On the statistical arrow of time

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

What is the physical origin for the arrow of time? It is a commonly held belief in the physics community that it relates to the increase of disorder, or entropy, as it appears in the statistical interpretation of the second law of thermodynamics. In this article, it is argued that the philosophical point of view that the arrow of time is a fundamental property of Nature is incompatible with the subjective information-theoretical interpretation of probability and entropy in statistical mechanics, where they arise due to the ignorance of the observer. This conclusion, however, relies on the premise that time's arrow has a statistical mechanical origin. If this assumption is abandoned, then there is no conflict between the belief in the fundamental character of time's arrow and the belief in the subjective information-theoretical interpretation.

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

Introduction

The story often told about the role of entropy in the second law of thermodynamics and the arrow of time typically goes as follows. Consider a gas of molecules that are constrained within a box of small volume in the corner of a room. As a door in the box is opened, the gas molecules will spread out into the room. That the gas spreads outward into the room, eventually filling all of it, rather than staying put in the corner, or coming back to the corner within a measurable passage of time, is argued to define a direction for the arrow of time. When the gas has spread out evenly in the room, the room is in thermal equilibrium with the gas molecules. Thus, it seems to be the case that systems, which are left to themselves, tend to evolve in time in such a way that they reach thermal equilibrium. From the statistical mechanical point of view, thermal equilibrium happens when all possible microstates are equally probable. The directionality in time is in this picture thus viewed as being due to the flow of probability from a non-uniform to a uniform distribution. The concept of entropy then enter as a type of measure of how far this process has come. For uniform probability distributions, the entropy is at its maximum value. The physical interpretation of this situation is that all states being equally probable means that the uncertainty experienced by an observer about the exact degrees of freedom of all gas molecules is at its maximum. Any given pair of molecules can be interchanged without the observer noticing it. The system is maximally disordered, in contrast to its initial condition where the observer was fairly certain about the locations of the gas molecules, as they were contained in the small box.

Consider another example. Imagine a neatly ordered stack of cards. If the cards are laid out on a table in an array, the system of cards has a well-defined and clear state. There exists only a single state in which the set of cards exist. Any disturbance to this state, such as interchanging the positions of a pair of cards, will yield a new state which is clearly distinct from the ordered initial state. Any human observer can easily tell that the cards have changed places. The entropy of the system of cards is at its minimum. Consider now that the initial condition is changed by throwing the set of cards up in the air and letting them fall randomly on the table. There is no clear pattern to how the cards are positioned on the table. Yet, whatever set of locations occupied by the set of cards, it is a unique state. If a pair of cards switch places, the physical state does change. However, since the initial configuration of cards on the table were random, with no apparent order, it is much more difficult for an observer to notice the change in state coming from interchanging the locations of a pair of cards. In fact, if all cards are completely randomly scattered in an even fashion on the table, it is said that the entropy for the system of cards is, from the viewpoint of the observer, at its maximum. This mean that it is in all practicality impossible for the observer to determine any changes in the state of the system due to interchanging locations of cards. Whatever the locations of the cards are, and however they are interchanged, they system will look the same from the perspective of the observer. When this is the situation, the system has a uniform probability distribution. All possible configurations of cards are equally probable. Entropy is thus a measure for the uncertainty, or ignorance, possessed by the observer about the physical state of the system. If the philosophical point of view that the set of cards have specific physical locations on the table is taken, i.e., that the system occupies a single physical state at any given time, independent on the ignorance of the observer, the concept of entropy must be subjective and should not be considered as a fundamental property of Nature. Based on this reasoning, it seems absurd to argue that the existence of time's arrow is due to entropy increase. At least it appears so if one chooses to believe that time's arrow is a property of Nature rather than a property of the observer's ability to

* andreas.henriksson@skole.rogfk.no

measure changes in the state of a system.

In this article, we will revisit the foundations of the theory of statistical mechanics. We will arrive at the conclusion that if we believe in time's arrow as being due to the increase in entropy, as it is encoded in the statistical interpretation of the second law of thermodynamics, while at the same time interpreting probabilities in terms of the information possessed by observers, then the belief that time's arrow is a fundamental property of Nature must at the same time be given up and rather be viewed as only existing in the minds of ignorant observers.

It should be remarked that the interpretation of probabilities, and hence entropy, as originating from the lack of knowledge by the observer of the system, is not universally accepted. There exist several other interpretations. The information-based interpretation, on which this article is built, is not necessarily reflective of the authors own personal view. That view has, frustratingly, changed many times over the years. The intention of the article is rather to show that if this viewpoint is taken, then the arrow of time must also be subjective rather than a fundamental property of Nature. The different probability interpretations are metaphysical in character, in the sense that there are, as of yet, no experiments to distinguish between them. They are, therefore, subject to personal taste. Independent on the chosen interpretation, the mathematics and theoretical predictions of the theory of statistical mechanics is the same.

Uncertainty and coarse-graining

The dynamical evolution of a system is quite complicated. Most systems of interest contain a vast number of particles that interact in complicated ways. For such large systems, it is usually extremely hard to track the individual evolution of each particle as the system evolve in time. The observer does not possess enough information to state with certainty the positions and velocities of all particles. Perfect knowledge about the position and velocity, or momenta, of each individual particle is lost. The observer loose information about the system over time. It is lost not because of a fundamental violation of information conservation in Nature but merely because of the difficulty for an observer to keep track of all the degrees of freedom. Therefore, from the perspective of the observer, there is an uncertainty Δq associated with the position of a state and an uncertainty Δp 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 the system occupy a state which lie within any given region Ω_j on phase space, whose volume V_{Ω_j} is given by the uncertainties Δq and Δp, i.e.

$$V_{\Omega_j} = \Delta q \Delta p. \tag{1}$$

The volume V_{Ω_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 Ω_j. Due to this lack of precision, the observer is unable to distinguish between states that lie within Ω_j. All states within Ω_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 Ω_j. Put differently, a mixed state ψ_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 Ω_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 ψ_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 evolves.

Probability conservation

Due to the ignorance of the observer, i.e., the observer's inability to distinguish the set of pure states within any given coarse-grained region Ω_j, it is necessary to introduce the notion of probability on phase space. Let P_j be the probability that the system occupies the region Ω_j and let P_α be the probability that the system occupies the pure state ϕ_α within Ω_j. If the observer knows with absolute certainty that the system occupies the mixed state ψ_j and not some other state $\psi_i, i \neq j \in [1, M]$, it is given that

$$P_i = 0, \forall i \neq j \in [1, M], \tag{2}$$

$$P_j \equiv \sum_{\alpha=1}^{N} P_\alpha = 1. \tag{3}$$

For continuous systems, the summation is replaced by an integral, i.e.

$$P_j \equiv \int_{\Omega_j} P_\alpha \, dV_\alpha = 1. \tag{4}$$

where $dV_\alpha = dq_\alpha dp_\alpha$ is the phase space volume of the pure state ϕ_α. If the knowledge possessed by the observer about the coarse-grained flow of the system is not lost over time, i.e., information is conserved, then the probability P_j is constant in time, i.e.

$$\frac{dP_j}{dt} = 0. \tag{5}$$

In other words, it is assumed that there is no loss of probability from Ω_j to any other coarse-grained region $\Omega_i,\ i \neq j$.

Written in terms of the probabilities P_α, the condition of no loss of coarse-grained knowledge become

$$\begin{aligned} \frac{dP_j}{dt} &= \frac{d}{dt}\int_{\Omega_j} P_\alpha \; dV_\alpha \\ &= \int_{\Omega_j} \left(\frac{dP_\alpha}{dt} + P_\alpha \; \vec{\nabla} \cdot \vec{v} \right) \; dV_\alpha \\ &= 0, \end{aligned} \tag{6}$$

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 Ω_j, the integrand must identically vanish, i.e.

$$\frac{dP_\alpha}{dt} + P_\alpha \; \vec{\nabla} \cdot \vec{v} = 0. \tag{7}$$

This is the continuity equation for probability flow within any given coarse-grained region Ω_j. It is referred to as the Liouville equation for the probability distribution within Ω_j.

Statistical equilibrium

The continuity equation can be rewritten, showing that probability is locally conserved within Ω_j. Using the total time derivative of P_α, i.e.

$$\frac{dP_\alpha}{dt} = \frac{\partial P_\alpha}{\partial t} + \vec{\nabla} P_\alpha \cdot \vec{v} \tag{8}$$

and the product rule

$$\vec{\nabla} \cdot (P_\alpha \; \vec{v}) = \vec{\nabla} P_\alpha \cdot \vec{v} + P_\alpha \; \vec{\nabla} \cdot \vec{v}, \tag{9}$$

the continuity equation become

$$\frac{\partial P_\alpha}{\partial t} + \vec{\nabla} \cdot (P_\alpha \; \vec{v}) = 0. \tag{10}$$

The term $\vec{\nabla} \cdot (P_\alpha \; \vec{v})$ represent the difference between the probability outflow and inflow for the pure state ϕ_α.

Consider a system which has been closed for a sufficiently long period of time such that the density of pure states within Ω_j, and hence M, do not change with time. In this situation, the probability distribution P_α has no explicit dependence on time. The continuity equation is then reduced to

$$\vec{\nabla} \cdot (P_\alpha \; \vec{v}) = 0. \tag{11}$$

This is the mathematical condition the system need to satisfy 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 implies 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 were not the case, the state points on phase space would lump together which would signify a violation of Liouville's theorem [1]. This imply that over the course of an extended 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 expectation, 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 [2][3][4]. Let n_α denote the number of times the system occupies the pure state ϕ_α. The total number of times, n, the system occupies the set of N pure states within Ω_j is then

$$n = \sum_{\alpha=1}^{N} n_\alpha. \tag{12}$$

The ergodic theorem then says that over an extended period of time, such that n is large, it is expected that the system occupies all pure states within Ω_j an equal number of times, i.e.

$$n_\alpha = n_\beta, \; \forall \beta \neq \alpha \in [1, N], \tag{13}$$

such that

$$n = N \cdot n_\alpha. \tag{14}$$

It is now possible to define the notion of a probability P_α for the pure state ϕ_α of a closed system from the notion of a relative frequency[1],

$$P_\alpha \equiv \lim_{n \to \infty} \frac{n_\alpha}{n} = \frac{n_\alpha}{N \cdot n_\alpha} = \frac{1}{N}. \tag{15}$$

Thus, all the pure states within Ω_j are equally probable. This imply that an observer has lost all information, down to the scale of V_{Ω_j}, about the system, since no distinctions can be made between the possible pure states within Ω_j. The uniform probability distribution given by equation 15 is commonly referred to as the microcanonical [5], or fundamental [6], probability distribution. Thus, given that the system satisfies the Liouville theorem, the microcanonical probability distribution satisfy the condition for statistical equilibrium.

---

[1] It should be noted 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.

Ergodicity breaking

There exist also non-uniform probability distributions. The non-uniformity arises 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 changes with time. If the interaction is uniform on phase space, the density changes 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 appears 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 possesses 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 assigns 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 because 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 consider 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 interacts 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 knows with absolute certainty that the system occupy a specific state ϕ_α, 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 implies that the observer possess zero partial knowledge which can be used to distinguish between some of the features of the set of states. Thus,

$$P_\alpha = \frac{1}{N} \; \forall \alpha \in [1, N] \; \rightarrow S(\{P_\alpha\}) = S_{max}. \tag{16}$$

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_{Ω_j} increase continuously.

v The entropy should satisfy the following composition law,

$$S(\{P_\alpha\} \cdot \{P_\beta\}) = S(\{P_\alpha\}) + S(\{P_\beta\}). \tag{17}$$

This composition law is understood as follows. Let Ω_j be divided into two subregions Ω_j^α and Ω_j^β such that $V_{\Omega_j} = V_{\Omega_j^\alpha} + V_{\Omega_j^\beta}$. The states $\phi_\alpha, \alpha \in [1, N_\alpha]$, belong to Ω_j^α and the states $\phi_\beta, \beta \in [1, N_\beta]$, belong to Ω_j^β, 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 Ω_j, i.e. $P(\Omega_j) = \{P_\alpha\}\cdot\{P_\beta\}$. The composition law thus state that the total uncertainty within region Ω_j is the sum of the uncertainties associated with the subregions of Ω_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_α, i.e.

$$\log P_\alpha = \lim_{n\to\infty} \log\left(\frac{n_\alpha}{n}\right) = \log n_\alpha - \lim_{n\to\infty} \log n < 0, \quad (18)$$

which, for a system in statistical equilibrium become

$$\log P_\alpha = \log \frac{1}{N} = \log 1 - \log N = -\log N < 0. \quad (19)$$

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.

$$S(\{P_\alpha\}) \sim -\left(\frac{n_1}{n}\log P_1 + \cdots + \frac{n_N}{n}\log P_N\right) \quad (20)$$

$$\sim -\sum_{\alpha=1}^{N} \frac{n_\alpha}{n} \log P_\alpha \quad (21)$$

$$\sim -\sum_{\alpha=1}^{N} P_\alpha \log P_\alpha. \quad (22)$$

This entropy function then satisfies conditions (iii) and (iv). With the proportionality constant identified with the Boltzmann constant k_B, it is referred to as the Gibbs entropy [5] and is, in the information theoretic language, identical to the Shannon entropy [7][8][9].

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

$$S(\{P_\alpha\}) = -k_B \sum_{\alpha=1}^{N} P_\alpha \log P_\alpha. \quad (23)$$

In statistical equilibrium, the Gibbs entropy reduce to the Boltzmann entropy [2][10],

$$S = k_B \log M. \quad (24)$$

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, ϵ, on phase space. Therefore, the state of the system is only known to lie within a finite region, Ω, of radius ϵ and volume V_Ω. 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 Ω, e.g. a and b, see Figure 1, might evolve differently over time. State a might evolve into either

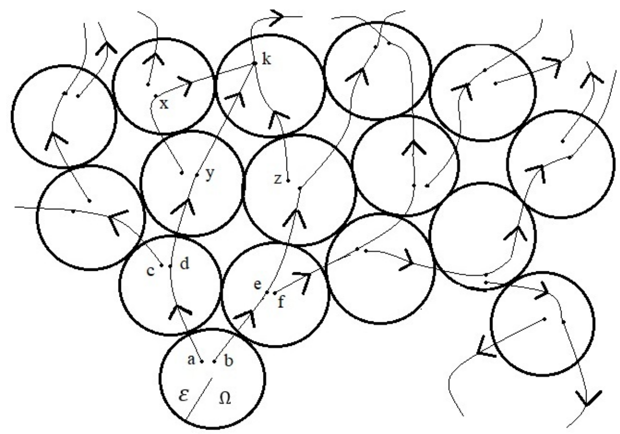

FIG. 1. Irreversible, entropy increasing, flow on phase space.

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 continues 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 lose information about the system because of not knowing the initial conditions of the system with infinite precision. It is also possible for the entropy to decrease over time meaning that the observer has gained information about the system. This corresponds 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_Ω which x, y and z could have evolved into. Thus, overall, the observer loose 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 system has reached statistical equilibrium where all states are equally probable since the observer is unable to make any distinctions between them. This tendency, of any given system, as viewed from an observer with limited knowledge of the initial conditions, to increase its entropy and evolve towards statistical equilibrium, is referred to as the second law of thermodynamics. In conclusion, it can be stated as follows:

Any given observer, whose knowledge about the initial conditions of any given system is limited, tend to lose information about the system at an exponential rate until there is none left.

It is important to emphasize that the apparent violation of determinism and reversibility, i.e., violation of the Liouville theorem, is not due to a fundamental character in the dynamical evolution of systems. The apparent irreversibility is only due to the ignorance of the observer.

Arrow of time

If the point of view is that it is the second law which dictate the directionality of time, then the following conclusion must follow: For an infinitely wise observer, who is able to determine the initial conditions and the laws of motion with infinite precision, the evolution of the system is completely reversible in time. For such an enlightened observer, there is no arrow of time. Time do not flow into the future from the past. The apparent unique direction in which time flow, i.e., toward the future, is merely a consequence of the fact that the observer does not possess infinite knowledge about the system under consideration. For such an ignorant observer, it is exponentially more probable that the system evolves in such a way that possible paths diverge on phase space. The diverging evolution define the direction, or arrow, of time as seen from the perspective of the observer. In the unlikely scenario that the possible paths converged at a quicker rate than they diverged, such that information on the average was gained, then the system would be observed to evolve backwards in time. The logical philosophical question to ask is then the following:

Can the seemingly universal feature of the arrow of time, flowing towards the future for all observers, really owe its existence to the inability of the observer to completely specify the state of the system with infinite precision?

To put it differently: presumably, a monkey is more ignorant, as compared to a human physicist, about the complete set of degrees of freedom characterizing e.g. the falling of a glass of wine. Yet, the human is more certain about what will happen to the glass of wine as it falls. The directionality of time does not seem less clear to the human despite having more precise information about the state of the glass of wine. This contradicts the logical consequence of the second law as stated above. There, the human is less limited than the monkey and therefore should loose information at a lesser rate and hence the monkey would have a greater sense of time's arrow. Thus, either the monkey does really have a much greater sense about the flow of time, or the directionality has nothing to do with the ignorance about the system possessed by the observer.

Conclusion

In the contemporary formulation of the thermodynamic arrow of time, as understood within the theory of statistical mechanics, the concept of entropy is employed to give a microscopic argument for the apparent directionality in time. It is commonplace in textbooks both at the upper-secondary and university levels to state time's arrow as a fact following directly from the tendency of systems to evolve towards statistical equilibrium where its entropy is at its maximum value. At the same time, a common interpretation for the concepts of probability and entropy in statistical mechanics is that they appear due to the lack of precise knowledge by the observer about the system under study. In this picture, probability and entropy are subjective concepts which depend both on the characteristics of the system and the observer.

In this article, it has been argued that the philosophy that time's arrow is a fundamental property of Nature is incomaptible with the subjective information-theoretical interpretation of entropy since the conclusion of the latter is that the appearance of time's arrow is due to the ignorance of the observer. This statement relies, ofcourse, on the premise that one believe that the increase in entropy, as it appear in statistical mechanics, is the physical origin for time's arrow. Thus, if one tend to believe in both the fundamental character of time's arrow and the subjective interpretation of probability and entropy in statistical mechanics, one is forced to abandon the idea that time's arrow has anything at all to do with the entropy of the system. In this case, it seems logical to consider the uncertainty in quantum mechanics as being the origin for time's arrow, given that one adopts the viewpoint that quantum probability is observer-independent.

[1] Henriksson, A., *Liouville's theorem and the foundation of classical mechanics*, Lithuanian Journal of Physics, **62**, no.2 (2022), 73-80.

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

[3] Birkhoff, G.D, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. USA **17** (1931), 656.

[4] Neumann, J.v., *Physical applications of the ergodic hypothesis*, Proc. Natl. Acad. Sci. USA **18** (1932), 263.

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

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

[7] Shannon, C.E., *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), 379.

[8] Jaynes, E.T., *Information theory and statistical mechanics*, Phys.Rev. **106** (1957), 620.

[9] Jaynes, E.T., *Information theory and statistical mechanics II*, Phys.Rev. **108** (1957), 171.

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