved with the use of Stirling’s formula

$$n! \approx n^{n} e^{-n} \sqrt{2\pi n} \approx n^{n} e^{-n} \tag{B.10}$$

where n is assumed to be large. This into the natural logarithm gives

$$\ln(n!) \approx \ln\left(n^{n} e^{-n}\right)$$

$$\ln(n!) \approx \ln\left(n^{n}\right) + \ln\left(e^{-n}\right)$$

$$\ln(n!) \approx n \ln(n) - n$$

$$\ln(n!) \approx n[\ln(n) - 1] \tag{B.11}$$

and this into the expression for entropy, replacing both factorials, gives

$$H = K_{e}\left[n[\ln(n) - 1] - \sum_{i=1}^{m} n_{i}\left[\ln\left(n_{i}\right) - 1\right]\right], \tag{B.12}$$

to a good approximation, where the change in base of the logarithm is compensated by an adjustment of the constant coefficient from K to K_e. Distributing the first term, and rewriting n as the summation over all n_i's yields

$$H = K_{e}\left[n \ln(n) - \sum_{i=1}^{m} n_{i} - \sum_{i=1}^{m} n_{i}\left[\ln\left(n_{i}\right) - 1\right]\right] \tag{B.13}$$

Similarly distributing the n_i's within the following term's summation gives

$$H = K_{e}\left[n \ln(n) - \sum_{i=1}^{m} n_{i} - \sum_{i=1}^{m} n_{i} \ln\left(n_{i}\right) + \sum_{i=1}^{m} n_{i}\right] \tag{B.14}$$

or more simply

$$H = K_{e}\left[n \ln(n) - \sum_{i=1}^{m} n_{i} \ln\left(n_{i}\right)\right]. \tag{B.15}$$

The general logarithm base is now restored, along with the corresponding restoration of the constant coefficient

$$H = K\left[n \log(n) - \sum_{i=1}^{m} n_{i} \log\left(n_{i}\right)\right]. \tag{B.16}$$

Again rewriting n as the summation over all n_i's, we obtain

$$H = K\left[\left(\sum_{i=1}^{m} n_i\right)\log(n) - \sum_{i=1}^{m} n_i \log(n_i)\right] \tag{B.17}$$

and distributing within that summation

$$H = K\left[\sum_{i=1}^{m} n_i \log(n) - \sum_{i=1}^{m} n_i \log(n_i)\right]$$

or

$$H = K\sum_{i=1}^{m} n_i\left[\log(n) - \log(n_i)\right]. \tag{B.18}$$

Combining the two logarithms produces the more concise expression

$$H = K\sum_{i=1}^{m} n_i \log\left(\frac{n}{n_i}\right), \tag{B.19}$$

and making use of another property of logarithms, this is rewritten as

$$H = -K\sum_{i=1}^{m} n_i \log\left(\frac{n_i}{n}\right). \tag{B.20}$$

Introducing a one in the form n/n and distributing the denominator gives

$$H = -nK\sum_{i=1}^{m} \frac{n_i}{n} \log\left(\frac{n_i}{n}\right) \tag{B.21}$$

and recognizing the ratios as probabilities, the appropriate substitution bears

$$H = -nK\sum_{i=1}^{m} p_i \log(p_i). \tag{B.22}$$

For the entropy of one symbol ($n = 1$), this becomes

$$H = -K \sum_{i=1}^{m} p_i \log\left(p_i\right),$$

in agreement with eqn. 6.8, given for this purpose: to express the entropy of a single symbol.