Arrow of time and observer information
Andreas Henriksson

To cite this version:

Andreas Henriksson. Arrow of time and observer information. 2021. hal-03265499v2

HAL Id: hal-03265499
https://hal.science/hal-03265499v2
Preprint submitted on 13 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Arrow of time and observer information

Andreas Henriksson
Stavanger Katedralskole, Haakon VII’s gate 4, 4005 Stavanger, Norway

In this article, the arrow of time as it appears in the statistical interpretation of the second law of thermodynamics is suggested to be of no relevance in understanding the true origin for the directionality of time. To arrive at this point of view, the theory of statistical mechanics is revisited, and restated in terms of the information content possessed by an observer of the system. By doing so, the statement that the arrow of time is due to the tendency of the entropy to increase with time, as encoded by the second law of thermodynamics, is argued to be equivalent to the statement that it is the increasing loss of information that causes time to flow towards the future.

Keywords: Information, Probability, Statistical equilibrium, Entropy, Second law of thermodynamics

INTRODUCTION

It is a commonly held belief that the arrow of time is connected with the concept of disorder, or entropy, as it appears in the statistical interpretation of the second law of thermodynamics. The tendency of the entropy to increase for macroscopic systems seem to define a directionality in which time flows. In this article, this belief is challenged. The key ingredient is to recognize the consequences of the well-established interpretation of entropy as a measure for the uncertainty, or information, possessed by an observer, about the state of the system. The second law of thermodynamics, from this point of view, thus give that the arrow of time is due to the tendency of the observer to loose information over time. This would mean that the asymmetry in time is not a fundamental property of nature but rather due to the non-existence of an enlightened observer. This seem absurd and is therefore suggested to be irrelevant in understanding the true origin for why time flows towards the future rather than the past.

UNCERTAINTY AND COARSE-GRAINING

The dynamical evolution of a system is quite complicated. Most systems of interest contain a vast amount of particles that interact in complicated ways. For such large systems, it is usually very hard to track the individual evolution of each particle as the system evolve in time. Perfect knowledge about the position and velocity, or momenta, of each individual particle is lost. It is lost not because of a fundamental violation of information conservation but merely because of the difficulty for an observer to keep track of the all the degrees of freedom. Therefore, from the perspective of the observer, there is an uncertainty \( \Delta q \) associated with the position of a state and an uncertainty \( \Delta p \) associated with the momentum of a state. For this reason, the observer is unable to determine with absolute certainty the state of the system at any given time. The observer can only determine whether or not the system occupy a state which lie within any given region \( \Omega_j \) on phase space, whose volume \( V_{\Omega_j} \) is given by the uncertainties \( \Delta q \) and \( \Delta p \), i.e.

\[
V_{\Omega_j} = \Delta q \Delta p.
\]

The volume \( V_{\Omega_j} \) is thus a measure of how ignorant the observer is about the details of the system, in the sense that the observer cannot locate an individual state to a greater precision than the size of \( \Omega_j \). Due to this lack of precision, the observer is unable to distinguish between states that lie within \( \Omega_j \). All states within \( \Omega_j \), with their individual sets of degrees of freedom, has, from the perspective of the observer, collapsed into a single state whose single set of degrees of freedom is given by \( q + \Delta q \) and \( p + \Delta p \). This so-called coarse-grained, or mixed, state is not a fundamental, or pure, state of the system. It is a description that average over all pure states within \( \Omega_j \). Put differently, a mixed state \( \psi_j \), \( j \in [1,M] \), where \( M \) is the number of mixed states on phase space, is a subjective representation, by an ignorant observer, of a collection of pure states \( \phi_{\alpha} \), \( \alpha \in [1,N] \), where \( N \) is the number of pure states within \( \Omega_j \). As the system evolve in time, the observer is only able to measure the coarse-grained flow, i.e. the jumping from one mixed state \( \psi_j \) to a different mixed state \( \psi_i \), \( i \neq j \).

It should be noted that due to the lack of perfect knowledge about all the relevant degrees of freedom, the observer is unable to predict a unique evolutionary path on phase space along which the system evolve.

PROBABILITY CONSERVATION

Due to the ignorance of the observer, i.e. the observers inability to distinguish the set of pure states within any given coarse-grained region \( \Omega_j \), it is necessary to introduce the notion of probability on phase space. Let \( P_j \) be the probability that the system occupy the region \( \Omega_j \) and let \( P_{\alpha} \) be the probability that the system occupy the pure
state $\phi_j$ within $\Omega_j$. If the observer know with absolute
certainty that the system occupy the mixed state $\psi_j$ and
not some other state $\psi_i, i \neq j \in [1, M]$, it is given that
\begin{equation}
P_j = 0, \forall i \neq j \in [1, M],
\end{equation}
\begin{equation}
P_j \equiv \sum_{\alpha=1}^{N} P_\alpha = 1.
\end{equation}
For continuous systems, the summation is replaced by an integral, i.e.
\begin{equation}
P_j \equiv \int_{\Omega_j} P_\alpha \, dV_\alpha = 1.
\end{equation}
where $dV_\alpha = dq_\alpha dp_\alpha$ is the phase space volume of
the pure state $\phi_\alpha$. If the knowledge possessed by the observer
about the coarse-grained flow of the system is not lost
over time, then the probability $P_j$ is constant in time, i.e.
\begin{equation}
\frac{dP_j}{dt} = 0.
\end{equation}
In other words, it is assumed that there is no loss of
probability from $\Omega_j$ to any other coarse-grained region
$\Omega_i, i \neq j$. Written in terms of the probabilities $P_\alpha$, the
condition of no loss of coarse-grained knowledge become
\begin{equation}
\frac{dP_j}{dt} = \frac{d}{dt} \int_{\Omega_j} P_\alpha \, dV_\alpha
= \int_{\Omega_j} \left( \frac{dP_\alpha}{dt} + P_\alpha \vec{\nabla} \cdot \vec{v} \right) \, dV_\alpha
= 0,
\end{equation}
where $\vec{v} = (\dot{q}, \dot{p})$ define the phase-space velocity of
the Hamiltonian flow. Since this should hold independently
on the size of $\Omega_j$, the integrand must identically vanish, i.e.
\begin{equation}
\frac{dP_\alpha}{dt} + P_\alpha \vec{\nabla} \cdot \vec{v} = 0.
\end{equation}
This is the continuity equation for probability flow within
any given coarse-grained region $\Omega_j$. It is referred to as the
Liouville equation for the probability distribution within
$\Omega_j$ [1]. Supposing that information is conserved within
$\Omega_j$, i.e. that the Liouville theorem is satisfied such that
$\vec{\nabla} \cdot \vec{v} = 0$ [2], then the probability distribution $P_\alpha$ is
conserved in time, i.e.
\begin{equation}
\frac{dP_\alpha}{dt} = 0.
\end{equation}

**STATISTICAL EQUILIBRIUM**

The continuity equation can be rewritten, showing that
probability is locally conserved within $\Omega_j$. Using the
total time derivative of $P_\alpha$, i.e.
\begin{equation}
\frac{dP_\alpha}{dt} = \frac{\partial P_\alpha}{\partial t} + \vec{\nabla} P_\alpha \cdot \vec{v}
\end{equation}
and the product rule
\begin{equation}
\vec{\nabla} \cdot (P_\alpha \vec{v}) = \vec{\nabla} P_\alpha \cdot \vec{v} + P_\alpha \vec{\nabla} \cdot \vec{v},
\end{equation}
the continuity equation become
\begin{equation}
\frac{\partial P_\alpha}{\partial t} + \vec{\nabla} \cdot (P_\alpha \vec{v}) = 0.
\end{equation}
The term $\vec{\nabla} \cdot (P_\alpha \vec{v})$ represent the difference between the
probability outflow and inflow for the pure state $\phi_\alpha$.
Consider a system which has been closed for a suffi-
ciently long period of time such that the density of pure
states within $\Omega_j$, and hence $M$, do not change with time.
In this situation, the probability distribution $P_\alpha$ has no
explicit dependence on time. The continuity equation is
then reduced to
\begin{equation}
\vec{\nabla} \cdot (P_\alpha \vec{v}) = 0.
\end{equation}
This is the mathematical condition the system need to
satisfy in order for it to be said to exist in statistical
equilibrium. In other words, a system is in statistical
equilibrium if there is no net probability flow on phase
space.

The incompressibility of the Hamiltonian flow imply
that the time the system spend in any single pure state,
before evolving to the next single pure state, is the same
for all pure states. If this was not the case, the state
points on phase space would lump together which would
signify a violation of information conservation. This im-
ply that over the course of a long period of time, the
total time spent by the system in any given pure state
is expected to be the same for all pure states. This ex-
pectation, which is due to a combination of the Liouville
theorem and the law of large numbers, is in this article
interpreted to be equivalent to the ergodic theorem of
statistical mechanics [3][4][5]. Let $n_\alpha$ denote the number
of times the system occupy the pure state $\phi_\alpha$. The total
number of times, $n$, the system occupy the set of $N$ pure
states within $\Omega_j$ is then
\begin{equation}
n = \sum_{\alpha=1}^{N} n_\alpha.
\end{equation}
The ergodic theorem then say that over a long period of
time, such that $n$ is large, it is expected that the sys-
tem occupy all pure states within $\Omega_j$ an equal number of
times, i.e.
\begin{equation}
n_\alpha = n_\beta, \forall \beta \neq \alpha \in [1, N],
\end{equation}
such that
\begin{equation}
n = N \cdot n_\alpha.
\end{equation}
It is now possible to define the notion of a probability
$P_\alpha$ for the pure state $\phi_\alpha$ of a closed system from the
notion of a relative frequency\footnote{It must be emphasized that this relative frequency is not possible to obtain from a set of repetitive experimental measurements, since the observer, being ignorant, is not able to distinguish between the set of pure states.}

\[ P_\alpha \equiv \lim_{n \to \infty} \frac{n_\alpha}{n} = \frac{n_\alpha}{N \cdot n_\alpha} = \frac{1}{N}. \]  

Thus, all the pure states within \( \Omega_j \) are equally probable. This imply that an observer has lost all information, down to the scale of \( V_{\Omega_j} \), about the system, since no distinctions can be made between the possible pure states within \( \Omega_j \). The uniform probability distribution given by equation (16) is commonly referred to as the microcanonical [6], or fundamental [7], probability distribution. Thus, given that the system satisfy the Liouville theorem, the microcanonical probability distribution satisfy the condition for statistical equilibrium.

**ERGODICITY BREAKING**

There exist also non-uniform probability distributions. The non-uniformity arise due to interactions that the system has, or have had in the not too far distant past, with an environment. In other words, the system is, or was recently, not isolated. Due to the interaction with an environment, the density of states change with time. If the interaction is uniform on phase space, the density change uniformly on phase space. However, in general, this is not the case. An interaction, characterized by a potential energy, do depend on the specific values for the generalized coordinates. In that scenario, the density of states is a local function on phase space. This has the consequence that the total time spent by the system within any given region on phase space is not necessarily the same as within any other equally-sized region. In other words, the ergodic theorem appear to be violated. Thus, not only is the probability distribution non-uniform when there is a non-negligible net interaction with the environment, it can also change over time. To put it differently, if there exist an interaction between the system and its environment, as seen from the perspective of an observer of the system, this imply that the observer possess knowledge, i.e. information, about the interaction. This information is used by the observer when assigning probabilities for the possible states of the system. The fact that the observer possess some amount of information mean necessarily that the probability distribution is non-uniform. It is only at statistical equilibrium, where all information is lost, that the observer assign a uniform probability distribution.

From the definition of probability in statistical equilibrium it is clear that the probability for any given pure state decrease as the number of pure states \( N \) increase, i.e. as the uncertainty volume increase. In non-equilibrium, where probabilities are not equal, it is the average probability which decrease as the uncertainty volume increase.

It should be emphasized that the apparent violation of the ergodic theorem is not of a fundamental character. It is only due to the fact that the degrees of freedom associated with the environment cannot be excluded when defining the degrees of freedom for the system. In other words, the environment should be included in the definition of the system. If that is done then there exist no environment and hence there cannot be any net transfer of energy and particles from, or to, the system. Then, this redefined system, which take into account all degrees of freedom, even those which the experimenter may think belong to an 'environment', do indeed conserve information and ergodicity is not broken. The probability distribution for the states of this redefined system is uniform, i.e. all mixed states for any given system, assuming the system has been defined such that no degrees of freedom are being forgotten, are equally probably. In most practical situations, however, there will always exist an environment to any system under study. The question is to what degree this environment interact with the system. The weaker the interaction, the weaker is the ergodicity breaking and the closer will the system come to a uniform probability distribution.

**ENTROPY**

A measure for the amount of information possessed by the observer, i.e. the amount of uncertainty in the determination of the pure state of the system, should depend on the probability distribution \( \{ P_\alpha \} \). This measure is denoted by \( S(\{ P_\alpha \}) \) and referred to as the entropy of the system. To obtain a specific form for the entropy as a function of the probability distribution, it is noted that this function should satisfy the following conditions.

i The entropy should be zero when the observer has complete knowledge about the evolution of the system. In other words, if the observer know with absolute certainty that the system occupy a specific state \( \phi_\alpha \), such that \( P_\alpha = 1 \) and \( P_\beta = 0 \ \forall \beta \neq \alpha \), the entropy must vanish.

ii The entropy should always be either zero or a positive number, i.e. \( S \geq 0 \).

iii The entropy should take a maximum value when the observer is maximally ignorant. This happen when the system is in statistical equilibrium. When all states are equally probable, it imply that the observer possess zero partial knowledge which can
be used to distinguish between some of the features of the set of states. Thus,

\[ P_\alpha = \frac{1}{N} \forall \alpha \in [1, N] \rightarrow S(\{P_\alpha\}) = S_{\text{max}}. \]  

(17)

iv The entropy should, in statistical equilibrium, be a continuously increasing function of the number of states \( N \). In other words, when \( N \) increase, the uncertainty volume \( V_{\Omega_j} \) increase continuously.

v The entropy should satisfy the following composition law,

\[ S(\{P_\alpha\} \cdot \{P_\beta\}) = S(\{P_\alpha\}) + S(\{P_\beta\}). \]  

(18)

This composition law is understood as follows. Let \( \Omega_j \) be divided into two subregions \( \Omega^\alpha_j \) and \( \Omega^\beta_j \) such that \( V_{\Omega_j} = V_{\Omega^\alpha_j} + V_{\Omega^\beta_j} \). The states \( \phi_\alpha, \alpha \in [1, N_\alpha] \), belong to \( \Omega^\alpha_j \) and the states \( \phi_\beta, \beta \in [1, N_\beta] \), belong to \( \Omega^\beta_j \) where \( N_\alpha + N_\beta = N \). The corresponding probability distributions, \( \{P_\alpha\}_{\alpha=1}^{N_\alpha} \) and \( \{P_\beta\}_{\beta=1}^{N_\beta} \), satisfy \( \sum_{\alpha=1}^{N_\alpha} P_\alpha + \sum_{\beta=1}^{N_\beta} P_\beta = 1 \) and, due to them being independent of each other, their product give the probability distribution associated with the region \( \Omega_j \), i.e. \( P(\Omega_j) = \{P_\alpha\} \cdot \{P_\beta\} \). The composition law thus state that the total uncertainty within region \( \Omega_j \) is the sum of the uncertainties associated with the subregions of \( \Omega_j \).

Conditions (i) and (v) suggest that the entropy has a logarithmic dependence on the probability distribution. Condition (ii) suggest that it is necessary to include an additional minus sign in the definition of the entropy. This is seen from the general definition of \( P_\alpha \), i.e.

\[ \log P_\alpha = \lim_{n \to \infty} \log \left( \frac{n_\alpha}{n} \right) = \log n_\alpha - \lim_{n \to \infty} \log n < 0, \]  

(19)

which, for a system in statistical equilibrium become

\[ \log P_\alpha = \log \frac{1}{N} = \log 1 - \log N = -\log N < 0. \]  

(20)

Since the entropy function should act as a measure for systems both in and out of statistical equilibrium, i.e. for both uniform and non-uniform probability distributions, it is required to take the statistical average of all logarithmic contributions to the entropy, i.e.

\[ S(\{P_\alpha\}) \sim -\left( \frac{n_1}{n} \log P_1 + \cdots + \frac{n_N}{n} \log P_N \right) \]  

(21)

\[ \sim -\sum_{\alpha=1}^{N} \frac{n_\alpha}{n} \log P_\alpha \]  

(22)

\[ \sim -\sum_{\alpha=1}^{N} P_\alpha \log P_\alpha. \]  

(23)

This entropy function then satisfy conditions (iii) and (iv). With the proportionality constant identified with the Boltzmann constant \( k_B \), it is referred to as the Gibbs entropy \( \mathbb{G} \) and is, in the information theoretic language, identical to the Shannon entropy \( \mathbb{S}_2 \).

In conclusion, the entropy of a system measure the amount of information within the system, and it is given by the Gibbs formula

\[ S(\{P_\alpha\}) = -k_B \sum_{\alpha=1}^{N} P_\alpha \log P_\alpha. \]  

(24)

In statistical equilibrium, the Gibbs entropy reduce to the Boltzmann entropy \( \mathbb{S}_2 \).

\[ S = k_B \log M. \]  

(25)

It is important to emphasize that entropy is not a physical quantity in the same manner as e.g. energy. It is determined by the probability distribution of the states of the system and as such it is a quantity which depend both on the specifics of the system and the amount of information possessed by the observer.

**SECOND LAW OF THERMODYNAMICS**

If the state of a system is known with infinite precision at some given time, and if the laws of motion are known to infinite precision, then any earlier or later states of the system can be predicted with infinite precision. In such a deterministic situation, information about the system is never lost. However, in practical reality, the experimental precision by which the state can be determined is limited. Instead of knowing the initial conditions with infinite precision they are known to some degree of error, \( \epsilon \), on phase space. Therefore, the state of the system is only known to lie within a finite region, \( \Omega \), of radius \( \epsilon \) and volume \( V_\Omega \). As the system evolve from the initial conditions it is not possible to predict the exact path on phase space. Any two neighboring states within \( \Omega \), e.g. \( a \) and \( b \), see Figure 4 might evolve differently over time. State \( a \) might evolve into either state \( c \) or state \( d \). Due to the limited precision, it is impossible to say which state it evolve into. State \( b \), on the other hand, might evolve into state \( e \) or state \( f \). This process of diverging paths continue as time unfold. Therefore, the number of states in which the system might exist increase over time. In other words, the amount of uncertainty, i.e. the entropy, increase with time. Alternatively put, over time, any observer will continue to loose information about the system as a consequence of not knowing the initial conditions of the system with infinite precision. Actually, it is also possible for the entropy to decrease over time meaning that the observer has gained information about the system. This correspond to the situation when possible paths converge at some point. For example,
the states $x$, $y$ and $z$ all converge into state $k$. The uncertainty of the system has thus decreased since there are now fewer possible states in which the system might exist. However, the probability that paths converge to a single state is much lower than the probability that they diverge to separate states. The reason for this is that the state $k$ is merely one possible state out of a large number of possible states within volume $V_\Omega$ which $x$, $y$ and $z$ could have evolved into. Thus, overall, the observer loses information exponentially over time. Eventually, all information has been lost. The observer has become maximally ignorant. The entropy has reached its maximum value. At this stage, the observer loses information about the system at an exponential rate until there is none left.

**CONCLUSION**

In this article, we have revisited some key aspects of the theory of statistical mechanics and shown that it led straightforwardly to the conclusion that the arrow of time is due to the tendency of an observer to lose information about the details of the system. Can the asymmetry in time really owe its existence to the degree of ignorance of the observer? Whatever the answer might be to the origin of time’s arrow, we feel that it is important to realize that the concept of entropy, and hence time’s arrow as it appears in the second law of thermodynamics, is subjective in the sense that it depends on the amount of information which the observer has accumulated.

---

[1] Bloch, F., *Fundamentals of statistical mechanics: manuscript and notes of Felix Bloch*, Imperial College Press and World Scientific Publishing, London, 2000.

[2] Henriksson, A., arXiv:1905.06185v3 [cond-mat.stat-mech] (2019).

[3] Boltzmann, L.E., *Vorlesungen über Gastheorie*, vol II, Verlag von Johann Ambrosius Barth, Leipzig, 1898.

[4] Birkhoff, G.D., *Proc. Natl. Acad. Sci. USA* 17 (1931), 656.

[5] Neumann, J.v., *Proc. Natl. Acad. Sci. USA* 18 (1932), 263.

[6] Gibbs, J.W., *Elementary principles in statistical mechanics*, Charles Scribner’s sons, New York, 1902.

[7] Khintchine, A.Ya, *Fundamental laws of probability*, Moscow State University, Moscow, 1927.

[8] Shannon, C.E., *The Bell System Technical Journal* 27 (1948), 379.

[9] Jaynes, E.T., *Phys.Rev.* 106 (1957), 620.

[10] Jaynes, E.T., *Phys.Rev.* 108 (1957), 171.

[11] Boltzmann, L.E., *Vorlesungen über Gastheorie*, vol I, Verlag von Johann Ambrosius Barth, Leipzig, 1896.