UNIVERSITY OF BELGRADE

FACULTY OF PHILOSOPHY

Jelena Z. Dimitrijević

PHILOSOPHICAL ASPECTS OF THE PROBLEM OF MAXWELL'S DEMON

Doctoral Dissertation

Belgrade, 2020

UNIVERZITET U BEOGRADU

FILOZOFSKI FAKULTET

Jelena Z. Dimitrijević

FILOZOFSKI ASPEKTI PROBLEMA MAKSVELOVOG DEMONA

doktorska disertacija

Beograd, 2020

Mentor:

Ph.D. Slobodan Perović

Professor, Philosophy Department, University of Belgrade

Committee members:

Ph.D. Miloš Adžić

Assistant Professor, Philosophy Department, University of Belgrade

Ph. D. Biljana Radovanović

Assistant professor, Philosophy Department, University of Niš

Defense date: ___________

Mentor:

red. prof. dr Slobodan Perović

Odeljenje za filozofiju, Filozofski fakultet, Univerzitet u Beogradu

Članovi komisije:

doc. dr Miloš Adžić

Odeljenje za filozofiju, Filozofski fakultet, Univerzitet u Beogradu

doc. dr Biljana Radovanović

Filozofski fakultet, Univerzitet u Nišu

Datum odbrane: ___________

PHILOSOPHICAL ASPECTS OF THE PROBLEM OF MAXWELL'S DEMON

Abstract

In this thesis, we will analyze Maxwell's thought experiment from few different points of view. First, we will analyze the relation of Maxwell's demon with the causality on the one side and indeterminism on the other. Then, we will examine it from the point of view of philosophy of time, thermodynamics, information theory, biology and quantum mechanics. The central part of the thesis will be the analysis between entropy and information in which we will conclude that information and entropy cannot be the same, and that any exorcism that counts on this relationship must be invalid. We will make several conclusions on different aspects of the problem of Maxwell's demon. We explore these different aspects of the problem of Maxwell's demon through chapters that can be read independently, but they also form a bigger picture by showing us the heuristic value that Maxwell's thought experiment bears for both philosophy and physics, but also biology, quantum computation, history of science, cosmology, etc. Instead of exorcising the demon, we can continue Maxwell's project of analyzing validity of second law through this thought experiment, by using it much wider, in many other fields to draw some new conclusions and pay attention to some unperceived aspects of old phenomena.

Key words: Maxwell's demon, the second law of thermodynamics, entropy, information.

Scientific field: philosophy

Scientific subfield: philosophy of science

UDC number: 167

FILOZOFSKI ASPEKTI PROBLEMA MAKSVELOVOG DEMONA

Rezime

U ovoj disertaciji analiziraću Maksvelov misaoni eksperiment sa nekoliko različitih tački gledišta. Najpre ću analizirati odnos Maksvelovog demona i pojma kauzalnosti na jednoj strani, kao i pojma indeterminizma, na drugoj. Nakon toga ću analizirati problem Maksvelovog demona sa tačke gledišta filozofije vremena, termodinamike, informatike, biologije i kvantne mehanike. Centralni deo teze biće analiza odnosa entropije i informacije u okviru koje ću zaključiti da informacija i entropija nisu i ne mogu biti isto, kao i da egzorcizam koji se oslanja na izjednačavanje ova dva pojma ne može biti validan. Doneću još neke zaključke koji se tiču raznih aspekata problema Maksvelovog demona. Istražiću ove različite aspekte kroz odeljke koji se mogu čitati nezavisno, ali koji takođe zajedno čine širu sliku, ukazujući nam heurističku vrednost koju ovaj Maksvelov misaoni eksperiment donosi kako na polju filozofije fizike, tako i na polju biologije, kvantne fizike, istorije nauke, kosmologije i tako dalje. Umesto da prognamo demona, mi možemo nastaviti Maksvelov projekat analiziranja validnosti drugog zakona termodinamike, koristeći ga mnogo šire, kao i u mnogim drugim poljima kako bismo došli do novih zaključaka i obratili pažnju na neke nove, ranije neopažene, aspekte starog fenomena.

Ključne reči: Maksvelov demon, drugi zakon termodinamike, entropija, informacija.

Naučna oblast: filozofija

Uža naučna oblast: filozofija nauke

UDK broj: 167

Contents

List of figures

1. Introduction

"The whole universe is in the
glass of wine, if we look at it closely enough."

Richard P. Feynman

In this thesis, we will analyze Maxwell's thought experiment from a few different points of view. We will analyze relationship between Maxwell's demon and causality on the one side and indeterminism on the other. We will also analyze it from the point of view of philosophy of time, thermodynamics, information theory, biology and quantum mechanics. Hence, we will make a several conclusions on different aspects of problem of Maxwell's demon. We explore different aspects of problem of Maxwell's demon through chapters that can be read independently, but they also form a bigger picture by showing us heuristic value that Maxwell's thought experiment bears for both philosophy and physics, but also biology, quantum computation, history of science, cosmology, etc. If we combine this with analysis from chapter 6, which shows us that Maxwell's demon does not violate second law of thermodynamics and there is no use in exorcising it, we can see how instead of exorcising it, we can continue Maxwell's project of analyzing validity of second law through this thought experiment, by using it much wider, in many other fields in order to draw some new conclusions and pay attention on some unperceived aspects of old phenomena.

Hereby, I will explain the structure of the thesis.

In Chapter 2 we will introduce some of the concepts that are crucial for further argumentation. In order to properly understand the influence of Maxwell's thought experiment with the demon we must explore the role that thought experiments can play in science. We will begin with explaining the concept of thought experiments and their characteristics and their relationship with scientific models. In order to explain thought experiments, we will characterize them as conceptual models. Hence, we will explain the concept of models. In the end of the chapter I will explicate my view of the function of thought experiments. Since we make solid base for understanding of the concept and importance of thought experiments we will turn to Maxwell's demon thought experiment and analyze its role in a particular scientific context.

In Chapter 3 we will explain in more details the experiment with Maxwell's demon. In order to explain why it is a problem and a challenge for both scientists and philosophers we will go through some examples from history of science and discuss some attempts to construct an engine which would behave the same way as Maxwell's demon. Thereby we will explain: Smoluchowski trapdoor, Feynman's ratchet and pawl, Gabor's engine, Feynman's trapdoor, Szilard's engine. Also, here we will explain Landauer's principle and discuss some of its critiques and some of its defenders. Landauer's principle has been used as one of the most common tolls for exorcising the Maxwell's demon and here we will show why it is inadequate. After it I will explain the concept of entropy, and separate a few different kinds of entropy and the second law because at least a part of the problem arises due to the confusion[1] between different kinds of entropy and the second law.

In Chapter 4 we will examine the role Maxwell's demon plays in our understanding of causation and indeterminism. In order to do this, we have to compare Laplace's demon and Maxwell's demon. The first step in this analysis will be separation of different kinds of determinism. Then we will discuss the symmetry of the universal laws. We will explain that deterministic laws are time-symmetric, (which means that there is no way to distinguish past from the future). We will briefly explore the role that determinism plays in theoretical physics, especially in special theory of relativity. Also, we will briefly address relation between determinism and free will. Finally, we will explain the notion of causation and explore the relationship between causality and entropy. In the very end, we will explain the notion of entropic arrow, and its relation to causality.

In Chapter 5 we will analyze the relationship between entropy and different arrows of time. First, we will explain why scientists used to reduce the so-called time arrow to entropic arrow; in other words, we will explain the importance of such a project and its historical background. The Key aspect for understanding this is asymmetric nature of time, hence that will be the next aspect of time we will analyze. We will also separate different arrows of time, both local and cosmic. Then, we will go through some attempts to explain future-past asymmetry that are not directly related to entropy. At the end, we will draw the conclusion on relation of entropy and arrow of time.

In Chapter 6 we will analyze the relationship between entropy and information in the framework of Maxwell's thought experiment. This will be the central part of the thesis because reduction of entropy to the lack or loss of information was part of most exorcist strategies, which we criticize. First, we will introduce different notions of information (Shannon, mutual, quantum, active, passive, inactive). Then, we will analyze the relationship between entropy and information. In order to do that, we have to go back to the Szilard's engine and Landauer's principle because that represents source of inspiration of various attempts to exorcise the demon by means of reduction of entropy to the loss of information. We will also analyze the case of intelligent demon. In the end of the chapter we will offer a solution of the Szilard's puzzle and finish our analysis with the conclusion that information and entropy cannot be the same, and that any exorcism that counts on this relationship must be invalid.

In Chapter 7 we will analyze the relationship between Maxwell's demon and entropy on the one side, and complexity of living systems and evolution, on the other. First, we will analyze notion of living systems and evolution. Afterwards, we will compare it to the notion of Maxwell's demon. Through the analysis which will include evolutionary dynamics and mechanisms of imitation and learning we will conclude that that Darwinian selection theory has similar structure as Maxwell's thought experiment with demon.

In Chapter 8 we will briefly discuss some solutions to the problem of Maxwell's demon offered in the field of quantum mechanics.

Most important conclusions we will draw are the following:

1. Maxwell's demon demonstrates that we cannot identify entropy with the arrow of time because of statistical nature of entropy.

2. Although information and entropy share the same mathematical form, they refer to the different physical concepts. Hence, they are not equivalent.

3. There is no need for exorcising the demon. Since he is subject to the weakened laws of thermodynamics, he could do nothing to violate them.

[1] Confusion between information entropy and thermodynamic entropy

4. At least part of the problem of Maxwell's demon arises because of the confusion between different kinds of entropy.

5. Darwinian selection theory has the same structure as Maxwell's thought experiment with demon.

2. Sympathy for the devil

"Please to meet you, hope you guess my name
But what's puzzling you is a nature of my game."
The Rolling Stones

In order to understand importance and the influence of Maxwell's thought experiment with the demon on various fields in science and philosophy, we will first outline the important role, both historical and epistemological, which thought experiments have had in philosophy and science in general. The outline of the exposition is as follows.

In Section 1 we will explain the concept and some characteristics of thought experiments and discuss some relevant accounts of thought experiments. In Section 2 we will explain relationship between models and thought experiments. In Section 3 we will explain functions of thought experiments and at the very end, we will consider what thought experiments cannot do.

Traditionally, demons were always related to the superstition and dark in both Western and Oriental traditions, while science has been, at least since the time of the Enlightenment, related to light and truth. Hence, the question is could science "cooperate" with demons? Could demons be used as a figure of thought that belongs to the realm of thought experiments? To what extent does it show that scientific arguments are "stories" or "fiction" (e.g., as argued by Alisa Bokulich[2])? Why demons should play this role?

First, when we talk about demons we are supposedly talking about beings with supernatural abilities that are both different from and greater than the human ones. Thus, they can give us valuable clues, ambient to "what if" scenarios, which are far from empirical facts but can redirect our focus to the less anticipated side of the phenomena we wish to analyze and research. They can obviously help us conceptualize the problems which are either not available in the natural (or even social) world by being prohibited or very improbable, or are available but are impossible to observe or verify; an example in the latter sense would be the fate of matter falling into a black hole, which happens all the time in the natural world and thus is eminently *possible*, but cannot be observed due to causal restrictions imposed by general relativity.

Still, the question is: where are their limits? What they could not teach us? What they could not show us about the empirical world? What is the role that demons, in sufficiently generalized sense, have in philosophical and scientific analysis of the external world?[3]

Demons have been used as both figures of speech and characters in philosophical and scientific analysis. They took part in the thinking process of many philosophers and scientists: Descartes, Bošković, Laplace, Maxwell, Loschmidt, Landsberg, Mendel, Nietzsche, Searle, and Freud are just

[2] Bokulich, Alisa. 2016, "Fiction As a Vehicle for Truth: Moving Beyond the Ontic Conception," *The Monist* **99**, 260-279.
[3] Weinert, F. *The Demons of Science*. Springer. (2016).

some of the examples. In somewhat looser terms, some aspects of the demon-related thought experiments could be associated with Einstein (for instance, his first thought experiment about chasing a ray of light could be interpreted in that manner, since obviously, no mundane agency could satisfy the precondition[4]) and Darwin (in his thought experiment about the wolves and the deer in a forest, he assumes that it is possible to track their population numbers over many generations, clearly unfeasible for humans of his times) as well. Throughout history of science important philosophical claims have been made in the name of demons. What, however, is the demons' general role? It is hard to escape the conclusion that their role is to question the possibilities and limits of human knowledge. Maybe the most important: they can provoke and challenge existing knowledge. In addition, they remind us that scientific knowledge has philosophical implications.

Laplace claimed that we can predict every event that will take place in universe if we have knowledge on initial conditions and mechanical principles. Laplace used demon as a tool with which he intended to show that our universe is a deterministic, clockwork Newtonian universe. He considered that world is contained of chain of events. The question is, what kind of the determinism was Laplace's determinism. Was it causal, metaphysical or scientific? The fundamental laws of physics are (with one possible microscopic exception, extremely small and only indirectly inferred) time-reversal invariant; they do not distinguish past from the future. This means that Laplace's demon cannot recognize time's arrow. In other words, to him all the events (those that have happened and those that are still to happen) are already present – there is no difference between prediction and retrodiction. The traditional problem is, of course, that if all events have (physical, i.e. mechanical) causes, question is whether there is a place for free will in such a view of the world. Laplacian determinism, however, is in principle compatible with the arrow of time, as perceived by highly imperfect observers and predictors *within the universe* (in contrast to the demon), such as ourselves.

In contrast to Newtonian mechanics, the science of thermodynamics as it emerged in the 19th century has been temporally asymmetric from the very beginning.[5] This is the consequence of generalization of our everyday experience of irreversibility into the second law of thermodynamics, as done by Helmholtz, Clausius, and Kelvin. As we shall see below, some of the concerns raised about the foundational role of the second law in contemporary philosophy of physics are still highly relevant for the interpretation of Maxwell's demon. If we put it in most general terms, Maxwell's demon challenged the claim that our universe is constantly moving from the state of order into the sequence of states of increased disorder, toward the heat death as (informally) the state of maximum disorder. The lesson that demon teaches us is that increase in entropy is not deterministic. It is probabilistic. It is not the case that entropy is increasing in every case and every time. From time to time there can appear decrease of entropy. The point is that this decrease is highly improbable, and it is unsustainable on long term timescale. The question which remains is the following: what does this tell us about the arrow of time?

And the large-scale question about the arrow of time is still not the logical endpoint. There is Landsberg's demon that forces us to focus not only on the universe, but on the hypothetical larger whole called the multiverse.[6] Landsberg's demon is located in "nowhere" and observes the multiverse and all the events that take place inside of it. Demons thus play an additional role on the border

[4] Pais, Abraham. *Subtle is the Lord: The Science and the Life of Albert Einstein: The Science and the Life of Albert Einstein*. Oxford University Press, USA, (1982), p. 131.

[5] Boltzmann, Ludwig, *Lectures on gas theory*, translated by S. G. Brush (University of California Press, Berkeley, 1964); Zeh, D. "The physical basis of the arrow of time." Springer-Verlag, New York (1992); Price, H. *Time's arrow & Archimedes' point: new directions for the physics of time*. Oxford University Press, USA, (1997).

[6] Landsberg, Peter Theodore, and Chaisson, E. J. "The enigma of time." *American Journal of Physics* 53.6 (1985): 601-602.

between science and philosophy: they should be able to show us what science can *in principle* discover about the world.

Allegedly, Maxwell's demon acts opposite to Laplace's demon. In the simplest setup of Maxwell's thought experiment, demon controls on which side is in the container with gas, separated on two parts, pass slow and on which side pass fast molecules. Thus, he creates temperature difference without expending work, which at first, might seem as a violation of the Second Law of thermodynamics. However, in most interpretations of the seemingly paradoxical conclusions, it turns out that it is not in fact the case (we shall consider philosophical criticisms of this claim in subsequent chapters). It "only" shows us that nature of the Second Law is statistic.

Our goal here is to analyze problems that science and philosophy share, and the relationship between these two demons is a very convenient testbed for a critical analysis of this topic. In order to do this, however, we must first analyze the relationship between thought experiments in science and concepts in philosophy.

In other words, we need to see how, although they represent supernatural beings, these demons involve notions that make impact on both scientific and philosophical problems, and have at particular epochs played key role in formation of our paradigms and worldviews. Laplace's and Loschmidt's demon[7] force us to rethink the problems of the Second Law of thermodynamics, determinism, indeterminism, causality, free will, arrow of time, and evolution (understood in a sufficiently generalized sense).

How can we *evaluate* a thought experiment in science and philosophy? The experiments that have happened only in the workshop of the mind have been making impact on progress in critical thinking since antiquity.[8] We can characterize thought experiments as a kind of conceptual models. Demons, as a part of thought experiments, are henceforth also conceptual models. Like all models in science, they use necessary abstractions and idealizations in order to test or represent the values, hypotheses, or scientific theories. The mere difference is that in thought experiments we do not test empirical parameters or values, but tend to analyze these values, hypothesis or scientific theories through the counterfactual scenario (usually in a qualitative manner, although in recent times there have been important examples of quantitative and even numerical thought experiments as well). Both scientific models and thought experiments have the role to explain target phenomena and point to some problems that we have not noticed earlier. In thought experiment we can ask: What if a demon could measure phase variables of individual molecules? What if it could manipulate individual molecules?

Thus, Maxwell's demon gives rise to discussion on statistical notions which leads to reconsideration of the notions of indeterminism. It is concerned with the second law of thermodynamics.[9] Hence, in this case, we are invited to consider a possible world in which entropy could decrease, even if only under the restricted set of circumstances. This instance is particularly apt for a generalization. Demons can help us to find out possible worlds. The demons of science test the laws to their very limit.[10]

We have many examples that demons have often been used as a methodological device. The best example of this is Descartes' demon.[11] Descartes used the demon as methodological device which aim was to defeat skepticism, while in the same time maintaining rational objectivity of philosophical reasoning.

---

[7] Loschmidt's demon is the one capable of changing signs of all particles' velocities (as in the eponymous Loschmidt's paradox). He would be able, for example, to bring molecules of perfume back into the bottle after spreading in the air.
[8] Brown, J. R. 1993, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences* (Routledge, London), (1993).
[9] Weinert, Friedel. *The Demons of Science*. Springer. (2016). introduction.
[10] Ibid., p. 56.
[11] Evil demon that deceives us about our perception of the world.

As we will see, Maxwell's thought experiment points out that nature of the Second Law of thermodynamics is statistic, not absolute. Increase in entropy is therefore probabilistic, not deterministic. Now, if we suppose that the world is indeterministic, what consequences will it have for the arrow of time? Does that mean in that case we cannot derive time's arrow from the entropy increase? Besides, question is: could we derive arrow of time from entropy at all? (latter question will be discussed in Section 5.

In order to grasp this problem, we need to clarify a lot of notions that held key for understanding of its foundations: determinism, indeterminism, entropy, arrow(s) of time, causality, and evolution.

2.1. Thought experiments

In this Section, we will explain the concept of thought experiment. In subsection 2.1.1. we shall describe some of its characteristics. In subsection 2.1.2. we shall explain what thought experiments represent. Further on, we will explain different accounts of thought experiments: Platonic view (in 2.1.3.), the argument view (in 2.1.4.), as well as the model-based view (in 2.1.5.).

What thought experiments offer to us is an understanding of the conceptual apparatus of a particular theoretical framework. It tends to remove the confusion and lack in logical part of theory. It also can make us notice some facts that were there all the time, but have gone unnoticed for various reasons. It can cause the reconceptualization which can lead to reorganization of knowledge. It can release tension between different outlooks on the world.[12]

As Weinert points out, thought experiment is, first and foremost, a conceptual model.[13] In a thought experiment, we tend to analyze situation that never took place or even cannot be realized in reality. Further, we explore some of the possible consequences of that situation. We can describe the aim of thought experiments as analysis of the hypothesis, argument, or scientific theory. They can also help us to create new hypotheses, draw up new conclusions, or come up with new questions or even whole research programs. Thought experiments can lead to modification of scientific theories or even their abandonment. Kant calls them experiments of pure reason.[14]

2.1.1. Some characteristics of thought experiments

1. **Thought experiments could help us draw the correct conclusion.** We can discover logical inconsistencies in our argument or hypothesis while performing a thought experiment. Being unsatisfied by the usual uncertainties about the initial conditions of the universe, Stephen Hawking made an assumption about no-boundary universe. In his assumption, he argued that space-time could be both finite and unlimited, if time is expressed in imaginary units described in imaginary time.

[12] Hacking, Ian, „Do Thought Experiments Have a Life of Their Own?“ Comments on James Brown, Nancy Nersessian and David Gooding, *The Philosophy of Science Association*, Vol. 1992, No., Volume two: Symphosia and Invited Papers, (1992): 302-308, pp. 304-305.
[13] Weinert, Friedel. *The Demons of Science*. Springer. (2016). introduction
[14] Kant, Immanuel. *Critique of pure reason*. Cambridge University Press, (1998).

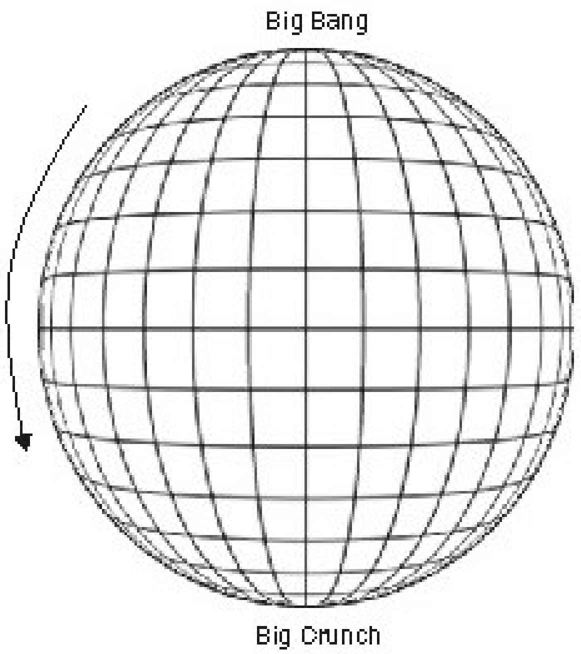

Figure 2.1. Hawking's no boundary proposal. **Adapted from *The Demons of Science* by Weinert, Friedel. (2016).**

2. Thought experiments may lead us to wrong conclusions. The so-called twin paradox of special relativity is an excellent example of this, since it offers a seemingly salient argument for the incoherence of the theory. Twin A cannot in the same time be younger and older from the twin B. The theory is sound, however, and the problem with the thought experiment is the "coherence gap" – omitting the states-of-affair corresponding to non-uniform motion of the spaceship. In order to be able to compare the ages of twins A and B, it is necessary to have prolonged acceleration at the turn-around, leading to conditions for application of special relativity not being satisfied. Non-obvious nature of such conclusion illustrates well how the "coherence gap" can be a serious difficulty in applying this method.

3. Thought experiments change our perspective. Thought experiments can move our focus from one part of the phenomena to other, which will lead us to change interpretation and understanding of that phenomena. Therefore, we can come to different and hitherto unsuspected conclusion. Since thought experiments cannot provide new empirical facts, they cannot provide empirical proof. Nevertheless, they held important place in the history of ideas, as testified by the historical record.

4. Thought experiments are fixed. They cannot change or mutate. When they are written, they are like an icon, or a character in the drama.[15]

[15] Hacking, Ian, „Do Thought Experiments Have a Life of Their Own?" Comments on James Brown, Nancy Nersessian and David Gooding, *The Philosophy of Science Association*, Vol. 1992, No., Volume two: Symphosia and Invited Papers, (1992): 302-308, p. 307.

2.1.2. What does a thought experiment represent?

What is the epistemic function of thought experiment? Perhaps the most interesting aspect of thought experiments is that although their lack of new empirical results, they still teach us something new about a real world.[16] Now, we will analyze the role they play in reasoning.

Some thinkers consider thought experiment as some kind of limiting case of real experiment. There is some kind of continuity between the real and thought experiment. Foremost early adherent to this view was Ernst Mach, who defended it against implicit criticism of contemporaries such as Poincare and Boltzmann. Machian view has been recently represented by James McAllister. The main point here is that within the context of the thought experiment one gets benefits without the loss. Thought experiments certainly have some similarities with real experiments. First of all, in thought experiment we accept some claims and hypotheses about external world. These claims that are used in thought experiment have some kind of empirical grounding.[17] However, it is important to note that empirical evidence on which it is grounded is actually the outcome of historical accomplishments.

Galileo (implicitly) argued that in cases where reduction of the effect of *particular* boundary conditions on exploration of the world is not possible. Therefore, important insight can come from thought experiments. His most famous thought experiment, proving the independence of gravitational acceleration on the mass of the falling bodies, played the crucial role in emancipation from the old, Aristotelian physics. So, in a very substantial sense, modern physics, and by extension science in general, began with a thought experiment.

In addition, thought experiments analyze the phenomena in "accident-free form." [18] Thought experiments explore occurrences of the phenomena that never happened, which actual appearances cannot do. Since Galileo, experimentalists held that thought experiments are continuation of real experiments or their predecessors. The goal of scientific experiment is to explore the phenomena through observation.

Thought experiments can also provide us with the level of abstraction and idealization that scientific experiments cannot reach. They share this characteristic with scientific models. Anyway, thought experiments may fail to capture ("detect" on empiricist-like views) the real phenomena.

The problem with Machian quasi-experimentalist view is that it ignores that results of thought experiments contain big amount of uncertainty. Some philosophers critical of views of Mach or McAllister point out that since thought experiments offer us conclusions that are not empirically proved, they will remain indeterminate.[19] From the other side, Mach claimed that we did not need to materialize thought experiments in order to get some results. Although thought experiments cannot provide validation of conclusions or proof they can provide idealizations.[20]

In a more conventional vein, Max Planck argued that thought experiment can have value even if it is not based on any measurements.[21] He claimed that thought experiment use abstractions, which are valuable for science as well as results of scientific experiments. In Planck's criticism of the quasi-experimentalist view, he stated that it would be mistaken to claim that a thought experiment has

[16] Brown, James R., *The Laboratory of the Mind: Thought Experiments in the Natural Sciences* (Routledge, London), (1993).
[17] McAllister, Janet, "The evidential significance of thought experiments in science". *SHPS,* 1996, 27, pp. 233-250, p. 233.
[18] McAllister, Janet, "Thought experiments and the belief in phenomena". *Phil. Sci., 71*, (2004): pp. 1164-1175, p. 1168.
[19] Norton, John A. On thought experiments: Is there more to the argument? *Phil. Sci., 71,* (2004): pp. 1139-1151.
[20]Mach, Ernst. "The science of mechanics: A critical & historical account of its development, (after the 9th German edition)." *Open Court.[JVB]* (1883). chapter 1.
[21] Weinert, Friedel. *The Demons of Science.* Springer. (2016).

importance only if it can be validated by measuring processes. This requirement will make entire classes of arguments, e.g., geometrical proofs, impossible. In thought experiments, to state it metaphorically, the spirit of researchers is lifted above the measurement tools and that is what help them to formulate new questions and new hypothesis. Planck claimed that thought experiment is a necessary abstraction. It has a same value for research as much as assumption for the real world.

2.1.3. The Platonic view

The Platonic view on thought experiments, as exposed in recent times most strongly by the Canadian philosopher James R. Brown, stands in strong contrast to the experimentalist view. Here, thought experiments are considered able to provide knowledge on empirical facts. We can describe the knowledge they provide in terms of Kantian philosophy, as *synthetic* a priori knowledge.[22]

We can use thought experiments as a testbed for the consequences of the theory. We can classify thought experiments as either destructive or constructive. Destructive thought experiments play the role of highlighting the problems in theory, while the constructive ones aim to lead us to some positive conclusion. As a prototype of this kind of thought experiment, we can take Einstein's thought experiment with the falling elevator, which played crucial role along the road toward the general relativity. This experiment indicated the curvature of spacetime by showing that light rays necessary move along curved trajectories in the gravitational field. Effectively, Eddington's observations of bending the light of background stars in the Solar gravitational field (exactly 100 years ago) tested the positive claims of this particular thought experiment.

On the other hand, some thought experiments are both destructive and constructive at the same time. For example, consider Galileo's thought experiment of the free-falling bodies.[23] Galileo destroyed Aristotelian claim that heavier bodies fall faster than lighter ones. But, at the same time it established the new explanandum – the invariant rate of falling objects. Galileo concluded that all objects fall at the same speed of 9.81 m/s^2 no matter how massive they are (and independently of their chemical composition and all other secondary properties). This conclusion was, on one hand, empirically confirmed by all kinds of mechanical experiments – e.g., measuring the period of a simple pendulum – but on the other hand *remained unexplained* in the theoretical framework of the classical Newtonian mechanics. It was only with the advent of general relativity (and, more generally, metric theories of gravity) that this insight following from Galileo's thought experiment has received an adequate physical explanation. The adequate explanation is contained in what Einstein dubbed the principle of equivalence. According to this principle, there is no physical difference between acceleration and local gravitational field.

The perception of abstract laws of nature is of specific kind. As Brown puts it:

Scientific experiment leads from senses to propositions.

vs.

Thought experiment leads from intellectual perceptions to propositions.[24]

[22] Weinert, Friedel. *The Demons of Science*. Springer. (2016). introduction
[23] Ibid., pp. 18-20.
[24] Brown, James Robert. "Peeking into Plato's heaven." *Phil. Sci.* 71.5 (2004): 1126-1138., chapter 3.

This parallelism is not in itself without problems. “Intellectual perception” is somewhat mysterious concept.[25] When Brown uses the locution “intellectual perception”, he means this in a very limited way. Nevertheless, as Planck observed idealization is necessary for scientific reasoning.[26]

Albert Einstein rejects the inductive view that scientific principles came from experience. Einstein argued (as did many others before him, including Darwin and Boltzmann) that if we only collect empirical facts, we could never develop a theory.[27] Compare Darwin’s famous quote (from the letter to Henry Fawcett in 1861): “How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service!”[28] Arguably, this is a fortiori true when theory includes a high level of mathematical complexity. There can be no human intuition leading to such mathematical complexity. On the other hand, rational thinking can help in formulation of mathematical equations like $E = mc^2$. However, in order to confirm that a proposed law is more than mere conjecture, we need to provide specific instances of empirical evidence that confirms that law, formulated in a convenient form. We need a special kind of Brown’s “intellectual perception” to understand the general scientific laws.[29]

A thought experiment includes idealizations and abstractions and can eliminate all the unnecessary empirical facts and make conclusions clearer and easier to make. Science, unlike introspection, must respect empirical constraints. Another possible function of the thought experiment is the exploration of the invariant relations. Invariant relations are the lens trough which scientist observes the world of scientific work in order to avoid peculiarities and quirks belonging to particular observers which have no general validity.[30]

We have already mentioned that thought experiments and models share abstractions and idealizations. Now we will analyze a more moderate position, according to which thought experiments are a special kind of arguments.

2.1.4. The argument view

For the thinkers who held this position, thought experiments represent special kind of arguments. They cannot provide us with a priori knowledge, in Kantian sense.[31] They rather infer consequences which can be tested in principle. For one to explore thought experiments he needs to imagine counterfactual situations.

---

[25] Norton, J. A. On thought experiments: Is there more to the argument? *Phil. Sci.*, 71, (2004): pp. 1139-1151.
[26]Weinert, Friedel. *The Demons of Science*. Springer. (2016), p. 21.
[27] Einstein, Albert. "Autobiographical notes (1949)." *Albert Einstein: Philosopher-Scientist* (1963), p. 89; Weinert, Friedel. "Einstein and the Representation of Reality."*Facta Philosophica,* 8 (1-2), (2006): 229-252.
[28] Weinert, Friedel. *The Demons of Science*. Springer. (2016), p. 23.
[29]Norton, John D. "Are thought experiments just what you thought?" *Canadian Journal of Philosophy* 26.3 (1996): 333-366. , Norton, John A. On thought experiments: Is there more to the argument? *Phil. Sci.*, 71, (2004): pp. 1139-1151.
[30] Mach, Ernst. "The science of mechanics: A critical & historical account of its development, (after the 9th German edition)." *Open Court.[JVB]* (1883).
[31] Norton, John. "Thought experiments in Einstein’s work." *Horowitz and Massey* 1991 (1991): 129-148; "Are thought experiments just what you thought?." *Canadian Journal of Philosophy* 26.3 (1996): 333-366., Norton, John A. On thought experiments: Is there more to the argument? *Phil. Sci., 71,* (2004): pp. 1139-1151., Hempel, Carl G. "Fundamentals of concept formation in empirical science, Vol. II. No. 7." (1952).*In Aspects of Scientific Explanation* (pp. 155-171). New York: Free Press/London: Collier Macmillan (1965).

Notably, John Norton claims that thought experiments are like ornaments that can be constructed into arguments.[32] While constructing arguments from thought experiments we are making hitherto implicit assumptions explicit. There are two conditions that such an argument needs to satisfy:

1. To display counterfactual situations.

2. To eliminate details that are not relevant for the goal of research. Due to the elimination thesis, these details can be ignored.[33]

Similarity that thought experiments share with the conclusions of arguments is that all the empirical knowledge in the conclusions has already been contained in its premises. They have only been made manifest by intervention of the "thought experimenter". The conclusions are either inductive or deductive inferences and they always have a certain degree of probability. Hence, thought experiments are inferential devices. There are two ways on which thought experiment can fail: first is if they are based on assumptions that are false, second because of fallacious inferences.[34] However, we can consider thought experiments as reliable because they are developed by deductive or probabilistic inferences.[35]

Norton attempts to explain Einstein's thought experiment with the elevator in free-fall as an argument:

1. An observer sees free bodies fall at equal rate.[36] (We set aside here the historical fact that this has been first established by Galileo's thought experiment.)

2. Inductive step: this will hold for all the phenomena (including propagation of light rays).

3. Uniformly accelerating reference frame and frame at rest in the homogenous field are observably the same. Nevertheless, they are not theoretically identical. Hence, this contradicts to the rule of construction.

4. We should not make theoretical difference between states that do not have differences on observational level.

5. Accelerating frame in empty is same thing as frame rest in the homogenous gravitational field.

If we need counterfactual and hypothetical reasoning in science, then we need thought experiments. Not any critical and hypothetical reasoning can be part of thought experiment. The problem is that thought experiment cannot be reduced on the patterns of logical argument.[37] This is the case with more imaginative thought experiments, of which thought experiments with demons are examples. Demon represents a tool for testing implications of the knowledge that we possess at present; these implications might change as science advances, as the example of Galileo's free falling bodies shows. This view might give impression that thought experiments can offer more conclusions than they usually do. Conclusions of thought experiments are not indeterminate or inexact. Nevertheless, two or more scientists can come to the different conclusion starting from the same thought experiment; as we shall see in further text, this exactly has been the case with Maxwell's demon.

---

[32] Norton, John. "Thought experiments in Einstein's work." *Horowitz and Massey* 1991 (1991): 129-148. Norton, John D. "Are thought experiments just what you thought?." *Canadian Journal of Philosophy* 26.3 (1996): 333-366. , Norton, John A. On thought experiments: Is there more to the argument? *Phil. Sci.*, 71, (2004): pp. 1139-1151.
[33] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[34]Norton, John D. "Are thought experiments just what you thought?." *Canadian Journal of Philosophy* 26.3 (1996): 333-366, p. 335.
[35] Norton, John A. On thought experiments: Is there more to the argument? *Phil. Sci., 71,* (2004): pp. 1139-1151, p. 1140.
[36] Norton, John. "Thought experiments in Einstein's work." *Horowitz and Massey* 1991 (1991): 129-148, p.137.
[37] Cooper, Rachel. "Thought experiments." *Metaphilosophy* 36.3 (2005): 328-347, p. 332.

We can use thought experiments for different purposes, and they can be modified according to one's purposes in each specific context. The thought experiments have one more function, and that is to challenge, change, or justify our understanding of concepts.[38]

Thought experiments are, of course, very different than real experiments from a practical standpoint. First of all, they are based on counterfactual reasoning. Second, because they contain large amount of employ abstraction and idealization. It is important to note that fruitfulness of a thought experiments does not come only from reasoning, either of deductive or inductive kind, but also from imagination. Strength can be gained either by reason or intuition. However, there are kinds of counterfactual reasoning that do not belong to the realm of thought experiments, notably those usually used in historical disciplines or in the context of legal studies. There are kinds of hypothetical and counterfactual reasoning that do not fall into category of thought experiment.[39]

2.1.5. The model-based account

This account held that the aim of thought experiments is to represent models of possible worlds.[40]

We already mentioned that thought experiments tend to construe counterfactual state of affairs with help of what if questions. In answering such "what if" questions we try to construct a coherent and consistent model. This model or template can illustrate both physical and logical possibilities.

Consider the issue of the thought experiment failure. This can happen in two ways:

1. Answers that are given as reply to the "what if "questions are not correct.

2. Model that is construed is not coherent or implications that have been derived are inexact.[41]

The strength of conclusions of the thought experiments depend on the exactness of the data that we use in it. They are powerful tools for they made possible inquiry on both possible and impossible worlds.

The question remaining is what exactly is the nature of the counterfactual reasoning and its use in the thought experiments? If thought experiment as such, should be considered as kind of conceptual model what is its relationship with the models that we are using in science?

2.2. Models and thought experiments

Here we will try to briefly address some of the questions, which will be, as will be seen later, relevant for the philosophical analysis of Maxwell's demon thought experiment. A few most pertinent are:

1.What is the relationship between models and theories?[42]

2. What kinds of models exist?

3. In which manner they represent external world?

---

[38] Hacking, Ian, „Do Thought Experiments Have a Life of Their Own?" Comments on James Brown, Nancy Nersessian and David Gooding, *The Philosophy of Science Association*, Vol. 1992, No., Volume two: Symphosia and Invited Papers (1992): 302-308.

[39] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

[40] Cooper, Rachel. "Thought experiments." *Metaphilosophy* 36.3 (2005): 328-347.

[41] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

[42]Ibid., p. 33.

2.2.1. Scientific models

Models are primary entities of modern science and scientific thinking. Carnap, Hempel, Braithwaite, and Nagel identified scientific model with mathematic model in mathematical theory of model. They thought that model is nothing but an interpretation of a calculation derived from theory. Soon it become clear that this conception of model is too narrow. This definition does not shed much light on the role which model play in process of scientific inquiry.[43]

Hesse and Einstein stresses role of analogy in forming of model. In contrast, Bunge stresses the role of background theory. According to him, any model has two components:

- A general theory;
- A special description of the relevant object or system.

We can characterize model as abstraction or representation of some particular real system, idea or object. Model in management or engineering or science has many different forms. Some models are prescriptive, they determine optimal practical activity, which is obviously of key interest in applied sciences, economics, and engineering. Linear programming models are prescriptive, because the optimal solution for the linear programming task suggests the best direction of operation. Other models are descriptive, which means that they describe relations which provide the information necessary for the evaluation.

2.2.2. Models as mediators

This view tends to look on models as mediators between theories and phenomena.[44] Theories such as Boltzmann-Gibbs' statistical mechanics or Einstein's general relativity are abstract and general. Their principles apply (allegedly) to some part of the world of phenomena. What distinguish models is that they are more concrete from the point of view of any specific situation under investigation. Aims of any model in science are manifold. They could be used both as tool for development and testing. Their function can be representational as well. Besides that, models provide understanding in both quantitative and qualitative manner. Complex models usually lead to quantitative prediction.[45]

Theories provide formal and mathematical framework in order to explain the phenomena. Models help us to understand the working of some particular system, including experimental setups we use to discriminate between theories. If a theory happens to be false, it cannot provide understanding, because it does not give us explanation. Model, however, might still give us coherent account about the system under the study, even if it fails to explain the phenomena: predictions derived in celestial mechanics before 1915 from the models then available are still considered coherent and valuable, in spite of the fact that their underlying theory (Newtonian gravity) has been shown to be, strictly speaking, false. This applies, among other items, to LeVerrier's model of the motion of Uranus, which enabled him to discover a new planet, Neptune, creating perturbations in Uranus's motion;

---

[43]Hartmann, Stephan, „The World as a Process: Simulations in the Natural and Social Science“, *Theory and Decision Library, Dodrecht*, (1996): 77-100, pp. 79-80.
[44] Cartwright, Nancy. "Models and the limits of theory: Quantum Hamiltonians and the BCS models of superconductivity." *IDEAS IN CONTEXT* 52 (1999): 241-281; Morgan, Mary S., and Margaret Morrison, eds. *Models as mediators: Perspectives on natural and social science*. Vol. 52. Cambridge University Press, (1999): 241-281.
[45]Hartmann, Stephan. "Models and stones in hadron physics." *Models as mediators: Perspectives on natural and social science* 52 (1999): 326. Extremely important example in this respect are, obviously, climate change and other ecological models.

such great discovery of a hitherto unknown planet of our Solar System is in no way, historical or epistemological, demeaned by the fact that his model was based on, strictly speaking, false theory of gravity.

The question is what does the concept of understanding really mean in connection with the notion of "intellectual perception" mentioned above? Hereby model gives us coherent account of the external world or data and makes it possible for us to understand it.

Even though it is not necessary for the models to provide exact account of the external world, they must explain empirical data that we have gained so far through observation and experiment. In order to reach their goal model, apply numerous techniques: abstraction, idealization, factualization, and systematization. This is the key similarity they have with thought experiments.

Abstraction is process during which certain parameters that are the part of the modelled system are removed.[46] *Idealization* is a process of simplification of properties in order to easier manipulate with parameters. *Factualization* is a process of approximation of a model to a real system through adding previously disregarded components and relaxing the idealized assumptions. *Systematization* is process during which model recombines some of the factors into coherent system.

From the formal point of view, models can represent a topological or algebraical structure. There are also models that can combine both structures, for example, structural models in ecology or traffic planning.[47]

2.2.3. Types of Models

From a general standpoint, there are various kind of models: phenomenological, computer, explanatory, testing, heuristic, didactic, fantastic, imaginative, substitutive, formal, analog, instrumental, etc.[48] We will briefly analyze some of them here, since particular issues will be recognizable in the debates on the meaning of entropy and the best ways of interpreting Maxwell's demon thought experiment.

First of all, let us notice that along the same lines one can distinguish between models of representative and interpretative kind. as well as between representative and interpretative models.[49] Also, Hartman considers model as set of assumptions about particular system. It can be static or dynamic. It is static if it only makes assumptions about the system. It is dynamic if it includes assumptions about the evolution of system through time.[50]

Models could be deterministic or probabilistic, as well. In deterministic models, we know all data if we know the initial conditions and the rules of dynamics ("laws"). In contrast, probabilistic models have probabilities (either epistemic or physical) governing transitions between the states of the system; in practice, there is not much sense in asking: what type of probabilities are those used in the model? This point is highly contentious exactly in the subset of statistical models most relevant for explaining the alleged paradox in Maxwell's demon thought experiment. Models also could be discrete or continuous, depending on the type of variable used in model.

---

46 Weinert, Friedel. *The Demons of Science*. Springer. (2016).

47 Weinert, Friedel. "Theories, models and constraints." SHPS *Part A* 30.2 (1999): 303-333.

48 Frigg, Roman and Hartmann, Stephan, "Models in Science", SEP (2012 Ed.), p. 1

49 Cartwright, Nancy. "Models and the limits of theory: Quantum Hamiltonians and the BCS models of superconductivity." *IDEAS IN CONTEXT* 52 (1999): 241-281.

50 Hartmann, Stephan, „The World as a Process: Simulations in the Natural and Social Science“, *Theory and Decision Library, Dodrecht,* (1996): 77-100, p. 81.

Weinert distinguished between a few further kinds of models – analog, scale, functional, hypothetical and structural models.[51] Still, since this classification has not included thought experiments, we can add them as a kind of conceptual models. This will have obvious advantages for our further discussion.

In the category of conceptual models, we deal with both analog models, as well as thought experiments.[52] Analog models function through analogies, by representing unfamiliar with familiar.

It is important to emphasize the limitations of the analog models from the outset. If models should represent real phenomena and their basic properties, mere analogy is not enough. In this case, we must develop a more detailed model if we want to comprehend sufficient part of the complexity of the phenomenon itself. (This is not to downplay or neglect the key role analog models play in heuristics, education and even public outreach of science.)

Most thought experiments could usually be thought of as a manner of representation of the conceptual models. Such conceptual models often have quantitative nature. They create *conceptual systems*, in order to test the facts or ideas. Hypothetical models incorporate idealizations and abstractions. They represent target system with only most significant relations and parameters. Examples of this kind of models play a significant role in social sciences such as economics or demographics, for instance, since they are usually the only kind available for realistic systems studied there.[53]

Scale models represent real system under investigation. The only thing that is changed is its size. Scale models are usually three-dimensional models representing configuration space of our mundane three-dimensional Euclidean space. They require from us to know the details of that system and, for obvious reasons, are of crucial importance for engineering and other practical purposes. The tradition of scale models go back to the early civilizations of ancient Mesopotamia and Egypt where architects and builders often constructed small models of temples, palaces, pyramids, etc. before embarking on the real construction work.

Functional model represents the functional relationship between some of the parameters and as such is of paramount importance in the fields in which explanation is achieved through functionalist theories. By far the best-studied examples of this are models used in evolutionary biology, for many reasons the main being that since Lamarck, through Darwin and Wallace, to the Modern Synthesis, the field has been completely dominated by functionalist explanations.[54] Here, the base of the representation goes from the topological to the algebraic structure: we notice that in the same area there exist different actors in a particular functional relationship, and then we try to ascertain how does this topology impacts their population numbers by setting up algebraic relations.

In structural models, we have both algebraic and topologic structure. They tend to explain some real state of affairs via mechanism or structure. These models can be used for representation of some macroscopic system, for example planetary system, where the output is both literally topological (e.g., predictions of eclipses or transits) and algebraic (e.g., the amount of the perihelion shift of Mercury due to other bodies which *cannot* be reconciled with observations).

---

[51] Weinert, Friedel. "Theories, models and constraints." *SHPS Part A* 30.2 (1999): 303-333.
[52] Weinert, Friedel. *The Demons of Science*. Springer. (2016) Note, also, that this is somewhat different than (but still *related to*) the meaning of "analog" used in electronics and computer science as the opposite of "digital".
[53] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[54] Gould, Stephen Jay, 2002, *The Structure of Evolutionary Theory* (Belknap Press, Cambridge, Massachusetts).

The question is, how models represent their target system? When we build a model, we create a kind of representative structure.[55] This representative structure, outside of purely logical and mathematical fields, is an important part of the model itself.

Models share a particular similarity with thought experiments, namely that they are subject to some constraints if they should represent some aspects of external world. Note that representation can also be relationship between a symbolic construct and a phenomenon. Representation here is not similarity or resemblance. It is not even a structural isomorphism of the model. In a thought experiment a counterfactual scenario is enacted, experimenter creates the possible world. Hence, we cannot compare model world and the real world.

Morgan stress that we learn about models in two points: in their construction and in their manipulation. There is no set of fixed rules for building models. At the initial phase in which the model has been made, we do not learn about it when we study it (since it is hard to see how could we infer any new information, unless perhaps we are especially strong Platonists); this occurs at a later stage, when we manipulate it.

Both construction and manipulation of models may change according to different activities which require different methodology, as function of the type of the model we are dealing with. Some models, like material models might seem unproblematic in the general concept of experimentation.[56]

With functional models, things are different. Since one significant class of models is mathematical, it is often not possible to solve it in an analytic way, but with the aid of numerical simulations.[57] An example of this is the problem of a travelling salesman which is conceptually simple, but intractable in real time due to non-polynomial increases in the number of operations required.

2.3. The function of thought experiments

So, what kind of benefit or value we derive from thought experiments? Let us briefly review some of the historically important perspectives. Mach stressed importance of instructive experience and even education in thought experiment. He classified them as intermediaries between the accumulation of facts and the reasoning. Besides, Mach emphasized that they had to stay close to empirical facts.[58] His logical positivist pupils largely held to his prescriptions.

Planck stressed the heuristic function of thought experiment. He claimed that they can provide new knowledges and new perspective on relationships in nature. Besides, even though there is no request for strict precision of thought experiments, they should not contain contradictions (in contrast to, say, the twin paradox thought experiment).[59]

---

[55]Morgan, Mary S., and Margaret Morrison, eds. *Models as mediators: Perspectives on natural and social science*. Vol. 52. Cambridge University Press, 1999. p. 33; Hartmann, Stephan. "Models and stones in hadron physics." *Models as mediators: Perspectives on natural and social science* 52 (1999): 326. chapter 2.

[56] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

[57] Frigg, Roman and Hartmann, Stephan, "Models in Science", SEP (2012 Ed.), p. 10-11. Also, one should keep in mind that, for instance,

[58]Mach, Ernst. "The science of mechanics: A critical & historical account of its development, (after the 9th German edition)." *Open Court.[JVB]* (1883).

[59] Ibid.

Thomas Kuhn claimed that their role demand much more than mere absence of contradictions.[60] They make us question concepts through which we understand the empirical evidence or even the external world itself. Thought experiments can be used as powerful analytic tool which can help us to overcome the crisis in science. While in the crisis, anomalies accumulate and the old paradigm is not able to solve them, they can help us redefine old concept and develop new ones.

Let us return to the example of Einstein's elevator in order to illustrate this; it postulates correlation between acceleration and gravitation. Still, it is important to note that although thought experiments can discover anomalies, they cannot solve anomalies, but only can help us to redefine concepts within them.

What exactly is the difference between what-if questions on the one side and thought experiments on the other? Some thinkers propose that beyond their counterfactual and hypothetical character thought experiments should fulfill one more: they must have strong relationship with both empirical evidence and theory beyond it. [61]

An additional role of thought experiment is investigating empirical consequences of theories. Kuhn claimed that their most important role is to validate if there is internal consistency, coherence as well as simplicity and adequate explanatory power.[62] Einstein's thought experiment with elevator establishes the equivalence principle; but is also great example of *explanatory power*, for equivalence principle will replace concept of gravitation with concept of the curvature of spacetime.

What exactly is the difference between what-if questions on the one side and thought experiments on the other? Some thinkers propose that beyond their counterfactual and hypothetical character thought experiments should fulfill one more: they must have strong relationship with both evidence and the theory beyond it (notably in light of the Duhem-Quine thesis).[63] Such relationship is obviously impossible for simple what-if questions if they are not motivated by deeper theoretical reasons. Critical rationalism of Karl Popper[64] distinguishes the critical and heuristic function of thought experiments from the apologetic one. According to him, thought experiments should be able to refute theories, since they can prove that the theory has an internal inconsistency. Without going into further detail, it is important to keep in mind for further purposes that, among major methodological views, it is Popper's approach which actually gives most epistemological latitude to thought experiments.

---

[60] Kuhn, Thomas S., A function for thought experiments. In I. Hacking (Ed.), *Reprinted in Scientific Revolutions (1981),* (pp. 6-27). Oxford: Oxford University Press; Humphreys, Paul. "Seven theses on thought experiments." *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grunbaum* (1993): 205-227.

[61]Irvine, Andrew D. "Thought experiments in scientific reasoning." *Thought experiments in science and philosophy* In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (1991): 149-165, p. 150. Humphreys, Paul. "Seven theses on thought experiments." *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grunbaum* (1993): 205-227.
pp. 220-221.

[62] Kuhn, Thomas. "Objectivity, Value judgement, and Theory Choice." *Thomas Kuhn (ed)* 76 (1973): 320-339.

[63]Irvine, Andrew D. "Thought experiments in scientific reasoning." *Thought experiments in science and philosophy* In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (1991): 149-165, p. 150. Humphreys, Paul. "Seven theses on thought experiments." *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grunbaum* (1993): 205-227.
pp. 220-221.

[64] Popper, Karl, *The logic of scientific discovery.* London: Hutchinson. (1959), appendix XI.

2.3.1. What thought experiments cannot do

We have already mentioned that Brown differentiates destructive from constructive thought experiments. The main function of the destructive ones is to discover problems of conceptual and logical nature within some theory.

Platonic experiments are constructive because they produce new knowledge. The problem with this constructive account is that it possesses a spirit of a paradox. What positive result could they establish if they cannot establish empirical claims?[65] If we distinguish understanding from knowledge, we can add more roles to thought experimentation – the role of understanding natural and social phenomena,[66] and even understanding of constraints that model needs to satisfy.[67]

Werner Heisenberg claimed that it is exactly our ability to express complex physical theories with general and basic notions represents a measure of our understanding. More general concepts would allow us to make more relations between different phenomena. Besides, it is important to note that new phenomena that scientists discover also require new notions. Heisenberg, like Bohr, emphasized that Einstein in 1905 developed new concepts of space and time in his special theory of relativity. However, this was just the beginning of the conceptual revolution of the 20th century physics. The problematizing of notions like causation came as a reply on new insights from the field of quantum mechanics.[68] Heisenberg, as perhaps the most prominent member of the Copenhagen school of thought after Bohr himself, held that understanding represents the ability to develop new notions when we are faced with new empirical evidence.

This, however, was not universally accepted, even among quantum physicists. For Erwin Schrödinger the ability to understand *is* actually the ability to develop specific conceptual models.[69] He strongly rejected the instrumentalism and antirealism of the Copenhagen school, which ultimately led him to abandon physics in favor of work in biophysics, origin of life and philosophy. His unease has been shared not only by Einstein, but by many other 20th century physicists, including Arthur Eddington, John Wheeler, Freeman Dyson, David Bohm, John S. Bell, Murray Gell-Mann, and many others. In general, according to these more realist thinkers, the primary aim of conceptual models is to be *assigned* to the observable phenomena. In fact, these observational phenomena make them understandable.

All in all, understanding of the thought experiments as a conceptual model, contributes to our understanding without adding empirical knowledge to it. Einstein, Mach, and Planck insisted that thought experiment in science must stay related to empirical evidence. Popper went even farther in arguing that thought experiments can overturn theories previously well-supported by empirical evidence. Hence, there is no need to limit them to what-if questions.[70] This is the main reason why we should construe them as conceptual, rather than some kind of mental model. They are modeling the physically possible worlds. Nowhere is that more obvious than in the Maxwell's strange thought experiment with a gatekeeper demon.

Now, when we have achieved better understanding of meaning and role of thought experiments, we can explore Maxwell's thought experiment.

---

[65] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[66]Weinert, Friedel. *The scientist as philosopher: philosophical consequences of great scientific discoveries*. Springer Science & Business Media, 2004, 3.1. chapter.
[67] Humphreys, Paul. "Seven theses on thought experiments." *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grunbaum* (1993): 205-227, p. 220.
[68] Heisenberg, Werner. "Nonlinear problems in physics." *Physics Today* 20 (1967): 27.
[69] Schrödinger, Erwin. "Conceptual models in physics and their philosophical value." *Science theory and man* (1957): 148.
[70] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

3. Maxwell’s demon

“Every bit in a computer is a wannabe Maxwell's Demon, separating the state of "one" from the state of "zero" for a while, at a cost. A computer on a network can also act like a wannabe demon if it tries to sort data from networked people into one or the other side of some imaginary door, while pretending there is no cost or risk involved.”
Jaron Lanier, *Who Owns the Future?*

“In the description of matter as a collection of molecules instead of a continuum, questions related to reversibility are presented for the first time in the invention, almost as a joke, of what is now known as “Maxwell’s demon”.”
Carlo Cercignani, *Ludwig Boltzmann: The Man Who Trusted Atoms*

“We all behave like Maxwell’s demon. Organisms organize. It sometimes seems as if curbing entropy is our quixotic purpose in the universe.”
James Gleick, *The Information: A History, a Theory, a Flood*

In Chapter 2, we introduced the all-important concepts of thought experiments and scientific model. This will help us explain in more detail how the problem Maxwell’s demon raises. In order to understand the problem, we will go through some examples from the history of science and we will discuss some attempts to construct an engine which would behave the same way as Maxwell’s demon. In Section 1 we will explain Smoluchowski trapdoor. In Section 2, we shall explain Feynman’s ratchet and pawl, as well as Gabor’s engine. In Section 3, we will explain Feynman’s trapdoor. Finally, in Section 4 we will explain Szilard’s engine.

Section 5 deals with Landauer’s principle. This principle has been used as a most common tool for exorcising the Maxwell’s demon and here we will show why it seems inadequate in philosophical terms. In subsections of Section 5 we will discuss Bennett’s version of Landauer’s principle and its critiques. Also, we will discuss some indirect proofs of Landauer’s principle and their accompanying problems.

After it, in Section 6, we will explain the concept of entropy, and separate a few different kinds of entropy because at least a part of the problem arises due to the confusion[71] between different kinds of entropy. In Section 7, we shall discuss different versions of the Second Law in order to see which one is the most relevant for the problem of Maxwell's demon. Now, we will start with explaining the Maxwell's thought experiment.

James Clerk Maxwell (1831-1879) has been long ago recognized as one of the greatest physicists, and the importance of the massive body of work achieved in his short lifetime has just been monotonously increasing during the last century and a half.[72] Maxwell's demon appears for the first time in a letter that he wrote to Tait on December 11, 1867. – The outstanding intellectual history of this problem is witnessed by the fact that cutting-edge research papers on the topic are still regularly published.[73]

Maxwell analyzed thermal phenomena from the perspective of atomic physics. The Scottish physicist was the first to realize that if we accept the atomic theory as a grounding for thermodynamics, then the validity of the Second Law of thermodynamics is only statistical.[74] The same insight later immensely bothered Ludwig Boltzmann and prompted him to introduce boundary conditions and cosmology as explanatory devices. There were admissible mechanical processes that violated the Second Law in special contexts and for a brief amount of time. Variations on very small scales were indeed observed in fluctuations phenomena, an example of this being the Brownian motion of a pollen grain (or any similar granular system) visible under the microscope. Apparently, a macroscopic motion of a grain arises without any macroscopic cause, leading – by the way of conservation of energy – to decrease of temperature of the fluid. More complex examples include behavior of systems such as spin glasses and their metastable states in condensed-matter physics. The key question is if there is a way for these microscopic violations of the Second Law to accumulate and produce some macroscopic violations.[75]

[71] Confusion between information entropy and thermodynamic entropy

[72] Mahon, Basil, *The Man Who Changed Everything – the Life of James Clerk Maxwell*. Hoboken, NJ: Wiley. (2003).

[73] Cottet, Nathanaël, et al. "Observing a quantum Maxwell demon at work." *Proceedings of the National Academy of Sciences* 114.29 (2017): 7561-7564.

[74] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.)

[75] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

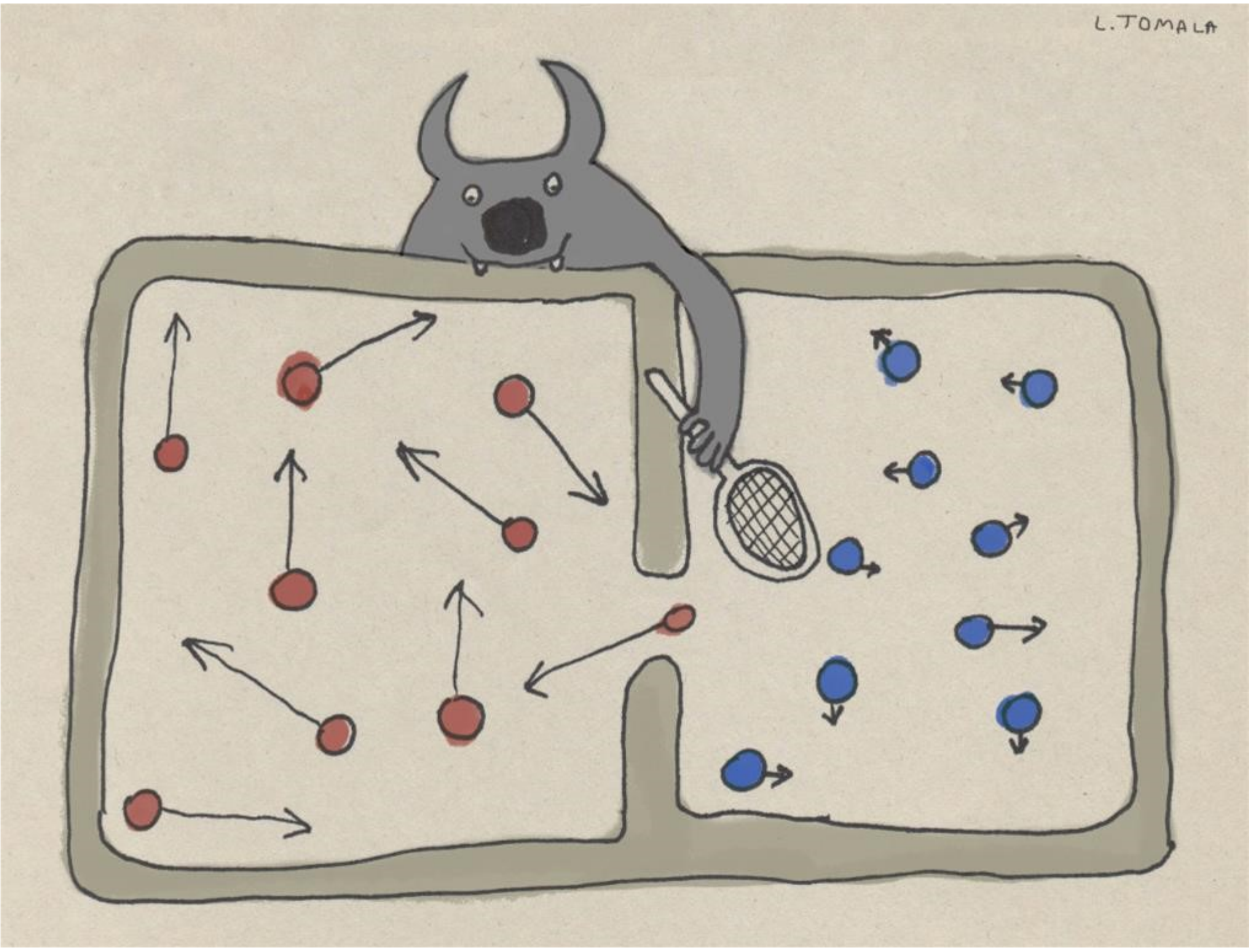

Figure 3.1. An illustration of the action of theoretical entity - Maxwell's demon. Adapted from https://bilimfili.com/maxwellin-cini-foton-kullanilarak-canlandirildi/ **by Ludwika Tomala.**

The demon can make the difference in temperature within a gas. Besides, demon can do it without any work. The container that contains a gas is insulated and divided in two with partition which is also insulated. The gas is in an insulated container, separated on two parts by an insulated partition. (The role of insulation will be considered later.) In the partition there is a hole, its size is that of a single molecule, so it can pass through. The mean kinetic energy per molecule (3/2)kT, hereby T is temperature and k is Boltzmann's constant. This conclusion is valid for gas in equilibrium at well-defined temperature T. The demon has a small shutter with which he could block the hole without any friction or other energy dissipation. If a molecule that comes to the hole from the left side is moving fast, the demon closes it. ("Fast" here can be operationally construed as "faster than the average", which then further provokes the question how the averaging procedure is conducted. In general, we are interested in the so-called *mean quadratic* velocity of molecules. While it is an important practical issue in statistical physics, the conclusions of the thought experiment do not hinge on it.) Therefore, this fast molecule would be reflected back to the left. If a slow (= slower than the average) molecule approaches from the left side, the demon leaves the hole open. Therefore, the molecule proceeds through to the right-side of the container. When another molecule that comes to the hole from the right side of the container is slower than average, the demon closes the hole; when the molecule that approaches the partition from the right side is faster than average, the demon leaves the hole open.[76] Faster molecules will accumulate on the left, while the slower molecules will accumulate on the right side of the partition. As a consequence, gas in the left side will be hotter (greater average kinetic energy per molecule, on any kind of averaging) and the right side will be

[76] Maroney, Owen, "Information Processing and Thermodynamic Entropy". SEP (2009 ed.)

cooler. No work in the classical sense is performed by the demon because of the assumptions that collisions with the shutter are elastic, and that shutter moves without friction.[77]

Therefore, a finite temperature difference will develop without work being performed. This temperature difference could, in turn, be exploited to extract work via any number of classical heat engines (Carnot's engine or whatever similar). This will result in a violation of the Second Law of thermodynamics which is not limited any more to brief fluctuations or very special boundary conditions. From the possibility of its violation, we can conclude that validity of the Second Law is not universal. It does not have universality of the First Law (or other conservation laws like the conservation of momentum or charge). It *must* be a statistical law.

In other possible formulations of the paradoxical conclusion of the thought experiment with the demon, the conceived situation would amount to the construction of the perpetuum mobile of the second kind, or – even more contentiously – to a decrease in entropy of a closed system.

There is an important philosophical issue to be tackled here: Are these various interpretations of the same thought experiment equivalent or not, and if so, to what extent? In what follows, we will first approach the interpreting task without invoking the concept of entropy, to see how far this could lead us in linking physical and computational processes. Entropy will be introduced later on (section 3.6) in our treatment, and we will investigate the issue whether it introduces substantial novelty in the description of the problem situation and in various ways of "resolving" the paradox.

The first point to be made is rather obvious (especially nowadays, after the advent of quantum physics, but it was not so in Maxwell's time): laws that holds on the macroscopic level may not hold on the microscopic level. Maxwell's view on probability leaves us with knowledge about averages over micro-states. Still it does not tell us anything about their individual properties. Since we do not observe particular molecules with any realistic laboratory apparatus, we must adopt statistical method and discard strict dynamical method.[78]

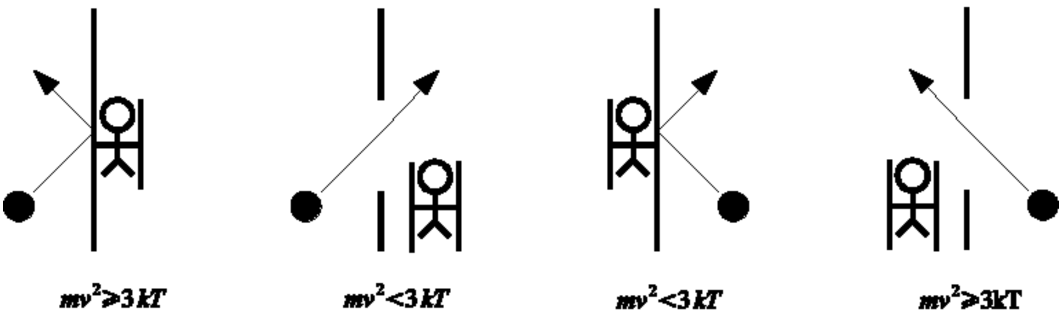

Figure 3.2. Modus operandi of a classical Maxwell's demon. Adapted from "Information Processing and Thermodynamic Entropy" by Maroney Owen. SEP. (2009 ed.). In this image, we see Maxwell's demon as separating molecules with velocities greater than the root-mean-square speed (corresponding to the mean translational kinetic energy in the gas at temperature T) to the left side of the partition, from those below that value (accumulating these "colder" ones on the right side).

There is a simpler version of Maxwell's demon that is always reflecting molecules that come from the left and never reflecting those that come from the other direction. It will make the difference in *pressure* between the two parts of the container. In language of thermodynamical variables, the gradient of chemical potential will arise instead of the gradient of temperature in the original

[77] Ibid.

[78] Maxwell, James., 1867, *Letter to P.G. Tait, 11 December 1867*, in *Life and Scientific Work of Peter Guthrie Tait*, C.G.Knott (author), Cambridge: Cambridge University Press, (1911): pp. 213–215.

Maxwell's formulation. Hence, a conventionally operating engine (a gas turbine, say) could use this difference to derive work. Intention of this thought experiment has been to show that gas could evolve contrary to the Second Law of thermodynamics (i.e., in Planck's statement, from a higher entropy state to a lower one). We should notice that the problem with simplified demon is that it becomes harder and harder for him to operate the shutter, thus necessitating the application of force and dissipation of energy, in contrast to the classical formulation of Maxwell, in which the shutter moves without friction (irrespectively of the temperature difference created hitherto) and therefore without expending of work at all times. It is rather intuitive to understand why it is so: pressure exerts direct force to the shutter, in contrast with the *average* kinetic energy of molecules.

Arguments in favor of the demon – or, more precisely, in favor of the demon's capacity to operate in the manner described – seem persuasive. The relevant question becomes then: why we have never seen such systems arise spontaneously? Could we cause such a reversible thermodynamic change (decrease in entropy in the conventional terminology, which is often confusing, as will be demonstrated below) to occur at the systematic level, and not as a mere fluctuation? Originally, Maxwell's demon was postulated to have powers of perception and perhaps agility/motoric coordination that are far greater than our own. However, some physicists and philosophers have historically held the view that devices which could exploit fluctuations in individual atomic/microscopic velocities are possible. The magnitude of difference between capacities of Maxwell's demon and realistic human capacities is not so extreme as is the case with Laplace's demon, for example. Hence, the former seems intuitively more acceptable than the latter. Especially in the epoch of advanced technologies like miniaturization or nanotechnology, the *practical aspect* of the problem should not be entirely neglected. Probabilistic arguments do not, it turns out, decisively prove that work cannot be extracted from the demon-like contraptions. For instance, as the effect of statistical mechanical fluctuations, we have Einstein's account on Brownian motion in 1905. This opened our way for exploration of this phenomenon.[79] Besides that, there were efforts to limit range of violations of the Second Law of thermodynamics. These efforts appeared due to the rising of the kinetic theory and recognition of the phenomena of fluctuations.

There have been many attempts to exorcise the demon on the basis of his inability to decrease the entropy, because demon is the subject to the Second Law.[80] In a scenario in which such a device were possible in practical terms, machines would produce work with no batteries needed and the world would have been a very different place indeed.[81] Therefore, there has to be a loophole in the paradoxical conclusion of the thought experiment.

3.1. Smoluchowski trapdoor

What will happen if we try to design Maxwell's demon as a physical device? Polish physicist Marian Smoluchowski in 1914 was the first to try to replace the demon with a physical device. In Smoluchowski's device one has an insulated container with a partition that separates the gases that have the same temperature and pressure on both sides. Inside the partition which is separating the gases there is a spring-loaded trapdoor installed. The molecules would strike the door from one side, would open them and the molecule will pass. After such an event, the door will slam shut, therefore

[79] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

[80] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471; "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.

[81] Bennett, Charles H. "Demons, engines and the second law." *Scientific American* 257.5 (1987): 108-116, p. 108.

preventing the passage of molecules from the other side. Spring that is holding the trapdoor is required to be weak, however, and the trapdoor must be light, since it is the only way molecular collisions would be able to open it. Operation of the trapdoor will, at first glance, lead to work being done by microscopic motions alone. The difficulty which has been revealed by subsequent analysis is hidden within the fact that spring has internal kinetic and potential energy. Therefore, it will absorb energy from the collisions of the molecules – and the spring is, presumably, also made of normal matter, i.e., molecules. As a result, it will begin to oscillate. At one point in time spring's energy will obtain the same temperature as the gas.[82] This is unavoidable: even if we suppose that there are internal degrees of freedom within the spring which could keep it from reaching the equilibrium with the gas for some time, this will only postpone the onset of oscillations.[83] At that point, the spring will begin to randomly flap back and forth, incapable of further providing useful work. It seems as if the entire work produced in the course of its operation could be explained as the consequence of the initial conditions: the spring mechanism being out of thermodynamical equilibrium with the gas in the container.

Perhaps this is a good place to consider an item of epistemological importance which is relevant for practically all versions of the demon thought experiment and yet is only rarely explicated: the role of physical idealizations in reaching the conclusion. To say that the trapdoor is light-weight in the specific context of molecules implies quite a strong idealization, since it would mean that it is significantly lighter (or has much smaller inertia, which is essentially the same) than the molecules – what is it made of then? Similar are the requirements of frictionless motion of the shutter in Maxwell's original setup, perfect insulation, etc. All these requirements are physically implausible, to say at least, even when taking into account possible wonders of future technology; one might speculate that the trap door could be made of localized electromagnetic force field, etc. However, for all these possible contraptions it would be necessary, strictly speaking, to show that no work is expended in the process, or at least that less work is expended than can be recovered by using the demon as the perpetuum mobile of the second kind.

Does this present an insurmountable difficulty for reaching the paradoxical conclusions? Not really, since there is no sense in which idealizations necessary for Maxwell's demon to operate are *bigger* or more epistemologically offensive than those which are used in other well-known instances in physics, e.g., when we are talking about point-like charge of an electron or the parallel field lines in a capacitor or even perfect insulation in almost every thermodynamical experiment. The latter is intimately linked to the entire classical thermodynamics and all our theoretical conclusions in it. We do not doubt the validity of conclusions on the thermodynamical efficiency of the Carnot cycle, for example, on the basis of the fact that perfect insulation is unattainable there either. So, while we will briefly consider the relevance of insulation in the discussion of the arrow of time, there is no forceful epistemological reason to assume that paradoxical nature of the thought experiment (if any) is brought about by such invalid idealizations, either individually or combined.[84]

Thus, the trapdoor will seemingly violate the second law. However, soon it will become clear that it is only over short periods, the Second Law is not violated on a longer timescale. Smoluchowski himself speculated that a new *modified* Second Law ought to postulate that device is unable to reduce entropy *continuously*, in contrast with reductions occurring in mere fluctuations of the system. More on the modified Second Law below.

---

[82] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

[83] Arguably, if the spring were to have an *infinite heat capacity*, it could have never reached equilibrium, although it would hardly be a smaller miracle, in the sense of Hume, than any number of violations of the Second Law.

[84] Again, the example of failed thought experiment of the "twin paradox" in Special Relativity is instructive, since the unphysical idealization there (a possibility that the spaceship turns around so slowly that it remains an inertial system all the time) is crucial for the outcome – it cannot be thought of as a limiting case of better and better approximations to the ideal case. The limiting process in the "twin paradox" would include radius of the trajectory curvature to go to infinity, which is obviously incoherent with the finite path needed for spaceship to return to Earth.

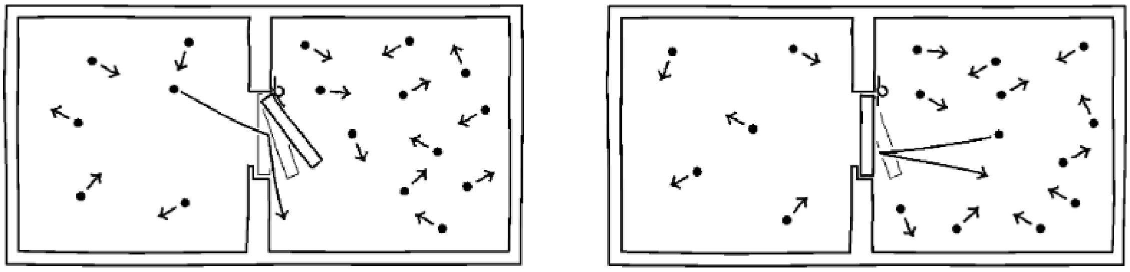

Figure 6 Smoluchowski Trapdoor

Figure 3.3. Smoluchowski trapdoor, a schematic presentation. Adapted from "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon" by Norton, John D. SHPMP. (2005).

3.2. Feynman’s ratchet-and-pawl and Gabor’s engine

Feynman stress that Maxwell’s demon could seemingly be displaced with a simpler device: ratchet and pawl. He construes it to explain why Carnot principle is true. This engine contains an isolated heat reservoir where everything is at the same temperature and it attempts to generate a work from it.[85]

Recall that, according to Carnot’s principle, there is a limit on amount of work that can be extracted in changing temperature of the system from one value to another. This principle is grounded upon another axiom of classical thermodynamics, which claims that heat cannot be converted into work through *any* cyclic process if every component of the system is at the same temperature. Maxwell’s demon should, in principle, be able to circumvent Carnot’s principle; along these lines Feynman conceived his famous ratchet.

The ratchet works like this, first: shaft turns only one way due to specific profile of the toothed edge. We have a box of gas inside is an axle with vanes in it.

[85] Feynman, Richard P., R. B. Leighton, and Matthew Sands. "Ratchet and pawl." *The Feynman lectures on physics* 1 (1963): 46-1.

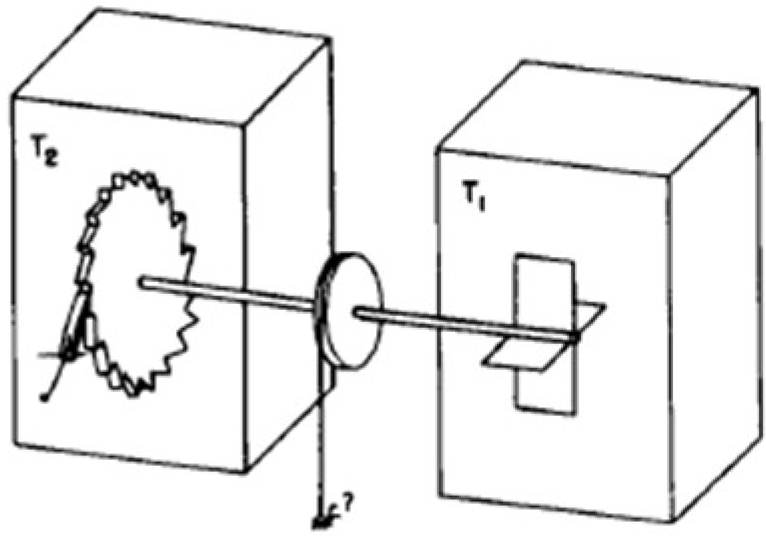

Figure 3.4. The ratchet and pawl machine, as described by Feynman. **Adapted from *The Feynman lectures on physics* by Feynman, Richard P., R.B. Leighton, and Matthew Sands. (1996).**

The molecules of gas will bombard the vane; therefore, the vane will begin to oscillate and jiggle. This will still be on the level of random fluctuations, similar to the Brownian motion. On the other end of the axle, the wheel is hooked and it can turn only one way – ratchet and pawl. Molecules of air are moving around at supposedly the room temperature, having the classical distribution of velocities. Since there are so many of them and the distribution is a very broad, there will be occasional collisions strong enough to move or jiggle the vans one way or another. If the shaft tries to jiggle, it could turn only one way. The wheel will turn and we could tie something on the string that is hanging from the drum and lift it, thus expending useful work. It is supposed to be turning extremely slowly (on cosmological timescales perhaps), but it will seemingly turn, converting chaotic heat motion of molecules into useful work. If it could work this would be another example of the perpetuum mobile of the second kind, since the gas in the reservoir with the vans will gradually get colder by the same amount of energy converted into the mechanical work.

Besides, if we consider Carnot's hypothesis, we can clearly conclude that it cannot work. However, when we consider the contraption, it seems quite possible, or at least it is not obvious why it should be impossible. Where is the "catch"? There are several problems with the whole setup, as Feynman carefully explains. First where is a pawl, there is string in it, because the pawl must come back after coming off any tooth. Second, if the material from which it is made was too elastic, tooth could come under when pawl is up, causing the wheel turn the other way around. Hereby, what is making this process irreversible is a *damping* mechanism that have stopped to bounce. Damping is necessarily a dissipative process, similar to friction, and hence antithetical to the Maxwell's-demon-like constructions which insist on frictionless nature of shutter sliding and similar moves. The effects of damping constitute in the energy that was stored as elastic energy in the pawl going into the wheel and showing up as heat. The wheel will get hotter and hotter. Some of the heat could be decreased if we put a gas in. But, if the temperature of gas continues to increase, both pawl and the wheel will reach the temperature where they exhibit from the Brownian motion. The limit on the process occurs because both the pawl and the wheel at some temperature T exhibit Brownian motion. The Brownian motion of the vanes would act to turn the axle *backwards*; hence, the pawl will occasionally "spontaneously" lift itself over a tooth.

This "conspiracy" tends to occur more and more frequently as things gets hotter and the average (or root-mean-square) velocity of molecules increases. Therefore, we found the reason why this machine does not work in perpetual motion, extracting work from microscopic chaos. Feynman concludes that it is necessarily to do work against the spring (or any other damping contraption in different realizations of the setup) if one wishes to rise the pawl upon the tooth.[86] In more than a metaphorical sense, one asymmetry (that of the shape of the ratchet) is cancelled by the other asymmetry (that of the heat flow, or the second law of thermodynamics).

Feynman analyzed how could ratchet ever function as an engine and concluded that it will go exactly in the opposite direction from the one for which it was intended. Even in the unlikely case someone tries to really construct such a lopsided design, it would not work in practice. The reason for this lies in a fact that if these temperatures are the same, it would be impossible to make it turn in one direction rather than in other.[87] This will present the endpoint of any work extracted, similar as in the Smoluchowski trapdoor "engine". The heat goes from the hotter body to the colder one, but when the temperature of the bodies in contact reaches the same value, there is nothing to determine further direction of the heat transfer. The system just oscillates around a position, similar to what we expect to have with a system in equilibrium. It is not possible to construct a machine which will turn only on one side on a long-time scale when it is entirely isolated, without any external perturbation.[88] This is consequence of the fact that the laws of classical mechanics are time-reversible.

An added virtue of Feynman's ratchet and pawl is in that it provides simpler explanation of why Gabor's engine cannot work. Explanation that Gabor gave depends on quantum theory of radiation.

The function of Gabor's engine (as conceived in 1964) is to catch the molecule at the one end of cylinder. In order to do this, engine use optically triggered mechanism. When the molecule is trapped, it will provide kTlnX[89] work due to isothermal expansion. As it turns out, there is no entirely irreversible mechanism that can trap molecule (not even optical one).[90]

We saw that Feynman points out when he analyzes case where pawl comes off the tooth even if it is construed to turn only in one direction. This mechanism has a probability $\exp(-E/kT)$ to run in the opposite direction from the one that is planned in the construction. Whatever is the work hypothetically done by this engine, it shares similarity with a Feynman's ratchet and pawl. These similarities are of mathematical kind. This is so because in both machines energy E is $E > kT \ln X$ if it runs in the intended direction, but energy will be $E < kT \ln X$ if it runs in the opposite direction. In both cases, mechanism will run in the direction opposite to the one that is intended on the long timescale. Therefore, in cases of these machines, the Second Law of thermodynamics would not be violated, no matter how strange or counterintuitive their operation might look.[91]

3.3. Feynman's trapdoor

Feynman has also argued that any other attempt to build such a finite-size demon will end up with result that demon gets so warm, that he cannot see after a while. We notice a clear sign of Brillouin's

---

[86] Feynman, Richard P., R. B. Leighton, and Matthew Sands. "Ratchet and pawl." *The Feynman lectures on physics* 1 (1963): 46-1.
[87] Ibid.
[88] An additional condition here is that the temperature is above the absolute zero. Strictly speaking, this condition (the Third Law of thermodynamics or Nernst's Law) is a consequence of the *quantum*, not classical mechanics, although we cannot discuss it here in more detail.
[89] X is expansion ratio.
[90] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.
[91] Ibid., pp. 509-510.

thinking about *perception* here. The simplest possible demon is a trap door. This mechanism is contained from spring that is holding the trapdoor over the hole. The faster molecule could lift it and get through, while the slow ones cannot get through. This mechanism is similar to Feynman's ratchet and pawl. Still, the problem here is that lifting the trap door is *necessarily* an inelastic collision – part of the fast molecule's energy must be expended to do the work against the spring keeping the trap door shut. That energy has to go somewhere – if we assume that the whole assemblage is insulated (again!) it is stored in the trap door + spring as heat. On a longer time-scale trap door would heat up. However, on this point we will consider that heat of the demon could be only finite. Therefore, the demon will heat up. (Infinite specific heat would imply various infinities, or "divergences" in thermodynamical variables, which would have been an example of a solution worse than the problem!) The demon will start to oscillate due to the Brownian motion, and could not tell if he is coming or going to the equilibrium position, much less if the molecules are incoming or outgoing. Therefore, such a mechanism would not work in intended way. (Feynman's idea from the beginning was, of course, that such "machine demons" must fail anyway and he goes into details for purely pedagogic purposes.)[92]

3.4. Szilard's engine

Does all this mean that Maxwell's demon cannot exist in a physical world? Could we construct an engine that could accumulate microscopic fluctuations and cause violation of the Second Law? [93] Here it is not enough if we just notice that devices as trapdoor and ratchet-and-pawl fail. Something more complicated might work. Purely mechanical laws tell us nothing about practical possibility of such a device.

Even if we go outside of the domain of mechanics, a perpetual motion machine of the second kind does not exist, be it based on electromagnetic, nuclear, solid-state or any other principles whatsoever. Still, question arise: would such a machine exist if intelligent being would operate it? Here we come across a very difficult problem of how to physically operationalize intelligent behavior, which led historically to the emergence of the entire field of physics of computation and its subsequent derivative fields. We will discuss this in more detail in Chapter 6.

In 1929, Leo Szilard[94] investigate this special case of intelligently operated Maxwell's demon in a very important model, often dubbed *Szilard's engine*. He did so by considering a box with a single molecule. We can imagine, for instance, that all other molecules are just "frozen" and observe the motion of a single one of them. If we want to reduce the entropy, we must presuppose intelligent being that will acquire knowledge about the fluctuations that occur and perform a measurement. Measurement has its compensating cost, so the Second Law would not be threatened.[95] While one molecule cannot constitute a "gas" in the thermodynamic sense (which seems to be misunderstood

---

[92] Feynman, Richard P., R. B. Leighton, and Matthew Sands. "Ratchet and pawl." *The Feynman lectures on physics* 1 (1963): 46-1 ; Zeh, Dieter. "The physical basis of the arrow of time." Springer-Verlag, New York (1992), subsection 3.3.

[93] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." *SHPMP* 36.2 (2005): 375-411.

[94] Szilard, Leo, "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings", Zeit. Phys. 53: 840–856. (english translation in The Collected Works of Leo Szilard: Scientific Papers, B. T. Feld i G. W. Szilard (Cambridge, Massachusetts: MIT Press, 1972), (1929): pp. 103–129.

[95] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

in many popular accounts), it certainly is a conceivable system in both classical and quantum mechanics.

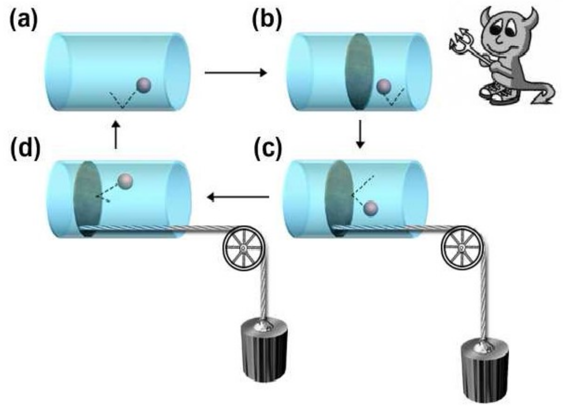

Figure 3.5. Szilard's engine with its major stages. Adapted from http://www.eoht.info/page/Maxwell%E2%80%99s+demon. Initially (a), we do not know the location of the molecule inside the chamber. Maxwell's demon inserts a barrier in the middle (b) and detects the molecule as being, say, in the right-hand part of the chamber. This information is recorded in demon's memory. In the same time, the molecule exerts a "pressure" on the partition, which supposedly can move leftward or rightward with no friction. Depending on the knowledge on the position of the molecule, demon now attaches a weight on which work could be exerted when the partition moves either leftward or rightward (c). In the situation shown, demon will attach the weight on the right side, so that the "isothermal" expansion of the "gas" performs useful work (d), which Szilard calculated to be $k_bT \ln 2$. We should keep in mind that the information required to be kept by demon is *exactly 1 bit*: independently from the side on which is the molecule.[96]

In this picture, we see the box with a molecule that is in contact with the heat reservoir on a given temperature and a partition. Due to the random fluctuations, energy would be transferred between the molecule and the heat reservoir, since the thermal contact transfers energy. With this thermal energy, the molecule bounces throughout the box, randomly.[97] The partition – of negligible mass – can be inserted into the box without exerting any mechanical work. By doing so, the partition is dividing box into two separate volumes, and slides, frictionlessly, toward left or right. Insertion of the partition in the box will result in collisions of the molecule that will imply a kind of "pressure" on the partition. (A qualification is necessary since the real thermodynamic pressure is collective, statistical phenomenon, just like the temperature, so in strict sense it cannot be exerted by a single molecule. However, collisions with the partition will happen nonetheless at this level of idealization and the conservation of momentum ensures that the massless partition will move identically as if the "gas" was expanding isothermally.) When the partition moves toward the pressure, this force will lift a weight. When it moves away, it will drop a weight.[98]

When the partition is inserted in the middle, there will be the same probability that molecule will be trapped on one of the sides– note that it is logically equivalent to statement that we *do not know* where the molecule is. If it becomes known – through a measurement performed by demon which

[96] Maruyama, Koji, Nori, Franco, and Vedral, Vlatko, Colloquium: The physics of Maxwell's demon and information. *Reviews of Modern Physics*, *81*(1), (2009): p.1-23.
[97]Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[98] Ibid.

could in any case be imagined completely analogous to the situation in the original Maxwell's thought experiment – which side the molecule is on, it is possible to extract work. If we make a calculation based on the ideal gas law, PV=NkT, where N = 1, the maximal amount of work that we can derive when partition slides towards one of the sides of the box is kT ln2. The factor "2" under the natural logarithm originates in the obvious fact that two equal volumes exist upon the insertion of the partition. Since the molecule is in contact with heat bath its kinetic energy is (3/2)kT. Work that we will extract would be drawn from the heat bath. The cycle would be completed when the partition reaches the side of the box. What are the net effects of such operation of Szilard's engine? An amount of heat which was derived from the heat bath has been turned into work. And it is not just a brief fluctuation. This cycle can be repeated indefinitely. If this could be the case, then even Smoluchowski's modified Second Law (which relies on thermodynamic behavior on *longer timescales*) appears to be violated.[99]

In order to extract the work, it is necessary to know on which side the molecule is. Without knowing it, we could not conclude in which direction the partition should be moved. Hence, *demon's knowledge matters*. Szilard's argument was that the Second Law would not be violated if the demon could not acquire the knowledge without paying the entropy cost. However, it is not clear if it is necessarily to acquire this kind of knowledge in order to operate the engine. Second, it is not clear that such a knowledge must come with entropic cost. So, Szilard's model was just a first step toward establishing the deep link between thermodynamics and information processing.

3.5. Landauer's principle

Consider a general class of systems ("machines") capable of computation. It has been assumed that the set of computing machines necessarily include machines that operate logically irreversible functions. Hereby, logical irreversibility of functions means that logical functions they perform do not have one value as output (since, if they were always single-valued, they would have been logically reversible). Where does this logical irreversibility come from, then? It has often been argued that it is related to some kind of physical irreversibility and that it requires dissipation of energy as heat. Engineer and physicist Rolf Landauer in his crucial 1961 paper analyzes two simple, but representative mechanical models of so-called bistable devices. He analyzed relationship between speed and energy dissipation, trying to calculate errors caused by thermal fluctuations.[100]

About the same time, Brillouin[101], Gabor[102] and Rothstein[103] have argued that if we want to acquire information through a measurement it will cost us not less than kT ln 2 energy per bit of information. The great pioneer of computing, John von Neumann[104] also followed the work of his erstwhile Budapest neighbor Szilard and suggested that processing of information causes dissipation of energy. Parenthetically, what this overview shows is that in the mid-20th century the lack of the physical account of information processes has become a major issue in which some of the best minds of science and engineering have been involved. And all of this, and what came later, originated in Maxwell's letter to Tait and the demon thought experiment.

---

99 Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

100 Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.

101 Brillouin, Leon. "Maxwell's demon cannot operate: Information and entropy. I." J. App. Phys. 22.3 (1951): 334-337; Science and Information Theory (New York: Academic Press), (1956).

102 Gabor, Dennis, "Light and Information", *Progress in Optics* 1, (1964): 111–153.

103 Rothstein, Jeffrey, "Information, Measurement, and Quantum Mechanics", *Science* 114: (1951): 171-175 .

104 Von Neumann, John. "Theory and organization of complicated automata." *Burks (1966)* (1949): 29-87.

Landauer also considered that any act of processing of information generates some amount of heat. Computing, like all processes coinciding to final rate must involve some dissipation. He argued that some minimal heat generation must take place, independently from the rate of that process.[105]

It should be noticed that the dissipation has its function. Landauer will tighten the concepts involved in dissipation. Binary function must have at least one degree of freedom related to the information. Usually, degree of freedom has been related with kT of thermal energy. This is energy that is required for signal that should pass from one device to other in order to overcome the noise.[106]

Arguments on the process of measuring, do not succeed to define this process adequately. Moreover, these arguments do not address the question: What are the conditions that measurement must meet to make it possible for system A to perform a measurement when it is coupled with system B? The mere fact that these two systems are coupled does not imply dissipation, however.[107] While this question has become extremely controversial and hotly debated in the domain of *quantum* mechanics, it is clear from Landauer's discussion that it is not entirely obvious or trivial in the classical domain either.

Simple binary engine that consists of a particle in a bistable potential is represented in figure 3.6 below.

[105]Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.
[106] Ibid.
[107] Ibid.

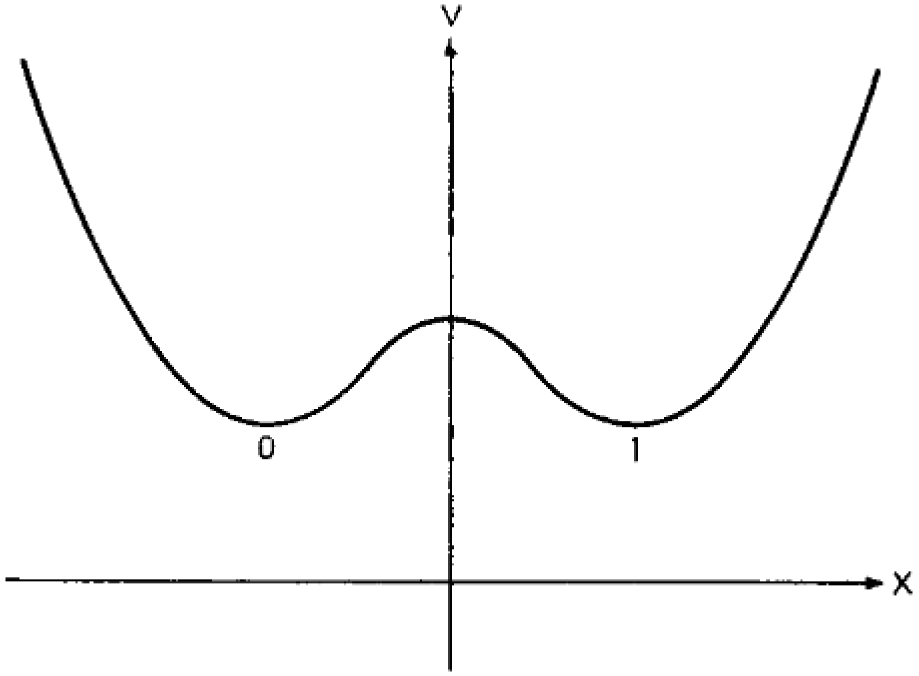

Figure 1 **Bistable potential well.**
x is a generalized coordinate representing quantity which is switched.

Figure 3.6. Bistable potential used as a model by Landauer in his seminal 1961 study. Adapted from "Irreversibility and heat generation in the computing process." by Landauer, Rolf. (1961).

Particle in the left minimum represents the state 0 (this is arbitrary label, of course). Right is the state 1. “Restore to one” operation will leave particle in the state 1. At first it seems like it is possible to move the molecule from 0 to 1 *without spending energy* because we extracted energy from the particle in its decline along the potential curve. A realistic computer does not operate in this manner. The way computer operates on any input information is independent on the exact data that the input contains, according to the classical prescription of Turing. In most cases, computer operates on information in a same way, according to its programming, no matter on the kind of data it is operating on.[108] For example, even the old mechanical adding machine simply adds its inputs, irrespectively whether they are odd or even numbers, primes or squares, etc.

Landauer classifies devices according to the manner in which they hold information that has not being processed or currently interacted with. The simplest kind of device holds information without

[108] Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.

spending any energy. In such a device, Brownian motion in the box is slower. The second kind of device will be the one that is in a steady but dissipative state.[109]

The third class is “catch all” or “wild card”: the class comprising the devices in which time variation is crucial for the recognition of information. Example of these kind are phase-bistable system studied by von Neumann. The lower frequency signal clusters around two values for the phase and this is a source of bistability. In which direction will information flow depend on the losses.[110]

Landauer firmly maintained that there must be physical states which correspond to the logical ones of the adequate physical system. All subsequent successes of physical implementations of computing devices have clearly justified this contention of his. He clearly made a distinction between logically reversible operations from the irreversible ones. We will consider any device as part of the former category when it is not possible to completely determine (or reconstruct) the input given the output. We saw that he considered logically irreversible operations as essential for computing; given the parallelism between logical and physical operations, logical irreversibility would, according to Landauer, imply physical irreversibility which is, in turn, causing energy dissipation of at least kT ln2 per single bit of information.[111]

For one operation to be logically reversible it needs to be like 1:1 map. Landauer argued that such operation could be performed without the compression of the physical state space. In contrast, if operation is logically irreversible, such operation would perform a reduction of the physical state space. The compression would be followed by an increase in entropy via dissipation of the heat.

Landauer’s argument for the logical irreversibility has three different levels.

At **the first level** of the argument he points out that most of the machines depend on steps that are logically irreversible. In addition, if some of the machines happen to copy their logical structure, they will also inherit their logical irreversibility. Therefore, they will inherit their physical irreversibility, as well.

The second level of Landauer’s argument analyzes a particular kind of computers that have logical functions of just a single variable or a pair of variables (analogous to the elementary Boolean logic calculus). On **the third level** of Landauer’s argument, a small “special purpose” (i.e., not a universal Turing machine) computer with three bit positions (three-input, three-output device) has been studied.[112]

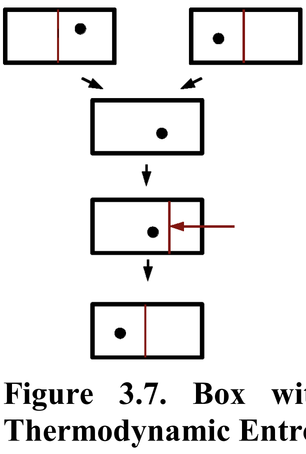

Figure 3.7. Box with single molecule. Adapted from "Information Processing and Thermodynamic Entropy" by Maroney, Owen. SEP. (2009 ed.).

[109] Ibid.
[110] Ibid.
[111] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[112] Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.

In order to do the quantification of the heat generation, he analyzes what happens when we reset the bit. He chooses this operation for it is the most basic of all the logically irreversible ones. It has only one output, state 0, for two possible input states: 0 and 1.

To understand it better, we will analyze device that is very similar to Landauer's. Imagine the box with a molecule in it. A partition will divide the box into two parts just like in Szilard's model. The box is in contact with external heat bath at temperature T. The molecule being on the left side will represent (logical) state 0, while the molecule being located on the right side will represent (logical) state 1. When the partition is removed, the molecule will be free to move chaotically inside the box. What will happen next is that collisions with the molecule will cause the "pressure" (with all the qualifications already considered) exerted on the partition. In principle, this will produce work. Energy from the work will be transformed into heat via the molecule; ultimately, it will be transferred to the heat bath. This process will require at least kT ln2 of work, as per the same reasoning as above.[113]

Alternatively, we can consider the original Landauer example: Imagine situation where we already have implicit restore operation acting on each bit of the logical assembly. We start from a single initial state for the entire collection of the bits, the one that corresponds to zero entropy. The initial entropy could increase by NkT ln 2 when the initial information is thermalized.

Landauer's argument is independent from connections that have usually been made between entropy and information. When we conceive a bit of information, we think about a bit located (as a particular state) in physical system with many additional degrees of freedom.[114] Therefore, Landauer presents a strong generalization of the Szilard model and answers the questions which remained open at the end of the previous section and which Szilard was unable to address.

In the interval between Szilard and Landauer, the very concept of the *computer* as a physical device has first emerged following the leads of Turing, Zuse, and von Neumann. Any computer requires an input in order to operate on it. It is highly unlikely that computer will operate on random data (which would be informational equivalent of the maximum entropy or the thermodynamic equilibrium). The erased bits may not carry the maximum of information. The fact that bits contained in the initial state did not have maximal diversity, implying that when we reset them the process may reduce the entropy increase (comparing to the entropy increase that reset or erasure operation will have on completely random data, i.e., those which have maximal diversity). This occurs only if we take advantage of our *knowledge* about the structure of the inputs. As any computer scientist or an applied mathematician knows all too well, a large fraction of the task of practical computing is exactly contained in precise structuring of the input information.[115]

Landauer repeatedly poses the key question whether entropy could be reduced by a logically irreversible operation. His conclusions are often expressed in the famous "Landauer's principle" in this way***: we could not reset a beat to zero without transforming less than kT ln2 work to heat.***

Although, many authors subsequently referred to it as *erasure*, Landauer, originally, referred to it as a *resetting*. A "resetting" and "erasing" do not, in fact, constitute the same operation, since the erasure need only to destroy information and does not necessarily leave the system in the zero state. In contrast, the resetting operation does return the system into the zero state. With that proviso, Landauer's principle introduces the entire, highly non-trivial, field of thermodynamics of computation. In particular, there is a firm connection between logical functions and any physical realization of these functions in any conceivable thermodynamical system.[116]

---

[113] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[114] Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.
[115] Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.
[116] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

The question of logical reversibility/irreversibility was directly addressed by Landauer in his 1961 paper. His aim was to question if the notion of a measurement is well defined. Landauer came to the conclusion that logical irreversibility is the essential part of any computation whatsoever. He considered this property of logical irreversibility to be the cause of necessary heat generation of minimum kTln2 in information processing.[117]

3.5.1. Landauer's argument in a nutshell

Now, let us stress the key points of Landauer's argument. Computer, with its information that bears certain degrees of freedom interact with the thermal reservoir. The interaction between computer and thermal reservoir act like a sink, it dissipates energy involved in computation. This is the first role of its interaction and it is happening due to the fact that computer performs irreversible operations.[118]

The second role of their interaction is a role of a source of noise that is causing errors. There is a small probability that switched element will remain in its initial state due to thermal fluctuations.

This is an *irreducible* property following from the definition of noise. Even if we believe in a perfectly deterministic world of classical physics, as stipulated by Laplace's demon, there is no way of accounting for the origin of thermal fluctuations on the epistemic level. While this could be interpreted as moving the focus to our imperfection as observers – and in particular as observers of phenomenological thermodynamics – there is no reasonable alternative here. It is for this reason, among others, that the comparative discussion of Maxwell's and Laplace's demon will be given in the next chapter.

In the specific context, we have two dominant sources of error:

1. Time allowed for switching is inadequate, therefore we have incomplete switching.

2. Thermal fluctuations that cause change in the stored information.

In the present-day actual computers, the thermal and the requirements for energy dissipation are calculated as absolute minimum. Actual devices will need much more energy to erase the information from the computer's past history.[119]

Bennett[120] continued argumentation that Landauer started. However, the difference is that he claimed that we could avoid logical irreversibility in computation. Later, in 1982, he argues that measurement can be operated with a logically reversible process. This way, we will avoid need for heat generation.[121] This is quite different than arguments of Neumann and Brillouin. However, most accepted interpretation Landauer's principle was the interpretation that Bennett gave.

According to Norton, all that this principle demonstrates is that particular erasure process causes increase of thermodynamic entropy by kT ln2. The ultimate origin of this entropy increase is in initial

[117]Ibid.
[118] Ibid.
[119] Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.
[120] Bennett, Charles H. "Logical reversibility of computation." *IBM journal of Research and Development* 17.6 (1973): 525-532.
[121] Bennett, Charles H. "The thermodynamics of computation—a review." *International Journal of Theoretical Physics* 21.12 (1982): 905-940.

step where we remove the partition. This happens because after we remove the partition it is not possible to know on which side is the molecule. It does not give us a proof that all possible erasure processes are of such a kind that they will necessarily create thermodynamic entropy.

However, no matter on which side is the molecule, the erasure will succeed inevitably. Still, this analysis does not show that we can satisfy robustness condition only via thermalization. This analysis cannot show robustness condition in the standard framework, since it contains processes of erasure which do not result in thermalization.

In Boltzmann's statistical physics thermodynamic entropy of a system was explained through its relation to the accessible volume of the phase space that it occupied.[122] Therefore, entropy for the system S will be S=kln (accessible phase space).

Before erasure the memory device contains data that molecule is either on left or right side of the box. For that reason, a molecule is related to phase volume on both parts of the box. If we erase memory from the device, molecule will be in the left part and phase space will be reduced to the half of its size. This will reduce entropy for kln2. This reduction would be compensated when the phase space of its surrounding double its size. It would also lead to entropy increase of at least kTln2.

It should be noticed that the molecule, since it was located in only one half before the process of erasure took place, was not related to phase space volume that includes the whole box. Although we do not know which of the halves it was, we know that it will always be one of them. Thus, there is no need that the erasure reduces phase space volume, it only has to replace the part of phase space to which molecule has the access.

3.5.2. Bennett's version of Landauer's principle

As already mentioned, the distinguished contemporary physicist Charles Bennett (in 2003) has proposed an interpretation of Landauer's principle. This interpretation is more general and in many ways explanatory superior. This version claims that not only erasure has entropy cost, but merging of computational paths that appears after the demon's intervention and does not include erasing also has entropy cost. New principle could be formulated as follows:

> "Any logically irreversible manipulation of information, such as the erasure of a bit or the merging of two computational paths, must be accompanied by the corresponding entropy increase in non-information-bearing degrees of freedom of the information-processing apparatus or its environment."[123]

Landauer's principle is considered as grounding principle in both thermodynamics and information processing. We already explained that it holds that any logically irreversible operation, must have its entropy cost.[124] An example of such an operation is erasure and, Bennett argue, merging of the computational paths.

Nevertheless, it is considered that logically reversible processes, can be (at least in principle) possible to accomplish in physical system, in a thermodynamically reversible manner.[125] Bennett refutes some arguments against Landauer's principle.

---

[122] Norton, John D. "Waiting for Landauer." SHPMP 42.3 (2011): 184-198

[123] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510., p. 50.

[124] Ibid.

[125] Ibid.

Landauer aimed to analyze digital computers from the thermodynamical point of view. He analyzed what distinguishes macroscopic from microscopic degrees of freedom. From this analysis, he draws the conclusion that via some degrees of freedom the logical state of the computation is encoded and this information bearing degrees of freedom must be designed robustly against random perturbations from the environment ("noise"). Therefore, we can determine computer's logical states from the initial values. While computer as a whole can be considered as an isolated system that is subject of reversible laws of motion, it is also subject of logical states that often evolves irreversibility.[126]

Landauer's principle thus could be formulated in the language of mechanics: Hamiltonian dynamics conserves entropy, entropy decrease due to information while a logically irreversible operation needs to be compensated by at least same amount of entropy increase, elsewhere (usually in the environment or degrees of freedom that do not carry information). Typically, entropy takes form of energy that is converted to heat, but entropy could be passed differently, for example, we could randomize those degrees of freedom which contain information about position of particles in the environment.[127]

Will logically irreversible operation be thermodynamically reversible, depends on the kind of data on which it is applied. If we apply it on random data, operation might still be thermodynamically reversible. If it is applied on known data, it will be thermodynamically irreversible. Bennett support this claim with the following argumentation: it will be thermodynamically irreversible due to impossibility to compensate entropy of the environment increase by lowering the entropy of the data.[128]

3.5.3. Objections to Landauer's principle

No matter how rational and commonsensical Landauer's principle is, there has been many criticisms in the last more than half of century. It is interesting that, while the principle has been largely accepted in physics of (classical) computation and even engineering, it encountered strong resistance in philosophy of science. Main objections to Landauer's principle were raised by Earman and Norton in a series of papers. For instance, in their 1999 paper it has been argued, that because it depends of the second law, Landauer's principle is either unnecessary or insufficient as a tool for exorcising the Maxwell's demon.[129]

In general, there have been three kinds of objections:

1. Landauer's principle is considered to be false. Arguments for this statement were that there is no necessary association of thermodynamic quantities, such as heat, entropy, and work on one side, and mathematical properties of abstract computing like logical reversibility. Therefore, connecting them is like comparing apples and oranges. This could be called the *category* objection.

2. It is false because operations that process the data need to pay the cost of at least kT ln2 of energy, whether process is logically irreversible or not; this could be dubbed the *redundancy objection*.

3. It is false because it is in fact possible to perform logically irreversible operations in a thermodynamically reversible fashion, using sufficiently sophisticated and subtle procedures. We can call this *algorithmic objection*, since it has important points of contact with the controversies

[126] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.
[127] Ibid., p. 503.
[128] Ibid.
[129] Ibid.

surrounding the Church-Turing thesis, hypercomputation, algorithmic information theory and other intriguing topics in applied mathematics and computer science.

The first objection sounds most „philosophical" and at least superficially touches on relationship between mind and matter, physical and cognitive properties of an entity. It seems similar to the challenges posed by the definition (or the lack thereof) of "intelligence", "memory" and other cognitive phenomena to day-to-day work in, say, psychology. On one hand, it is appealing to consider questions about demon's intelligence and the general issue how is possible to characterize an intelligent being in purely physical terms. Linking Maxwell's thought experiment, even if indirectly, with those important and evergreen topics cannot be accidental; we shall return to some of these issues in the Chapter 6. However, one should not fall into the trap that relying on limited – and to a large degree administrative and anthropomorphic – categories could impact the validity of a physical description like the one given by Landauer. If we accept physicalism, there is no *de re* difference between statements of thermodynamics and statements about some physical system embodying some computation in its registers, memories, etc. In brief, the first claim is wrong because entire universe is subject to Hamiltonian or unitary dynamics if they are considered closed autonomous system with classical properties (like the identity of parts) and under physicalism.

The second objection was refuted by explicit models of physical mechanisms that operates reversible computations with zero cost (for example ballistic computers[130]) or Brownian computers per cost tending to zero for slow operations.[131] Also, one should note that the value of kT ln2 is still extremely small value from the point of view of dissipation of energy in practical computing.

Almost all data here are processed on the macroscopic apparatus, real or theoretical. Processing of information like transcription/reverse transcription between DNA and RNA has some stages that are allegedly performed via chemical reactions which are reversible.[132] Additional difficulty here is that these biochemical processes include molecules with so large molecular weight that they are "neither here nor there", not belonging firmly in either macroscopic realm of classical mechanics, not in the microscopic realm of particles and atoms ruled, as we know now, by quantum mechanics. Physicists call such systems *mesoscopic*, denoting an intermediate scale, where parts of any problem situation could be represented by quantum models and other parts with classical models. While it is impossible to deal adequately with the topic here, it is somewhat ironic to note that, contrary to naive expectations based on popular descriptions of quantum correlations (like in the EPR pairs), quantum computation seems so far to be still *more* reversible than its classical counterpart.

Reversible measurement is one example of significant example of logically irreversible operation that is related to the Maxwell's demon and Szilard's engine. It applies reversible path from the initial state of memory about the location of the molecule to the two of the possible states, depending on side on which the molecule in the Szilard's model engine resides.[133]

Third position was asserted by Earman and Norton[134] as well as Shenker.[135] Earman and Norton hold that demon's memory could be in two states (conventionally denoted with R and L), similar to computer program having two logically reversible subprograms. They hold that gas and demon are

---

[130] Fredkin, Edward, and Tommaso Toffoli. "Conservative logic." *International Journal of Theoretical Physics* 21.3 (1982): 219-253.
[131] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.
[132] Ibid.
[133] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.
[134] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond, SHPMP 30.1 (1999): 1-40, pp. 16-18; Shenker, Orly R. "Logic and entropy." (2000). http://philsci-archive.pitt.edu/115/1/Shenker_Logic_and_Entropy.doc.
[135] Shenker, Orly R. "Logic and entropy." (2000). http://philsci-archive.pitt.edu/115/1/Shenker_Logic_and_Entropy.doc.

returned to their initial states by the modified Szilard's engine. If this radical conclusion were really true, the second law of thermodynamics would be unavoidably violated.

Bennett argues that although both L and R are logically reversible, we cannot claim the same for their combination. The reason is that it will include a merging in the flow. This is example of logical irreversibility similar to the data erasure. Each time it begins a new process, it must pay the thermodynamic cost according to Landauer's principle. Merging of the flow of control constitutes a logical irreversibility that is illustrated in Figure 3.8 (from Bennett's 1982 paper[136]). We shall see below that it is essentially the same thing as merging of phase-space volumes.

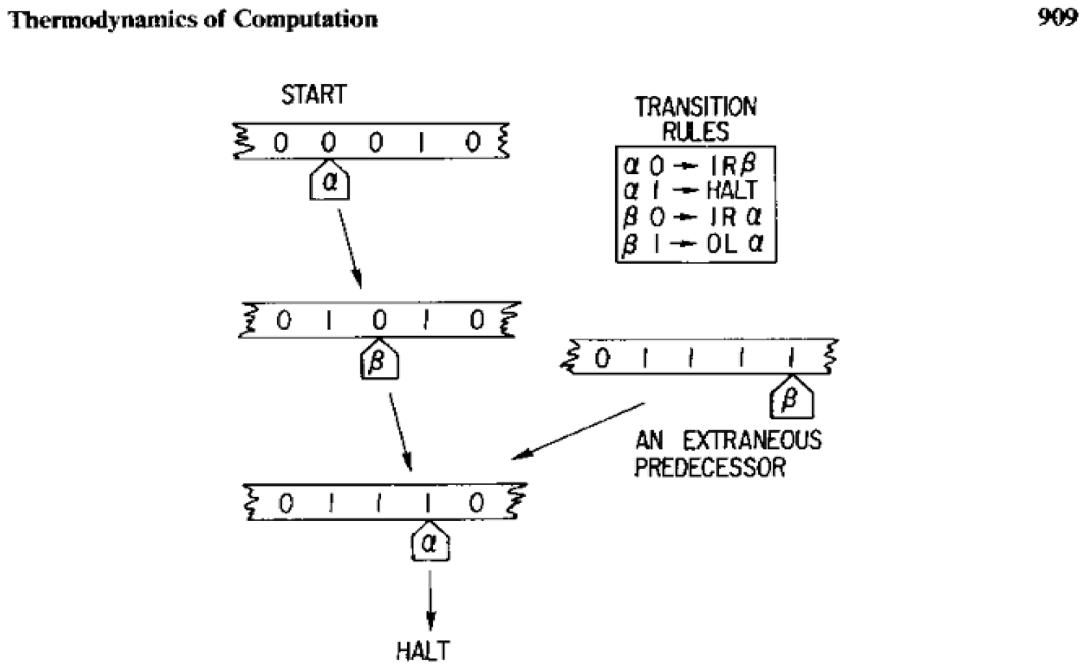

Figure 3.8. Merging of the flow of control instructions constitutes a logical irreversibility (topological structure entails irreversible behavior). Adapted from "The thermodynamics of computation—a review." By Bennett, Charles H. (1982).

The very same mechanism is presented by Shenker in her figure 5 in a more "engineering" setting.

---

136 Bennett, Charles H. "The thermodynamics of computation—a review." *International Journal of Theoretical Physics* 21.12 (1982): 905-940.

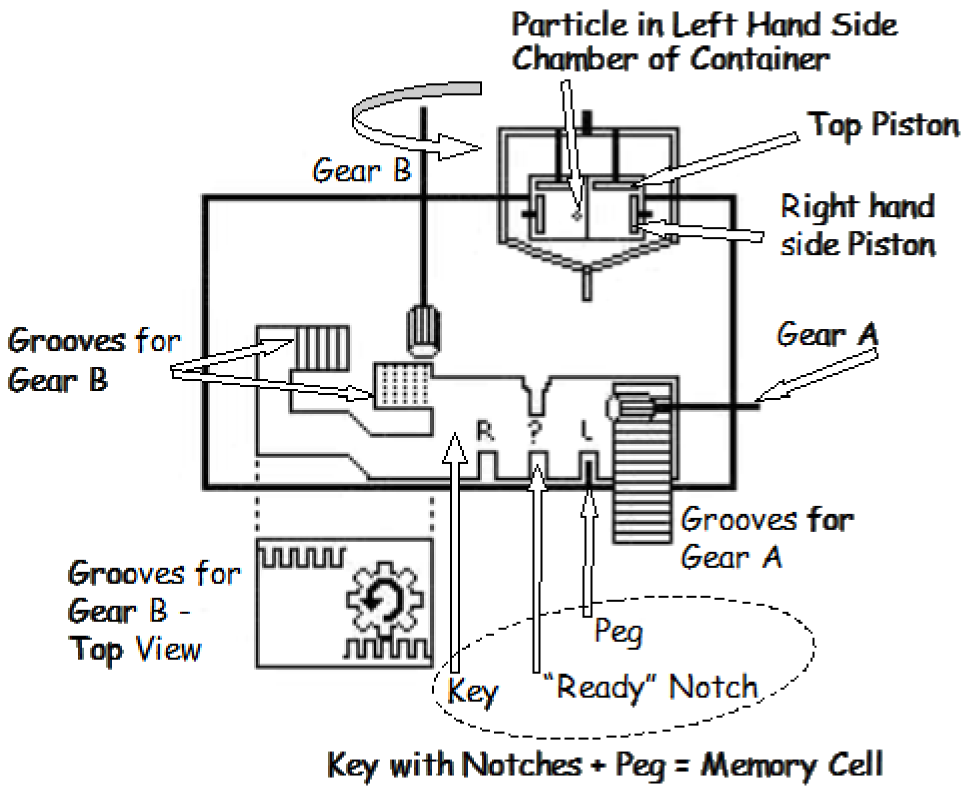

Figure 3.9. Final Turing machine configuration. Adapted from "Logic and entropy." by Shenker, Orly. (2000).

There we have rotation which is not thermodynamically reversible. Although it might not be obvious, there are two degrees of freedom present in setup's mechanics:

1. angle by which the pinion is rotating;

2. horizontal shift of key with its two attached half-racks.

At stage 1, information bearing coordinates are confirmed in one of two merging paths. In the second stage, the barrier is removed and we have a situation in which the information-carrying coordinate gains access to range which is now doubled. This process entails an irreversible entropy increase of kT ln2.[137]

The key insight is that a logically irreversible operation could be thermodynamically reversible. Bennett considers that if we apply logically irreversible operation to random data, and if bit has same probability of being distributed between two states, then the process will be thermodynamically reversible; in other words, it will decrease the data entropy. This entropy decrease is compensated with at least same amount of the entropy increase, however. This is exactly what is happening when we apply Landauer's principle to the analysis of Szilard's engine: the bits that have been erased are *random*. Therefore, erasure of the bits in this process is a reversible transfer of energy, according to Bennett. This entropy is transferred to the environment and it exactly compensates for the earlier entropy decrease (where entropy passed from the environment due to isothermal expansion). The total amount of work is zero and any conclusions we may draw from the analysis are coherent with the Second Law of thermodynamics.[138]

[137] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.
[138] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.

Schneider's critique of Landauer's principle introduces some new elements, in particular some that are applicable to the case of biological information. According to this critique, the energy that is spent while the biological information has been processed does not originate in the bit erasure. It is result of process that has two stages. In the first stage, energy is added, while at the latter stage it has been dissipated when organism falls into one of the stabile states.[139] Bennett argues that this does not contradict to Landauer's principle. In fact, this kind of processing of information is not common to all types of organisms. Moreover, we can interpret DNA transcription to RNA as kind of logically reversible process if we focus on removing of the reaction pyrophosphate as main part of copying process.[140]

As already mentioned, Landauer's principle could be applied in various ways to multiple types of physical systems that are claimed in the literature to exhibit reversible computation. Three interesting and well-studied types of such systems are the following:

1. **Ballistic computers**, such as Fredkin's billiard ball computer. These macroscopic systems are not able to merge the phase-space trajectories and, therefore, the only way we can program them is to do logically reversible computations, at constant velocity (of the balls) and while doing so they do not dissipate the energy.[141] These devices need to be isolated from external heat baths, which can occur only by total neglect of all dissipative forces, therefore this example is irrelevant to the problem of Maxwell's demon.

2. **Brownian machines which are externally clocked** – these devices control parameter is externally varied, in order to force the system to operate in a cycle, instead of moving toward the equilibrium state. Rest of the parameters could move randomly. Earman and Norton's realization of Szilard's engine was of this kind. Also, most of the proposals for quantum computers are of this kind. The thermodynamic cost that such a machine has to pay equals the work that external agency has done (appropriately averaged over time). Hence, they clearly do not violate the second law.

3. **Fully Brownian machines**, such as Feynman's classical ratchet-and-pawl, Bennett's enzymic computers, Brownian motors used on the nanometer-scale to control some chemical reactions in nanotechnology and even some biological enzymes such as RNA polymerase.[142] The entire controversy over universality of Landauer's principle could be translated into question whether at least some devices of this type are capable of systematic violation of the second law. We saw that Feynman's ratchet-and-pawl does not work, but there is still some controversy over the issue whether more complex and sophisticated devices of this type could indeed work.

Thermodynamic price for operation of this device is determined with the weakest spring that is able to start the movement. This kind of machine could go backward in order to explore the previous states and its logical characteristics. In this process, a merging costs less than in an externally clocked Brownian device.[143]

---

[139] Schneider, Eric D., and James J. Kay. "Life as a manifestation of the second law of thermodynamics." Math. Comp. Mod. 19.6-8 (1994): 25-48.

[140] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.

[141] Fredkin, Edward, "An Introduction to Digital Philosophy", *International Journal of Theoretical Physics* 42 (2003): 189-247.
Physics 42: 189-247.

[142] Feynman, Richard P., R. B. Leighton, and Matthew Sands. "Ratchet and pawl." *The Feynman lectures on physics* 1 (1963): 46-1; Bennett, Charles H. "The thermodynamics of computation—a review." *International Journal of Theoretical Physics* 21.12 (1982): 905-940. ; Hänggi, Peter, and Fabio Marchesoni. "Artificial Brownian motors: Controlling transport on the nanoscale." *Reviews of Modern Physics* 81.1 (2009): 387.

[143] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.

Earman and Norton argued that Landauer's principle is either unnecessary, or insufficient to exorcize the Maxwell's demon. Owing the fact that, if the demon is considered to be a thermodynamic system, it is already subject of second law, and there is no need for further arguments on relationship of entropy and information. So, it will not be able to detect molecules, moves the partition, etc. without dissipating energy and increasing the total entropy. Therefore, there is no need to save the second law of thermodynamics from it. Demon's internal degrees of freedom are "responsible" for balancing the balancing the books, as far as entropy is concerned. On the other hand, if we do not consider the demon as a part of a thermodynamic system that obeys the Second Law in the first place, then there is no supposition that could protect the Second Law from the demon (since, logically speaking, we have already condoned its violation in the setup of the thought experiment).[144]

Bennett argues that the importance of Landauer's principle lays mainly in its pedagogic purpose. It helps students to avoid a common misconception: that every act of processing the information must have intrinsic cost of kT ln2, nevertheless, whether it is irreversible or not.[145]

Landauer's principle explains that processing of information does not have thermodynamic cost that is intrinsic and cannot be reduced, although it seems that the operation of information destruction (popular "erasure" or "bit erasure") requires a cost that is enough to protect thermodynamics from Maxwell's demon. Thereby, measuring and copying are intrinsically irreversible only in case they are memorized over some previous information.[146]

3.5.4. Norton's critique and Bennet's version of Landauer's principle

Landauer's principle states that erasure of n bits necessarily requires cost of at least k ln n in thermodynamical entropy. This holds only for an erasure processes that imply phase space expansion (that are thermodynamically irreversible). The source of the entropy cost on this view is in the expansion of the phase space. From the opposite point of view, Norton argues that it has not been proved that this is a *necessary* step in the process of erasure.[147]

According to this line of criticism, arguments which tend toward establishing that law holds universally crucially depend on particular statistical interpretation of the system. In particular, they depend on formulation of a canonical ensemble holding random data. Critics such as Norton argue that this formulation of canonical ensemble is illicit. To be really successful, exorcism of demon needs to prove that all possible devices would fail to violate the second law of thermodynamics, which has obviously not been done. There are very many, perhaps an infinite number of such devices, a really clever thinker could conceive. Landauer's arguments are related to tightly to particular examples, therefore it needs general validity for general exorcism. Extended version of Landauer's principle, that Bennett has offered seems to be subject to the same criticism.[148] One of the problems is that Bennett uses Landauer's principle as a direct consequence of the Second Law.

It was already mentioned, that Bennett asserts that strength of Landauer's principle is significant because it locates the entropy cost correctly, in the process of resetting ("erasure" of information), correcting an earlier common misconception (e.g., Brillouin) that located the entropy cost in

---

[144] Ibid.
[145] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.
[146] Ibid.
[147] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.
[148] Ibid.

information *acquisition* ("perception" or "observation"). I will discuss this in somewhat more detail in Chapter 6.

Also, Norton argued that Landauer's principle properly located increase of entropy in the erasure of the memory. Kind of entropy that we have here, is thermodynamic. This principle is frequently cited in terms of heat, instead of entropy. For example: erasing of a bit of information from memory of the device has to come with entropy cost of at least kT ln2 of the heat transferred to environment at temperature T.[149]

The familiar slogan: that erasure of one bit of information will increase the entropy of the environment[150] by at least kT ln2, captures the basic justification of the principle: performing the erasure will either reduce the number of states, or compress the volume of the occupied phase space, or (as Bennett would claim) it effects a many to one mapping.[151] Other locutions are sometimes used, as mentioned above, like the merging of trajectories (in phase space), etc.

Shizume[152] and Piechocinska[153] provided some proofs of Landauer's principle. Still, none of them seems to satisfy conditions Norton considers necessary for establishing a truly general proof. These proofs provide analysis of the source of the entropy created in the erasure process, but the properties that are essential for this process and memory devices, as well as a solid reason why it must hold in every possible erasure procedure, remains unclear in all these.[154]

If we want to exorcise the demon, however, we must establish whether all possible erasure processes share that same characteristic that is essential for the principle to hold. One should know how to distinguish essential properties of the memory device from the incidental ones, as well as which physical laws could be referred to in order prove the principle. For instance, the using of dissipative processes such as friction is usually considered an effective erasure pathway in realistic physical systems, although this is unfortunate since fine-grained information is preserved in such processes (and hence cannot address the underlying philosophical issues). In other words, we are facing limitations on our physical imagination and intuition in trying to analyze *all physically conceivable* erasure pathways.

This is not necessarily just a negative claim. Norton asserts that phase-space expansion is included in the erasure process in which entropy is created. He believes that compression of phase space cannot create entropy due to its thermodynamical reversibility. The only thing that compression of phase space is actually achieving in all these cases is moving the entropy that has already been created to the environment.

Hence, Landauer's principle – Norton argues – does not prove that the entropy cost will need to be paid each and every time we erase the information. The cost will be incurred only if the first step of the erasure is present: removing of the partition. This first step is thermodynamically irreversible and it corresponds to the dominant term in the total balance sheet. The situation is often mispresented and unclear, because the removing of the partition is represented (erroneously) as a thermodynamically reversible (= adiabatic, isentropic) process. Instead, it is exactly the step which produces entropy.[155]

---

[149] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.
[150] Leff, Harvey S. and Rex, Andrew F, *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (Philadelphia: Institute of Physics Publishing). (2003): p 27.
[151] Bennett, Charles H. "The thermodynamics of computation—a review." *International Journal of Theoretical Physics* **21**.12 (1982): 905-940, p. 307.
[152] Shizume, Kousuke. "Heat generation required by information erasure." *Physical Review E* 52.4 (1995): 3495.
[153] Piechocinska, B. 2000, "Information erasure", Physical Review A, 61: 1–9.
[154] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.
[155] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

It is in part a consequence of misleading presentation of the puzzle. An additional problem perceived by Norton is that Landauer's principle did not indicate that the thermodynamic entropy is created at all through the removal of the partition, this part of the entire process being ignored in the "canonical" presentation of the principle. Norton claims that it is exactly the thermodynamical irreversibility of this step (removing of the partition), which represents the main reason why the entropy cost equal to kT ln2 is created. Without this step, there would be no entropy cost.[156]

The underlying causes of the confusion could be formulated in the language of statistical mechanics. We need to ask ourselves: How would this process look statistically, in terms of its canonical distribution? If we have many subsystems ("elements") that form a canonical ensemble, then canonical distribution will be the distribution of energy that is most probable and the energy function E(x) will express the energy of component at point x (represented, in general, by a 6N-dimensional vector in the phase space of the system). Norton explains that insofar we wish to generate the canonical distribution from an ensemble for one component, as in Szilard's model, we must acknowledge that, when phase space is sampled, its energy function E(x) has to stay the same.[157] To generate canonical distribution in a different manner would result in forming an illicit ensemble, violating the law of energy conservation. It would be as we have neglected conservation laws on microscopic level at some points in our sampling and accounted for them at other points. Such approximation may appear in discussions of some physical problems, but according to Norton's overall view, they cannot be applied here, since it is exactly the consistency of microphysical description which is at stake in the problem of Maxwell's demon.

Finally, Norton notices that even if we have an ensemble of canonically distributed (sub)systems, we can treat them as clones of a single (sub)system only if the energy function E(x) is the same for every individual member of the ensemble. Otherwise, to consider them clones of this system would be erroneous.[158] On one side, this does sound intuitive and rational: after all, many bank accounts might have the same interest rate and the same procedures for deposits and withdrawals, but they are certainly not the same, since some people have millions, while others keep like $50 in their accounts. And even if we find people with the same total balance of deposits minus withdrawals, their accounts would not be the same in any particular moment of time. We would be neglecting the entire evolutionary trajectory which led to the observed state. This error is common in the literature on Landauer's principle.[159]

---

156 Ibid., p. 388.
157 Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.
158 Ibid.
159 Rex, Andrew and Larsen, Ross , "Entropy and Information for an Automated Maxwell's Demon," Workshop on Physics and Computation, pp. 93-101. URL = <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=615503&isnumber=13433>, (1992).

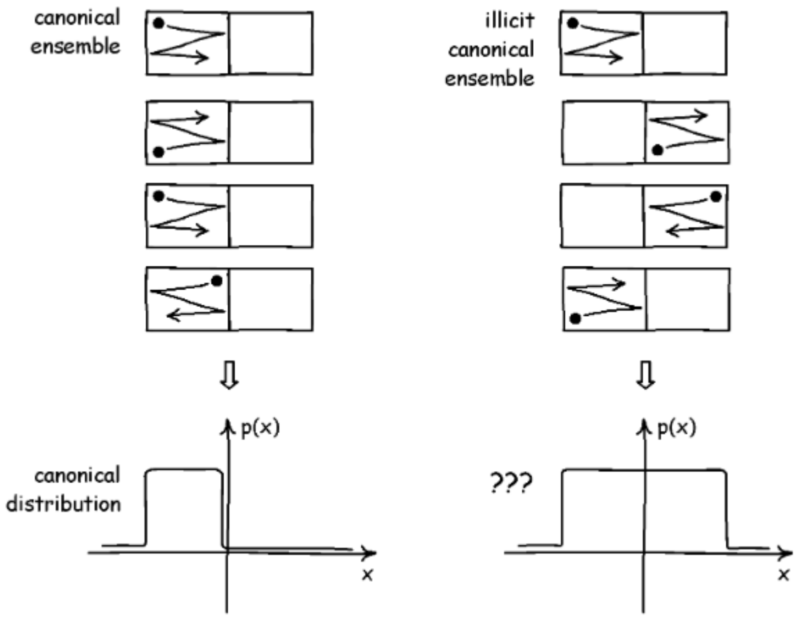

Figure 3. Licit and illicit canonical ensembles and the distributions they produce

Figure 3.10. Licit and illicit ensembles and their distributions. Adapted from "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." by Norton, John D. (2005).

Leff and Rex argue that although the process of erasure is logically irreversible in general, there is a single particular case where it is not. It is the case in which exactly half of the members of the ensemble under consideration are in state L, and the other is in state R. They claim that we can conclude this from the fact that removing of the partition leaves system in the initial thermodynamical state. Subsequently, Leff and Rex analyze a reversed process. According to them, any subsequent inserting of the partition has no entropic effect, for it is same probability that half of the ensemble will be in one of these two states, which counts in the end.[160]

In other words, they consider that entropy of those cells where the partition is removed is equal to those of set of cells with random data. This is incorrect. Norton argues that we cannot consider thermodynamical entropy as a property of the set of cells that we classified by some criteria; instead, it has to be (in at least some viable sense) a property of the individual cell. Thermodynamic entropy must be an attribute of the cell and its physical state; it does not depend on its relations toward other cells. Norton claims that Leff and Rex treated the collection of cells were carrying random information as a canonical ensemble. A set of cells with just random information does not represent a canonical ensemble in the established sense of statistical mechanics. (However, this semantic argument is quite weak, especially since the term "ensemble" has so many different application within physical science.) Probability distributions for each individual cell cannot be taken as representative for the entire phase space and one cannot expect to produce a distribution with properties that all of them will have.[161]

---

[160] Leff, Harvey S. & Rex, Andrew F. 2003, *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, Philadelphia: Institute of Physics Publishing, (2003): p. 21.

[161] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP **36**.2 (2005): 375-411.

Some philosophers were suggesting to Norton that the argument of Leff and Rex hinges upon another probability distribution that simulates entropic properties of canonical distribution. Anyway, the deeper issue there is that Leff and Rex use wrong sense of entropy.[162]

Confusion of the two kinds of entropy appears since the values of

(i) the entropy of cells that carry random data, and

(ii) entropy of the resetting of all the data in cells,

are the same. It is wrong to reduce entropy that appears to the information entropy change, because this way, we are mixing two kinds of entropy. The correct meaning of entropy at hand in exorcisms and potential exorcisms is the *thermodynamic* entropy. Although erasure process can be considered as compression of a phase space, it is not one of kind that would reduce thermodynamic entropy, because it will not reduce volume of phase space for a system that is canonically distributed. Leff and Rex have analyzed not the thermodynamic entropy in strict sense; instead, their discussion is ambiguous in the sense that it includes both thermodynamic entropy and information theoretic entropy $S_{info.}$ This dubbed "augmented entropy" by Norton. As he explicates: "[T]he thermodynamically irreversible process of the removal of the partition turns out to be a constant augmented entropy process."[163] Arguably, this is the core of the problem with all exorcisms that does not take the initial step into account (removing of the partition).

When the demon initial removes the partition, it is a process with constant augmented entropy ("augmented adiabatic"). Still, it is an irreversible process in the thermodynamical sense, because it creates ("mere") *thermodynamic* entropy. During this process, the entropy is passed to the environment as a heat. Bennett is allegedly making a mistake of similar kind, his understanding of the canonical ensemble also being illicit.[164] Notably, Bennett analyzes erasure of a bi-stable ferromagnet, where a piece of ferromagnetic material is in one of the two states at the beginning. The ferromagnet resets when we change the external field (while keeping the other parameters constant). If we do not know initial state and possible to describe with a probability that is equally distributed between the two minima, then he considers the erasure as thermodynamically reversible. In a case that we know the initial state, then it is both logically and thermodynamically irreversible.

As we have already mentioned, Bennett later[165] redefined concept of thermodynamic reversibility. He argues that: " [if] the logically irreversible operation is applied to known data, the operation is thermodynamically irreversible because the environmental entropy increase is not compensated by any decrease of entropy of the data."[166]

His idea was that thermalization and randomness can be considered as equal. Thus, a collection of cells (no further physical specifications) with random data is considered a canonical ensemble, contrary to Norton's interpretation. The problem with this argument is that presupposition that whether we know data or not, it has impact on its additivity which leads us to problematic consequences. For example, if person A knows certain data, while person B does not, so the data are random for person B, but not for person A. From this, we must conclude that the thermodynamic entropy for memory consisting of N devices is NkT ln2 less for A than for B.[167] This is both intuitively implausible and leads into all kinds of strange, even mystical consequences. Furthermore,

---

[162] Ibid.
[163] Ibid., p. 393.
[164] Bennett, Charles H. "The thermodynamics of computation—a review." *International Journal of Theoretical Physics* 21.12 (1982): 905-940., p 311-312
[165] Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510. p. 502.
[166] Ibid, p. 506.
[167] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

when A tells B where some particular sequence is located, it will allegedly drop for B as well! While this is not paradoxical in itself, it runs counter to the entire conventional physical wisdom from the times of Boltzmann (if not Galileo) to this day. What one observer has or has not said to another should not influence physical properties of the system under study.[168]

The usual view among philosophers of physics is that the domain of the Landauer's principle is much wider than what is considered by the critics. We noticed that supporters, such as Bennett, corroborated this argument with the claim that entropy that generates from the erasure process arises from the mapping of many physical states to one. In many applications, e.g., in physics of computing, in cosmology, even in bioinformatics, the validity of Landauer's principle is usually just (often tacitly) *assumed*. So, independently of particular resolution of this controversy, this circumstance clearly demonstrates the *relevance* of philosophy of science for resolving important explanatory and heuristic issues in science itself.

When we analyzed Bennett's interpretation of many-to-one mapping argument, we saw that his argument was based on the assumption that state of the memory cells before the erasure occupies more phase space (has larger phase volume) than after the erasure. Norton argues against this claim. It would be a mistake to hold that compression of the phase space and the reduction of a canonical ensemble are the same.[169] This is the main reason he rejects many to one mapping argument.

But does erasure have to be thermodynamically irreversible always and everywhere? This is the case in the examples that we are familiar with. We do not have general and definite proof that this holds for all the erasure processes. If we want to conclude that Landauer's principle is sufficient to exorcise all kinds of Maxwell's demons, we need to provide a general proof that will hold for all the processes.

Distinguished South African physicist and philosopher of physics Jeffrey Bub claimed that any such principle should operate independently from the state that has been erased.[170] Still, it has not yet been proven how this particular requirement makes all the possible erasure processes thermodynamically irreversible. It is clear why in erasure processes we are familiar with, there is step that assures that these processes are thermodynamically irreversible. Still, it is not clear how can we generalize and expand these conclusions which stand for the standard erasure procedure to all the possible erasure procedures and pathways? How to be sure that some future Smoluchowski or Feynman will not construct much more complicated molecular engine in which the erasure process will not be subject to the standard constraints?

Therefore, the imperative is to ask: what assures us that all erasers must be of the standard type? An eraser can possibly operate even if it does not record states in a memory device. Even if it does record, it is not clear why cannot it use the state that is under the pressure for tracking of the procedure that has been followed?[171] Note that the idea of recording the states is the exact *opposite* of what we consider to be an erasure in the vernacular sense of the word. This is an important instance showing how much is fine-grained microscopic view of information and entropy removed from our common-sense intuitions.

Since we cannot be sure at present that all erasures are of this type, Norton argues that Landauer's principle cannot exorcise all demons.[172] If we want to successfully overcome this challenge we should take notice that these attempts of exorcising the demon presuppose these claims: processing

[168] The exception might be made for some interpretations of quantum mechanics which are often ascribed this "subjectivist" property. Without going into this deep philosophical morass here, it is enough to note that such situation has never arisen in classical physics, including classical statistical mechanics.
[169] Ibid.
[170] Bub, Jeffrey. "Maxwell's Demon and the Thermodynamics of Computation." SHPMP 32.4 (2001): 569-579, p. 573.
[171] Leff, Harvey S. and Rex, Andrew, F*, Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* Philadelphia: Institute of Physics Publishing, (2003): p. 21.
[172] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

of information must be done computationally, information needs to be stored in memory device, erasure is necessary part of the operation and it needs to be done in standard manner (removing of the partition and the compression of phase space). For Landauer's principle to universally hold all of the above conditions should be met. Therefore, there are many ways on which it can fail.

Svedberg[173] proposed a device whose goal was to violate the Second Law by going around one of these conditions. This engine neither acquires nor works on (processes) information. It is composed from colloid in which we have particles and lead casing that surrounds the colloid. The particles emit their thermal energy, this thermal energy is increased by their thermal motions. The lead casing absorbs the radiation that particles produce and heats up, while causing the colloid to cool down. This is an alleged example of direct violation of the Second Law.

In this case, we do not need the erasure of information, since there is no information that has been acquired or processed. Landauer's principle does not put any constraints on this device, since there is no sense in which it can be relevant for it, since there is no information processing nor erasure. At least there is no evident *macroscopic* sense; since colloids are actually quite complex systems on the microscopic level, one can hardly claim that the violation there is *obvious* or self-evident. After all, components in a colloid could have different chemical potentials, and some of the known Brownian motors are constructed to exert useful work against the differences in chemical potential.

Beside that there are demons – or demon-like setups – that do not perform any computation or information processing. These are the demons that do not observe or compute, but still operate the system according to a well-defined procedure nevertheless. An example in this category would be Smoluchowski's trapdoor. Even if we in practice construe a sort of Smoluchowski's trapdoor as some kind of computer which will have memory, there would not be any two-step erasure process. Hence, Landauer's principle would not seemingly be relevant, or at least it would require a detailed theoretical model to demonstrate its relevance.[174]

Finally, we can have computational demons that are not standard in other respects. Since obviously, *some* Maxwell's demons are computational devices, Norton presupposes that they have the standard architecture. By the standard architecture, he means a central processing unit and a memory device, which has finite capacity, and therefore must be erased from time to time – the concept of a computer due to pioneers such as Turing and von Neumann. How can we know that device, without distinct memory which require erasure as a step in the operation, cannot exist? Besides, we must remember that device does not have to be a universal Turing machine (UTM). In contrast to UTM, the demon has a *special purpose*, it has only one operation.[175]

Going further afield, another problem is that there might be devices which have non-standard erasure protocols. Is the two-step (removing of the partition and phase space compression) erasure process the only possible erasure process? As we have seen, the critics have already raised the possibility of alternative erasure procedures that could, at least in principle, be thermodynamically reversible.

Also, we can impose the question: must entropy cost always be the same as the entropy gain? In the case of Szilard's single molecule gas engine we see the entropy which has been reduced in its operation is equal to the entropy that will be induced by the erasure of demon's memory. However, this is just one case, albeit historically the most influential one. We do not have any special insurance

---

[173] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471, pp. 443-444.
[174] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.
[175] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

that entropy reduction in the device will be at least equal, if not lower, than entropy increase of the erasure procedure?[176]

And what if we can have demons that need not to perform erasure? Even in the case we presuppose that Maxwell's demon can be *only* a computational device and that his design could be only standard (includes the standard two-step erasure process), it will still be possible that this demon need not necessarily to perform this kind of erasure operation. We have example of no-erasure demon in Szilard's single molecule gas engine. Today, it is mainly held that the detection of a molecule could be thermodynamically reversible process, which is historically opposite to the considerations in 1950s and even in Brillouin's and Feynman's time.[177]

Memory device that has two states, our conventional L and R, will record location of the molecule and which program is to be run. In the beginning, the device is in the initial state L. When the device measures position of the molecule, it switches to the program L or R, depending on the side where is the molecule. Suppose that the first program leaves memory device unaltered, and second just switch it from L to R, so there is no erasure. It has been assumed that the memory device could be switched from L to R or otherwise, without paying in thermodynamic entropy, because the process of switching is thermodynamically reversible.[178]

Now, let us go back to the Bennett's version of Landauer's principle. We saw that Bennett supported his version of principle with many to one mapping argument. We may ask again the same foundational question: what is the *exact* content of Landauer's principle? Where is its general proof? What constitutes a computational path? We have seen that many-to-one argument is invalid when it comes to the thermodynamical price paid in the memory device. Could it be valid if it locates thermodynamical cost in merging of computational paths instead?

It seems that it would be invalid there, as well. In the erasure process of our memory device, we have expanding of the phase space; subsequently, we have a compression of the phase space. It might seem that computers, when computational paths merge, do not have a particular state that corresponds to it. Consequently, these two processes are different.

Let us check on a further example. Suppose we have no-erasure demon that implies two programs on one memory device. A molecule represents our memory device. Chamber has two parts, L and R. A molecule could be trapped in one of this two parts via two pistons. Function of the program L is to keep the molecule in the region L. Function of the program R is to read R. While doing it, the molecule will move to R. This process will be thermodynamically reversible and the phase-space volume accessible will stay constant.[179]

Since the process that switches the system to L from R is the same constant-volume process, it is thermodynamically reversible. Volume of the phase space remains unchanged during this process.

So, the thermodynamic entropy cost is not incurred. We should notice that these processes are fundamentally different from the expansion and compression of phase space of the system; energy is not transferred from the device to the environment in the thermodynamically reversible manner.[180]

---

[176] Ibid.
[177] Ibid.
[178] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.
[179] Ibid.
[180] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

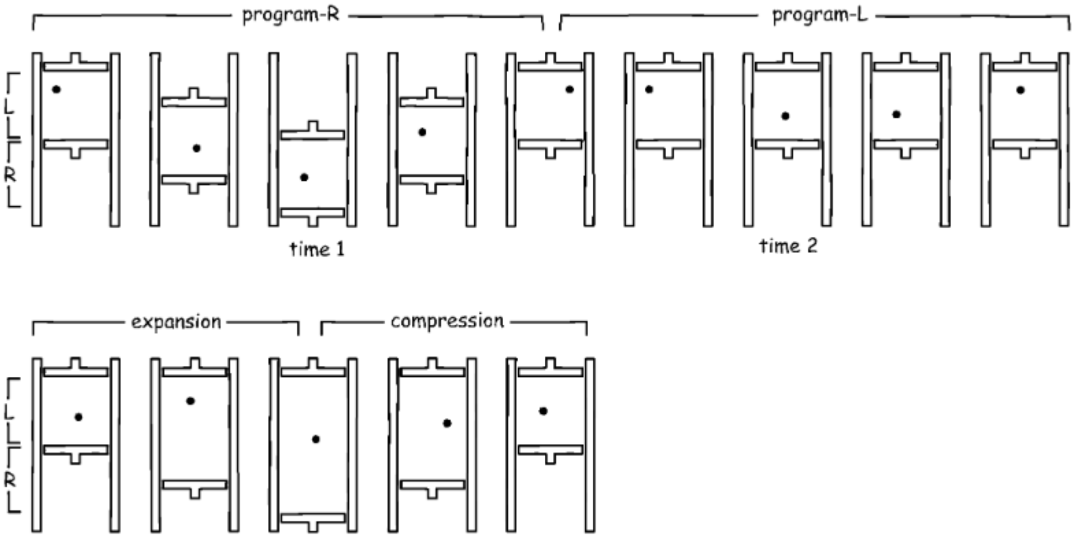

Figure 8 Time development of the no erasure demon's memory device and a different device that expands and contracts the phase space

Figure 3.11. Norton's depiction of time evolution of the no erasure demon's memory device and a different device that expands and contracts the phase space. Adapted from "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." by Norton, John D. (2005).

Clearly, Bennett understands the no-erasure demon in a different manner. He takes it as a reversible process with two parts, where the phase space is compressed and thermodynamical entropy of kT ln2 is passed from the device to the environment.

Hereby, we have two states, time 1 and 2. If we combine it to make a phase space volume double and consider it as a canonical ensemble, that would make this combination illicit, as per Norton's argument discussed above. Energy E (x) is different for times 1 and 2, one of which is finite just in the L part, and the other just in the R part.[181] Hereby, it looks as if we have a phase-space compression that is not in the same time a compression of the canonical ensemble. So, we cannot apply formulae like:

$$S = \frac{\overline{E}}{T} + k \ln \int_{\Gamma} \exp(-E/kT)dx \qquad (3.1.)$$

for the entropy of a canonical distribution, since the integral is not necessarily well-defined. Consequently, we cannot infer that its thermodynamic entropy is kTln2 greater than the default L state.

Norton describes this situation as follows: "If both expansion and compression were effected as thermodynamically reversible processes, any thermodynamic entropy passed to the environment by the compression would be balanced exactly by entropy drawn from the environment in the expansion phase."[182]

Therefore, this reasoning cannot provide strong proof that there is thermodynamic entropy cost arising from the merge of the computational paths in the example of demon that does not perform the erasure. Furthermore, this cannot provide grounds to accept neither Landauer's principle, nor

[181] Ibid.
[182] Ibid., p. 408.

Bennett's extended version of it.[183] We seems to be back where we started with the original criticism of Earman and Norton.

All in all, fundamental weakness of exorcising Maxwell's demon with Landauer's principle is on methodological level. There is no general proof, there are just many examples and thought experiments and illustrations.

Does erasure of one bit of information necessarily cause generation of thermodynamic entropy? This question remains open. It is also not clear if all the arguments against possibility of the existence of Maxwell's demons could be collected exorcise all the Maxwell's demons.[184] Exorcism of Maxwell's demon, in order to be successful, requires more general proof.

3.5.5. An indirect proof of Landauer's principle and its problems

We saw that the problems with Landauer's principle are mainly incorrect assumptions that compression of the phase space is an inevitable part of erasure process or that uncertainty of random data causes the rise of thermodynamic entropy. Now, let us focus on the indirect proof of Landauer's principle, which will not rely on assumptions such as these.

Ladyman, Presnell, Short, and Groisman[185] (henceforth LPSG) made further advance in this area. LPSG attempt to provide more proof for more general version of Landauer's principle. They seek to demonstrate that information entropy that is usually placed in the step of mixing of macrostates have thermodynamic significance.[186]

Earlier attempts to prove Landauer's principle used direct approach, through direct examining of erasure process. LPSG applied a different, indirect strategy. This is direct examining of the erasure process. LPSG try to determine the entropy cost of erasure indirectly. Their strategy is to relate the erasure process with a process that *surely* causes thermodynamical entropy reduction. They presuppose a general statistical form of the Second Law.[187]

After they assume the Second Law, they attempt to give proof for that Landauer's Principle. They choose the 'sound' rather than the 'profound' horn of the dilemma, identified by Earman and Norton.[188] LPSG tend to establish relationship between logical and thermodynamic irreversibility by generalization of Landauer's Principle.[189]

Every logically reversible operation can be applied in both logically reversible and irreversible manner, but the crucial question is: are there any logically irreversible operations that could be applied in logically reversible manner?

---

[183] Ibid.

[184] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

[185] Ladyman, James, et al. "The connection between logical and thermodynamic irreversibility." SHPMP 38.1 (2007): 58-79; Ladyman, James, Stuart Presnell, and Anthony J. Short. "The use of the information-theoretic entropy in thermodynamics." SHPMP 39.2 (2008): 315-324.

[186] Norton, John D. "Waiting for Landauer." SHPMP 42.3 (2011): 184-198.

[187] Ibid.

[188] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." *SHPMP* 30.1 (1999): 1-40; "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471

[189] Ladyman, James, et al. "The connection between logical and thermodynamic irreversibility." SHPMP 38.1 (2007): 58-79., p. 72.

Table 3.1. The relationship between logical and thermodynamical reversibility.[190]

Table 1
A table representing the different possibilities for logical and thermodynamic reversibility

Possibilities	Thermodynamically reversible	Thermodynamically irreversible
Logically reversible	✓	✓
Logically irreversible	?	✓

Our paper addresses the issue of whether any logically irreversible transformation can be implemented thermodynamically reversibly.

They conclude that it is not possible. LPSG will form complete relationship between logical and thermodynamic irreversibility. They address Landauer's Principle in following form: "If L is logically irreversible, then every L-machine is thermodynamically irreversible."[191]

The logical operation reset represent the simplest logically irreversible transformation, LPSG analyze it for its logical irreversibility (reset operation maps logical states to another logical states). They impose a question if Landauer's principle would have been true, if we were concerned only with deterministic operations.[192] They define logically irreversible operations as equal to logically reversible ones with addition of one or more resetting operations. In other words, every irreversible operation incorporates reset.[193]

Besides, LPSG also analyze relationship of known and unknown data and computation process, but they use it in other sense. They assume that device is the one that "knows" or "do not know" the data. They conclude that since the memory device is not changed in the case of known data, we cannot have logically irreversible operations run on them.[194]

Arguments that were offered earlier were often limited to examples of L-machines. LPSG will turn to generalizing without rigorous proof. On the other hand, Norton and other critics are concerned about the conjunction of information entropy with the thermodynamic result, especially in cases where part of the process (like the removal of the partition in Szilard's model) is ignored.

LPSG defined the (information) entropy with the equation below:[195]

$$S = -k \sum_{n} \lambda_n \mathrm{In}\, \lambda_n \tag{3.2.}$$

It is equal to the thermodynamic entropy in the case where the system has a canonical probability distribution. LPSG presuppose that what the second law of thermodynamics state for entropy, also

[190]Ibid., p. 60.
[191]Ladyman, James, et al. "The connection between logical and thermodynamic irreversibility." SHPMP 38.1 (2007): 58-79.
[192] Ibid.
[193] Ibid.
[194]Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.
[195] Ladyman, James, et al. "The connection between logical and thermodynamic irreversibility." SHPMP 38.1 (2007): 58-79.

holds for information entropy. Norton[196] argues that we cannot attribute entropy to probabilistic mixtures, as long as the phase space does not comprise all microstates. LPSG accepted that this is reasonable objection. They look toward extending thermodynamic to probabilistic mixtures of macrostates.

LPSG define L-machine as thermodynamically irreversible. They show that the process px is thermodynamically irreversible if x could not be determined on the basis of L(x). It seems that LPSG proved the generalized qualitative form of Landauer's principle. They state that all the L engines would possess, if L is logically the trait of thermodynamical irreversibility that if L is logically irreversible.[197] In the former proof, they rely on numerous assumptions about idealized physical systems (commonly accepted). These assumptions are such as: that molecule is treated as ideal gas; that it is possible to move the partition frictionlessly and that the Second Law of thermodynamics is statistically true.[198]

Problematic point of their proof is that they assume the source of the entropy cost of erasure. However, LPSG can only conclude entropy cost if these suppositions hold.[199] Since LPSG presuppose that violation of the Second Law holds, and consider only cases that do not permit its violation, it turns out their proof cannot exorcize all of the Maxwell's demons. They do not try to base the law on the underlying physical traits of the system that should be analyzed, but presuppose that physical traits of the system can subsidize statistical form of the law. This presumption is mistaken.[200] The step in which they generalize the concept of reversibility is not compatible with the thermodynamic entropy of the state.[201]

LPSG aim is to give proof of generalized form of Landauer's principle. In other words, they tend to show that each and every physical implementation of logically irreversible process is thermodynamically irreversible. Their proof is based upon a model which has two key elements: a memory device M and a single molecule G.[202] G is single molecule of the gas that is used to randomize the state.

The proof has 4 steps. In the first step, a partition is inserted in the middle of the chamber in which is molecule G. Chances that molecule is trapped on either of the sides of the chamber are equal. In the second step, memory device M is set in one of the two states (L or R), after the measuring and locating on which side of the chamber the molecule G is. Nevertheless, of the states in which device M is set, entropy will be the same. Therefore, no heat will be passed to the environment.

In the step three, reversible isothermal expansion has been performed. G has come back to initial state. In step four, the erasure is performed. Memory device has returned to the initial state. In the end of their analysis, LPSG give the statistical version of the Second Law of thermodynamics in form of a postulate. Therefore, after a rather grandiose detour, we are back to what is essentially Smoluchowski's modified Second Law. The main result of LSPG is that the step 4 produces the thermodynamic entropy. LPSG clearly states that the proof is based on the statistical form of the Second Law.[203]

---

[196] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.
[197] Ibid.
[198] Ladyman, James, et al. "The connection between logical and thermodynamic irreversibility." SHPMP 38.1 (2007): 58-79.
[199] Norton, John D. "Waiting for Landauer." SHPMP 42.3 (2011): 184-198.
[200] Ibid.
[201] Ibid.
[202] Ibid.
[203] Ladyman, James, et al. "The connection between logical and thermodynamic irreversibility." SHPMP 38.1 (2007): 58-79, p. 59.

However, the fact that this proof depend on the statistical form of the Second Law is the reason for its failure.

3.6. Understanding of entropy

The concept of entropy has always been highly contentious. In 19th century, in the age of Clausius and Kelvin, entropy has been understood as an ever-increasing measure of disorder – and it was identified with arrow of time. There are two more dominant understandings of entropy:

1. entropy as a loss of information;

2. entropy as a function of phase-space volume.

The latter, following Boltzmann and Gibbs, prompts a question if operations in physical systems are reversible or not. To answer it, we will need help of Loschmidt's demon. This demon – introduced in the famous Loschmidt's objection to Boltzmann's statistical interpretation of thermodynamics – makes trajectories of motions within a mechanical system reversible. While doing this, demon could not turn back time itself, only reverses motions of particles in a selected small part of the world. Lesson we could learn from Loschmidt's thought experiment is that if trajectories of systems are reversible, we could claim that world is indeterministic. The present state of affairs could have alternative futures.[204]

This analysis will, in the further course of this dissertation, lead us to several consequences. First, we will need introduction to the *conditional* notion of causality. The notion of causality becomes less deterministic and more probabilistic in the process. The notion of entropy has been used for the determination of direction of causality, the distinction between past and future and the time's arrow. On the top of that, Maxwell's demon shed a light to the statistical notions.[205]

We will now, briefly analyze entropy as a measure of disorder, entropy according to Kelvin, Carnot, and Clausius and entropy in terms of phase-space volumes, while I will leave the analysis of entropy as loss of information for Chapter 7 where I will analyze relationship between entropy and information in some further detail.

3.6.1. Entropy as disorder

The notions of order and disorder seem to be subjective and are rarely explicitly defined. What is content of these notions? We could imagine we have cube with black and white molecules. We can put molecules on many ways so that black and white ones would be on opposite sides. This number of ways, on which we can organize system on the inside (internal degrees of freedom) while it does not change outside appearance, is what we understand by physical notion of disorder. This immediately makes clear at least the difficulties we have in talking about the entropy of the universe or heat death, and related topics, since it is obviously impossible to look at the universe from the "outside". The logarithm of the combinatorial number of ways in which the internal constituents of the system could be arranged is entropy, as per Boltzmann's famous formula.[206]

---

[204] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[205] Ibid.
[206] Feynman, Richard P., R. B. Leighton, and Matthew Sands. "Ratchet and pawl." *The Feynman lectures on physics* 1 (1963): 46-1.

We can understand entropy in these terms and consider it as a measure of disorder. As entropy always increases, entropy of the universe increases, as well. Therefore, the universe goes from order to disorder. However, in contrast to the 19th century views, this does not *necessarily* mean that disorder in the universe is approaching its maximum – since we now know that additional degrees of freedom can arise, for instance through expansion of the universe.

Where does the reversibility of the other physical laws came from, then? All the fundamental laws of physics (for example, Newton's laws of motion or Maxwell's equations of electrodynamics) are reversible. The question is: what is the origin of the irreversibility of the Second Law? We can say that it comes from order that is decaying into disorder. This begs the question, however, namely how can we know this if we do not understand the origin of the order? Why are all everyday situations out of equilibrium?[207] Theoretically it is possible that mixed molecules get organized by simple fluctuations in the same order again. One theory is that present order in our world is just question of luck. According to this theory, irreversibility will be just accidental event (one of the "lucky accidents" of life). This goes back to Boltzmann, his debate with Zermelo in 1895 on irreversibility, and his bold cosmological proposal that the origin of the thermodynamical irreversibility or the entropy gradient is a huge fluctuation which occurred in otherwise dead universe of maximal entropy.[208]

Gradient of what, exactly? The concept of entropy is anything but unambiguous. There are many approaches to thermodynamics.[209] We can find the traditional approach in the work of Carnot, Kelvin, and Clausius which we need before giving the final assessment of this Chapter.

3.6.2. Carnot, Kelvin and Clausius on entropy

The only way that closed thermodynamic system is correlated with the world is through the exchange of work and heat. Work that is performed can, for instance, lower the weight via gravitational potential, as in Feynman's sketch of his ratchet contraption. Work that is extracted could be used to lift weight. Heat is exchanged between the heat baths and the system. There can be many baths, and their temperatures can be different. In a cycle that is closed, we have operations that are implied; in the end, all these operations leave the system in the same thermodynamical state as in the beginning of the cycle. Nevertheless, a weight in the gravitational potential can change its position. Experiments have showed that in every closed cycle of production of heat, Q_i, in heat bath at temperature T_i, (requiring the *total* work $W=\sum_i Q_i$, in accordance with the First Law of thermodynamics), the Clausius's inequality:

$$\sum_i \frac{Q_i}{T_i} \geq 0 \qquad (3.3.)$$

Kelvin's formulation of the second law of thermodynamics claims that it is not possible to accomplish a cyclic process where the net result is only that heat is extracted from the heat bath, and some mechanical work is performed (e.g., a weight is lifted as in the examples above).

Clausius' formulation is similar to Kelvin's. He claims that the only way to perform a cyclic process is to get as a result heat transferred from a lower-temperature heat bath, to the one with higher temperature. This perfectly corresponds with our everyday experience and our intuitive

[207] Ibid.

[208] Steckline, Vincent. S. 1983, "Zermelo, Boltzmann, and the recurrence paradox," *Am. J. Phys.* **51**, 894-897.

[209] Uffink, Jos,. 2001, "Bluff your way in the second law of thermodynamics", *Studies in the History and Philosophy of Modern Physics* 32, (2001): 305–394.

understanding of temperature. It is often popularly phrased as "the heat goes only from a warmer to a colder body".

Kelvin and Clausius, following Carnot, argued that the inequality holds generally, for all the closed cycles.[210] This tallied well with the phenomenological insights into not only functioning of steam engines and other then-popular heat engines, but also to the everyday experiences. The thermodynamic entropy S_Θ, is:

$$S_\Theta(A) - S_\Theta(B) = \sum_i \frac{q_i}{T_i} \qquad (3.4.)$$

For any other process,

$$\sum_i \frac{Q_i}{T_i} \geq \sum_i \frac{q_i}{T_i} \qquad (3.5.)$$

Thermodynamic entropy, by this definition, is dependent on cyclic processes which could reach equality in these relationships, if they are performed under the idealized conditions (e.g., perfectly isolated systems, etc.). These processes are named reversible processes, for obvious reasons.

This kind of processes consists of those in which all variables that are related to states of the system undergo only small changes. This change can go either direction. Usually, these are processes which occur very slowly, so that heat has enough time to be homogeneously distributed throughout the system.

The Clausius inequality showed above limits all of quasistatic reversible paths from A to B to the same value. A quasistatic reversible path is an idealization that can be reached only in the limit extremely slow processes (the rate of processes goes to zero).[211] However, this idealization does not look *more* harmless than those we have already encountered above in discussing various demon-like contraptions like the ratchet-and-pawl and others; note that *any* diagram plotting changes of temperature in, say water in a kettle is similar quasistatic approximation, since the systems with continuously changing temperature cannot be in equilibrium and therefore cannot satisfy the requirements of the Zeroth Law of thermodynamics to have strictly well-defined temperature in the first place. We should notice that only if the Clausius inequality holds, we can define thermodynamic entropy as a consistent single valued function of a thermodynamic state. If a demon exists in any of its forms, it will open the possibility that Clausius inequality will not be valid (such as in the case of Szilard's engine).

[210] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[211] Ibid.

3.6.3. Entropy as a function of phase-space volumes

A better and more adaptable way of understanding entropy as a function of phase space volumes, originating with Gibbs: entropy of the system rise as the micro-states of a spread into the phase space that is available.[212] Such spreading will very probably continue until the system enters the state of equilibrium. This perspective on entropy will be a statistical mechanical generalization of the age-old procedure of the counting of states or degrees of freedom. To be able to use the apparatus of statistical mechanics, we should analyze a microscopic space of states and its evolution. In such an analysis, it is usually taken to be the phase space, in which N-body system possesses 3N configuration (position degrees of freedom) and same number of momentum (or velocity, which is equivalent in classical physics) momentum once, giving 6N phase variable all together. In this phase space a single point is corresponding to the physical state of the *N* components (or subsystems), presuming that the bodies are particle-like, i.e., have no internal degrees of freedom. Of course, real molecules do have internal degrees of freedom, for instance vibrational or rotational, but this complication is irrelevant for our present purposes. We can easily imagine the system in question being a monoatomic gas, like helium.

Liouville's Theorem states that, whenever a system evolves through Hamiltonian evolution (essentially by following the established conservation laws of energy and momentum), the set of states that occupies phase space does not change. Penrose stresses that theorem does not prescribe what shape phase volume should have. It can be generalized. We could imagine it, as, for example, volume of a phase tube. It can travel through the phase space and evolve in time. Usually, phase volumes evolve in great extent, so they end up reaching larger part of the phase space.[213]

Boltzmann entropy is defined as $S_B = k_b \ln W$, where W is volume of the region of state space in which is the microstate (because S_B is a property of a microstate). It is important to note that all the microstates that belong to a given region share the same Boltzmann entropy.[214] In that *a posteriori* sense we could talk about entropy as a property of *macrostates* instead. We can explain entropy with its corresponding to the "unconstrained state-of-affairs" – the less constrained it is, the more accessible microscopic states for values that are given for the constraints that exist.[215]

Here, W is understood as thermodynamic probability (other labels are used in the literature, somewhat but not entirely interchangeably, like "weight", "statistical weight", even "likelihood", etc.). It describes the various ways on which macroscopic state could be realized in terms of its underlying microstates.[216] We can understand this with analogy to the spreading of perfume from the bottle. The molecules are microscopic elements and the pressure they made on the bottle is macroscopic state. Hereby, we see that there is one macroscopic state that corresponds to the numerous microstates.[217] A macro-state like temperature or pressure could consist of numerous microstates. The larger W is, the more microstates, and the greater amount of entropy will be.[218]

W is related to the ways in which we can arrange microstates of the system. Usually entropy is described as disorder, but it is more precise to understand it as an increase in W.[219] In reality, there are temporal spreading and spatial spreading. We can explain temporal spreading as the expansion

[212] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[213] Penrose, Roger. *The road to reality*, Alfred A. Knopf (2006).
[214] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[215] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[216]Ibid., p. 161.
[217] Ibid.
[218] Ibid.
[219] Leff, Harvey S. "Entropy, its language, and interpretation." *Found. Phys.* 37.12 (2007): 1744-1766.

of states in the phase space that is available by the microstates, hence the Boltzmann statistical definition of entropy.[220]

Boltzmann's statistical definition is referring to the expansion of microstates in the phase space that is available. Boltzmann's entropy is a measure of the number of micro-states in a macro-state. Molecules of the gas spread in the phase space that is available, since there are more manners to populate larger space.[221] This is, among other things, why expansion of the universe is so important when we are discussing the prospect of heat death. Terms such as "phase space diffusion" are often used in the literature to denote the same process, by analogy with the usual concept of diffusion (e.g., percolation of coffee or tea particles in water), but the analogy should not be taken too literally, since 6N-dimensional phase space is radically different from the Euclidean space of our everyday life.

We can define the division of the phase space into distinct regions which would be amenable to further analysis on many ways. The criterion going back to Boltzmann and which is commonly accepted is to group together such microstates that are impossible to distinguish macroscopically. Thus, we have "compartments" within phase space where different microscopic states correspond to the same macroscopic state. This is also known as *coarse-graining*. In addition, it is important to note that if we talk about Boltzmann entropy of a macrostate, it is implied that we consider only microstates that are within that particular macrostate. It is not guaranteed that Boltzmann entropy, SB, would not decrease. Decreasing of SB can happen due to classical reversibility (Loschmidt) and recurrence (Zermelo) posed by Boltzmann's H-theorem. This may occur theoretically, but this does not happen (often) in practice in the real world. Of course, if we limit ourselves to very particular macrostates, it is true that the system could evolve from a high to a low volume macrostate, as it happens with those fluctuations which seemingly defy the Second Law, like the Brownian motion. Still, Liouville's Theorem limits microstates from larger macrostate that can move into smaller macrostate to be a small fraction.[222]

Suppose we have microstates that are in an initial macrostate and they are evolving into one final macrostate. Then, according to the Liouville's Theorem this final microstate cannot have less entropy than initial. This process is deterministic on macroscopic level? Penrose[223] analyzed the change in Boltzmann entropies for such processes. He argued that for macroscopically indeterministic process, Boltzmann entropy can decrease. But he also shows that the corresponding decrease is limited. He argued that if we use a statistical entropy it would not decrease even for macroscopically indeterministic processes.[224]

3.6.4. Which sense of entropy is relevant for Maxwell's demon?

Different senses and meanings of entropy create much confusion in discussions not only of Maxwell's demon and related thought experiments (and some real experiments!), but also in applications to other fields of science, notably molecular biology or cosmology. For our purposes here, the strict choice is quite clear. Thermodynamic entropy is the kind of entropy that should be considered in relation to Maxwell's demon. The formula is:

$$\delta S = \frac{\delta Qrev}{T} \qquad (3.8.)$$

---

[220] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[221] Ibid.
[222] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[223] Penrose, Oliver, *Foundations of Statistical Mechanics*, Oxford: Pergamon Press, (1970)chapters V, VI.
[224] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.), paragraph 2.1.

Clausius formula defines changes in entropy in some processes, in particular in classic heat machines (e.g., Carnot cycle and other less efficient cycles which are used in internal combustion engines).

One would modify Maxwell's original scheme could be made if we add a heat engine which will convert heat from hotter to colder side and produce some work. This will cause that heat is taken from a heat bath and used to produce work. This problem is usually first perceived in connection with the Maxwell's demon and is tantamount to the construction of the perpetuum mobile of the second kind. This will represent a violation of the Thomson's formulation of the Second Law:

"It is impossible by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects."[225]

Thomson's formulation is often quoted as simple ban on all perpetuum mobile engines of the second kind. Clearly, if we supply energy from elsewhere, we could achieve the cooling, as is done in any refrigerator – but the thermodynamic cost is simply transferred via electric power network to somewhere else, say burning of fossil fuels or Solar fusion, etc. The following is a compatible formulation as well:

"Every physical or chemical process in nature takes place in a such a way as to increase the sum of the entropies of all the bodies taking part in the process."[226]

3.7. Understanding of the the second law

Microphysics in principle allows evolutions into states which have a lower thermodynamic entropy. Since it is possible for thermodynamic entropy to go down, on microscopic level, this has not been the motivation behind the exorcism of Maxwell's demon. As we have mentioned above, Smoluchowski argued that Maxwell's demon cannot violate a *modified* version of the Second Law. Smoluchowski has not proposed exorcism of Maxwell's demon. Essentially, Smoluchowski's proposal performs intervention on the level of macroscopic, phenomenological thermodynamics – as such, it differs dramatically from the subsequent "schools" of Szilard-Landauer-Bennett (which turn to *processing* of information) as well as Brillouin-Feynman's view (that *gathering* of information, or perception, is the problem).

In Smoluchowski's version, it is stated that demon cannot produce work *continuously*, on macroscopic timescales. Such a demon could be regarded as "tamed". A tame demon can make "straight" violations of the Second Law.[227] However, demon cannot make "embellished" violations. By embellished violations is meant straight violations that will be used to produce the work continuously. As we shall discuss later, this emphasis on longer timescales will be related to the problems of the arrow of time and applications of phenomenological thermodynamics in areas such as cosmology or biological evolution.

There is also somewhat modified stronger form stating that the demon cannot operate a cycle (if it should be done in a finite time), in which work which we expect it to produce will be positive. This is aimed at preempting effects such as appear in Szilard's engine, since it could produce work close to one.[228] This would also help cut the "Gordian knot" of confusions surrounding numerous mystifications about entropy being "hidden in the environment" in each repetition of a cycle. As

[225]Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004)., p 13.

[226] Planck, Max, *Treatise on Thermodynamics*. Longmans; reprinted New York: Dover. (1926) p 103.

[227] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471

[228] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

Carnot was well aware, cyclic motion does not imply that everything is exactly reset; on the level of statistical mechanics, it means that we are dealing with the grand canonical ensembles (those which exchange both energy and particles with the environment) all the time.

These two formulations leave open space for violations of other kinds. A demon can produce work in continuity (supposedly asymptotically decreasing, but still finite at any chosen moment in time) but never complete the cycle, or a demon can complete the cycle, but not continuously. Can we consider these options realistic – or can they be given a physical operationalization? This question has not yet been decisively answered by physicists even in the constrained setting of classical physics. There is, however, an interesting task for the philosophy of science here. We need to clarify what is exactly meant by the "completed cycle". Must the cycle be completed with certainty or just with probability? And does the system come back to initial state or equivalent state at the end of the completed cycle?[229] Therefore, the issues surrounding the general concepts of a "cycle" and a "state" come into focus, giving another important opening for the philosophy of science.

Furthermore, there is an eminently epistemological issue: what is the desired resolution of the "problem" of multiple and often confusing formulations of the second law? Task that needs to be required here is not to define a modified law, since there have been many proposed modifications so far. Sometimes these are useful – we shall discuss an application to the domain of cosmology when dealing with the arrow of time. The task at hand is to show that modified law that is given, is indeed something that we should consider if we wish to have ordered thermodynamics, without "untamed" demons. What kind of modification is appropriate? A violation of such *optimally modified* version would prove that untamed demons exist. Its proof, on the other side, will show that all demons – including those asymptotic ones lacking operationalization *at present* – are tamed.[230]

Constrained violation will be the one in which the (unmodified) Second Law in, for example, Carnot's or Clausius' formulation, is violated. A cycle that will complete in a finite time, with certainty and whose only effect would be conversion of heat into work will constitute an unconstrained violation.[231]

Now, when we have achieved understanding of the problem of Maxwell's demon, we can analyze other philosophical aspects of it. In order to analyze those other aspects, we have to take a look on whole problem from a new perspective, from the perspective of determinism. Let us see what happen when Maxwell's demon meets a Laplacian demon.

[229] Ibid.
[230] Ibid.
[231] Ibid.

4. Maxwell’s demon and Laplace’s demon

“There is in certain ancient things a trace
Of some dim essence
More than form or weight;
A tenuous aether, indeterminate,
Yet linked with all the laws of time and space.
A faint, veiled sign of continuities
That outward eyes can never quite descry;
Of locked dimensions harboring years gone by,
And out of reach except for hidden keys.”
— H.P. Lovecraft

In the previous Chapter, we explained Maxwell’s thought experiment with demon. Now, when we understand aim and role of Maxwell’s demon we can compare it to the role of Laplacian demon and draw some conclusions.

In this Chapter, we will first clarify Laplace’s thought experiment and his deterministic viewpoint. After it, in Section 1, we will explain different kinds of determinism and discuss which of them we can consider Laplacian. In Section 2 we will briefly analyze time reversal invariance. This will prepare us for discussion about asymmetry of time, which we will continue in Chapter 5. Here we analyze it from the point of view of determinism; the complementary discussion in Chapter 5 will be given from the point of view of thermodynamics.

In Section 3 will analyze determinism within the scope of special theory of relativity, which bear some relevance for information processing and the connection between information and entropy. Section 4 will give a brief account on the old philosophical problem of the relationship between determinism and free will.

In Section 5 we will introduce indeterminism and in Section 6 we will analyze its relationship with free will. After it, near the end of Section 7 we will introduce the entropic arrow. This introduction will prepare us for Chapter 5 in which we examine the relationship between entropy and time arrow(s).

Pierre-Simon, marquis de Laplace, following Newton and Bošković, imagined universe as a clockwork, system that is predictably accurate. Laplace's understanding of determinism has two components:

- Present state of physical world is the *consequence* of any state which came before, and the *cause* of any future state.
- A hypothetical superior intelligence could, in the case it knows all the natural forces and laws, predict from the one point of time all the past and the future states.

However, the argument itself does not make the distinction between past and future. Even though in this conception of the world it will be possible to predict both past and future, it would be impossible to distinguish it. There is no distinction between prediction and *retrodiction* in the Laplacian world. The fact that we rarely use the term "retrodiction" in the vernacular is just the particularity of human culture and language, not something related to the real physical world. The thing with the clockwork universe is that in it we cannot recognize the arrow of time. Future is predictable, and past is retrodictable with arbitrary precision.[232] Note, however, that it need not necessarily be retrodictable in real time, since the computational complexity of even perfectly Laplacian world might diverge. This point will be of importance later.

Aim of the Laplace's thought experiment was to demonstrate that universe, on the then best available account of dynamic of physical objects, is a deterministic system. Meaning of determinism is related to our understanding of dynamical laws. The better understanding of natural laws we have, the better characterization of determinism we have available. Besides that, our understanding of natural laws depends on physical theory we accept in the accordance with the Duhem-Quine thesis.[233]

4.1. Several versions of determinism

One way of defining determinism is determination from the current state of affairs, its previous or future state. Suppose we know laws on which the evolution of the system is based; from that knowledge we are able, in principle, to predict or retrodict all past or future states of the system. This kind of determinism is *predictive determinism*, because the focus on ability of the subject to predict the future from the current state. The subject needs to know in which state is the system at the time of prediction. Second, subject need to know fundamental laws, which will direct evolution of system toward its future. Laplace's determinism is *causal determinism*. It is not about claiming that we can make accurate predictions, it is about causal relations that connect all the events in the universe.

Laplace derives determinism, which in this case means the capacity of intelligence/computational systems to predict events in the future. From this principle, we assume that Laplace follows Leibniz's principle of sufficient reason.[234] It is not, strictly speaking, a part of the classical mechanics, but is required for the kind of causal account Laplace was striving for. Present events are linked with past events that caused them, as stands in Leibniz's principle nothing can exist without being caused.[235]

Laplace's goal is to begin from the knowledge on current state and derive from it retrodictability of past and predictability of future states of the universe. In order to achieve this, Laplace makes an auxiliary hypothesis: we consider the current state of the universe as consequence of any previous

---

[232] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

[233] Earman, John*, A Primer on determinism*, Reidel, Dodrecht, (1986), p. 4.

[234] Kožnjak, Boris, Who let the demon out? *SHPS*, 51, (2015): 42-52, Laplace, Pierre, "*La probabilité des causes parles événements*." Œuvres Complétes de Laplace 8 (1774): 27-62.

[235] Weinert, Friedel. *The Demons of Science*. Springer. (2016).p . 66.

state, as well as the cause of the following state. Instead of the present state, one could in principle use any other completely specified state. Note that this immediately excludes states which are in any way causally preferred, for instance, the states near the Big Bang in most modern physical cosmologies. Modern cosmologists usually freely admit that the relevant lawful regularities, or at least some aspects of them, were different or at least incompletely known, at those early epochs.

In brief, Laplace introduces the Leibnizian necessity postulate into his determinism to create his demon. The demon needs to have knowledge on the current state of affairs and the law which govern them.[236] As Laplace himself put it:

> "An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as of the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis, to it nothing will be uncertain, the future as well as the past would be present to its eyes."[237]

This is example of extremely mehanicistic position. Same position held Newton, Galileo, and Boyle. Ruđer Bošković also used a demon to describe his deterministic position:

> "Now, if the law of forces were known, and the position, velocity and direction of all the points at any given instant, it would be possible for a mind of this type to foresee all the necessary subsequent motions and states, and to depict all the phenomena that necessarily followed from them." [238]

We can identify causation and determination only if we presuppose that the cause will always lead to the same consequences. Clearly, there are several kinds of determinism:

1. **Metaphysical:** physical systems have one pathway from the past toward future. This pathway does not branch or bifurcate.

2. **Causal:** Universe has its unique structure in which all the events are causally related.

3. **Predictive:** events in deterministic world are, at least in principle, both predictable and retrodictable.

While metaphysical determinism originates with ancient atomists, causal determinism introduced in the Enlightenment is the strongest kind. It was held by Bošković, Laplace, and d'Holbach.[239]

Some thinkers, like Maudlin or Price, consider relationship between cause and effect as sufficient for meaning of determinism, hence we must analyze it carefully. Others, like John Earman, consider relationship between causality and determinism clear and unproblematic. Moreover, the notion of causality is even more obscure than determinism, so we should not try to explain it by notion of causality.[240]

Other authors argue that Laplace's determinism is predictive and criticized it as such.[241] It has often been the case that philosophers confuse determinism with predictability. Popper consider Laplace's determinism as predictive and show that Laplacian predictability cannot be achieved. He characterizes predictive determinism by using the idea of predictor as a real, physical thing (a machine or being). Predictors are able to determine previous and future states from the conditions

---

[236] Ibid.

[237] Laplace, Pierre Simon. *Théorie analytique des probabilités*. Courcier, (1820). Paris; Nagel, Ernest, and David Hawkins. "The structure of science." *American Journal of Physics* 29.10 (1961): 716-716, pp. 281-282.

[238] Kožnjak, Boris, Who let the demon out? *SHPS*, 51, (2015): 42-52

[239] Earman, John, *A Primer on determinism*, Reidel, Dodrecht, (1986), pp. 4-6.

[240] Ibid., *pp.* 5-6.

[241] Popper, Karl, "Indeterminism in Quantum Physics and in Classical Physics, Part I" *BJPS* **1**, (1950): 173-195.

that are present and knowledge on scientific laws.[242] On this basis, he tries to show that the notion of perfect prediction encounters paradoxes even in a deterministic world.

If the cognitive abilities of observer such as ourselves could be treated as abilities of a predictor, then they should be able to predict future states. What will he need in order to be able to do so? One should know boundary conditions and dynamical laws for particles. Still, it would not be possible to calculate every following state. One of the reasons, they cannot do so is because information on its own state would cause changes. Therefore, value of the information would drop.[243]

The famous French physicist and one of the fathers of quantum mechanics, Louis de Broglie, claimed that there are three problems of Laplace's determinism:

1. It is not realistic to expect that precise predictions are possible. Objects in space interfere on many levels.
2. Both observations and the measurements can be subject to error.
3. This precision is limited by the microscopic ("atomic") realm. We cannot know all the parameters of microscopic particles simultaneously.

Now we can see in more detail why causal determinism is much stronger than predictive. One can make a mathematical calculation that could predict events without knowing/understanding causal chain behind it.[244] The universe could be predictable and, at the same time ontologically indeterministic. Predictability is neither a sufficient, nor a necessary condition for determinism.

One way out of the dilemma is to acknowledge that there are two kinds of predictions: deterministic and probabilistic (which is visible exactly in the example of Maxwell's demon). The universe could be deterministic, yet not predictable via computations, especially not in real time.[245]

Even determinism in classical physics is limited. If we take quantum mechanics seriously, we see that indeterminism may give better explanation of reality. One premise of Laplace's argument can be denied if the evidence shows that the claim that every event has its cause or effect, is false. One such an example might be the famous two-slit interference experiment which is often used as a pedagogical introduction to quantum physics and its indeterminacy: electrons pass through slits and arrive at the screen, but their pattern of arrival is not unique consequence of the manner of their passage through the slits.[246] This is what Feynman called "the only great mystery" of the physical world. Laplacian causal determinism turns out to be false in the microscopic, quantum world.[247]

Mechanistic view benefits from the association of determinism and predictability. But the question is, what is the real relationship between determinism and predictability? Predictive determinism does not have to include complete predictability of all the previous and following states in universe, it only describes an idealization. We cannot expect to know the boundary conditions perfectly in the real world. Also, the physical laws use abstraction and idealization to great extent. Based on boundary conditions and universal laws, we can make only limited predictions.[248]

Astronomy provides standard examples for both prediction and retrodiction. In many cases of interest, we need to use the same underlying dynamical mechanism to ascertain events which are

---

[242] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[243] Popper, Karl, "Indeterminism in Quantum Physics and in Classical Physics, Part I" BJPS 1, (1950): 173-195, p. 189.
[244] Weinert, Friedel. *The Demons of Science*. (2016) . So-called Monte Carlo simulations, originating by John von Neumann and quite popular in recent years, are a glaring examples of this.
[245] Penrose, Roger. *The Emperor's New Mind: concerning computers, brains and the laws of physics.* (1989): p. 220.
[246] Feynman, Richard P., R. B. Leighton, and Matthew Sands. "Ratchet and pawl." *The Feynman lectures on physics* 1 (1963): 46-1.
[247] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[248] Norton, John A." On thought experiments: Is there more to the argument?" *Phil. Sci.*, 71, (2004): pp. 1139-1151.

either in our past or future (eclipses of the Sun and the Moon being historically crucial examples); in others, astronomers are trying to predict events which have already occurred, although the information about them, in accordance with the special relativity have not yet reached us (e.g., explosions of supernovae or transits of planets across the disk of distant stars). Successes of such research programs testify that, in particular well-defined circumstances and contexts, the lack of complete knowledge of boundary conditions is not a problem for reaching scientifically meaningful prediction/retrodiction.

Another important point to which examples from astronomy direct us is that there is no distinction *on the procedural level* between verification of predictions and retrodictions. Inferences from cosmological theory are often billed as predictions, although they by definition pertain to the distant past, e.g., statements about the Big Bang or the origin of the microwave background radiation. This will become more important later in considerations linking Maxwell's demon to the arrow of time.

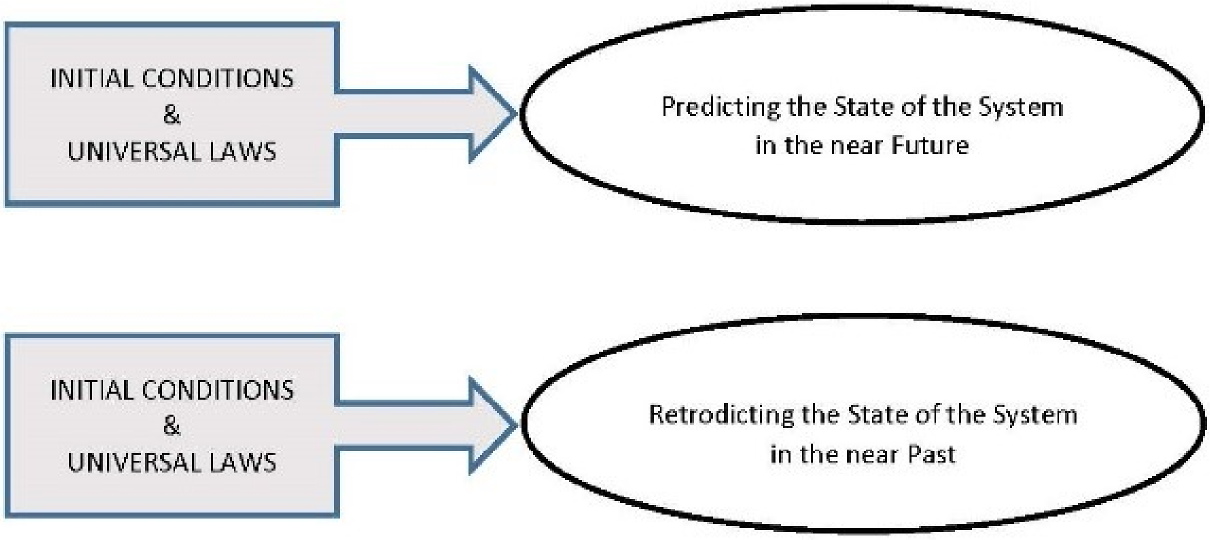

Figure 4.1. Heisenberg's description of predictive determinisM. Adapted from *The Demons of Science* by Weinert, Friedel. (2016).

Laplace's demon teaches us the important lesson that concepts such are determinism and causality are not independent, the precept which is all too often forgotten. For example, in evolutionary biology there are causal explanations of the branching of lineages in the past, but there has been no prediction (in general, with exceptions for particular laboratory experiments with evolution of bacteria, such as those conducted by Richard Lenski) about splitting of lineages in the future.

We have already said that Laplace's determinism could be considered as predictive or causal. A reason for considering Laplace's demon, as example of predictive determinism, lay in his superhuman intelligence that can predict both past and future. Bošković considered that ability of subject to observe lay in the core of predictive determinism. The problem is, if Laplacian demon knows all the events, independently from when they have happened, does this mean that for the Laplacian demon the whole history has already happened? This remains an interesting topic for philosophy of science – and we shall return to it at a later point.

4.2. Symmetry of universal laws

Preempting some of the discussion of the next chapter, we need to briefly consider the time reversal invariance. The distinction between past, present and future has not been made in fundamental physical laws. Hence, fundamental laws do not indicate an arrow of time being time invariant instead. In some form, this has been known since Aristotle and other ancient Greek thinkers.[249] That is, in some sense the notion of time reversal *symmetry* predates the concept of the physical law itself.

However, as Newton first clearly established, the dynamical behavior of physical system depends on both its initial conditions and dynamic laws. The laws that determine the system could be time symmetric, while the system could be both time symmetric and asymmetric.[250] For example, the direction of planetary orbits (prograde or retrograde) is dependent on the initial conditions in the epoch of planetary formation, although the laws of motion of planets– Kepler's laws – are time symmetric.

This is a key point of difference between a universe where we have time asymmetric processes and the Laplacian universe. In the universe that has time symmetric processes, we have time's arrow. The Laplacian universe has deterministic laws which are time symmetric. If based (solely) on these laws it is not possible to distinguish past from future, can determinism make place for time's arrow?

If we consider, possible end points of the dynamical evolution, we can notice that the end of the time or universe would still be conjecture in spite of all the discussions starting from the 19th century controversy on the universe's heat death. From this we can conclude that determinism does not exclude time asymmetric world. Since deterministic world has dynamic evolution and evolution is *orientable* in topological sense. Laplacian world has previous states that cause proceeding states. Determinism has its limits.

4.3. Determinism in special theory of relativity

In special theory of relativity (henceforth STR), the Kant's statement that judgments on time and space have origin in our perception is defeated. This ontological turn introduced by Einstein is often unnecessarily downplayed. A principle of relativity was already stated in the Renaissance in form of the Galilean principle. However, it was limited to the phenomena of classical, Newtonian mechanics. This principle tells us that all coordinate systems (systems of reference) should be regarded as physically equal.[251]

It seems like STR is not compatible with the arrows of time and does not leave space for free will. Since STR followed up classical mechanics in terms of both theory-building and in most of its ontology, many view it as deterministic. It seems that compatibility with the time arrow(s) and free will requires some kind of probabilism that would allow for an open future.

If we want to add probabilism to the STR, it will require proving the possibility that the initial data inside past light cone, would not limit events to realization of an event E1 now. Everything that happened in the past light cone must be compatible with an alternative event E2 being actualized now. If STR is truly deterministic, this would be impossible. However, earlier we showed how

[249] Kutrovátz, Gábor. "Heat death in ancient and modern thermodynamics." OSID 8.4 (2001): 349-359.
[250] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[251] Ibid.

determinism leaves place for different possibilities.[252] Besides, if we modify initial conditions while the law stays the same, it could lead us to alternative possibilities.

Probabilism leaves space for alternative futures. According to this view, fundamental laws are probabilistic, instead of deterministic. However, if probabilism presuppose that there is strict difference between past that is fix and future that is open, then it is not coherent with STR. It is often considered as a problem for STR explanation of decidedness of the same event for different observers. Could one and the same event be decided for one observer, but not for another? In fact, however, in STR there are no same events which are different for different observers, because the same event has the same coordinates in spacetime for all observers (although their time frames and synchronization might, of course, vary). The difference only appears if two observers travel on different trajectories through spacetime, but then they cannot go through same event.

However, events that are separated in space (unlike events that are separated in time) have a conventional time order. Events that are related in space do not need to be causally connected (some of them, separated by the so-called timelike intervals, could indeed be, while those separated by spacelike intervals cannot be). For events that are separate in space, it would not make sense if we ask which of these two came first.[253]

Observers that are time-like related could disagree on answering the question if two events took happened in a same time, but cannot disagree on temporal order of these events. In all these examples, we limit ourselves to flat or Minkowski spacetime of STR; the inclusion of curved spacetime in general relativity opened some stranger possibilities, such as closed timelike curves mentioned above.

The question is does STR imply static interpretation of spacetime? We can have a strong version of determinism if we limit on parts of spacetime that belong to the Cauchy surface. The problem is that if we clame that S is the Cauchy surface, it is not equvivalent to the claim about whole spacetime. However, we can avoid this by limiting on Minkowski spacetime. Then we assume that Cauchy surface always exist.[254] Still, that is just one of the interpretations of the Minkowski space-time.

Another possible interpretation would be *evolving block universe*.[255] Based on this interpretation there is place for the open future and as a new event happen, four-dimensional spacetime unfolds. Popper concludes that STR cannot be considered as deterministic, because a Laplacian demon working within the constraints of STR could only make retrodictions, but not predictions. All of the future states cannot be predicted by predictors because the interferences that have not been planned or expected could too place.

Besides that, there are problem for determinism in STR on particle level. In special relativity particles acting one on another at distance do it with the delay corresponding at least to the light travel time. This delay could make all attempts to localize determination impossible. Mechanism with delay action cannot be deterministic.[256] So, Laplacian determinism has limits both in special relativity and in classical mechanics.

If the number of particles increases, the effective indeterminism increases as well. Even demon as conceived by Laplace could not compute such systems with non-polynomially diverging number of required operations (what is known in computer science as NP-hard problem; more on this below). Hence, determinism and computability may diverge at some point. Also, there are events which are

[252] Maxwell, Nicholas. "Are probabilism and special relativity incompatible?." *Phil. Sci.* 52.1 (1985): 23-43.; "Are probabilism and special relativity compatible?." *Phil. Sci.* 55.4 (1988): 640-645. , Penrose, Roger. *The Emperor's New Mind: concerning computers, brains and the laws of physics*. (1989).

[253] Penrose, Roger. *The Emperor's New Mind: concerning computers, brains and the laws of physics*. (1989), p. 371.

[254] Earman, John. *A Primer on determinism*, Reidel, Dodrecht, (1986).

[255] Ellis, George Francis Rayner. "The arrow of time and the nature of spacetime." SHPMP 44.3 (2013): 242-262.

[256] Erman, John. *Primer on determinism, Reidel, pp*. 66-68.

indeterministic yet computable – random events due to chance. It is important to notice that concepts of indeterminism and non-computability are not interdependent. Events that happened truly randomly do not have – by definition – a pattern. We cannot predict it nor compute it.

To summarize this section, omniscience of the Laplacian demon can be:[257]

1. computational omniscience;

2. dynamical omniscience (can determine evolution of the system in time);

3. observational omniscience.

The demon can identify time's arrow in a case of an evolving block universe, even in the case of deterministic underlying physics. This will become relevant in our subsequent discussion about the degree the entropy gradient determines the arrow of time.

4.4. Determinism and free will

We can define free will as capability of subject to choose course of its own actions. So, the question is how to make such a characterization of free will compatible with scientific determinism.

However, if the 20th century physics teaches us anything new about the relationship of determinism and free will it has done so through the relationship between time and determinism. Deterministic theories are usually time symetric. In other words, past states determine future states, but future states also fixate previous states. Now the science has taught us that our separation of "time" on "tenses" is just an illusion. The main characteristic of time is that it does not belong to one particular event more than label "here" belongs to it. Characteristics of time depend on the physical world and events in it.[258] Yet indeterminism on the quantum level is not within the scope of subject's actions.

According to Stephen Hawking in his celebrated *Short History of Time*:

> "[I]t seems that we are no more than biological machines and that the free will is just an illusion. Molecular biology shows that biological processes are governed by the laws of physics and chemistry. Therefore, they are as determined as the orbits of the planets. Recent experiments in neuroscience show that our physical brain follow the known laws of science that determines our actions and not some agency that exists outside those laws."[259]

The "catch" here is that nowadays we are aware of the degree to which even orbits of planets are unpredictable on sufficiently large timescales; those timescales are not present, at least not explicitly, in the dynamical laws themselves. They are long from the human standpoint for planetary motions, which is why we have not been aware of them until the last century. For systems such as human bodies and minds they could be much shorter, which is the reason why what Hawking writes has not been obvious long time ago.

However, we saw that Laplacian demon fluctuates from scientific to metaphysical determinism. We also mentioned that since Laplacian determinism is causal, it needs to accept the static block universe. According to this view, the past, present and future are on the equal level of reality.[260]

---

[257] Frigg, Roman, et al. "Laplace's demon and the adventures of his apprentices." *Phil. Sci.* 81.1 (2014): 31-59.

[258] Arsenijević, Miloš. *Vreme i vremena,* Dereta, Beograd, 2003, pp. 153-155.

[259] Hoking, Stiven. *Kratka povest vremena*, Sfinga, (1988), p. 106.

[260] Weinert, Friedel. The Demons of Science. Springer. (2016).

STR and even classical mechanics leave place for indeterminism built in the initial conditions. If one changes the initial conditions, it leads to alternative future. However, one may also think about the initial conditions which are inherently uncertain within some small margin of error. Anyhow, the universe of real physics is not deterministic in the Laplacian manner. Besides that, even a deterministic world can represent time's arrow, because of its dynamical evolution. We shall return to this important point later in this dissertation.

Maxwell himself has written about the difference between the dynamical and statistical kind of knowledge. This leads us to make the difference between causes and reasons. He claimed that it would create the difference whether we take the relevant research as historical or predictive. In other words, it is not the same if our aim is to determine the past or future state from the present state of a system.[261] Laws of nature describe what is happening (in compact manner), rather than prescribing what is going to happen. Hence, they are compatible with free will.[262] Maxwell's view on the role of the laws of physics persists to this day, in spite of the periodic challenges on part of both physicists (e.g., Heisenberg) and philosophers (e.g., Maudlin).[263]

It goes much further, however. Statistical method enabled time's arrow in Maxwell's view of thermodynamics. Besides that, according to him it saved the free will. Since determinism is limited, we should turn to indeterminism which will lead us to Maxwell's demon. And he introduces a possibility to distinguish between local and cosmic time's arrow.[264] These ramifications of Maxwell's view remain highly controversial to this day.

Maxwell's demon has the ability of computing motion of every single molecule in a gas, which would be impossible task for a human or even the most powerful present-day computer, although their motions are determined by physical laws. Maxwell's demon takes simpler task than Laplace's demon. He is manipulating every single molecule in the gas container (which is much smaller than the entirety of the world, and could in fact be quite small – as in Szilard's model) and calculating their movements and velocity, nothing more.[265]

All in all, on the very basic level Maxwell's demon contradicts the idea of Laplace's demon, since the former is conceived as a being capable of free decision-making on the basis of available data. As we have seen in previous chapters, demon's task is to test the Second Law and *eo ipso* its philosophical consequences. Thus, indirectly but not less strongly, Maxwell's demon challenges us to reconsider our notions of causality, time direction and indeterminism.

4.5. Indeterminism

Nowadays, indeterminism is usually associated with ideas and concepts of quantum mechanics. The simplest example from quantum mechanics that addresses indeterminism we can find in following historically all-important experiment: a beam of electrons is emitted to the magnetic field of Stern-Gerlach apparatus; subsequently if electron has 50% probability to be deflected upward or downward.[266] Most interpretations of quantum theory – including the historically dominant

[261] Maxwell, James Clerk. "Essay for the Eranus Club on science and free will. Does the progress of physical science tend to give any advantage to the opinion of necessity (or determinism) over that of the contingency of events and the freedom of the will." *The scientific letters and papers of James Clerk Maxwell* 2 (1995): 814-823.

[262] Earman, John, *Primer on determinism, Reidel,* p. 238.

[263] Hajzenberg, Verner, 1989, *Fizika i metafizika,* Nolit, Beograd; (1989); Maudlin, Tim. *The Metaphysics Within Physics*, Oxford University Press, Oxford. (2007).

[264] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

[265] Ibid., p. 98.

[266] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

Copenhagen interpretation – promote the view that this indeterminism is *irreducible*, i.e., it cannot be removed by better observations or better technology of the future, nor can the involved probabilities be explained away in terms of some "deeper", sub-quantum theory.

This should be compared with indeterminism in classical physics, as discussed above. For instance, such is the indeterminism in the evolution of the Solar system on the long-time scale, which is clearly a classical system by virtue of its size and the huge number of particles involved. It should be noticed that indeterminism could be interpreted in both ontological and epistemological terms, and the two are not at interchangeable. Indeterminism in the epistemological sense is the one we see in motions of celestial bodies, which are influenced by nonlinear, chaotic dynamics. Essentially all orbits in the Solar system are chaotic, when observed on long enough timescales. They merge in proximity of the attractors and are essentially impossible to separate in the phase space, creating characteristic "noise bands" of effectively, if perhaps not mathematically, merged trajectories. However, in neither sense is indeterminism to be confused with total randomness. If we know present state of affairs, it is possible to predict if it is probable for a certain event to occur or in which region (i.e., finite volume) of the phase space it is likely to be at some future time.[267]

From the point of view of causal time's arrow, indeterminism tells that on the basis of knowledge on present state of system we could predict only probability of future event. Therefore, to embrace indeterminism does not mean to deny the possibility of prediction. This was clear even to the late Epicurean atomists, such as Lucretius, although they admitted an uncaused and indeterminate "swerve" in mechanical motion of their atoms.

Indeterminism is not statement that supports complete unpredictability. The successes of quantum mechanics, both as experimentally supported theory, and even more as a source of properly working technologies, would be impossible otherwise. Statistically, it is possible to predict future development of events from present affairs. Still, the question is could we apply this on retrodiction? Could we establish *all* causes of current state of affairs based only on the knowledge on the present state?

In an indeterministic system, it could be impossible to find single cause, but only the most probable one. We could see it on the example of evolution of the solar system that we already mentioned. Let us analyze how the Solar System evolves: inside the chaotic zone, the planets stay within this zone. Inside this zone and on sufficiently long timescales their trajectories are indeterministic. Therefore, determining the past state of the planets will be impossible.[268] On the other hand, statements such as that the planets will remain bound to the Sun, and will remain within a particular interval of their orbital parameters are clear-cut truths.

Insofar as we presume that this epistemological sense is the only one relevant, a Laplacian demon could complete his task in spite of the indeterministic long-time evolution of, say, planets in the Solar system. Here we assume, as Laplace apparently did, that the information processing by the demon is perfect at all times. If the demon is able to perform an arbitrarily large (but finite) amount of computations per unit time, he will be able to overcome the effects of chaos in the motion of particles and their aggregates. On the other hand, we might be able to perceive the increase in the computational load as the time elapses. Indeterminism could make room for time's arrow, because in an indeterministic system we could distinguish future from the past. It could represent a solid basis for time asymmetry, but what about free will? What effect does it have on causality?

[267] Ibid.

[268] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

4.6. Indeterminism, determinism and the possibility of free will

The possibility of free will is often held as incompatible with determinism and related to indeterminism. Since indeterminism leaves possibility of open future through its calculating of probabilities for every outcome, it was considered as base for possibility of free will.

History branches and therefore leaves place for a choice in acting one way or another. The problem is that although it leaves place for free will, it does not prove it. As Eddington has pointed it out as one of the first thinkers on the subject of mind and quantum physics: "indeterminacy of a few atoms does not guarantee free will."[269] Proof of indeterminism on microscopic level, disappears on macroscopic level because of small size, large number of their degrees of freedom.

4.7. Causation

Laplacian demon helped us to notice requirement to distinguish causality from determinism. The notion of indeterminism helped us to realize that cause could have numerous effects of different probabilities. We could understand notions of cause and effect in terms of antecedent and consequent conditions. That would help us describe relationship between causality and entropy.[270] Kant considered the very concept of causality as a priori category. Influences by Kant, Niels Bohr wrote:

„Causality may be considered as a mode of perception by which we reduce our sense impressions to order."[271]

Heisenberg and Planck claimed that examples from quantum mechanics proved that causation is not a necessarily a priori category. If we identify causality with determinism, it would be impossible to explain indeterminism on microscopic level. De Broglie's thought experiment points out that on microscopic level we have both causality and indeterminism.[272] Can philosophical model of causation describe adequately notion of scientific discovery? There are numbers of causal conditions and variables. Of these conditions some are dependent, while others are independent. However, consequent conditions will depend on the antecedent ones, conditionally.[273]

4.7.1. Causality and entropy

Attempts to define notion of causality usually did it in terms of antecedent and consequent conditions. However, causality should be defined beyond its relation to a sequence of events. Between the consequent condition and the antecedent, should exist relationship of dependency. Reason lies in fact that there should be difference between correlation and causality. What are entropic connotations of causality? Some thinkers proposed an association of direction in causality with the one we found in entropy. The temporal relation between cause and consequence dependent on temporal asymmetry, or time's arrow. So, the question is, where does this asymmetry come from if not from entropy?

---

[269] Eddington, Arthur Stanley. "Relativity theory of protons and electrons." *Cambridge: Cambridge University Press* (1936): pp. 86-91; Earman, John, *A Primer on determinism*, Reidel, Dodrecht, (1986). chapter 12.
[270] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[271] Bohr, Niels. *Atomic theory and the description of nature: four essays with an introductory survey*. Vol. 1. Cambridge University Press, (2011), p. 116.
[272] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[273] Ibid.

Causality could move from disorder to order as well as from order to disorder. Although entropy of system can decrease as well as increase, according to the Second Law, it will increase statistically until it reaches the maximum value. There is no ambiguous link between the direction of causation and entropy. We could explain entropy in terms of lack of information, still it would not make relationship between entropy and causality less complex. In fact, it would only make it more complex since entropy is associated with loss of information in some cases, but cannot be considered as equal to it, which we will explain in more detail in chapter 6.

Let us remember the example with molecules of perfume: if we let them out of the bottle information about their location would be lost. Now if we analyze Maxwell's thought experiment from the informational point of view, we could say that demon's task is to get the information. Task of the Maxwell's demon can be to separate air from perfume molecules, instead of separating fast from slow ones. Hence, he could bring them all back into the bottle. In this manner, the course of causation would allegedly be reversed.[274]

Still, the question remains: if we define entropy in terms of phase space volumes, will it change our considerations of its relationship with causality?

4.7.2. The entropic arrow

Entropic and causal arrows do not need to have the same direction. If we analyze relationship of entropy and causality in the terms of phase space volumes, we can see that the phase space volume which corresponds to causal conditions needs to take less space than the phase space volume corresponding to effect conditions.[275]

Applying of force will dissipate energy, hence it will result in entropy increase. This increase will result in a destructive interference (causal) unfolding in the local system, still it would not *necessarily* affect biological and constructive causal inference. As we will point out, later in the chapter 7, in the case of evolution, the entropy rise could result in the increase in order as well. This has not been seriously doubted in either scientific or philosophical circles; what is debatable and indeed has been fiercely debated for decades is whether known physical laws (or a well-defined subset of them, like the "laws of the Newtonian universe", "laws of Maxwell's electrodynamics", etc.) are *sufficient* for such an increase in order, or there are additional, as yet undiscovered laws playing a role in this phenomenon.

According to the Second Law of thermodynamics, the entropy of a future state is higher than the one in an antecedent, previous state, at least statistically. This is coherent with entropy decrease on local level because of the constructive causal interference. Still, if entropy's arrow could be related to causality, could it also be related to the time's arrow. In other words, could it indicate direction of time, the same way it indicates direction of causality? We will examine this in next Chapter.

[274] Ibid.

[275] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

5. Maxwell’s demon, entropy and arrow(s) of time

“It would be possible to describe everything scientifically, but it would make no sense; it would be without meaning, as if you described a Beethoven symphony as a variation of wave pressure.”
Albert Einstein

“As far as mechanics is concerned, we could also remember events in the future.”

Hemmo/Shenker, The Road to Maxwell’s Demon 2012, chap. 10.6

In the previous chapter, we already mentioned entropic arrow which give us solid base to start. We will analyze the relationship between entropy and different arrows of time.

In Section 1 we will explain motivation to reduce time arrow on entropic arrow that lay deep in historical background. Key aspect for understanding this is asymmetric nature of time, so that will be next aspect of time we will analyze in Section 2.

In Section 3 we shall separate different time’s arrow, both local and cosmic. Then, in Section 4 we shall go through few attempts of explanation of the future-past asymmetry that are not directly related to entropy. Finally, in Section 5 we will analyze the relationship between entropy and arrow of time.

5.1. Historical background

If we make certain statements about the external world or the way in which the world function, we can extract temporal asymmetry.[276] The first attempt of this kind of extraction was Boltzmann’s famous H-theorem, the first attempt to do it armed with the then new tool of statistical mechanics. Boltzmann has constructed a function of phase variables only (through the energy *E* of each particle) which he claimed to have unidirectional behavior:

[276] Sklar, Lawrence. "The elusive object of desire: in pursuit of the kinetic equations and the second law." In Savitt, Steven F. *Time’s arrows today: recent physical and philosophical work on the direction of time*. (1998): 191-217, p. 193.

$$H(t) = \int_0^\infty f(E,t) \left(\ln \frac{f(E,t)}{\sqrt{E}} - 1 \right) dE. \tag{5.1.}$$

In his 1866 paper, Boltzmann develops his H theorem and claims to have an analytical (in mathematical sense) and general proof of the Second Law of thermodynamics.[277] Namely, the function *H* seems to be monotonically decreasing, so that its negative could be associated with entropy (at least up to a multiplication constant). The results of this research program were ambiguous. On one hand, Boltzmann's result confirms that the gas in equilibrium is well represented by the Maxwell's distribution. On the other hand, the attempt to prove or derive the thermodynamical temporal asymmetry on the micro-level through the H-theorem was a failure, due to "sneaking in" an asymmetrical assumption in the first place. The problem was Boltzmann failed to ask the right question. He realized that instead of asking for the cause of the entropy increase, he should have been asking for the reason of the low entropy in the beginning.[278]

Boltzmann's H-theorem had time-asymmetric conclusion, but it did not really provide the explanation for the origin of the asymmetry, as Boltzmann hoped. In contrast, it was a kind of mistake: he imported a time-asymmetric assumption into his reasoning, making it somewhat circular. If he did not import time asymmetric assumption, he would not get the time-asymmetric conclusion, so the apparent asymmetry would have remained as puzzling as before. The assumption he imports is the so-called "hypothesis of molecular chaos". Actually, that is Maxwell's hypothesis that the probabilities of velocities of colliding particles are independent. This is asymmetric for we expect that the velocities of particles will correlate because of collision between them, thus enabling clear distinction between the time before the collision ("the past") and the time after the collision ("the future"). A truly symmetric assumption did not and could not ever provide for distinguishing between the past and the future. Therefore, the asymmetry Boltzmann got out was just the asymmetry he presupposed.[279]

About 20 years after the attempt with the H-theorem, Boltzmann returned to this problem in 1890s with an entirely different approach (although, of course, completely consistent with the statistical mechanics developed by him and Gibbs). While the H-theorem was essentially an attempt to locate the asymmetry in dynamical laws, with its failure, Boltzmann concluded that it had to be located in the boundary conditions instead. In other words, the solution of the puzzle needs to be found in cosmology. In modern terms, the correct perspective necessary for the attempts to explain temporal asymmetry, is to point out high improbability of the low entropy initial state within the collection of all the possible initial states. For example, Roger Penrose suggested a new law of nature restricting so-called Weyl curvature to zero for the "sources" of outgoing matter such as hypothetical white holes or the Big Bang initial singularity. D. Hugh Mellor, a distinguished contemporary philosopher of science, has noted that the whole question is imposed in methodologically problematic manner. What arguments do we have to presuppose *any* probability of the initial state of universe?[280]

However, cosmologists who discuss these issues tend to make some mistakes, noticed by philosophers of science such as Mellor's doctoral student Huw Price. The most common mistake is to fail to recognize that certain key arguments are not sensible to temporal direction. Every conclusion that holds for one direction in time, holds as well for the opposite one. It is often neglected that

---

[277] Uffink, Jos, "Boltzmann's Work in Statistical Physics", SEP (2017 Ed.).

[278] Price, Huw. "Cosmology, time's arrow, and that old double standard." In Savitt, Stevan F. *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time* (1998): 66-96, p. 66.

[279] Price, Huw. *Time's arrow & Archimedes' point: new directions for the physics of time*. Oxford University Press, USA, (1997).

[280] Sklar, Lawrence. "The elusive object of desire: in pursuit of the kinetic equations and the second law." In Savitt, Steven F. *Time's arrows today: recent physical and philosophical work on the direction of time*. (1998): 191-217, p. 206.

statistical arguments, when properly analyzed and without sneaking in asymmetric assumptions, such as the H-theorem, are insensitive to temporal direction. Hence, they cannot explain temporal asymmetry, at least not without introducing additional assumptions.[281]

Temporal asymmetry that is necessarily introduced is the cosmological asymmetry owns its cosmological nature to the fact that entropy is low in the vicinity of the Big Bang. The puzzle on the origin of the temporal asymmetry thus becomes related to explaining this particular feature of the early universe. In this case, we need to ask: why is the universe so low-entropy near the Big Bang? Smoothness is the equivalent of low entropy condition since it corresponds to the state of minimal gravitational entropy; any clumpiness in the distribution of matter increases gravitational entropy.

Roger Penrose asked what is the fraction of possible universes that would have such a high degree of smoothness in their early stage. He stresses that smooth Big Bang is highly improbable, equally improbable as the "Big Crunch". Hence, he avoids double standard fallacy. In other words, he does not imply argument on the future state if he does not imply it on the past state, as well.

This kind of argumentation is problematic. As Price points out: "...nothing in the universe tells us that one end of the universe is objectively the start and other end objectively the finish."[282] Of course, the "objectivity" here is prejudicated on the possibility of having an "Archimedean point", or the observers capable of being "outside" the universe, which is problematic in itself.

Thus, the basic dilemma of cosmology and time asymmetry remains. We have two options and we can accept only one of them:

1. that entropy will decrease towards both future and past singularities;

or

2. that the temporal asymmetry, as well as the low entropy in the region of the Big Bang, cannot be explained via time-symmetric physics.[283]

If we accept first horn of dilemma we will allow that universe might have low entropy at both ends. This would-be time-symmetric law, similar to other time-symmetric laws we know from the rest of physics (Newton's laws, Maxwell's equations, etc.). Presumably, such a law or effective law would be a product of yet nonexistent theory of quantum gravity, explaining physics very close to singularities.

However, Penrose thinks that there is strong argument confronting the claim that entropy will decrease towards every singularity. Penrose considers that, if we want to save the temporal symmetry, we have two options, we can either reject black holes in future, or accept an increase in number of white holes in the past. He claims that the first option demands improbable "conspiracy" which is physically unacceptable, since there is no reason whatsoever why, for example, a massive star could not collapse in a black hole tomorrow. The only way to avoid this is to have especially fine-tuned, teleological initial conditions, preventing such collapse which is expected under a range of typical or "natural" initial conditions. Problem with the second option is that it contradicts with the smoothness observed in the beginning of universe, before the formation of any structure, as seen in the microwave background radiation. However, this show us that Penrose made the mistake by embracing the double standard (he succeeded to avoid it at first). He accepted the naturalness argument toward the future, while rejecting it toward the past (because he allows black holes in the past but not in the future). In this case, there is some unknown factor that disallows the natural behavior of gravitational collapse

---

[281] Price, Huw. "Cosmology, time's arrow, and that old double standard." In Savitt, Steven F. *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time* (1998): 66-96.
[282] Price, Huw. "Cosmology, time's arrow, and that old double standard." In Savitt, Steven F. *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time* (1998): 66-96.
[283] Penrose, Roger. *The Emperor's New Mind: concerning computers, brains and the laws of physics*. (1989). chapter 7.

– without any proper justification in the known physics. Therefore, we have no non-questionable grounds to exclude the assumption that the same mysterious factor could appear in the future.[284]

So, could we show that despite the highest-level physical laws are symmetric, the universe that they determine at a lower level is not? Conventional statistical analysis does not show that this solution is more probable or more plausible than the time-symmetric universe. If we give up on double standards, the statistical arguments we discuss here are incompatible with the hypothesis that the Big Bang itself is only a statistical coincidence. So, the puzzle remains, where does this asymmetry come from? Could the observed asymmetry come from entirely symmetric premises? Price claims that in order to be able to solve this puzzle, we must accept an atemporal viewpoint. He calls it Archimedean point, or the view from "nowhen".[285]

5.2. The past-future asymmetry

Here, we will briefly go through some aspects of asymmetry of the time's arrow:

1. A traveler through the time could not enter the time-machine and come back to the past.[286] Moreover, he could not go even as observer, because in that case he will change the amount of entropy that should stay the same for the past. Time travelers could not change the circumstances that would affect the spreading in the phase space that has taken place between one year and another.

Reason why it is impossible lies in the difference of time traveler's accessible phase space from the one of the "normal inhabitant" of time. Difference lays in the distribution of the states. Temporal distance between grandmother and granddaughter is not a path in space that could be travelled numbers of time. They are separated by the energy dissipation that could not be reversed, due to the fact that this dissipation has changed entropy of the state. This is analogous to the case of Loschmidt's demon (the one which reverses directions of particles' motions).

Even if Loschmidt's demon could reverse the energy spreading, he would not be able to bring us back to the past. He would only be able to bring us back in the copy of the past, but not in the past, as long we retain the notion of objective physical time.[287]

2. While the past is unchangeable, the future is open, which is represented in the model of branching tree. We can influence the future, since we could choose freely in which direction we would spread energy and how we would use it.

3. Only memories and records on past exists, not on the future. Hereby, the concept of empirical record (including results of all scientific experiments upon which our scientific understanding of the world is based) refers to manifestation of the past entropic states.

The main reasons why we do not have any record on future are following:

(A) Spreading of the energy depends on past, not future;

(B) energy spreads progressively and depends on details of system's history.

---

[284] Price, Huw. "Cosmology, time's arrow, and that old double standard." In Savitt, Steven F. *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time* (1998): 66-96, pp. 78-86.
[285] Price, Huw. *Time's arrow & Archimedes' point: new directions for the physics of time*. Oxford University Press, USA, (1997).
[286] Weinert, Friedel, *The march of time*. New York: Springer. (2013) chapter 4.7.
[287] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

Hereby we should notice that time's asymmetry is based on dynamics that could be related to the Second Law of thermodynamics. Time's asymmetry is objective and the reason for this is contained in the energy balance that differs in the past and future stages of the system.[288]

5.3. Arrow(s) of time

Uncertainties like those mentioned above leave space for freedom of will and independency of mind. As Weinert claims, that is the reason for compatibility of determinism and time anisotropy.[289] Different models of universe are time-orientable in the general case (exotic exceptions like the Gödel rotating universe with closed timelike curves exist, but are nowhere near the realistic case, conflicting with almost everything we know from observational cosmology). Hence, they are coherent with the existence of a truly universal, cosmological time's arrow.

What is the relationship between physical and phenomenal time? We have subjective sense of the time's passage. Does this phenomenon tells us anything about the physical time? We can differ about various arrows of time, although it would be obviously philosophically preferable to have various arrows of time unified by the same underlying processes.[290]

5.3.1. Local and cosmic time's arrows

Cosmic time's arrow represents global flow of time on the universal level. It can be also described as physical arrow of time, which we already mentioned, but put in the wider context and made independent of the spatial location of the observer. We could not derive cosmic time's arrow from the local arrows for several reasons, among which the one of most significance for modern cosmology is that the expansion of the universe (esp. the accelerated expansion, discovered in 1998) makes local regions causally disconnected in the course of history. In particular, the "horizon problem" which arose in 1970s even in the context of the then popular matter-dominated models, shows that the early universe was orders of magnitude smoother in causally disconnected regions already at the time of recombination, which occurred about 400,000 years after the Big Bang. Subsequently, and especially in the cases of dark energy-dominated models popular nowadays, we have disconnection of everything which is outside of observer's event horizon. All this means that the global asymmetry in one important sense takes precedence over any local asymmetries.

There are numerous local time's arrows. For example, we can have psychological one, which represents sense for time passage. The psychological arrow tracks phenomenal time, it tracks past one remembers and future one anticipates.[291] The phenomenal time is not the same as empirical one. Our perception of time changes as we grow old, the fewer new experience we have, the less is left to remember, so it might seem like years are passing faster.[292]

Other kinds of time's arrows are unfamiliar, but it should be noticed that they all share one characteristic: irreversibility. One kind is related to the measurement in quantum mechanics. In this process state of system is reduced on its result, but exists only in superposition before the measurement. This makes the difference between past and the future in microworld an objective one, since the individual outcomes of measurements are always much simpler than a previously existing

[288] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[289] Ibid.
[290] Ibid.
[291] Ibid.
[292] De Bovoar, Simon. *Starost I, II. BIGZ, Beograd* (1986).

superposition.[293] Another is related to the emission, propagation, and absorption of electromagnetic waves in classical electrodynamics. We also have causal time's arrow. One more example would be historical kind of time's arrow that refers to evolution of the systems from lower to higher level of complexity.

What is the most adequate relationship between these arrows? Can we consider some of them as more fundamental? What is relationship between thermodynamical and cosmic time's arrow? In comparison with others, the arrow of thermodynamics seems quite unique. It appears to lead to reduction of order and the amount of information, which is characteristic it has in common with the arrow of quantum measurement.[294]

5.4. Accounts on the past-future asymmetry that are not based on entropy

So, after all we are again faced with same questions: what is the reason that the past not exist? Why the present exists? Every event seemingly goes from being part of the future to present and subsequently fades to past. Whether the notion of entropy can provide a satisfactory answer? The question is could we explain our experience of time's asymmetry with entropy increase?

Here I will offer some other explanations of the time's asymmetry on the part of contemporary philosophers of science such as Michael Lockwood, David Albert, and Storrs McCall:

1. Lockwood uses conditional model of causation. On this account, causation of the events consists of necessary and sufficient conditions. These conditions are capable of explaining a particular event, for each event depends on its cause.[295]

Reichenbach observed that we could presuppose the complete cause from a partial outcome, while it is impossible to presuppose the complete outcome based on just partial cause.

Partial effect → total cause → past (records)

Partial cause ↛ total effect ↛ future

Total cause = common cause of multiple effects. (Reichenbach 1956: 180)

Key to symbols: → (inference permitted); ↛ (inference not permitted)

Figure 5.1. Partial effects and total causes. Adapted from *The Demons of Science* by Weinert, Friedel. (2016).

[293] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

[294] Ibid.

[295] Lockwood, Michael. *The labyrinth of time: introducing the universe*. Oxford University Press on Demand, (2005).

However, knowing only partial causal conditions is not enough to conclude the complete outcome. Past conditions of the events are usually localized, while future conditions are not.[296]

Now I will stress the basic points of Lockwood's account on what distinguish past from the future:

1. Prevailing partial effects we observe in present overdetermine past outcomes. All of the partial outcomes determines the complete cause.

2. Future conditions of the events are not highly localized.

2. Albert[297] distinguish retrodicting of the past from retrodicting of the *records* of past. Processes of performing predictions and retrodictions have common characteristic: they both begin from the present state and then draw the conclusions about different epochs of past and future.

If we try to get knowledge on past states by process of retrodicting, we would not be able to infer all the characteristics of the past state from it. The only way to get complete knowledge on the past state is via records.

Instead of explaining the old one, Lockwood introduces a new notion of asymmetry. Unlike Lockwood, Albert indicate relationship of past and the Second Law of thermodynamics. He described our sensing of the past as confirmation of a lower entropy state in the early universe.[298]

McCall claims that what distinguish future from the past, is that while the past is one and fixed, future possess openness for different possibilities. He argues that any account of the past that relies on whether time travelling is possible, or whether we could change the entropy amount cannot be adequate. What make the past unique is its *unchangeability*.

If we take into account that the universe is indeterministic, according to him, it will require that all of the possible futures would be in a different branch. Therefore, degree of probability of some alternative future would be determined by the proportion of branch on which it is located. It should be noticed that this account is not compatible with the thesis that time's flow is purely subjective phenomenon. McCall's model of time is tree-shaped, where the trunk represents the past, while every branch represents alternative future. The present is represented by the first branch point.[299] It is similar to the so-called Everett's interpretation of the quantum-mechanical wavefunction, which is (ironically) usually taken as deterministic.

[296] Lockwood, Michael. *The labyrinth of time: introducing the universe*. Oxford University Press on Demand, (2005), p. 253.
[297]Albert, David Z. *Time and chance*. Cambridge, Massachusetts, London:Harvard UP. (2001): 1285-1286., chapter 6.
[298] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[299] Ibid.

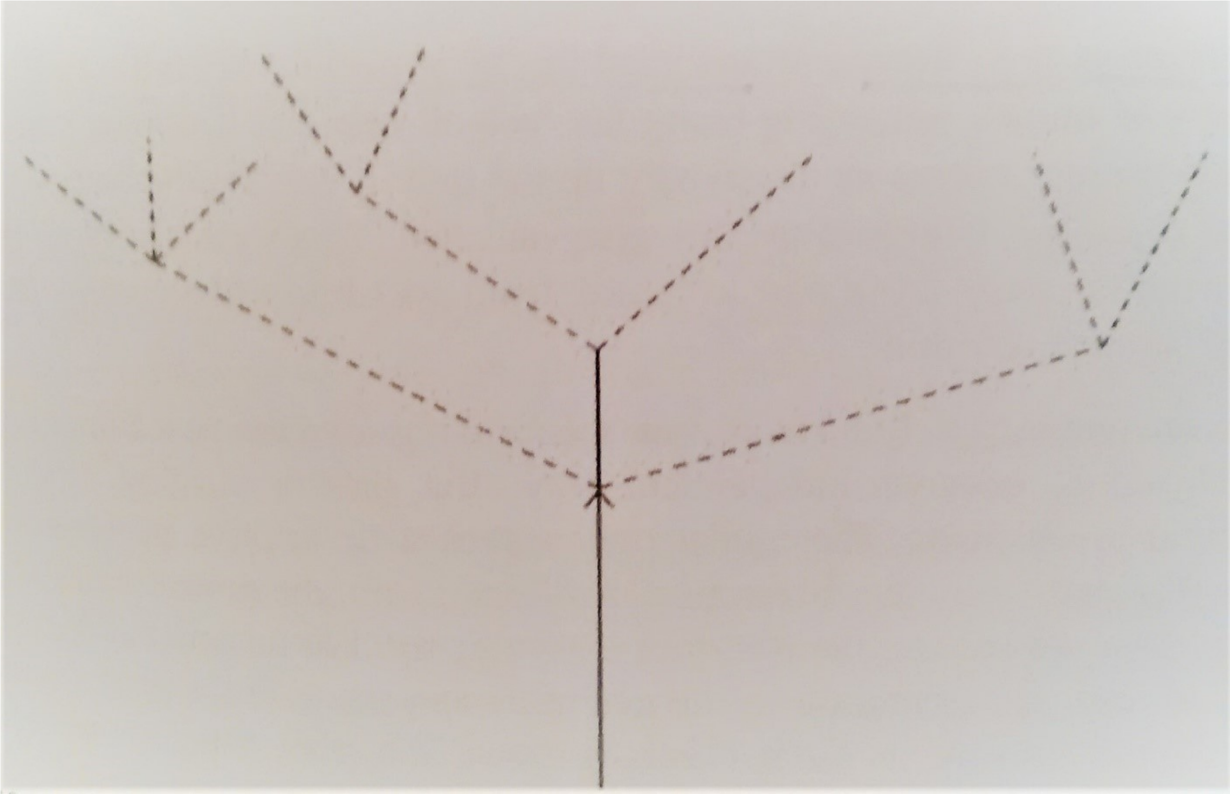

Figure 5.2. Branching tree model of time. Adapted from *The Demons of Science* by Weinert, Friedel. (2016).

Time passage is presented as turning of branches into trunk, as potential future becomes present, and falling off branches if the possible future fails to actualize into the present.

However, McCall's model lacks sufficient and precise method. He proposes the concept of decoherence as a method for explaining his model of a branching tree with set of histories that are probable, equations of motion from classical mechanics.

In the domain of quantum mechanics, the concept of decoherence seems particularly useful for conveying the relevant meaning of McCall's account. Decoherence refers that due to measurement of the environment, we have the emergence of classical macrostates, from the basis of quantum states. It can lead to various possible histories of our world, to dependency of histories from the branches and fixing of the past. When we say that histories are branch-dependent it should be understood as contingency of possible history which took place.[300]

The process of decoherence could be explained as a loss of phase information, that leads to noise increase, and quantum measurement that leads to entropy decrease, it could be considered as a physically irreversible process. However, since this is the characteristic it shares with the Second Law of thermodynamics, McCall considers entropy should be related to the time's asymmetry.

The problem is that account on entropy that is concerned with quantum states not as adequate as statistic notion of entropy, since human's sense of time's asymmetry deals with macroscopic systems. We could use the concept of entropy for describing the asymmetries in our surrounding. Hereby, we are concerned with local time's arrow. Another way is to use it for describing asymmetries in universe, in which case we deal with the cosmic time's arrow. Within cosmological context, the

[300] Ibid., p. 157.

notion of entropy is concerned not only with the origin of time asymmetry but also the whole state of cosmos.[301]

There are significant difficulties in any attempt of measuring or quantifying the entropy of a given state. Usually we infer it from macroscopic parameters, such as temperature, pressure, work. We cannot ever be sure, though, that we have captured all relevant degrees of freedom which contribute to entropy; the example of degrees of freedom associated with the gravitational field discovered only in 1970s. However, it should be remembered that concept of entropy is not irreplaceable, and that the Second Law of thermodynamics can be expresses without referring to it. For example, the original Clausius formulation is expressed in terms of transformation of work into heat, without reverse transforming complete amount of heat into work. The introduction of entropy has been highly useful in the sense of organizing our knowledge of various physical systems and their evolution, though.[302]

5.5. Entropy and arrow of time

Mach was first to propose the thesis that we can reduce or explain, time asymmetry on the basis of the entropic asymmetry of the physical processes that was established by the Second Law of thermodynamics. Reason for this lays in fact that amount of entropy in isolated system (observed on the long-term time scale) can only rise toward the future. There is only one time direction towards which entropy can rise on the long-term time scale.[303] The question is: why should we make such a reduction? Would it help us to understand difference between future and the past better or only drive us to false assumptions that would obscure our understanding on nature of the time?

Mackie says about it that our notion of time is grounded in an empirical experience of causal chain of events. The direction of events shape our notion of time's direction. There were suggestions that if we should differentiate the future from the past in the movie that we watch, we could do so only by the means of entropic arrow.[304] Therefore, we need this reduction for the aims of differentiating beginning from the end.

The kind of reduction that has been proposed for this aim was scientific one. It reduces macroscopic matter to arrays of atoms. It reduces light to electromagnetic radiation. The thesis of this kind of reductionism is that we realize that time arrow and entropic arrow are identical through our empirical experience.[305] Boltzmann claims that it resembles a way we make difference between directions that lead downward and upward. We can distinguish these directions in space and make conclusions on gravitational force indirectly.[306]

What are the consequences of this reduction? If we analyze the Boltzmann's analogy we can ask: can we reduce time asymmetry to the behavior of systems in it like we reduce space asymmetry to the behavior of the objects that obey gravitational force? This analogy will fail.[307] Besides, there were

[301] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[302] Uffink, Jos,"Bluff your way in the second law of thermodynamics", *Studies in the History and Philosophy of Modern Physics* 32, (1853): 305–394; Leff, Harvey S. "Entropy, its language, and interpretation." *Found. Phys.* 37.12 (2007): 1744-1766.
[303] Sklar, Lawrence. "Time in experience and in theoretical description of the world." In Savitt, Steven, *Time's Arrows Today. Recent physical and philosophical work on the direction of time* (1998): 217-229.
[304] Ibid., p.217.
[305] Ibid., p. 219.
[306] Boltzmann, Ludwig, *Lectures on gas theory*, translated by S. G. Brush (University of California Press, Berkeley, (1964) pp. 446-7.
[307] Sklar, Lawrence. "Time in experience and in theoretical description of the world." In Savitt, Steven, *Time's Arrows Today. Recent physical and philosophical work on the direction of time* (1998): 217-229.

lots of attempts to explain this, starting from Reichenbach, David Lewis and others.[308] Still, the attempt of explaining the temporal asymmetry by means of entropy has not been completed.[309]

Reichenbach asks why should we not distinguish between time of perception and time of physical universe?[310] Impossibility of absolutely simultaneous events that happen in different space, postulated by STR drive some philosophers to deny that these two senses of time are identical. Kurt Gödel argued that time of the perception and time of the physical world are not identical. He held the possibility of closed timelike causal loops construed in his cosmological model as final proof of his argumentation. (Empirically, we cannot yet be sure whether closed timelike curves, or loops, exist in nature; if they do, it is likely that they are associated with exotic astrophysical objects, such as black holes, white holes, cosmic strings, etc., which are not readily available for our inspection.).[311]

Sklar argues that we should not replace realism with representationalism. If we deny the identity of time of the physical world and time of perception we would advocate against realism. Now is the place to recall the problem which was stressed long ago, by Kant.[312] If causality holds only for the perception, but not for the physical world that we perceive, how can we explain the relation between perceived and the actual world? If we reduce asymmetry of time on the asymmetry of entropic arrow, then we must conclude that entropic asymmetry is the only asymmetry of the world.[313]

The concept of entropy plays an important role when we consider any arrow of time, but it still should not be overrated. Both Boltzmann and Eddington at first identified the time's arrow with the entropy increase (or entropy gradient), on the basis of the Second Law of thermodynamics. However, they were both reserved on the nature of this identification. Boltzmann, for one, argued in 1890s that one could understand the validity of the Second Law and the Heat Death of the universe without considering its irreversible transition from initial to final state.[314] Boltzmann accepts local time's arrows, but not the global, cosmic one. The reason for this lies in that he considered that the complete universe exists in an equilibrium state and we are just inhabiting a fluctuation.[315]

On the other hand, in his early works, Eddington accepted total equality between time arrow and entropy, claiming that: "time's arrow is a property of entropy alone."[316] Later, Eddington changed his statement and did not held any longer that the entropy increase is equal to time's arrow.

Despite the Second Law of thermodynamics having a statistical nature and its straightforward identification with time's arrow would be wrong, but it would be useful as criteria for time's anisotropy. Eddington distinguished between local and global time's arrows. He offered a construal of the global cosmic arrow of time, that is unrelated to the increase in entropy: expansion of the universe. He alludes to a position which will much later be developed by David Layzer, Paul Davies, and others: that expanding of the universe generates entropy in the universe regarded as a statistical ensemble, by the very fact that it enables larger configuration space for all particles in the universe.[317]

Boltzmann and Eddington both left their previous position of identity of entropy increase and time's arrow in favor of identification of entropy increase and anisotropy of time. They left that position

---

[308] Reichenbach, Hans. *The direction of time*. Berkeley: University of California Press. 1956. (1991): 198. chapter 4.
[309] Sklar, Lawrence. "Time in experience and in theoretical description of the world." In Savitt, Steven, *Time's Arrows Today. Recent physical and philosophical work on the direction of time* (1998): 217-229.
[310] Ibid.
[311] Ibid.
[312] Kant, Immanuel. *Critique of pure reason*. Cambridge University Press, (1998).
[313] Sklar, Lawrence. "Time in experience and in theoretical description of the world." In Savitt, Steven, *Time's Arrows Today. Recent physical and philosophical work on the direction of time* (1998): 217-229.
[314]Uffink, Jos, "Boltzmann's Work in Statistical Physics", *STA (2017).*
[315] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[316] Eddington, Arthur. The Nature of the Physical World: Gifford Lectures (1927). *Cambridge University Press,* (2012)., p. 74.
[317] Layzer, David. 1976, "The Arrow of Time" *Astrophysical Journal* **206**, pp. 559-569.

because of the realization that the Second Law became construed as a statistical law. This turn happened, at least in part, because the analyses of the problem of Maxwell's demon.[318]

Irreversible processes are processes of high complexity. Their reverse to initial state is highly unlikely. Still, theoretical possibility of it exist. If they ever happen it will not violate micro processes, neither will it violate the Second Law of thermodynamics.

Weak T invariance should not violate the laws that are concerning the most fundamental processes.[319] Because of this T invariance is coherent with asymmetric solutions, under consideration of suitable boundary conditions.[320] Future quantum field theory will shed some further light on the small violation of T invariance in low-energy weak interactions – for the moment, it is impossible to connect it with any other observed asymmetry.

If the cold cup of tea left to itself, become hot again, at some point in time, it would not represent the violation of the Second Law– but such a behavior has never been observed, nor is it expected to be. Why is this the case? Poincare recurrence theorem asserts that a wide class of systems will return to the state that is similar to their initial state, after a sufficiently long (but finite!) time. Poincare recurrence time is a measure of how long will it take and it only exists for isolated systems. It system can return to its initial state for $10^{10^{25}}$ years for a gram-mole of gas. [321]

As Eddington realized an increase in entropy is not identical with time's arrow because recurrence is theoretically possible – expressed symbolically by Loschmidt's demon thought experiment.[322] The empirical world exhibits *de facto* irreversibility.

Identification of entropy increase with the arrow of time is mistaken. It is mistaken because time's arrow has one direction, but entropy has two directions, since at least Maxwell's demon could decrease the entropy. However, statistical nature of the Second Law, is not an obstacle, for considering thermodynamics probability was one of the criteria for inferring the time's anisotropy[323]

A lot of parallel processes in our empirical world exist (for example, the emergence of classical systems from quantum). All of them indicate time's asymmetry. As Popper's analogy illustrates, it must be recognized that boundary conditions of the world and the initial low entropy need to be considered in our quest for an explanation of the time's arrow(s).

Boundary conditions in realistic cosmologies are mainly asymmetric. Boundary conditions are no more merely *stipulated*, as it was the case in earlier times, since modern cosmology is concerned with explanations of events such as the Big Bang and its initial low entropy starting from still deeper theories. Here, one might be seemingly justified to ask: in terms of what? The answer lies beyond the scope of the present dissertation, in fields which have arisen in the last quarter of century, like quantum cosmology and string cosmology. However, we do not have evidence that entropy in universe will be low in the future. It seems, on the contrary, that the universe become more disordered and less capable of extracting useful work as we move toward the future.

Nevertheless, these solutions might be hard to find – and we anyway do not understand all dynamical laws at present (e.g., we do not understand gravity on the microscopic, quantum level). So, instead of investigating time's symmetry of the solutions of fundamental dynamics, we can only take entropy

[318] Weinert, Friedel. *The Demons of Science*. Springer. (2016), p. 122.
[319] Landsberg, Peter Theodore, and Eric J. Chaisson. "The enigma of time." *American Journal of Physics* 53.6 (1985): 601-602.
[320] Price, Huw. *Time's arrow & Archimedes' point: new directions for the physics of time*. Oxford University Press, USA, (1997), pp. 77-79.
[321] Griffiths, Robert B. *Statistical irreversibility: classical and quantum*. Cambridge Univ. Press, (1996): pp. 149-150.
[322] Loschmidt's thought experiment in which demon would be able to bring molecules of gas back into the bottle.
[323] Weinert, Friedel. *The Demons of Science*. Springer. (2016), p. 122.

as an indicator of the time's arrows. One will experienced flow of time even in models universes that close back on themselves, as in the old-fashioned oscillating universe models.[324]

Now we need to examine relationship between entropy and information.

[324] Ibid.

6. Maxwell's demon, entropy and information

"Von Neumann told Shannon to call his measure entropy, since "no one knows what entropy is, so in a debate you will always have the advantage."
— Jeremy Campbell, Grammatical Man: Information, Entropy, Language, and Life

"Thinking generates entropy."

— James Gleick, The Information: A History, a Theory, a Flood

"Information is physical."

Rolf Landauer

"Demon lives!"

Norton and Earman, "Exorcist part II"

In this Chapter, we will explore relationship between Maxwell's demon, entropy and information. In order to do it, first we must introduce formalizations of information, which we will do in section 1.

In section 2 we will start to examine relationship between entropy and information. In order to do so, we need to explain that in history of science, it was often considered that entropy and ignorance are more or less the same, which can nowadays be shown as wrong. Here, we will also recall Szilard's engine, previously introduced in Chapter 3; now we can look at it from a fresh perspective of the previous two chapters. We shall try to explain this important model again in order to stress the puzzle it represent for relationship between information and entropy, which we will try to resolve in later Sections.

Section 3 comes back to discussion on Landauer's principle in order to emphasize importance of the role of information erasure for any analysis of the relationship between entropy and information. In section 4 we will discuss Brillouin's information exorcism.

In section 5, we will try to explore entropy-information relation without including demons in picture. Hence, we will analyze particular counter-intuitive engines without demon. In section 6 a few lines of argumentation which claim that demons could exist in principle will be addressed.

In section 7 we will pose the question – if entropy and information are same, would it make any difference if demon is intelligent or not? In section 8 we will criticize exorcising the demon by information cost, which is, as briefly mentioned above, a major issue in contemporary philosophy of physics.

In section 9, a solution to Szilard's puzzle will be considered. In Section 10 we will summarize the relationship between Shannon information and thermodynamic entropy.

6.1. Information

In this section, we will explain various formalizations of the notion of information. In order to achieve that we will discuss some of the old problems related to the information. We will analyze following claims:

1. Information should be represented on a basic physical level like mass or energy.

2. Theory of information solves measurement problem in quantum mechanics.

3. Thermodynamic entropy and information are equal.[325]

As we have seen, acceptance of the third claim has led many philosophers to the conclusion that Maxwell's demon can be exorcised solely by the means of information.

Information is the notion that we usually do not explicitly define. The definition is needed for purpose of different usages of information. In the information theory, we are dealing with various kinds of information: actual, algorithmic, Shannon, Fisher, quantum, etc. The concept of information we will use here is essentially that introduced by Shannon.

6.1.1. Shannon information

Since notion of information was primarily used to describe sending of different signals through messages, as well as capacity for carrying the signals, the definition of Shannon information first came as a solution for the following question: what is the shortest way to code the message. For example, if a subject is about to send a message a that is contained into a set of a bit strings (1's and 0's) with probability P_a, the one who receives message would decode it back from the bit string into message. Shannon's theorem expresses the shortest bit-strings from which message could be decoded without error[326]:

[325]Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004)., p. 10.
[326] Ibid., p. 13.

$ISh = -\sum_{a} p_a \log_2 p_a$

(6.1.)

Shannon's information is a measure of Shannon's entropy. It represents the shortest way to express a message. We have two possible interpretations of this claim:

1. Shannon 's information represents measure in which content of message is unknown (before reading the message).

2. Shannon's information represents the measure of information we get after reading the message.[327]

It is important to notice that Shannon's information is not related to the meaning of the message, but is rather related to its probability. It is related to the probability of being sent. This is what $-\log 2\ P_a$ measures. The less probable message will contain more information.[328] Probability of the message being sent, or probability of the truth of the statement, drop with the increase of information or elements it contains. For example, it is more probable that subject would send information A, than both information A and B. Reason for this is merely logical. If A and B is true, then A must be true. But if A is true it does not follow that A and B would be true. Like in the Linda paradox, it is more probable that Linda works in bookshop than that Linda works in bookshop and is feminist.[329] Most of the people would reason that second statement is more probable, because of the Linda's personal history (she was feminist earlier). This would be a mistake because we measure probability regardless of the Linda's personal history, since second statement contains the first. If second statement is true, so would be the first. This example show us why it is important to distinguish meaning from probability. Meaning cannot undergo measurement.

6.1.2. Mutual information

The mutual information is information that a receiver gained from the message that has been sent. We can regard the amount of mutual information as a symmetric function that expresses the information that is common to both parties that are included in communicating. We can also express its function as correlating states of the both parties included in communication.[330]

We should notice that the measure of information probability increase with the greater degree of its improbability and the smaller degree of knowledge. If message is transmitted and decoded reliably, both subject who sent and the one who receives will be correlated to maximum degree.[331]

---

[327] Dugić, Miroljub, *Osnove kvantne informatike i kvantnog računanja*, Prirodno matematički fakultet univerziteta u Kragujevcu, (2009), p. 3.

[328] Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004)., p. 14.

[329] Kaneman, Danijel, *Misliti brzo i sporo,* Heliks, 2015.

[330] Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004)., p. 16.

[331] Ibid.

6.1.3. Quantum information

We saw that first definition on information was the one given by Shannon and it was represented by shortest bit string per signal that can transfer the message adequately. Later, Schumacher generalized it, in order to make it compatible with quantum theory.[332]

Classical information is characterized by possibility of distinguishability of information which is based on classical reality of physical states of the system on which the process is running.[333] Analogously, quantum information conveys the information about the state of a quantum system, with all the restrictions and constrains following from the nature of quantum theory (superposition, entanglement, unitary evolution, etc.).

Table 6.1. Comparison of quantum and classical information

Quantum information	vs.	Classical information
Quantum informatic limit (relation of uncertainty, noncomputability, quantum uncertainty) $\forall\ \vert\psi, \exists \hat{B}: \Delta\hat{B} \neq 0$		Existence of unique values of all the variables and all the states of system in any moment $\forall B, \exists b_{,} \forall t$
Indistinguishability of nonorthogonal states - no-cloning theorem		Indistinguishability of classical states (value of variables) is a consequence of the metrological mistake.
$a_i \xleftrightarrow{nonlocalitz} b_j\ [\hat{A}, \hat{B}] \neq 0$		$a_i \xleftrightarrow{locality} b_j$

[332] Ibid., p. 18.

[333] Dugić, Miroljub, *Osnove kvantne informatike i kvantnog računanja*, Prirodno matematički fakultet univerziteta u Kragujevcu, (2009), p. 96.

Quantum entanglement		All the states of a classical system are separable – every subsystem has a particular state in every given moment.
$\mid \Psi\rangle = \sum C_i \mid i\rangle_1 \mid i\rangle_2 \neq \mid\bullet\rangle_1 \mid\circ\rangle_2$		$(\bullet)_1(\circ)_2$
Quantum nonlocality (quantum holism) Quantum entanglement- Bell's nonequality does not hold $\mid \Psi\rangle A = \sum C_i \mid i\rangle_1 \mid i\rangle_2 \xrightarrow{\text{measure in time t}} \mid k\rangle_1 \mid k\rangle_2$		Operation on one subsystem does not have to affect other subsystems of a complex system, if they are remote in space. Bell's theorem is always valid.
Measured in bits.		Measured in *qubits*.

[334]

Without a priori knowledge, the measured quantum information has no value. For example, if we take a measurement in the basis |0in0|,|1in1|, results will depend on the ensemble on which we measure. If it is ensemble 1, results of the measurement will objectively represent state of the system, while if the ensemble 2 is in case it will cause wavefunction collapse, and all the records on the system would be destroyed. If we do not know which ensemble we measure, we would not be able to interpret measuring results adequately. The problem is that one could never know which was the actual ensemble. However, if we want to apply information to quantum system we need to have well-determined process of measurement.[335] Here, the definition of quantum information touches upon the most important ontological and epistemological problems of philosophy of quantum mechanics.

When we perform any classical measuring, we can partition phase space to finer degree, until we reach the probability density as distributed over the complete phase space. In a case subject observes probability distribution for the states that is not correct, he can correct it via application of Bayesian rule again and again. In this manner, information he gains about a system will become objective characteristic of the ensemble.[336] This is a different way of reaching the same conclusion as in Chapter 5 above, that the information about the initial superposition of states is lost in the course of

---

[334]Dugić, Miroljub, *Osnove kvantne informatike i kvantnog računanja*, Prirodno matematički fakultet univerziteta u Kragujevcu, (2009), p. 97-98.
[335] Ibid., p. 29.
[336] Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004), pp. 23-24.

measurement (or the "collapse" or "reduction" of the wavefunction). The initial superposition of the decayed or non-decayed atomic nucleus collapses into the state corresponding to a living cat or a dead cat in Schrödinger's famous thought experiment – so the information of the other component of the wavefunction is irretrievably lost.

6.1.4. Active, passive, and inactive information

It is important to acknowledge to what extent our ontological commitment in the domain of interpretation of wavefunction influences our understanding of quantum information. This is especially the case with notions such as active and passive information. Suppose that we observe the conventional two-slit experiment with quantum interference. While quantum measurement is concerned with deeper properties of a system, active information, provides a consistent interpretation of the interferometer within the framework of Bohm's interpretation. It clarifies correlation of path measuring and interference. If we want to consider role of information in the system, we should differentiate the concepts of active, passive, and inactive information. The best way to do so is through a specific example.

Consider a particle moving in an external solvable potential. We have solved the dynamical (Schrodinger's) equation for such a system and obtained various possible solutions in form of wave packets. *Active* information would be the one that is related to wavepacket. From the other side, the same information is *passive* for the wavepacket with which it is not associated. In case where they overlapped, passive information became active.[337] Of course, if we reject the very concept of particle trajectory, which is done in the orthodox Copenhagen interpretation, the difference between active and passive information becomes irrelevant.

Now, let us summarize. Shannon information can be expressed as lack of information on system's state. According to standard quantum theory we cannot interpret measurement as a measure of a previous state of affairs.

From the other side, notion of active information enable interpretation of these measurements. In a case, we measure trajectory of the particle, information on other wavepocket will be inactive. In case we do not measure it, both wavepackets will be active whenever interference occurs, and the trajectory is defined by information located in both sides of interferometers. However, when we perform a measurement and acquire information, the information associated with other wavepackets will become *inactive*.

6.2. The relation between information and entropy

We saw that we can understand the Second Law either as a decrease of order or a decrease of information. To analyze the relation between information and entropy we must go back to Szilard's thought experiment, which places entropy in process of acquiring of information via measuring process. His argument is the prototype of "informational exorcism" and informational explanation of the nature of entropy and its gradient. However, that is not the first time that entropy is taken into

[337] Ibid., p. 33.

account on the basis of its relation to information. There have been some earlier attempts to correlate the lack of information and entropy. If we dig deeper through the history of ideas, we can even find that a concept of ignorance was rooted into understanding of entropy, although it was not considered equal to it.

6.2.1. Entropy and ignorance

The notion of entropy is one of the most fundamental, yet arbitrary notions. This has led many thinkers to reduce entropy increase to lack of knowledge on the system's microstate. Maxwell himself wrote: "The idea of dissipation of energy depends on the extent of our knowledge ...[it] is not a property of things in themselves, but only in relation to the mind which perceives them."[338] Since the second law of thermodynamics is not to be considered as absolute law, but only statistic it opened a space for further relativization of the notion of entropy, especially by relating it with the ignorance of the current state of the system.

Similarly, Feynman interpreted entropy increase in terms of its relationship to lack of information: "What has happened is that my knowledge of the possible locations of the molecule has changed. The less information we have about a state, the higher the entropy."[339]

Clearly, the association of a physical quantity with subjective knowledge is problematic from the standpoint of classical physics and predominant scientific realism. Therefore, the primary question is should we take the "lack of knowledge" as an objective or a subjective characteristic? If we lack knowledge for it is difficult to determine exact microstate of a body, it could be considered as an objective characteristic.[340] In this manner, the entropy increase could be understood in terms of complexity of the interactions between numerous bodies. An argument against this interpretation was that numerous irreversible processes appear only due to our ignorance on microstates of the systems.

There are many problems related to the link between knowledge and entropy. Perhaps the most pertinent to the "real" scientific issue is, how is it possible that we empirically observe both the amount of information and the entropy increasing as the time passes?[341] It seems indubitable that our knowledge about the empirical universe that surrounds us increases. We gain new information every time we observe. If we identify entropy with the lack of information, we seemingly come to paradoxical situation to claim that entropy reduces every time we gain new information about the system we observe. How can we relate changes in entropy of the system which we are measuring with information we acquired through it?

6.2.2. Szilard thought experiment and its influence on the discussion

Szilard in his 1929 paper was first to draw the attention to the entropy cost on information processing, and his work has inspired most of the discussion on the relationship between information and entropy. He proposed that we should exorcize Maxwell's demon by means of entropy cost incurred via

[338] Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004), p. 65.
[339] Feynman, *Lectures on computation*, Penguin (1999), p. 20.
[340] Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004), p. 66.
[341] Ibid., p. 66.

demon's acquisition of information. Hereby, the demon is considered in terms of "information-processor."[342] Before we analyze this relation let us go briefly to the original Maxwell's thought experiment in order to stress the importance which information has in it.

Acquiring of knowledge through measuring of a system is possible only if it can have at least two possible results. In a thermodynamic ensemble, content of the measuring will be the selection of sub-ensembles. Maxwell suggests that this selection can have anti-entropic nature. Before the demon separates molecules, their distribution in the box is random and unknown. The demon would achieve higher degree of order by separating them. While increasing the order, demon will increase the amount of knowledge on the molecules location. On this example, we could easily point out the link between order and information.

Szilard's own answer to the dilemma (where does the entropy come from?) was the assumption that demon cannot operate the engine both continuously and reliably.[343] In order to support this assumption he analyzed where and why does it go wrong. After he imposed validity of second law, he removed every other source of the entropy increase. After it, Szilard concludes it must take place during the measurement.

He argued that entropy is produced because the demon needs to determine location of a molecule via measurement. Without knowing the location of a molecule, the demon cannot connect the weight to piston adequately. Therefore, while the demon could reduce entropy if he measures current location of the molecule, the sole act of measuring would produce at least as much entropy.[344] According to this line of argumentation, the demon cannot reduce entropy, unless he produces some. The Second Law that was postulated was a modified one, similar to the proposal of Smoluchowski. It required that the average production of entropy during measurement must be equal to the reduction of the entropy that took place as the action of the demon, understood as a product of the same measurement.

Szilard, himself did not offer explicit definition on entropy, but from the context we can conclude that it was entropy of macrostates. Szilard's argument is based on the claim that as long as statistical mechanics is not coherent with the existence of untamed demon, there must be an entropic cost. Origin of that cost is related to the acquisition of information.[345]

He offered an *example* of a measurement process where this kind of entropic cost is demonstrated. Nevertheless, he did not give a general argument which will prove that all measuring processes require entropic cost, which is – underneath a superficial disguise – the very same situation we encountered in Norton's criticism of Landauer's principle. What Szilard claimed is only that *if* such a measurement would be possible, so would the untamed demons. He concluded that it is not the sole act of measuring, but the erasure of measuring result produced the entropy.

Now, we must make a short digression in order to elaborate upon the notion of the untamed demon. The untamed demon is a demon that would be able to make straight violation of the Second Law. According to Norton and Earman there are two ways to violate the Second Law:

(**Straight Violation**): achieving reduction of entropy in isolated systems.

---

[342] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471.
[343] Detailed explanation of Szilard's machine is provided in subsection 3.4.
[344] Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004), p. 68.
[345] Maroney, Owen. "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

(**Embellished Violation**): achieving reduction of entropy that could be used for producing useful work.[346]

A tamed demon would be able to make embellished violations.

Now, let us go back to the Szilard's argument. When Shannon found that the measure p ln p[347] was significant for information theory, many physicists and philosophers of science started to build upon Szilard's argument, adding to it the suggestion about connection of entropy and information.[348] In order to develop idea, both Gabor[349] and Brillouin[350] constructed models of dissipative measurement. This particular model of measurement consisted in shining a beam of light to the part of the device in order to discover if molecule is there. Obviously, interaction between light and the molecule needs to be taken into account, including transfer of energy, momenta, etc. Their analysis aimed to establish the conclusion that process of measurement created at least as much entropy as it is allegedly removed by obtaining better information.

Gabor and Brillouin generalized from these results and claimed that acquisition of information necessarily leads to energy dissipation and production of additional entropy. In particular, Brillouin held that the fact that Shannon's information and Gibbs's entropy share similar mathematical structure indicates that entropy is equal to lack of information. He addressed information as negentropy.[351]

Furthermore, Brillouin accepted the most general sense of entropy as being equal to the Gibbs entropy of a system. He made a distinction between bound information (information that is related to the some kind of physical system) and free information (information that is present only in someone's mind). He accepted that measuring can cause decrease of the system's thermodynamic entropy. However, this holds only if it creates at least equivalent quantity of bound information within the same device that performed measurement. Although, it is not clear whether he argued that it is the sole act of creation of the bound information causing the entropy production, or is it bound information entity that we must add to thermodynamical entropy in order to protect the generalized Second Law.[352]

6.3. Information erasure

We have already discussed Landauer's principle, its criticism and its defense, but we need to quickly go back to it in order to stress some points which could improve our understanding of the link between entropy and information.

---

346 Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471.

347 According to Shannon, the amount of information I in message x is given by: $I(x) = -\log px$; We can interpret this formula as inversion of Boltzman's entropy. Characterization of the comunication entropy by Shannon is entropy of a system of messages. This is equal as Gibbs's entropy in physics. In p ln p, p is probability of information.

348 Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

349 Gabor, Dennis, "Light and Information", *Progress in Optics* 1, (1964): 111–153.

350 Brillouin, Leon. "Maxwell's demon cannot operate: Information and entropy. I." J. App. Phys. 22.3 (1951): 334-337.

351 Brillouin, Leon, *Science and Information Theory*, (New York: Academic Press), (1956).

352 Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

Although Landauer never explicitly tried to exorcise Maxwell's demon, or criticized Szilard's engine, his ideas caused criticisms of the latter, as well as provided an argumentation to be further developed by subsequent exorcists. Landauer argued that logically irreversible operations were the source of heat generation; also, he believed that they were the necessary part of computation.[353] It is important to notice that, even though there is a possibility of simulation of logically irreversible operations with logically reversible operations, it will produce information that needs to be kept in a memory of the device. To complete a thermodynamic cycle, without storing information, it will be needed to reset the memory to zero. Still, the very operation of resetting to zero will come with the corresponding entropy cost.

Demon must store the information which he has acquired via that measurement. Subsequently, he can extract useful work. Demon keeps the information on the measured location of molecule when the cycle ends. While the demon performs the process many times, his memory is filling up, until it unavoidably runs out of space, which must eventually happen since demon's memory is finite. When that happens, demon will either be unable to operate further or he will reset his memory. If the demon resets his memory, such an act would lead to the increase of Boltzmann entropy elsewhere.

Bennett[354] argued that logically reversible computation does not need to store the additional information.[355] He also demonstrated physical model which is able to perform this kind of measurement. Key point here is that the act of measuring creates correlation between the system that is performing the measurement and the state of the system which it is measuring. There were no such correlations previously (notice an analogy to Boltzmann's "molecular chaos" hypothesis).

Nevertheless, Bennett needed to implement this measurement process into physical device. Hence, he made a kind of Szilard's engine where the molecule is diamagnetic. The measuring device in this engine is a one domain ferromagnet that has initial polarization that is fixed. Bennett's idea was that it is possible to correlate ferromagnet's polarization to the location of the diamagnet. He considers that this could be done by careful manipulation which will perturb the magnetic field y using diamagnet. Nevertheless, resetting of the the polarization of ferromagnet will lead either to usage of correlated location of diamagnet either to heat generation of kT ln 2.[356] It is important to notice that ferromagnet in this engine should be considered as equivalent to demon's memory. Thereby, this represents the Bennet's argument against the Szilard's and Brillouin's claim that measurement must be dissipative. Bennett also claimed that resetting of demon's memory is the step in which the heat generation must occur, due to necessarily logical irreversibility of this step.

All of this, creates grounds for thesis on relationship between entropy and information. Basic steps of this argumentation are:

• We can regard entropy as a measure of ignorance on system's state.

• By performing a measurement upon the system one acquires information and reduces its ignorance on state of system.

[353] Landauer, Rolf. "Computation and physics: Wheeler's meaning circuit?." *Found. Phys.* 16.6 (1986): 551-564.
[354] Bennett, Charles H. "The thermodynamics of computation—a review." *International Journal of Theoretical Physics* 21.12 (1982): 905-940.
[355] Bennett, C. H. "Logical reversibility of computation." *IBM journal of Research and Development* 17.6 (1973): 525-532.
[356] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

• After it, entropy of the system decreases equally to the amount of Shannon information that one gains through the process of measuring.

• Still, information needs to be recorded in the memory (device).

• Since information should be recorded more than once, after a while erasure operation must be performed in order to make space.

• According to Landauer's Principle, the erasing process dissipates the amount of energy equal to the one necessary for erasure of Shannon's information. This creates at least the same amount of entropy that has been reduced via information acquisition from the measurement process.[357]

Problem with this argument is that it is circular. Landauer's Principle established kT ln2 (dissipated) energy cost by assuming the validity of the Second Law. As we already mentioned, Bennett used Landauer's principle for proving the validity of the Second Law. This is the circularity on which Norton referred in *Eaters of the Lotus* study.

6.4. Brillouin's information exorcism

Brillouin gives the simplest interpretation (proof) of Szilard's Principle.[358] Brillouin explains it as follows:

"Any experiment by which an (sic) information is obtained about a physical system corresponds in average to an increase of entropy in the system or in its surroundings. (...) [An] information must always be paid for in entropy, the price paid being larger than (or equal to) the amount of information received."[359]

Therefore, any reduction of entropy must be followed by production of at least as much entropy. In other words, information I would come with costs that must be paid with at least as much entropy as the gaining of the information reduce, in the first place. This explains the central problem of the Maxwell's demon. From this perspective, it can be represented as change from negentropy to information. Later, it turns back into negentropy. Demon gains information on system; he uses it to reduce the entropy of system. Information is then converted into negentropy. After reduction of negentropy, it comes gain of the same amount of entropy.[360]

Later, Brillouin will claim that "bound information" is just representation of the limiting case of "free information."[361] He divides the complete entropy of a system on entropy S and negentropy I. Negentropy I corresponds to bound information. The complete entropy of system would be (S I). According to Carnot's Principle it could not decrease in system that is closed.

Thus, Brillouin choose the sound horn of the dilemma. He argued that the very concept of information offers exorcism of Maxwell's demon, because it shows his inability to reduce the entropy. In fact, by

[357] Maroney, Owen. *Information and entropy in quantum theory*. quant-ph/0411172 (2004)., p. 71.
[358] Brillouin, Leon. "The negentropy principle of information." J. App. Phys. 24.9 (1953): 1152-1163., pp. 1152–1153
[359]Brillouin, Leon. "The negentropy principle of information." J. App. Phys. 24.9 (1953): 1152-1163., p. 1153.
[360] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[361] Brillouin, Leon, 1956, *Science and Information Theory*, New York: Academic Press. chapter 12.

choosing the sound horn of dilemma, Brillouin presupposes that demon and the system he is manipulating are both part of closed system that obeys to the second law. For this reason, created negentropy, must be neutralized by at least same amount of entropy that is created in some other part (or later) in the system. No other explanation for demon's exorcism is needed.[362]

When Brillouin refers to information, he suggests concept of information that we use in everyday sense, as a measure of knowledge of a system. This is anthropomorphic, but it does not have important place in exorcizing of the demon. This keeps the argumentation reasonable and comprehensible.

Denbigh, among others, criticizes Brillouin's argument that information and negentropy are interconvertible.[363] He argued this could be applied only under the special conditions. However, when these conditions are met, it would become trivial. This triviality comes from the fact that bound information is only label for the fluctuations of entropy. According to Denbigh, what Brillouin did, was simply turning negative entropy into concept of information.

Among other philosophically relevant counterarguments, Biedenharn and Solem[364] pointed out on contradiction between Brillouin's argument and the Third Law of thermodynamics. It is not possible to identify information with negentropy for the information is not temperature sensitive. Still, according to this law, systems should have 0 entropy on 0 temperature.

Biedenharn and Solem's criticism has not been entirely refuted to this day. Moreover, Brillouin's concept of bound information leads us to potentially more problematic results. Let us take case of one molecule gas that is trapped in half-volume placed gas. Take notice we do not have information on location of the molecule. How will Maxwell's demon decide to remove the shutter? For success of his operation, molecules inside of the box must be detected. Brillouin proposes a method for detecting molecules – lamp or torch emitting photons (quanta of light) from hot filament, like in a conventional bulb. In this manner, any molecule that approaches will be detected. The key prerequisite is that these photons must have sufficiently high energies to be detectible above the noise of the thermal background. (let us disregard other problems with perturbations from the outside world, some of which, like the cosmological microwave photons, cannot ever be entirely absent). The torch would, therefore, need to be powered by electricity from the power network, violating the isolation prerequisite, or its batteries would eventually run dry, stopping demon's work. This will result in entropy cost for quanta's energy dissipation which is greater than the previous entropy reduction.[365]

However, these demonstrations, models, realizations, etc. do not and cannot provide a general proof. If we consider demon as part of canonical system, he must fail. Then, we must consider demonic senses as part of canonical thermal system, as well.

---

362 Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.

363 Denbigh, Kenneth George, Jonathan Stafford Denbigh, and Zeh, Dieter. "Entropy in relation to incomplete knowledge." *Cambridge University Press* (1985)., pp. 112–115.

364 Biedenharn, Lawrence. C., and Johndale C. Solem. "A quantum-mechanical treatment of Szilard's engine: Implications for the entropy of information." *Found. Phys.* 25.8 (1995): 1221-1229., pp. 1227–1229

365 Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.

6.5. Engines without demons

In this Section, we will analyze if Szilard's engine could be construed without demons. Thinkers that have criticized informational exorcism, proposed this kind of engines in order to show that an engine could work without implying measurement process.[366]

For instance, Popper has rejected Szilard's understanding of mechanical entropy. The latter takes mechanical entropy as merely subjective quantity. In contrast, Popper's aim was to show that, in a system with demon, we can extract work even if the demon is not intelligent, which would suggest an objective grounding for entropy.[367] Nevertheless, do we really need the concept of information in order to understand Szilard's machine? Some philosophers argued that we do not. The engine can perform the cycle even without a demon. We might not need information as a description of the results of the measurement performed.

The simplest kind of the engine without demons has been described by Feyerabend.[368]

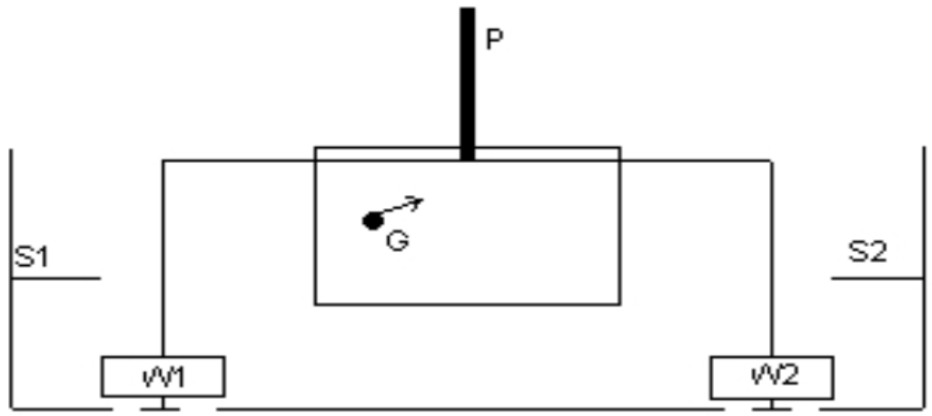

Figure 6.1. Popper version of Szilard's engine. Adapted from "Information and entropy in quantum theory." by Maroney, Owen. (2004).

Here we have one weight on both sides of partition. Weights are on the floor and they are connected to partition. For example, if molecule G is on the left when we insert piston, it will go to the right side and lift W1. If molecule is on the right, W2 will be lifted. Hereby, heat is used for lifting of W1 and W2. Which seemingly implies that in this case there is no need for a demon.[369]

It is not clear if this engine violates the Second Law. Feyerabend considers it as a perpetuum mobile. Popper[370] argues this engine works only in a case where it contains one atom, for it takes only little

---

[366] Feyerabend, Paul K."On the possibility of a perpetuum mobile of the second kind", in *Mind, Matter and Method: Essays in Philosophy and Science in Honor of Herbert Feigel*, in. P. K. Feyerabend and G. Maxwell (Minneapolis, Minnesota: University of Minnesota Press), (1966): pp. 409–412; Jauch, Josef-Maria and J. G. Baron. "Entropy, information and Szilard's paradox." *Maxwell's Demon: Entropy, information, computing* (1990): 160-172.
[367] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[368] Feyerabend, Paul K."On the possibility of a perpetuum mobile of the second kind", in *Mind, Matter and Method: Essays in Philosophy and Science in Honor of Herbert Feigel*, in. P. K. Feyerabend and G. Maxwell (Minneapolis, Minnesota: University of Minnesota Press), (1966): pp. 409–412
[369] Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004).
[370] Popper, Karl R. *Autobiography of Karl Popper*. In P. A. Schlipp, editor, The philosophy of Karl Popper, Open Court 1974, pp. 124-133.

place and hence avoids entropy production. However, only in case where we have more atoms could we discuss expanding gas (as would strictly make sense to speak about "heat" only in the case of macroscopic system composed of many atoms). Chambadal[371] argues we can apply thermodynamically related notions in systems that are contained of more than one object or body. Jauch and Baron[372] criticized validity of this argument on the basis of claim that laws of ideal gas will be violated by inserting of partition. The latter point is also a bit of begging the question, since the validity of the ideal gas laws is contingent upon having a large number of atoms or molecules.

It seems that these arguments accept ability of heat to lift a weight repeatedly without creating entropy. In that case Kelvin's formulation of the Second Law is violated, no matter how many atoms are in the engine or how little energy gain.

6.5.1. Objections to the Popper-Szilard Engine

There are two kind of critics of this engine that we will consider here. First, we will consider the one formulated by Leff and Rex.

Their criticism is grounded in Landauer's Principle. They considered that when one cycle finishes, the position of the piston and pulleys at the moment when the weight is lifted, plays the role of memory device. Further, piston should be removed to start a new cycle, and this step is equal to erasure and it requires a kTG ln2 dissipation.[373] A new puzzle emerges: how to start a new cycle without performing measuring of piston position? Maroney proposed one possible method.

He argues that since we have shelves S1 and S2 which will emerge on both sides when gas expands, they will support weight that is rising whether it is W1 or W2. If we rise W1, the piston will be the right side. If we rise W2, on the left. Therefore, piston could be removed from one side to the center, outside of container, regardless of the side on which it is. He concludes that it would neither create entropy, nor violate Landauer's Principle.

The second objection is Zurek's. He argues that quantum measurement prevents this kind of machine from working on predicted way – an objection which gains force in any situation in which a single particle or a small number of particles are considered. The engine without demon that operates it, exploits location of the molecule without performing any measurement. The aim is to show that it has the potential to work even without any measuring process being performed before it.[374] Situation is more complex if we take quantum objects into account. For in the case of classical gas, if the molecule's location is unknown, it could still be known that it is located on one side of the chamber, not both sides. Quantum molecules can, in principle, be "on both sides" (its wavefunction filling all available space). Only in case we perform measurement will it collapse in one of two states.

Zurek tried to prove that gas could lift a weight only in case it is not on both sides, but one of them and this requires measurement. However, since his proof was based on free energy, it must be assumed that statistical free energy is an adequate measure of work potential, and this will hold only

---

[371] Chambadal, Paul, and MF INGHAM. *Les Paradoxes en Physique. Paradoxes of Physics... Translated... by MF Ingham*. Transworld, (1973).
[372] Jauch, Josef-Maria, and J. G. Baron. "Entropy, information and Szilard's paradox." *Maxwell's Demon: Entropy, information, computing* (1990): 160-172.
[373] Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004), p. 74.
[374] Zurek, Wojciech H. "Reversibility and stability of information processing systems." PRL 53.4 (1984): 391.

if we presuppose validity the Second Law (which is to be proved). Biedenharn and Solem argued that observing itself performs work on gas. In the expansion that will follow, this work will be extracted. The problem is they do not demonstrate how this work is performed.[375] We can conclude that this kind of critic remains incomplete.

6.5.2. Maroney's version of the Szilard's Engine

Maroney modified Szilard's engine that seems to reduce the entropy.[376] In the first step (a) the piston is inserted in box with molecule. Second step (b) molecule press the piston from left side. In third step (c) shelves extend onto both sides, and hold the lifted weight. In fourth step (d) piston has been removed from chamber, but it is still correlated with the location of lifted weight. The fifth step (e) this correlation has been used for piston resetting.

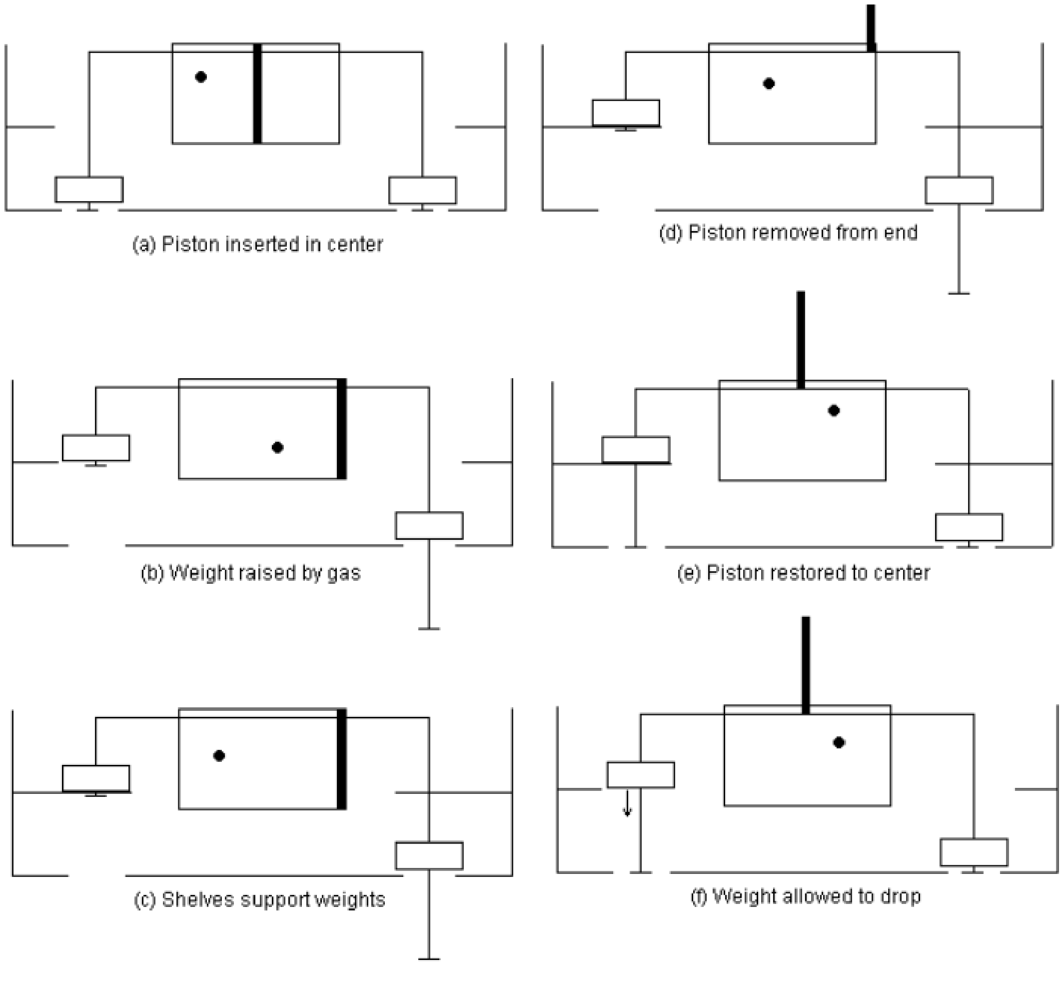

Figure 6.2. The complete cycle in Maroney's version of Popper-Szilard Engine. Adapted from "Information and entropy in quantum theory." by Maroney, Owen. (2004).

Function of memory device that records locations of atom, was performed by location of the raised weight. By removing the shelves, weight will fall down and expand the energy used for its lifting.

[375] Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004).
[376] Ibid.

Engine will be in its initial state, again. This might not be obvious, but we may suppose that the contact with the heat bath is achieved via perfectly conducting walls of the container. Hereby, if weight is related to hotter heat bath than the atom, then it will cool down causing the violation of the Second Law.[377] However, engine cannot function this way in the long term. The question remains: is Szilard Engine adequate or successful paradigm for the entropy-information link?

6.6. Argumentation that demons exist

David Albert went in the opposite direction in the entire controversy about understanding the work of Szilard's engine. Hemmo and Shenker went along and developed his argumentation.[378] They argued that demons could, in fact, exist in the relevant context.[379] The main claim, we have already encountered above, is that Boltzmann's entropy can go down in a macroscopically indeterministic processes. Albert adopted Boltzmann's entropy as the only adequate measure of the thermodynamic entropy. Therefore, if we insert the partition into the center of the box in Szilard's engine, it will reduce Boltzmann's entropy irrespectively of the location of the molecule, since the insertion is an allegedly indeterministic process. The purpose of the "demonless engine" is to show that it is possible to extract the work without the aid of intelligence and measurement.

We saw that Landauer's, as well as Penrose's and Bennett's position will have held that location of the molecule does matter if we want to extract work, but Hemmo and Shenker argue that this is not necessarily the case. In order to prove it, they perform an operation which achieves erasure via destroying information, on an auxiliary system. That operation is supposed to destroy information about molecule's location without any need to pay cost in thermodynamical entropy.[380]

Only in the course of a macroscopically indeterministic process entropy can be reduced. It is not possible to restore with certainty both the system and auxiliary to their initial state without dissipation of heat. However, Albert realized that it is still a constrained violation (in the sense discussed above), and we can have a tamed demon. Therefore, we still have the possibility of a modified Second Law, in the manner of Smoluchowski.

Zhang and Zhang [381] have conceived an example of engine that will function as untamed demon. In their example, the partition is placed into the center of the box, where potential cable, that depends on velocity, creates a pressure in gas. Hereby, Boltzmann's and Gibbs's entropy will be reduced and the phase space compressed. These will be the case even in the macroscopically deterministic processes.

---

[377] Ibid., pp. 76-79.
[378] Hemmo, Meir, and Orly Shenker. "Prediction and retrodiction in Boltzmann's approach to classical statistical mechanics." (2007). http://philsci-archive.pitt.edu/3142/1/Hemmo_Shenker_on_Boltzmann_23Jan07.doc
[379] Albert, David Z. *Time and chance*. Cambridge, Massachusetts, London:Harvard UP. (2001): 1285-1286.
[380] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[381] Zhang, Kechen, and Kezhao Zhang. "Mechanical models of Maxwell's demon with noninvariant phase volume." *Physical Review A* 46.8 (1992): 4598.

6.7. Intelligent demon

If information and entropy turn to be essentially the same, then would it make any difference if demon is intelligent or not? Some philosophers argue that question of demon's intelligence is a feature that does not affect his success nor failure. His success or failure is independent on demon's nature, in sense that he can be either living being or device. [382] On the other hand, characteristic of intelligence was attributed to demon from the start.

Notions of intelligence and knowledge played important role in the Maxwell's thought experiment. Maxwell expressed this in his letter to John William Strutt, devoted on an allegedly "easier" way to violate the Second Law. There Maxwell concludes: "The moral drawn was that:"[t]he 2nd law of thermodynamics has the same degree of truth as the statement that if you throw a tumblerful of water into the sea, you cannot get the same tumblerful of water out again."[383] From this, one can infer that energy dissipation could be reversed and that amount of energy we could use for work depends on degree of our knowledge.[384]

Maxwell explains importance of the demon's knowledge in his famous *Encyclopedia Britannica* article:

"Idea of dissipation of energy depends on the extent of our knowledge. Confusion- is not a property of material things themselves, but only in relation to the mind which perceives them. [...] Similarly the notion of dissipated energy would not occur to a being who could not turn any of the energies of nature to his own account, or to one who could trace the motion of every molecule and seize it at the right moment."[385]

We have already analyzed in detail many mechanisms that are proposed to function as Maxwell's demon. This pertains to the Smoluchowski trapdoor, as well as Feynman's ratchet and pawl and others. All of them must have failed because they overlooked fluctuation phenomena. Could intelligent being operating such a device change it?[386]

6.7.1. Smoluchowski and the naturalization of Maxwell's demon

Smoluchowski has analyzed the case of an engine where intelligent being would intervene.[387] Fluctuations prevent engines from operating on long-term scale, but the question is would they succeed if an intelligent being were to operate them. Such a being does not necessarily need abilities that Maxwell's demon has. It would be enough if we conclude when we should insert the partition into box.

---

[382] Leff, Harvey S. & Rex, Andrew F, *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* (Philadelphia: Institute of Physics Publishing) 2003; Hemmo, Meir, and Orly R. Shenker. *The road to Maxwell's demon: conceptual foundations of statistical mechanics*. Cambridge University Press, (2012).

[383] Strutt, RJ & Rayleigh, JW. *Life of John William Strutt Third Baron Rayleigh*. Arnold, Madison: University of Wisconsin Press. (1924)., p. 47.

[384] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471.

[385] Maxwell, James C., (1878), *Diffusion Encyclopedia Britannica*, 7,214, Reprinted in: Maxwell (1952): 625-646, p. 646.

[386] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

[387] Smoluchowski, Marian V. "Experimentally verifiable molecular phenomena that contradict ordinary thermodynamics." *Phys. Z* 8 (1912): 1069. §20.

Smoluchowski held that such naturalization of Maxwell's demon necessarily leads to his exorcism:

"[...] there is certainly no doubt that an intelligent being to whom physical phenomena are transparent could bring about processes that contradict the second law. [...] On the other hand, it is not to be excluded that the activity of intelligence, the mechanical operation of the latter, is connected with the expenditure of work and the dissipation of energy and that perhaps after all a compensation still takes place."[388]

This is the model of exorcism which has become common in the meantime, almost the textbook version. In first step, the demon is naturalized. Maxwell's demon is observed as a living organism, since he has a certain intelligence. Thus, being a living organism, he must pay an entropy cost (not to mention the implicit entropy cost of the preceding biological evolution leading to its emergence in the particular form; we shall return to related philosophically interesting questions in a subsequent chapter).

Szilard[389] accepted Smoluchowski's line of argumentation, and argued that as a physical system, demon is also determined by the Second Law. He also argued that we could preserve the Second Law (in its statistical form) if we place entropic cost in process of measuring the information. Hence, we need an intelligent being able to locate the molecules and record or remember gathered information.[390] Let us now go back to the basis of Szilard's argumentation.

Szilard accepted that fluctuations exist and presuppose they are subject to probabilistic law. This version of second law is weaker one. Besides, it is analog to Smoluchowski's version. This version does not allow cyclic processes whose results violates original second law. This is coherent with Smoluchowski's considerations.[391] However, these cyclic processes demand to be operated by intelligent beings in order to avoid entropic cost. Still, on the level of other kinds of exorcisms of Maxwell's demon, sometimes it is not clear whether their goal is to defend statistical or absolute form of the second law? Attempts of Smoluchowski and Szilard were directed toward defense of the statistical form of the second law.

There were various kinds of criticism, on behalf of the Szilard's engine. For example, Jauch and Baron complained on idealizations that engine employs are not legitimate, for process of shutting the door violates the gas law.[392] Norton and Earman argue, that they missed the point Szilard made. The variations in gas density and pressure that appear are just fluctuations. Szilard's point is in trying to analyze if intelligent being could accumulate these fluctuations into macroscopic ones.[393]

However, it is clear that both Szilard and Smoluchowski aimed to protect the Second Law from embellished violations, which means that no work could be continuously exploited from macro system. Still, there is one question that remains unanswered, did demon's naturalization serve as

---

[388] Smoluchowski, Marian V. "Experimentally verifiable molecular phenomena that contradict ordinary thermodynamics." *Phys. Z* 8 (1912): 1069 p. 1080.
[389] Szilard, Leo, "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings", *Zeitschrift fur Physik* 53: 840–856. (english translation in The Collected Works of Leo Szilard: Scientific Papers, B. T. Feld i G. W. Szilard (Cambridge, Massachusetts: MIT Press, 1972), (1929): pp. 103–129.
[390] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471.
[391] Smoluchowski, Marian V. "Experimentally verifiable molecular phenomena that contradict ordinary thermodynamics." *Phys. Z* 8 (1912): 1069.
[392] Jauch, Josef-Maria, and J. G. Baron. "Entropy, information and Szilard's paradox." *Maxwell's Demon: Entropy, information, computing* (1990): 160-172., Section 4.
[393] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471.

exorcism by means of classifying him as a thermal system that is governed by the Second Law, or exorcism by means of classifying him as information system, that has hidden entropy cost related to information acquisition?

Smoluchowski choose first option, while Szilard choose the second kind of exorcism. Is Szilard's argumentation just upgrading Smoluchowski's argumentation, or are they opposed to one another? One more question is relevant in this regard: is postulating the entropic cost of information gain independent from the Second Law (or grounded on it)? Szilard held the latter position.[394] We need to review the difficulties with this kind of position.

6.8. Disadvantages of exorcising of demon by information cost

Benefits of exorcising the demon on the account of information, seem to be illusory. This kind of exorcism imposes unsustainable equivalence between the increase in the amount of information and the decrease of entropy.[395] Norton and Earman express it in form of dilemma on exorcism of Maxwell's demon via information. The grounding presupposition of this dilemma is that the demon is a physical system, and thus subject to physical laws.[396] Once we allow that, however, it would be a problematic cherry-picking to separate the Second Law from other physical laws.

6.8.1. Sound versus profound dilemma

Earman and Norton classify all attempts of information-theoretic exorcism as parts of the central dilemma.[397] Demon and the system either form or do not form a canonical thermal system. The demon is either subject to limitations derived from laws of physics, or he is not. Earman and Norton suggest that those who want to exorcise the demon by means of information theory in order to protect the Second Law need to pick either "sound" or "profound" horn of the dilemma.

The first horn of the dilemma holds that both the demon and system are canonical thermal systems, therefore, demon cannot succeed. Still, reason for demon's failure lies in the initial assumption of his failure.[398] It is assumed that the demon is subject to the Second Law – which is assumed to hold for all physical systems.

The "profound" horn presupposes validity of the Landauer's principle. It also present it as an independent axiom, which is not possible to derive from statistical mechanics and from statistical mechanics alone. Further step is to derive the modified Second Law and deduce that untamed demons cannot exist. Does this mean that statistical mechanics is incomplete without Landauer's principle? If true, that would certainly be a deep, "profound" insight into the very foundations of physical

[394] Ibid.
[395] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471.
[396] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[397] Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471; . "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[398] Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

reality. Will untamed demons be possible only on the ground on statistical mechanics, without taking Landauer's principle into account? Norton and Earman argued that, if we do not have independent grounds on which we can base Landauer's principle, we cannot confidently claim that it is not possible to design a machine which will produce unconstrained violations. That a *particular* proposed device does not work as suggested, is certainly not in itself an argument against Norton's and Earman's case; it might be the case that our imagination is simply too limited and the time elapsed since Maxwell and Szilard is too short.

In this case, there is a need for law that could establish validity of the Second Law for the combined system. This law will require independent proof.[399] Such a proof is still been missing.

Any information theoretic exorcisms must fall into one horn of dilemma. It is impossible to belong to both horns at the same time, for demon and system that he manipulates cannot form canonical system and not form canonical system, at the same time. If we choose that they do form canonical system, in which case we choose sound horn of dilemma, it is assumed that combined system is subject to the second law, which prevent entropy decrease in system on long time scale. If information theory exorcises the demon and saves the Second Law, it also built its exorcism by naturalization of the demon. From this we can conclude that sound horn of dilemma does not add anything to the original (unmodified) Second Law.

In a case of informatic-profound exorcism, new principles are added to the Second Law. The problem is, it failed to provide proof of this new principle. Among the thinkers involved in this debate, there are representatives of both horns of dilemma.[400] It is also not clear why at all we need exorcism of the demon? Demon itself is unable of violating weakened laws of thermodynamics. As long as he stays outside of the scope of these laws he could nothing to violate them.

In general, there have been two categories While Szilard, Gabor, and Brillouin argued that this shows that the entropic cost lies in information acquisition, others like Landauer, Penrose, and Bennett held that it was the information *erasure* which has a necessary entropic cost.[401]

6.8.2. Norton's and Earman criticism of information exorcism

Norton and Earman argue that both of horns of this dilemma are ineffective, no matter if they accept Szilard's argumentation that places entropic price in gaining of information or Landauer's argument that places entropic price in erasure of memory. There is no need to connect entropy and information to protect the Second Law, since the hypothesis that the composite system is canonical ensemble (held by proponents of the sound horn), already protects it.[402] However, we should notice its heuristic value: it explains results of Maxwell's demon actions. Still, it could not be generally useful, information does not play role in every single case of demonic actions on system.

From the other side, if we consider demon as an entity separated from the canonical system, there is no way for saving the Second Law's universality.[403] There is no postulate or hypothesis that can be

[399] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[400] Ibid.
[401] Ibid.
[402] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[403] Ibid.

construed to save it from demon. Once we consider systems that are not canonical thermal systems, any confrontation to nonentropic behavior of system is under the question.

Skordos construed an engine that represents a kind of Maxwell's demon. It uses microdynamics dynamics of 2-dimensional system comprising disks and a membrane which is reversible in time and conserves energy.[404] Reason for which we consider it a Maxwell's demon is its production of a density difference. However, disks' dynamics leads to variance of a phase space volume. Skordos tries to exorcise the demon through relationship between information and entropy decrease. His mistake is to consider the demon as holding racket that can be moved to few determined positions in order to deflect the disks, which causes irreversible dynamic, for various paths become non-distinguishable.[405] The inverse of this process would lead to many various initial states.

Hence, Skordos argues that relationship between entropy and information exorcizes the demon:

„Because Maxwell's demon can only operate with finite information (we can think of it as a microscopic computer) it follows that the tennis demon cannot imitate the membrane reversibly. "[406]

This anthropomorphizing of demon is a mistake. The thickest membrane could be imitated by force field, there is no need for anthropomorphizing by adding an animate being. The more important issue that should be considered here, however, is could demon be exorcized via placing the entropic price in the process of gaining the information.[407]

From the other side, Szilard's strategy goes in the opposite direction. If we naturalize the demon, he would be governed by naturalized information theory. The type of information demon could gain as well as the price that has to be paid in entropy, would depend on physical system that contains the demon as well.[408]

However, even if we assume that a general law that postulate the relationship between entropy decrease and information gain does not exist, we can still consider that such a law could exist in *restricted* context of some particular theory. Some thinkers have argued that quantum mechanics could offered us that context. This will be discussed in more detail in the Chapter 8.

We saw that information exorcism could be separated into two categories. The information-theoretic exorcism of von Neumann places the entropy price on the process of information gain. Information gain which makes one able distinguish between different states with same probability, has entropy cost of kT log n entropy. We could call this statement "Szilard's principle", since it has been attributed to him most often. Another approach is based on Landauer's principle (as we have already mentioned) and places entropy cost in the process of erasure of information. By erasure of information stored in the memory device we pay entropy cost of kT log n entropy.

[404] Skordos, Panayotis Augoustos. "Compressible dynamics, time reversibility, Maxwell's demon, and the second law." *Physical Review E* 48.2 (1993): 777.
[405] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[406]Skordos, Panayotis Augoustos. "Compressible dynamics, time reversibility, Maxwell's demon, and the second law." *Physical Review E* 48.2 (1993): 777., p. 783
[407] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[408] Ibid.

If we abandon the demand to stay in scope of canonical thermal systems, we enabled demon to reduce the entropy. Since Brillouin's analysis (mentioned above, in section 6.4) on the quantum character of light, it is possible to hold that Brillouin tends to prove inevitability of quantum demon's failure.

Raymond accepts Szilard's account on relationship of information and entropy, but introduces a different notion of entropy.[409] He introduces notion of entropy which includes basic non-equilibrium systems. Raymond analyzes simple engine that is operated by a demon and has two chambers with the door between them. In the end, he concludes:

> "No observer yet considered has proved capable of storing information in any system without degrading an amount of energy sufficient to make the total entropy change in a system, including the observer, positive. The second law is therefore not in danger through the treatment of information as a form of negative physical entropy."[410]

Still, he did not establish a general result. Here is another example of somewhat more efficient program: Szilard's engine which contains *three* one-molecule engines. Hereby, in the case where every molecule is on the left, demon could use advantage. This state is not probable (but is not astronomically improbable either!), but demon's strategy is waiting for it to take place.[411]

On the long enough time scale, violation of the Second Law will happen spontaneously. Therefore, there is no need for the demon that will process information. Price that has to be paid for the erasure of information from the memory device will not prohibit its violation. Therefore, the main goal should be protection of the Second Law of thermodynamics, not from straight violations, but from the embellished ones. Recall that embellished violations are those in which entropy is continuously reducing and work has been produced constantly. Norton and Earman argued that this kind of protection could be reached by naturalization of the demon as a part of canonical thermal system. They also argued that it shows us notion of information has nothing to add to defense of the Second Law.[412]

The greatest strength of the information-theoretic exorcism is its generality. It is also a weakness, however. Correct understanding of the problem of Maxwell's demon must come from fundamental physical laws, not information theory. However, information-theoretic exorcism at least provides us with heuristic benefits. After all, we are left with question: why would we even have to try so hard to exorcise the demon, when Maxwell himself was on his side?

6.9. Solution of the Szilard's puzzle

Maxwell's original thought experiment has not included considerations about transforming free energy into useful work. Originally, as already mentioned, Maxwell's experiment has two demons: pressure demon and temperature one. The temperature demon operated trapdoor that separated two chambers which contained gas. Demon allowed fast atoms to pass on one side, but not slow ones.

---

[409] Ibid.

[410] Raymond, Richard C. "The well-informed heat engine." *American Journal of Physics* 19.2 (1951): 109-112., p. 141.

[411] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.

[412] Ibid.

Now, let us briefly analyze Maxwell's pressure demon in order to show its similarity to Szilard's Engine. The pressure demon opens the trapdoor to the atom that is approaching from the left, and close for the atom that is coming from the right. Demon operates trapdoor not by performing a work upon the system, but by involving its own internal degrees of freedom or an auxiliary system. As already mentioned, operation of the pressure demon becomes more and more difficult with the passage of time.

We now assume that demon has auxiliary states. Demon now measures the location of the atom and hold the trapdoor open or closed. Chances are equal that atom is located on either of the sides, at any given time. The atom then evolves and goes on the right side, while demon stays in the same state. Entropy of both atom and the demon has increased. We should notice that hereby the demon has a role similar to the one of piston in Szilard's engine.[413]

In this thought experiment, subensembles are correlated to auxiliary system, here it is the demon, and in thought experiment with Szilard's engine it is the piston. This operation can be performed on subensembles, but not on the entire ensemble. However, it is problematically to separate ensemble into subensembles.

Hereby, an attempt of separating ensemble into subensembles is attempt to reverse subensembles mixing and avoid the free energy loss. In other examples of "demonic" contraption, this loss of free energy was represented indirectly through work extraction. Pattern that could be noticed in all of these examples is attempt to increase free energy of ensembles through work on its subensembles.

The relationship between mixing and correlations imposes, but also resolves the puzzle given in Maxwell's thought experiment. In technical term, unitarity of phase-space evolution is causing the mixing entropy. It is only possible to reverse mixing and separate ensemble into subensembles if we include auxiliary system in it. If separation of ensemble to subensembles results in an increase of free energy, we would have at least as much entropy increase of the auxiliary system. This kind of resolution will be important for us later, when we consider how much some complex biological systems actually behave similar to some realizations of Maxwell's demon.

The relationship between subensembles and the auxiliary system should be under control. It would be a mistake if a wrong subensemble correlates to auxiliary system. This could result in compressing of system to 0 volume, instead of producing free energy. Therefore, the machine would break down.

In essence, the problem of Maxwell's demon has its origin in mixing of the subensembles which causes entropy rise. Maxwell's demon has the ability of knowing velocity of each atom, and hence, separate ensemble on subensembles. This reverses the mixing and case entropy decrease. Still, it is not possible for a demon to sort the molecules by unitary operation that would act only in space of the gas. [414]

The auxiliary system needs to be included here. Entropy of the auxiliary system rise at least as much as entropy of the gas reduces. We encounter the same problem considering the free energy. In order to gain free energy from subensembles, the auxiliary system must be included and its entropy rises. The rise of entropy is directly related to free energy acquisition. This is the first step of the soulution to Maxwell's and Szilard's puzzle.

There is no need to place the entropy gain in measurement process, since we already have entropy gain in the auxiliary system. Besides, we need to analyze work that should be derived from thermal

[413] Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004), pp. 171-173.
[414] Ibid., pp: 173-177.

fluctuations. It turns out that fluctuations would show gain in free energy equal to the ensemble's gain in free energy. The pertinent question is: how would such an increase in free energy be used in other systems, for example in any system that is lifting a weight?

Reckon the additional characteristic of weight's position in Szilard's engine. Hereby, the work that has been derived is compressing the weight in a different way. It depends on which one of the subensembles has been selected. Depending on the location of the gas, weight from the one side will be lifted. This is correlation that make imperfect resetting possible. How do the fluctuation probabilities enable that imperfect correlation which inhibits passing of the energy to the heat bath at higher temperature?[415]

The point is that machine needs to move less energy to the rising cycles than lowering ones, on the long-time scale, because of the average length of each cycle. However, probability fluctuation relationship prevents the violation of the statistical Second Law in every system that exhibits Brownian motion. For this relationship is the reason for imperfect correlations. If correlations were not imperfect, system would have implied entropy reduction.

Now, let us briefly consider the thesis that identify entropy with loss of information. Maroney argues that this identification has its roots in dissatisfaction with description of entropy and physical systems within the scope of statistical mechanics. Besides, part of the problem comes from confusing Boltzmann's and Gibbs's entropy and the manner in which they handle thermal fluctuations.

Boltzmann's entropy supposes partitioning of phase space on macroscopically different observable states, each of these having entropy $S_B = k \ln W$. Here, W is volume of the partition's phase space (analogous to the statistical weight discussed above. Thus, if we could refine observational states, we could lower the system's entropy, until because of a fine grained description, it disappears. In Gibbs entropy, we consider an ensemble of states prepared in the equivalent manner. Here entropy is klnp on average for the ensemble. Fluctuation here is represented by separations of ensemble into subensembles. Still, by "zooming in " to the level of particular states, the entropy of the subensembles would dissapear. We should notice that here one should not exclude ensemble description, because entropy would still be present in the mixing of subensambles. The conceptual problem would arise here, for the ensemble is not actually there, but we have only an individual system in a well-defined state. It appears that if it would be possible to completely and precisely define the actual state, entropy would disappear. This is a point where puzzle of the Maxwell's demon arises.

Answer lays in the fact that demon should obey to the laws that holds for the system he is manipulating, because he is an active participant in it. This means being subject to unitary evolution, for the demon cannot decrease the system's entropy if he does not increase its own. We saw through fluctuation probability relationship, that correlation with other system would not change this.

Even if we introduce an intelligent demon, he would be governed by the same physical laws as the system, since he is also part of the physical system. He is governed by unitary evolution and being described as a part of the whole ensemble. This is something which we should have done anyway within scientific realism and physicalism. The intelligent, information processing demon, has no extended and "special" role compared to the auxiliary in the machine without the demon.

[415] Ibid.

6.10. Can Shannon's information be considered equal to the thermodynamic entropy?

A vital part of this deflationary project is resolving the confusion in the realm of physics of information. Newertheless, Shannon-Schumacher information and Gibbs-von Neumann entropy have similar mathematical form, their concern are distinct physical notions. Hence, they cannot be held as equal. We will now explain why.

We explained influence of Landauer's Principle on the information theory in Chapter 3. This influence is important for the relationship of entropy and logical information. We will briefly analyze the nature of relationship between thermodynamical entropy and logical information. It will turn out that these notions are very different.[416]

Recall that the analysis of Landauer's principle above have brought us to conclusion that logically reversible operations are not necessarily thermodinamiclly reversible. Also, thermodynamically reversible operations are not necessarily logically reversible. In these conclusions, we find reason why it is impossible to reduce Shannon's information to Gibbs entropy based on Landauer's principle. We can come to the same conclusion through analyzing notion of Shannon's information and thermodynamic entropy.[417]

Shannon's information could be defined as representation of a system which is in one of a numerous states ρ_a, and such situation occurs with probability p_a, averaged over an ensemble. Shannon information quantify the knowledge that is acquired upon the discovery that the observed state is ρ_a, out of all possible states. In the process of sending the signal, transitions that it undergoes while transmitting is noise. The effect of this will be reduction of the received information.[418]

From the other side, thermodynamic entropy is not sensitive to transitions of this kind, the only condition that has to be met is that ensemble density matrix stays the same. The chance that the state ρ_a will appear is p_a. If we assume that system in the equilibrium state is related to the heat bath it is permitted to go through transitions that include every possible state. This will not change neither density matrix, nor thermodynamical properties of the system. This is the point in which it completely differs from the Shannon information.[419]

What entropy and information *do* have in common is functional form. This holds in both classic and quantum mechanics. This certainly has practical benefits, since some of the results acquired in information theory could be used in thermodynamics if we are careful enough to keep track about the meaning of symbols and quantities. Also, there are limiting cases where notion of Shannon information and Gibbs entropy would *appear* as same, but even though they are functionally similar, for the reasons we explained above, it would be a mistake to consider these two notions as equal. In fact, this careless behavior is responsible for much of the philosophical confusion surrounding information-theoretic exorcisms.

Now, after examining algorithmic complexity, we will consider another kind of complexity of great scientific and philosophical interest – the complexity of living systems – in order to see which role

[416] Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004).
[417] Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004).
[418] Ibid.
[419] Ibid.

information can play there, and do they behave similar to Maxwell's demon or not. We will also compare behavior of Maxwell's demon with the Darwinian evolution and its specific mechanisms.

7. Maxwell's demon, entropy and complexity of living systems

"How remarkable is life? The answer is: very. Those of us who deal in networks of chemical reactions know of nothing like it...How could a chemical sludge become a rose, even with billions of years a try?"

George Whitesides

"At first, it seems as if the existence of complex life forms on Earth violates the second law. It seems remarkable that out of the chaos of the early Earth emerged an incredible diversity of intricate life forms, even harboring intelligence and consciousness, lowering the amount of entropy. Some have taken this miracle to imply the hand of a benevolent creator. But remember that life is driven by the natural laws of evolution, and that total entropy still increases, because additional energy fueling life is constantly being added by the Sun. If we include the Sun and Earth, then the total entropy still increases."

Michio Kaku, Parallel Worlds: A Journey Through Creation, Higher Dimensions, and the Future of the Cosmos

In this Chapter, we explore relationship between Maxwell's demon and entropy on the one side, and complexity of living systems and biological evolution on the other. In order to do so, we will first give a brief overview of living systems and their evolution.

In Section 1 we will explain the concept of living systems. In Section 2, we shall outline the relationship of evolution and complex systems.

In Section 3 we will go back to the concept of information to examine its relation and role in living systems and through that analysis we will start to draw the similarities of evolution and Maxwell's demon. In Section 4 we will take one closer look on the Darwinian evolution and similarity in behavior that it shares with Maxwell's demon. We will continue to analyze this relation in Section 5 by means of evolutionary dynamics, and in Section 6 through mechanisms of imitation and learning. Finally, in Section 7, we will outline some conclusions.

The thought experiment with Maxwell's demon can be related to complexity of living systems and evolution in many ways. Maxwell's demon has inspired Mark Ridley to imagine "Mendel's demon" who works with genes. Mendel's demon is able to manipulate the gene inheritance. He decides if a gene would be passed down to the offspring or not, similar to the decision of Maxwell's demon whether to allow a molecule through the partition or not.[420]

At first, it might seem that evolution of species (as well as the cosmological evolution) contradicts to the rise in entropy and moving toward the state of equilibrium. At a glance, it looks as if the Darwinian evolution acts like Maxwell's demon. However, biological systems are living far from equilibrium of isolated systems. Observe the wider background: as we have seen in Chapter 4, there have been important entropy-increasing processes taking place since the Big Bang. In particular, gravitational structure formation occurred, creating galaxies, their groups, rich galaxy clusters, superclusters, and so-called "large-scale structure" consisting of sheet- and filament-like superclusters separated by huge voids – and all that uniformly. On a local scene, it has been continuous star formation which provides for chemical enrichment and creates habitable planets as sites for biological evolution. If our complete universe is evolving toward the state of heat death, which would make any kind of living being impossible, how come that we experience evolution on complex life on Earth seemingly progressing over time?

As long as we understand the Second Law as a statistical regularity, it does not mean that processes that decrease the entropy are prohibited, although as explained this will correspond to an *extremely* low-probability occurrence. Still, the forming of order in one subsystem, needs to be compensated with rise of entropy in another subsystem. Life forms on Earth are open systems that pay entropy price via exchanging the energy with their surroundings.[421]

In particular, life forms on Earth are far from equilibrium for their consumption of free energy from outside of their local ecosystem. They are possible, for the energy (high-energy optical photons), provided by the sun, and Earth, including its biosphere, re-radiates high-entropy form (low-energy infrared photons) back into the universe. If we consider the immediate local environment of a living creature on Earth as its system – of the kind discussed in previous chapters – then such systems can never be isolated; they *must* be open, instead. Formation of clusters and galaxies and stars is also compensated by entropy rise elsewhere in the universe, notably in its expansion and in formation of black holes as sites with extremely high entropy.[422]

An interesting question related to biological evolution is: can it be used as criterion for local time's arrow? If we take Dollo's principle into account, we assume irreversibility of evolution.[423] This principle asserts that if through evolutionary processes some function or species or organ is lost, it could never appear again (in the same form). However, since then it was proven in a special, laboratory context that evolution could, on occasion, be reversible. Thus, we can conclude that Dollo's principle is just an empirical generalization, not an expression of a deeper dynamical law.

There are further reasons why evolution is not a good criterion for defining an arrow of time. The most important is that best representation of evolution is by aims of evolutionary three in which branches are both currently living and extinct species.[424] Also, we now know that there are evolutionary mechanisms which do seem symmetric in terms of both time and complexity; the clearest example is the so-called *horizontal* (or *lateral*) *gene transfer*. While this mechanism played very important part in the history of life and is ubiquitous among microorganisms, it is also known to operate on all scales of complexity (e.g., gene transfer from bacteria to insects or from retroviruses

[420] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[421] Mainzer, Klaus. *Symmetries of nature*. Walter de Gruyter, (1996), p. 529.
[422] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[423] Ibid., p. 144. Epistemic status of Dollo's law is, as is the case with any other "laws" in biology, highly controversial. See, e.g., Gould, SJ 1970, "Dollo on Dollo's law: irreversibility and the status of evolutionary laws" J. Hist. Biol. **3**: 189-212.
[424] Weinert, Friedel. *The Demons of Science*. Springer. (2016).

to humans), thus short-circuiting attempts to use genomic complexity as unidirectional pointer. This topic has not been elaborated in the literature on philosophy of biology thus far.

7.1. What is life?

In his celebrated book, *What is Life?* Schrödinger tried to explain processes in biology and physics. He noticed that life includes two basic processes: “order from order” and “order from disorder”.[425] He took notice that gene (which is later discovered to consist of DNA) governed a process that produced order from order in species, this is depicted in fact that ancestors passed traits to their offspring. It seems that the other process (order from disorder) contradicts the Second Law of thermodynamics, because in closed systems, entropy should increase. Living systems, however, are the antithesis of such a disorder. Schrödinger offered solution from the perspective of nonequilibrium thermodynamics, in other words, he noticed that life forms exist in a physical world explicable (at least in principle) by statistical mechanics.

We should not expect that living organisms will obey to the Second Law of thermodynamics because it only applies on closed systems, that living beings certainly are not (they require free energy from their environment). If we isolate living beings from environment, they will die. Later, their remains will come to state of maximum entropy, in other words in the state of thermodynamic equilibrium. This is the counterargument most often given in discussions. However, despite its formal exactness, it gives solution for only one, philosophically less interesting, part of the problem, first formulated by Schrödinger. The really interesting part of the problem is the one related to the phenomena of life that does not present only local decrease of entropy (on the cost of entropy increase elsewhere), but systematic decrease of local entropy followed by increase in complexity, namely creation of new structures. How could it be the case that such a class of phenomena, even they are in accordance, ever come into existence?

7.2. Evolution and complex systems

David Krakauer has suggested a basic principle for every adaptive system (those systems whose better adaptability is passed dawn on next generations), the “demonic selection principle”, that states organisms could not be more complex than their environment. This basic principle has far-reaching consequences. Besides, processes of global increase or local decrease of entropy, could be used to define “past” and “future” independently from other considerations.

Is it possible to explain living forms on Earth on both micro- and macro-evolutionary levels by increasing complexity of biological lineages? What is complexity, after all? What do we learn by measuring it? By measuring it we learn to organize and classify systems. It makes us understand specific systems better and facilitate development of predictive hypothesis.[426]

Darwin had written a lot on the topic of diversity, but less on complexity. It seems like the phenomenon of stereotypy (morphological similarity of different species) that usually occurs in biological taxonomy undermines the relationship between diversity and complexity: if there were

[425] Schneider, Eric D. and James J. Kay. "Life as a manifestation of the second law of thermodynamics." Math. Comp. Mod.19.6-8 (1994): 25-48. , p. 25.
[426] Machta, John, Natural complexity, computational complexity and depth, *Chaos* 21, 037111 (2011).

billions of insect species with similar form or physiology, does it mean that the whole taxon has the same complexity as some other taxon with only a few species?

However, it is a challenge to explain basic evolutionary change in a few lineages that evolved highly complex mechanisms of reasoning or other cognitive phenomena. This would amount to illuminating the roots of complex information processing mechanisms.[427] There are two approaches to analysis of evolutionary complexity. One takes functional features and explores tendencies and correlations in various lineages over time, such are the increase in genome and cell size, or diversity. Second approach is a more neutral and reductionist one, it analyzes increase in numbers of components or interactions between genetic material.[428]

Both approaches came down to the same conclusion: complexity is often limiting itself. Besides, with number of components, grow number of their constraints, which limit their further development of diversity. Lineages should never fall below a minimal complexity needed for life maintenance (we do not consider alive organisms that are simpler than bacteria, like viruses or prions), there is no known upper limit for complexity.[429] In other words, the thermodynamic price (expressed in increase of global entropy) that needs to be paid for evolving toward complexity is still unknown and should be further examined by biologists, physicists, and philosophers.[430]

7.3. Information as language of life: similarities of evolution and Maxwell's demon

The question one may ask is: how can we increase the amount of information encoded in a physical system when we start from very simple initial state and have only a limited amount of free energy at our disposal? Formal structure of this problem is similar to the problem of undecidability in mathematics. This problem also shares its form with problem of Maxwell's demon – entropy increase and information loss in any coarse-grained isolated system. It is related to the key puzzle in biology: how do we distinguish living from non-living complex systems?[431] It is tantamount to ask for the really general definition of life, project which is notoriously difficult one in philosophy of biology.

Some philosophers, like Wiley, argue that speciation follow the same pattern as the Second Law, since evolution of the species goes further and further from the thermodynamic equilibrium.[432] Thus, these aspects of the living systems could be employed in a kind of descriptive definition of life. Others, like Krakauer, argue that it is analogous to the Szilard's interpretation of Maxwell's demon, where a molecule is placed in one or the other part of the container and corresponds with state "0" or state "1".[433] We have multiple examples of processes in which entropy is reduced; one of the most important is process of the cell differentiation that leads to morphological differences between cells

---

[427] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110.
[428] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110.
[429]Ibid.
[430]Ibid.
[431] Davies, Paul CW, and Sara Imari Walker. "The hidden simplicity of biology." Rep. Prog. Phys. 79.10 (2016): 102601.
[432] Depew, David J. "Nonequilibrium thermodynamics and evolution: a philosophical perspective." *Philosophica* 37.1 (1986): 27-58, p. 28.
[433] However, we must recall that key difference is that transfer entropy is based on the standard conditional (Bayesian) probabilities, while the information flow is based on interventional conditional probabilities.

in organs and tissues in eukaryotes.[434] We already saw, why it seemed at first that demon did not spend any free energy, but later we concluded that the thermodynamical price must be paid. Now, we need to couple this with the fact that – on the macroscopic level – living systems are far from the equilibrium. Therefore, in order to make the balance in entropy, metabolic processes that produce energy, dissipation of heat and dissipation of waste product must be involved.[435]

Life forms and their evolution in direction of greater complexity and increased order need energy from the Sun. This energy is carried by photons which have much higher individual energy (being in the optical and near-ultraviolet part of the spectrum) than those (mainly infrared) that the Earth sends back in universe. The difference in entropy enables life forms.[436] The complete amount of energy stays the same, but "degrades in quality", since it is increasing the entropy of the universe. Nevertheless, life forms on Earth get free energy and this open system tells us something relevant on the relationship between problem of Maxwell's demon and evolution of living organisms.[437] As Davidson formulates it: "Entropy reduction contributes arise from internal self-organization, information storage and transfer, at a single cell level, as the only way to reconcile this with laws of thermodynamics is by balancing these free energy changes with metabolic energy expenditures."[438]

Maxwell's demon sorts atoms by their velocity, the same way natural selection sorts genomes by their adaptive values. As we saw, the price in evolution is being paid through usage of free energy sources (such as Sun), which provide for the primary productivity of biosphere. What is not clear in analogy is: what has happened with obtained information which demon must save/memorize somewhere? Krakauer suggests that it is placed in more adaptable genes in biosphere.[439] Therefore, this is related to what is, in the context of philosophy of biology and astrobiology, known as the *coding concept*.

As is widely known, nucleic acids that transmit genetic information have two varieties: DNA and RNA. DNA is better for information transfer since it is more stable against perturbations from the environment. Through DNA and RNA all living organisms pass genetic information coded through the universal genetic code inherited from the common ancestor to their offspring. DNA is similar to information in the context of computer science in sense it has dual role: it must act both as a software program to be read out, and as hardware to be replicated. This was ingeniously understood by John von Neumann even before the role and chemical structure of specific nucleic acids were known. First, DNA has to transcribe coded instructions for making proteins (software). Second, DNA is copied during the replication process (hardware).[440]

---

[434] Davies, Paul CW, Elisabeth Rieper, and Jack A. Tuszynski. "Self-organization and entropy reduction in a living cell." *Biosystems* 111.1 (2013): 1-10.
[435] Ibid.
[436] Weinert, Friedel. *The Demons of Science*. Springer. (2016).
[437] Mainzer, Klaus. *Symmetries of nature*. Walter de Gruyter, (1996). chapter 4.44.
[438] Davies, Paul CW, Elisabeth Rieper, and Jack A. Tuszynski. "Self-organization and entropy reduction in a living cell." *Biosystems* 111.1 (2013): 1-10, p.1.
[439] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110.
[440] Davies, Paul CW, and Sara Imari Walker. "The hidden simplicity of biology." Rep. Prog. Phys. 79.10 (2016): 102601, p. 9.

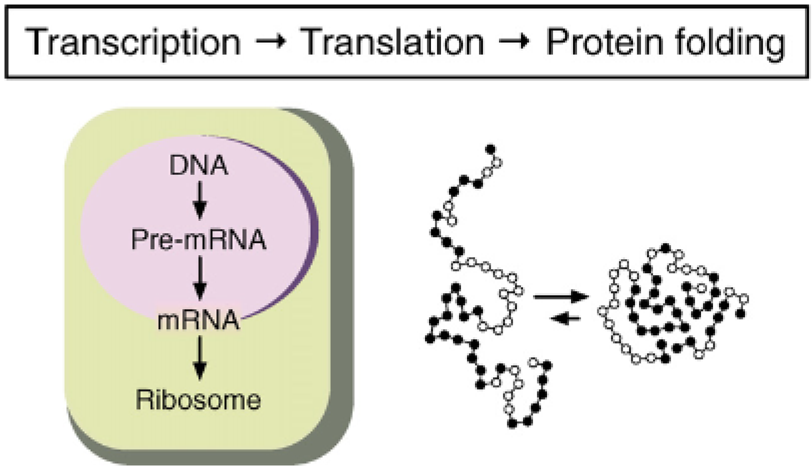

Figure 7.1. Schematic processing of genetic information leading to protein synthesis. Adapted from "Self-organization and entropy reduction in a living cell." by Davies, Paul CW, Elisabeth Rieper, and Jack A. Tuszynski. (2013).

In Figure 7.1, we see processing of genetic information. It represents classical "dogma of molecular biology" (named as such by Francis Crick): DNA carries information which undergoes replication and subsequent transcription into RNA. Then we have gene expression which consist in translation of information written on RNA. Then information is passed to mRNA and onto ribosome which performs the synthesis of proteins. (In rare cases, RNA can perform both functions in a limited context; this is the justification for the "RNA world" hypothesis about the origin of life.) On the right-hand side, we have represented folding of a protein. Complex motion of folding and unfolding of proteins are also interesting for the problem of Maxwell's demon, since these may be a result of Brownian motions.[441]

We have already mentioned similarity Maxwell's demon share with evolution is sorting molecules the same way evolution is sorting genomes. Genomes of eukaryotes contain instructions for building specific type of cells – neurons.

We can summarize the key points of Krakauer's argument as follows:

1. Maxwell's demon should not be closer to equilibrium than the system it manipulates. Selection mechanisms have filtering which limits the target level of complexity. These mechanisms also should be at least of same information richness as the genome they target.

2. The Darwinian demon, similar to Maxwell's demon who maximizes information on location of molecule (and/or their velocity), maximizes genomic information. Another similarity is the energy dissipation, which Darwinian demon does through morbidity.

3. Error correction remains a problem, especially with increasing complexity of the genome. The mechanisms for error correction are quite heterogeneous, some of them being surprisingly efficient (e.g., in extremophiles). The free energy budget has been used for error correction in genome.

[441] Ibid.

4. In evolution, it is a function of mutual information between genome and environment distributions which needs to be optimized, taking into account that the environment might consist of other genomes as well.

5. Constraints established by the environment could be overcome by mechanisms such as plasticity or learning.

6. According to the selection principle, the demonic filter should be determinable from the complexity of agent. Hereby complexity could rise through niche construction.[442]

Here we perceive that Krakauer's argument explicitly takes into account ecological component of evolution – something that is recently accepted in the neodarwinian theory ("eco-evo" paradigm).

Finally, relationship between entropy in cells and information can improve our understanding of the disease (especially those caused by genetic mutation). An example is cancer since it represents changes on the molecular level which we can understand in terms of a *local* gain in entropy.[443] Maybe future biological software engineer will be able to study organism's logical or informational structure and explain what is happening at the global level. This engineer may even be able to work out what might go wrong in, for example cancer and fix it by repairing the informational circuitry.[444]

It has even been discussed that predictable, general laws classical biology still miss (as mentioned with the spurious regularities such as Dollo's Law)[445] could come from the realm of information theory of information dynamics.[446] Since the biological processes differ, their rules differ also. Similarly, as components of biological processes change, rules change with them. This constitutes central part of the epigenetics. For example, cancer will radically change gene and it will cause changes in its trajectroy.[447] Now if we recall that DNA functions as both hardware and software, we can easily see why Davies and other philosophers of biology predict that concept of information might be the solution.

So, a straightforward question is: is natural selection relevant for both hardware (are both psychological and informational structure something that can undergo the process of natural selection? If it were not, software could never become so complex. Evolution was comprehended in the terms of information that has been accumulated. Now, we have realized that the reason for fitness of the organism can be understood through the mutual information that has been exchanged between the genome and the environment. Now, the question is: do the accumulated information affect fitness directly?[448]

Analysis of the role of information in evolution can help us to understand biological organisms better. Lizier and Prokopenko have used some computer simulations (using cellular automata) for the research of this topic. Results of their research were that we can understand behavior of living systems by changes in both structure of information and causation.[449] Their research and other studies

---

[442] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos* 21.3 (2011): 037110., p. 2.

[443] Davies, Paul, Lloyd A. Demetrius, and Jack A. Tuszynski. "Implications of quantum metabolism and natural selection for the origin of cancer cells and tumor progression." *AIP advances* 2.1 (2012): 011101.

[444] Davies, Paul CW, and Sara Imari Walker. "The hidden simplicity of biology." Rep. Prog. Phys. 79.10 (2016): 102601.

[445] Biology has some overt simplifications like Mendel's laws of genetics, the widespread appearance of fractal structures and the arrangement of leaves in Fibonacci series. Also, there is Darwinian evolution, which is not predictive in the general case (there are exceptions in simplified conditions).

[446] Davies, Paul CW, and Sara Imari Walker. "The hidden simplicity of biology." Rep. Prog. Phys. 79.10 (2016): 102601.

[447] Ibid.

[448] Davies, Paul CW, and Sara Imari Walker. "The hidden simplicity of biology." Rep. Prog. Phys. 79.10 (2016): 102601.

[449] Ibid.

performed in the emerging field of artificial life rise hope that same method might shed more light on topics we discuss.

7.4. Darwinian evolution

We can consider natural selection in Darwinian evolution as some kind of filter which eliminates species or individuals that are not well-adapted to the environment. Natural selection shares similarities with Maxwell's demon, which means it can gather information on generations of species and manipulate them, which will result in formation of better adapted genotypes and phenotypes in later generations.

Both Maxwell's and selective demons act iteratively and change the system they manipulate (or target agents in a case of selective demon). Also, we can relate the simplified thought experiments with particles to the notion of *evolvability*, a very important concept which can help us evaluate various methods of gaining adaptive value.[450] Evolvability measures something that can be intuitively understood as flexibility of the biosphere, and as such is a function of the coding concept mentioned above. We can understand evolvability as directly proportional to number of possible methods to gain an adaptive advantage or to "climb" the adaptive peaks in the landscape; this is equivalent to the number of partitions that demon closes or open, depending on his insight in the state of molecules.

In natural selection, genomes within the generations have the functions that enhance them. Evolution explains changing of genotypes on long timescales, via mechanisms of natural selection and genetic drift. Demon that manipulates complex correlations between individuals, populations, generations, and species, and calculates adaptive maps could have at least heuristic value.

7.5. Dynamics of evolution

Darwin and Wallace provided a mechanism that explains modification in form of living organisms as cumulative effect of various small changes that are taking places among generations. The features in organisms of populations change over time. Darwinian evolution explains how this happens by means of inheriting the features from ancestors and adaptability of these features in different environments. As it turns out, we could interpret principles of selection and evolution as distributions of probability, correlating them with Shannon's information.[451]

The principle of selection could be demonstrated this way: information that is gained to perform adaptive behavior stands in direct proportion to the life loss. In order to find the maximal adaptive value, all resources are channeled in one single genotype. In this manner, thermodynamics imposes limits on evolving of life forms. Ability to evolve to better genotype includes going further from the equilibrium. Entropic price that the Darwinian demon hereby has to pay is contained in loss of information on environmental states.

---

[450] Sober, Eliot, *Filozofija biologije,* Plato, Beograd. (2006).
[451] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110.

Also, we can say that Darwinian demon can "select against" and "select for" some feature. Darwinian demon select for complementary and against noncomplementary genomes.[452] Similarly, Maxwell's demon selects for fast molecules and against slow ones.

To increase the amount of information that an organism can carry beyond environmental limit, new mechanisms must be developed. The kind of mechanisms that could overcome these obstacles is *cognitive*: while cognitive abilities rise, whole complexity of organism rises, as well as that amount of information that could be safely stored, processed, and transmitted. Since, the environment is evolving, in contrast to the case of Maxwell's demon, the Darwinian demon has to change preferred traits with time. There are two reasons for this. First, there is dissipation in populations the demon selects. Second, demon is modified by these populations.

We saw that dynamics in evolution maximized information content in genome, to the amount that free energy can support. This means that aim of selection is to minimize conditional entropy. This requires further information, which is not already present in the genome; the proposed kind of selection is actually collecting all such information about the environment and its possible interactions with the organism in question. This indicates insufficiency of purely gene-centric view of evolution, such has been traditionally associated with the studies of Richard Dawkins, George C. Williams, and Daniel Dennett. Following Krakauer, we shall define system complexity as: "the maximum of the difference between the information in the genome and the equivocation."[453] This quantity represents maximal informational content that the demon could transmit from its memory into genome (in other words, it is mutual information). Demon holds in his memory all the information that genome contains.

The rise of information content of the environment is in direct proportion with complexity of the genome. The measure of complexity is based on an ensemble. As such, it requires measurement of genomes in different environments. One question arises on this account: since mutual information can be considered as an indicator of evolutionary complexity, why species vary in their complexity? Why some of the species become "living fossils" and be able to have nearly fixed configuration, while in some cases complexity is reduced over evolutionary "deep time" leading to extinctions?

1. **The diversity of complexity**: We could explain simplicity as low variability of environment. Organisms experience different environments, however. In some environments, there may be no information left to be extracted. Thus, if there are environments that differ by their level of intrinsic complexity, same should be held for organisms which inhabit them.

2. **The constant complexity**: If the environment is not changing on long timescales, so should not the organisms. This is probably an explanation for the empirical data on, say, living fossil fishes such as the *Coelacanth*, which has changed little in the last 400 million years.

3. **The complexity decrease**: Two cases can lead to decrease of complexity. In first place, it is decrease of states of the environment that are probable; an excellent ecological example of this is the simplification introduced in many ecosystems around the world by human use of intensive agriculture in the last 8000 years or so. The second case is rise of noise or a decrease of free energy that is accessible for error correcting.[454] This might have occurred in most of the previous mass extinction episodes in the history of life. Consequent macroevolutionary consequence might be switching to an entirely different ecological and behavioral pattern, for instance switching from K- to r-reproductive strategy or the so-called faunal overturns, noticed by paleontologists.

---

[452] Ibid.
[453] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110., p. 2.
[454] Ibid.

Now, we will outline mechanisms responsible for overcoming the initial complexity. The Darwinian demon could be internalized via cognitive mechanisms that are constantly evolving. Besides, the demon dissipates energy not only through mortality, but also through neural processes. Neural structure and neural plasticity enable species for learning that can be achieved not only on the population level, but on the individual level as well. However, population selection has its special task, of establishing mechanisms that make adaptation and plasticity possible.

7.6. Imitation and learning

Hereby, we will conclude that problems which the demon encounters while selecting are the same ones that cognitive systems meet. This insight shows the central importance of nested selection mechanisms, which could enable all of the adaptive processes. Still, it is important to notice that computational limitations that selective demon imposes, probably cannot be overcome by learning (this is analogous to a NP-hard problem in computing, since it is not clear whether obstacle in solving the problem appears due to our lack of knowledge at this stage of scientific and evolutional development or is problem itself inherently unsolvable in some acceptable time). It should also be noticed that what is learnt cannot be more complex than the reward signal.

The reward signal is indicating the best strategy that should be adopted. The more successful the strategy, the better the reward. Learning through imitation shares the architecture with evolutionary dynamics. Both of them have limited resources. In the case of evolution, individuals die for adaptation to be established among the population, while in imitation learning, strategies are being abandoned. Hereby we lose information not due to individual genome that transmits it. We lose information due to abandoned strategies.[455] The greater the reward behavior receives, the greater are the chances for its fixation in the form of a ("good") strategy.

Organism could be regarded as hypothesis on the current state of affairs in environment. If they survived, experiment successfully verified the hypothesis, and genotype. The demon will select for those genotypes and phenotypes that are best at locating sources of the free energy. This is valid until the environment itself changes, which in normal circumstances occurs quite slowly (and is thus an analogue of quasi-stationary changes of temperature in classical thermodynamics), on timescales longer than the duration of a generation of organisms.[456]

We have seen in previous chapters that Maxwell's demon is usually represented as computational entity. The Darwinian demon personifies selective pressures with complex ecological correlations. In the final analysis, this would have to encompass the entire terrestrial biosphere, and even beyond it, the cosmic environment of our planet, since no ecosystem could realistically be modelled as an isolated "closed box". Till recently, origin of selection has rarely been object of analysis, traits and limitations that selection imposes were far more often under the focus. However, in order to comprehend evolutionary complexity that constantly rises, it should be explained why selective filter becomes more and more narrow.[457]

Living organisms originated from high entropy. Ecosystems have various network with complex correlations and structure. One can ask how simple life forms have selected such a complex, adaptive mechanism? How rise in capacity to develop and carry a code appeared?

[455] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110.
[456] Ibid.
[457] Ibid.

In few past years, there were attempts to explain the emergence of selection mechanisms themselves. This effort centered on the developmental and ecological aspects of the evolutionary processes as much as the conventional mechanisms of such processes ("eco-evo-devo" approaches). While first step was to develop explanation of selective values and its effects on genotype distribution and establishing, further research requires explanation of its origin and distribution. This could be achieved through explanation of *niche construction*. It could be accomplished by determining the basic biochemical processes first and considering selection pressures as made from various biological interrelations.[458] It is important to notice that organism also construes selective bottleneck, hence causing changes on the population level (on local level). Thus, adaptation that is successful would establish itself.

It should be noted that niche construction has one more interesting trait. By imposing some kind of recursion in selective dynamics it could develop complexity of unbounded level. The top of construction of the niche in environment is the phenotype. The phenotype of one organism puts selective pressure on the genome.[459] Besides, we can even interpret developing of cell or cell division and differentiating of multicellular organisms as part of construction of niche. There are genotypes which are encoding rules for nest building or even exceedingly complex behaviors such as woodpecking.

Genomes that can construct demons could be the solution for rise of amount of exchanged information between organisms and their surroundings. Besides, not every lineage would have some other ability than to develop niche construction in order to survive. In this manner, we left open space for macro-evolution on the individual and population level. There are many directions to continue the research, for example, if niche construction is solution for explaining the complexity of the evolution, then we must have broader formulation of the construction algorithms that control development of organisms and their environment.

7.7. Conclusion

Darwinian selection has similar structure to the way Maxwell's demon operate. Like Maxwell's demon puts constraint on distance from thermodynamic equilibrium, Darwinian demon put constraints on amount of information that genome could transmit. Dynamics of evolution tends to maximize the information that genome could hold. Mutual information, in other words information shared between environment and organisms could be taken as measure of its evolutionary complexity. Amount of information that genome could carry could be raised via mechanisms of learning and plasticity.

Still, complexity of learning could not be greater than that of rewarding signal. Process of evolution as well as mechanism of learning have limited complexity established by environment. Environment could be changed by niche construction or by activity of organisms, thus extending learning capacity. Niche construction rises mutual information by rising accessibility to low entropy, fine grained resources from the environment. Thus, complexity could be measured more precisely ecologically, than structurally. It is clear that ideas related to the classical puzzle of Maxwell's demon have a large role to play in this new kind of macroevolutionary synthesis.

[458] Ibid.

[459] Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110.

After these brief remarks on the wider aspects of Maxwell's demon from the perspective of biology, we will once again return to the realm of physics, to see if *quantum*, instead of classical, mechanics can offer to us some new interesting solution to the old puzzle of Maxwell's demon.

8. Maxwell's demon and quantum mechanics

In this Chapter, we will briefly discuss some solutions for the problem of Maxwell's demon offered in the field of quantum mechanics and explore the role they might play. Since it is almost universally accepted that quantum mechanics offers the best description of the microscopic world, on the level of individual atoms and molecules, it is highly relevant for the problems such as Maxwell's demon which reference such microsystems.

In the systems of quantum mechanics Hilbert's space of states substitutes for the classical phase space. Dynamic evolution of quantum states generally unfolds as a unitary process (represented by a unitary operator); what happens next is, unfortunately, contentious depending on the accepted interpretation of wavefunction and its relation to physical reality. According to the standard ("Copenhagen") textbook interpretation, there is a doubly stochastic transition if and when wavefunction collapse ("measurement-like interaction" or simply "measurement") occurs. This significantly impacts the structure of quantum state space for composite systems, comprised of many subsystems. Volumes of chunks of the state space have dimensionality of the smallest subspace containing the chunk. This means that Hilbert's space of composite systems consists of subspaces of individual sub-systems or components. An analogue of Liouville's Theorem holds for unitary evolution, i.e., for everything that happens between measurements. A consequence is that any measurement can possibly increase, but not decrease, the state space volume. In this chapter we want to stress that the differences between classical and quantum account of Szilard's engine and Landauer's principle.[460]

8.1. The role of quantum mechanics

The main additional complexity in quantum theory were find in the field of relationship between measurement and irreversibility. Analysis of classical problems from the perspective of quantum mechanics led to some interesting claims.

Von Neumann[461] was the first to discuss irreversibility (of wavefunction collapse) in the measurement process. While doing so he was referring to Szilard's classical argument. Measurement procedures of both Brillouin and Gabor– that is, using light to detect molecule's position and velocity – by definition require the quantized treatment of light (i.e., oscillating electromagnetic field) to produce dissipation of energy. (After all, the very concept of "photon" makes sense only in the

[460] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[461] Von Neumann, J. *Mathematical Foundations of Quantum Mechanics* (Princeton, New Jersey: Princeton University Press). [1932] 1955, Chapter V.2

context of quantum physics.) Gabor claimed that measurement via classical electromagnetic fields can be dissipative. Therefore, it will lead to violation of the Second Law. For Bennett and Penrose there is no need for untamed demons, because there is no need to perform measurements to produce heat. Still, it is possible to argue that in multiple quantum measurements some heat must be dissipated. Nevertheless, the fact that some quantum processes do dissipate heat to produce measurement is not a valid argument against Bennett's claim that there are measurements that do not.[462]

Besides, thermodynamical irreversibility is often associated with wavefunction collapse, which might seem as contradiction, due to a fact that Szilard's engine should lead to possibility of microscopic entropy, while quantum mechanics should lead to opposite.[463]

8.2. Some solutions offered from the perspective of quantum mechanics

The attempt to exorcize the Maxwell's demon by means of quantum mechanics was persevering, but not widely accepted. Could on this line of argumentation principle, that Earman and Norton asked in their "profound" horn – a deeper, foundational principle that could establish Szilard's and Landauer's – be found?

From the other side, those who agree that quantum mechanics provides aims for exorcizing the demon, do not agree neither on strategy for exorcizing, nor on the aspects of quantum mechanics on which the exorcism should be based. The main weakness of those arguments is that they provide aims for exorcizing within the realm of quantum mechanics, but presuppose that such an exorcism is universal.[464] This might be unwarranted in the generalization to the quantum field theory, as well as in other aspects mentioned (e.g., in trying to account for the cosmological arrow of time, or the origin of biological information).

Slater argued that Heisenberg uncertainty principle blocked Maxwell's demon.[465] His interpretation of the demon does not include being that manipulates perpetuum mobile through manipulation of the molecules. He represented demon as a being that applies the reversibility objection of Loschmidt through the reversion of molecules velocities. Slater argued that the principle which holds that it is not possible to determine position and velocity of particle with same precision, simultaneously, blocked Maxwell's demon. Still, the existence of experiments in quantum mechanics (for example, spin-echo effect) that can bring the system back to its initial state, even if measurements on individual states of atoms have not been performed.

Gabor argued that even if we keep up to the tenets of classical (i.e. not quantized) electromagnetism, the entropic cost for gaining information on molecule location through the measurement in the Szilard's engine could be lower than we obtain in the course of isothermal expansion.[466] Hereby, the

---

[462] *Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004).*

[463] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

[464] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." *SHPMP* 30.1 (1999): 1-40.

[465] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." *SHPMP* 30.1 (1999): 1-40., See: Slater, 1939, Ch. 3, Sec. 4.

[466] Gabor, Dennis. *Electron optics at high frequencies and at relativistic velocities*. British Thomson-Houston Company, (1951).

quantum nature of radiation played the role of our savior, for the demon must use a sufficiently energetic photon to mark the molecule on the background of noise (created by the blackbody radiation in the cavity).

Leon Brillouin, as a proponent of information theory exorcism and Szilard's principle, denied quantum mechanical exorcism. According to him, the main problem of measurement is not the one find in the scope of quantum mechanics and uncertainty principle. The main problem, in his opinion is grounded on entropic cost and statistical thermodynamics.[467] In theory, all our discussions should be held in quantum terms, because we live in a quantum world, or at least a world in which quantum physics gives better description of the phenomena than any classical theory. From the other side, if we accomplish an explanation on classical level, as was the case with the Brownian motion experiments, the results will contain Boltzmann's constant *k* instead of Planck's constant *h*. Brillouin holds that this could be considered as proof of novelty and of his "Negentropy Principle of Information" and its irreducibility on quantum level.[468] On the other side, Bennett explicitly rejects the idea that a sound exorcism had to wait for quantum mechanics.[469]

Many tried to find relation between quantum measurement and the Szilard engine. One example of this is Zurek's attempt.[470] We already mentioned this in chapter 6. Zurek claims that "[i]t is not too surprising, for, after all, thermodynamic entropy is incompatible with classical mechanics, as it becomes infinite in the limit."[471]

Subsequently, Zurek argues that uncertainty of the molecule in a superposition in Szilard's engine is not subjective, but objective. Here we perceive the difference between views stemming from the orthodox, Copenhagen interpretation of the ontology of wavefunction and those more realistic approaches favored in modern quantum information and computation theory. In those approaches, the molecule occupies both sides of the chamber objectively. Work can be extracted only after a measurement is performed. Resetting operation was still considered as step that generate the entropy. Zurek bases the generation of entropy in this step on Landauer's principle. He justifies this by claim that demon is in the state of statistical mixture while observing each of outcomes, after the measurement has been performed and work extracted.[472]

The demon should reset its measuring system to previous state for making the cycle complete. Hereby, the operation reset is equal to erasing information of the size of one bit. It should be noticed that here, Zurek presupposes that Landauer's principle holds. Thus, he attributes k log 2 of entropy cost to reset operation. Presupposition of truth of this principle should protect the Second Law of thermodynamics.[473]

There are two ambiguous points in Zurek's arguments. First, it is not clear that Zurek has argued that the demon is in the superposition of the measurement outcomes. Second, it stays unclear to what extent an outcome of the measurement must be a reduction (or collapse) of the system wavefunction, violating unitarity of evolution. There is a whole family of interpretations of quantum mechanics in

---

[467] Brillouin, Leon. "Maxwell's demon cannot operate: Information and entropy. I." J. App. Phys. 22.3 (1951): 334-337; Leff, Harvey S. and Rex, Andrew F. 1990, *Maxwell's Demon: Entropy, Information, Computing* Princeton, New Jersey: Princeton University Press. (1990): p. 137.

[468] Brillouin, Leon. "The negentropy principle of information." J. App. Phys. 24.9 (1953): 1152-1163, p. 1162.

[469] Bennett, Charles H. "Demons, engines and the second law." *Scientific American* 257.5 (1987): 108-116.

[470] Zurek, Wojciech H, "Maxwell's demon, Szilard's engine and quantum measurements", in *Frontiers of Nonequilibrium Statistical Physics*, G.T. Moore and M.O. Scully (eds.), New York: Plenum Press, (1986): pp. 151–161.

[471] Zurek, Wojciech H. "Reversibility and stability of information processing systems." *PRL* 53.4 (1984): 391., p. 250.

[472] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).

[473] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.

which the collapse of wavefunction never occurs (the most famous being Hugh Everett's "many worlds interpretation", but this also pertains to various "consistent histories", "many minds," and many others). Hereby, whether wavefunction collapses is considered crucial for the existence of demons, which is somewhat awkward epistemical position.[474] It does not help Zurek's position if the wavefunction collapse is only apparent or an illusion.

It should also be noticed that all previous attempts of finding the proof for Landauer's principle were grounded in classical processing of information and classical information theory. This includes all arguments about the thermodynamic cost of computation and/or erasure, etc. One may argue that a lower bound can be derived for quantum processing of information. Still, there is no available value suggested for such a quantum bound so far. It appears that even an analog logically reversible operation on quantum level (such as Bennett's[475] procedure) may have additional entropy cost. Others, like Allahverdyan and Nieuwenhuizen argue that derivations of the lower bound have assumptions that quantum theory can violate in the low temperature regime.[476] This is a very active area of research in the context of quantum information theory.

It is not clear if quantum mechanics provides any specific mechanism for exorcism that could not be construed with the classical. For example, having a molecule in the cylinder is idealized as a boundary condition of the problem of solving Schrödinger's equation for an infinite square potential well. A thermalized particle at the given temperature T is presented via an ensemble of energy eigenstates distributed as a canonical ensemble of the classical statistical mechanics. (An implicit difference is that a high temperature approximation is used in order to avoid specifically quantum effects with no classical analogs, like Bose-Einstein condensation, etc.) The mechanism of the cycle is similar to those in the Szilard's engine.

Only difference is part that represents insert the partition again in the cylinder center. Hereby, the role of insertion is to divide the system wave function into two distinct parts that would be persistent on both sides. Then comes the quantum measurement operation that will collapse the wave. After it, it will fill only one part of the cylinder. Thus, re-compression of the gas would be completed.

Norton and Earman have insisted that quantum mechanics does not have central importance in neither analysis, nor solving of the problem of Maxwell's demon.[477] They argue that in case of quantum mechanics demon is tend to be exorcized via same mechanism which has been already used in classical system by Bennett. Zurek's quantum exorcism and Bennett's classical share the same pattern: they use step of memory erase to introduce Landauer's principle that would defend the Second Law of thermodynamics. In both cases Landauer's principle is presupposed, not proved.

Biedenharn and Solem offered an analysis of Szilard's engine from perspective of quantum mechanics that is similar to Zurek's.[478] These authors pointed out that step of measuring process as central point in argumentation against possibility of Maxwell's demon. Zurek demonstrated that measuring effect is gain in free energy, $F = U - TS$, which hides changes in entropy. Biedenharn and Solem, argued that the measuring process would change gas energy, and that by observing additional

[474] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[475] Bennett, Charles H. "Logical reversibility of computation." *IBM journal of Research and Development* 17.6 (1973): 525-532.
[476] Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.).
[477] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[478] Biedenharn, Lawrence C., and Johndale C. Solem. "A quantum-mechanical treatment of Szilard's engine: Implications for the entropy of information." *Found. Phys.* 25.8 (1995): 1221-1229.

energy has been inserted into the system. They conclude that this energy is exactly the one that has been obtained through the expansion, not the energy from the reservoir that is transformed into work.

Beghian gives another method for demon's exorcism in the domain of quantum mechanics.[479] He gives example where demon manipulates particles of a Bose gas, which consist of indistinguishable bosons (quantum particles with integer spin that can have same energies and are capable of having identical energies and other quantum numbers). The demon should be able to distinguish molecules by their velocity and operate the trapdoor. However, in order to do this, the demon must mark the molecules, this will cause the change in their entropy. For Beghain, the entropy hides in process of marking the molecules, thus saving the second law of thermodynamics.

The problem is it is grounded on anthropomorphic and classical premises. It is not clear that demon should be able to distinguish the particles by velocity to sort them. As long we are presupposing that demon is living being, we are naturalizing him. We already showed that demon can be mechanical device. For example, Zhang and Zhang's demon is only a field with specific features.[480] This field can be described as quantum, although that has not been strictly proved so far. However, all these attempts remain incomplete and unsuccessful in exorcising the demon.

[479] Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.
[480] Ibid.

9. Conclusion

The goal of Maxwell's thought experiment was to show, in accordance with the prevailing mechanicist views of classical physics and the Newtonian worldview, that validity of the Second Law is only statistical. The original demon was first described as a living being, but soon it has been reduced to a valve or some other contraption. Subsequently, the role and nature of the demon has been evolving trough the time. Smoluchowski formulated the modified version of the Second Law and tame the demon, which made violation of the Second Law impossible (although original Maxwell's considerations permit violations of the Second Law, since its validity is taken as limited).

Later on in this exciting intellectual history, the question of demon's intelligence came into focus. Szilard analyzed whether an intelligent demon could defeat the Second Law. His conclusion was that intelligence is not enough for this purpose, since measurements that must be performed have inevitable entropic cost. Questions of demon's perception and intelligence were followed by the question of nature of that intelligence. In the end, for the aims of discussion, the concept of the demon's intelligence has been reduced to processing of information. In the same time, this development led to the emergent field of physics of information.

Great names of 20th century physics such as Von Neumann, Brillouin, and Gabor analyzed whether information processing has intrinsic entropic cost. They concluded that even that the possession of information can decrease entropy, its acquisition result in entropy increase. Further on, the analysis has led to the provocative question is the lack of information equivalent to entropy?

Landauer's seminal analysis has showed that measurement does not necessarily lead to entropy increase. Nevertheless, he addressed another source of entropy increase: *erasure of information*, or more precisely, removing of the partition which is considered as necessary part of the erasure process. Since then, has been argued, that, if demon's memory is not infinite, it will require erasure of information which will lead to entropy increase. Later, Bennett has showed that we can avoid entropic cost in logically reversible computation. Still, he argued that there was logically irreversible step of erasure which cannot be avoided in Szilard's engine. The latter might be relevant for some models, in particular for the "demonic machines" such as "Brownian motors", etc., but is of limited philosophical importance.

We conclude that the origin of the problem of Maxwell's demon lies in entropy that is caused by mixing of subensembles. However, it also rises and solves the problem. In Maxwell's thought experiment, while inspecting every atom, demon is also sorting the ensemble to subensembles. This resulted in mixing that reversed the entropy, seemingly violating the second law. The problem was that this operation could not be unitary one only in the gas space. It can be shown that this sorting could not be performed by any unitary operation acting just on the state space of the entire ensemble (gas). An auxiliary system or systems need to be included, which will effectively increase the entropy that would compensate the entropy loss in the gas.

In order to gain free energy from each of the subensembles, the postulated auxiliary system needs to be employed. Entropy of auxiliary system will rise, for it is correlated to acquiring of free energy. Hereby, we conclude that the measurement could not lower system's entropy since, rise of auxiliary's system entropy compensates it. This has been the first part of solution.

The second part of the solution is that demon, as being the part of system, must be governed by its laws. The demon cannot reduce entropy of the system without inducing its own entropy, for he is an object of unitary evolution himself.

We have also seen that while the information theory approach would naturalize the demon via its intelligence and then use it for his defeat, this does not add any crucial feature to the scenario. A careful analysis showed that any intelligent being needs to be treated as a physical system. In the end, it all comes down to the same conclusion – the demon is subject to normal unitary evolution, and as such he must be part of a canonical ensemble in the sense of standard statistical mechanics. The concept of demon's intelligence adds nothing to the concept of Maxwell's demon. We would come to the same conclusion, regardless of demon's intelligence.

After all, the basic underlying question had been from the beginning: what Maxwell's demon cannot tell us about the world? The thought experiment introduced to us indeterminism and irreversibility. It imposed Laplacian demon in his identification of causality and determinism. We saw that physical systems could indeed be indeterministic, without *necessarily* lacking the causal order. The notion of indeterminism has challenged and ultimately changed notion of causality. It went from being a deterministic concept to probabilistic.

What about the arrows of time? As a causal relation demands energy, it might seem that entropy could be related to both causality and time's asymmetry. It might seem that demon could return trajectories to their initial conditions, at least theoretically, but beside time's irreversibility we must take into account phase space spreading and energy transfer. With this considered, we conclude that returning to initial conditions is highly improbable. After all, Maxwell's demon demonstrates that we cannot identify entropy with the time's arrow, for its statistical nature. While showing us that entropy has statistical nature, the demon from Maxwell's thought experiment proves that time's arrow and entropy's arrow cannot be identified.

This does not mean that direction of entropy gradient could not be used to induce various time's arrows. The very fact that different criteria – which are of undoubted physical relevance – could bring us to the various interpretations of the time's arrows indicates that the analysis of the direction and flow of time (if any) remains one of the crucial philosophical tasks of wider importance. In particular, the worldview imposed by the best scientific theories is of necessity incomplete without accounting for various arrows of time, which demonstrates the continued relevance of philosophy of science.

In this thesis, we have also concluded that the Darwinian selection theory could be cast in a form which has essentially the same structure as Maxwell's thought experiment with demon. Hence, we can conclude that a particular worth of Maxwell's demon is at least in the domain of heuristics. After all, there is no need for exorcising the demon. Since he is governed by the laws of thermodynamics, he cannot violate them.

Bibliography:

Albert, David Z. *Time and chance*. Cambridge, Massachusetts, London:Harvard UP. (2001): 1285-1286.

Arsenijević, Miloš, *Vreme i vremena,* Dereta, Beograd, (2003).

Bennett, Charles H. "Logical reversibility of computation." *IBM journal of Research and Development* 17.6 (1973): 525-532.

Bennett, Charles H. "The thermodynamics of computation—a review." *International Journal of Theoretical Physics* 21.12 (1982): 905-940.

Bennett, Charles H. "Demons, engines and the second law." *Scientific American* 257.5 (1987): 108-116.

Bennett, Charles H. "Notes on Landauer's principle, reversible computation, and Maxwell's Demon." SHPMP 34.3 (2003): 501-510.

Biedenharn, Lawrence C., and Johndale C. Solem. "A quantum-mechanical treatment of Szilard's engine: Implications for the entropy of information." Found. Phys. 25.8 (1995): 1221-1229.

Bohr, Niels. *Atomic theory and the description of nature: four essays with an introductory survey*. Vol. 1. Cambridge University Press, (2011).

Bokulich, Alisa. 2016, "Fiction As a Vehicle for Truth: Moving Beyond the Ontic Conception," The Monist 99, 260-279.

Boltzmann, Ludwing, *Lectures on gas theory*, translated by S. G. Brush (University of California Press, Berkeley, (1964).

Brillouin, Leon. "Maxwell's demon cannot operate: Information and entropy. I." J. App. Phys. 22.3 (1951): 334-337.

Brillouin, Leon. "The negentropy principle of information." J. App. Phys. 24.9 (1953): 1152-1163.

Brillouin, Leon, *Science and Information Theory,* (New York: Academic Press), (1956).

Brown, James Robert, *The Laboratory of the Mind: Thought Experiments in the Natural Sciences* (Routledge, London), (1993).

Brown, James Robert. "Peeking into Plato's heaven." *Philosophy of Science* 71.5 (2004): 1126-1138.
Bub, Jeffrey. "Maxwell's Demon and the Thermodynamics of Computation." SHPMP 32.4 (2001): 569-579.

Cartwright, Nancy. "Models and the limits of theory: Quantum Hamiltonians and the BCS models of superconductivity." *IDEAS IN CONTEXT* 52 (1999): 241-281;

Chambadal, Paul, and MF INGHAM. *Les Paradoxes en Physique. Paradoxes of Physics... Translated... by MF Ingham*. Transworld, (1973).

Cooper, Rachel. "Thought experiments." *Metaphilosophy* 36.3 (2005): 328-347.

Cottet, Nathanaël, et al. "Observing a quantum Maxwell demon at work." *Proceedings of the National Academy of Sciences* 114.29 (2017): 7561-7564.

Davies, Paul, Lloyd A. Demetrius, and Jack A. Tuszynski. "Implications of quantum metabolism and natural selection for the origin of cancer cells and tumor progression." *AIP advances* 2.1 (2012): 011101.

Davies, Paul CW, Elisabeth Rieper, and Jack A. Tuszynski. "Self-organization and entropy reduction in a living cell." *Biosystems* 111.1 (2013): 1-10.

Davies, Paul CW, and Sara Imari Walker. "The hidden simplicity of biology." Rep. Prog. Phys. 79.10 (2016): 102601.

De Bovoar, Simon. *Starost I, II. BIGZ, Beograd* (1986).

Denbigh, Kenneth George, Jonathan Stafford Denbigh, and Zeh, Dieter. "Entropy in relation to incomplete knowledge." *Cambridge University Press*, (1985).

Depew, David J. "Nonequilibrium thermodynamics and evolution: a philosophical perspective." *Philosophica* 37.1 (1986): 27-58.

Dugić,Miroljub, *Osnove kvantne informatike i kvantnog računanja*, Prirodno matematički fakultet univerziteta u Kragujevcu, (2009).

Earman, John, *A Primer on determinism,* Reidel, Dodrecht, (1986).
Earman, John, and John D. Norton. "Exorcist XIV: the wrath of Maxwell's demon. Part I. From Maxwell to Szilard." SHPMP 29.4 (1998): 435-471.

Earman, John, and John D. Norton. "Exorcist XIV: The wrath of Maxwell’s demon. Part II. From Szilard to Landauer and beyond." SHPMP 30.1 (1999): 1-40.

Eddington, Arthur S. *The Nature of the Physical World: Gifford Lectures (1927)*. Cambridge University Press, (2012).

Eddington, Arthur S. "Relativity theory of protons and electrons." *Cambridge: Cambridge University Press* (1936): pp. 86-91.

Einstein, Albert. "Autobiographical notes (1949)." *Albert Einstein: Philosopher-Scientist* (1963).

Ellis, George Francis Rayner. "The arrow of time and the nature of spacetime." SHPMP 44.3 (2013): 242-262.

Feyerabend, Paul K. “On the Possibility of a Perpetuum Mobile of the Second Kind”, in *Mind, Matter and Method: Essays in Philosophy and Science in Honor of Herbert Feigel*, in. P. K. Feyerabend and G. Maxwell (Minneapolis, Minnesota: University of Minnesota Press), (1966): pp. 409–412.

Feynman, Richard P., R. B. Leighton, and Matthew Sands. "Ratchet and pawl." *The Feynman lectures on physics* 1 (1963): 46-1.

Feynman, *Lectures on computation*, Penguin (1999).

Fredkin, Edward, and Tommaso Toffoli. "Conservative logic." *International Journal of Theoretical Physics* 21.3 (1982): 219-253.

Fredkin, Edward, “An Introduction to Digital Philosophy”, *International Journal of Theoretical Physics* 42 (2003): 189-247.

Frigg, Roman and Hartmann, Stephan, "Models in Science", SEP (2012 Ed.), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2012/entries/models-science/

Frigg, Roman, et al. "Laplace’s demon and the adventures of his apprentices." *Philosophy of Science* 81.1 (2014): 31-59.

Gabor, Dennis, *Electron optics at high frequencies and at relativistic velocities*. British Thomson-Houston Company, (1951).

Gabor, Dennis, “Light and Information”, *Progress in Optics* 1, (1964): 111–153.

Gould, Stephen Jay, 2002, The Structure of Evolutionary Theory (Belknap Press, Cambridge, Massachusetts).

Griffiths, Robert B. *Statistical irreversibility: classical and quantum*. Cambridge Univ. Press, (1994).

Hacking, Ian, „Do Thought Experiments Have a Life of Their Own?“ Comments on James Brown, Nancy Nersessian and David Gooding, *The Philosophy of Science Association*, Vol. 1992, No., Volume two: Symphosia and Invited Papers, (1992).

Hajzenberg, Verner. *Fizika i metafizika* (Nolit, Beograd; in Serbian). (1989)

Hänggi, Peter, and Fabio Marchesoni. "Artificial Brownian motors: Controlling transport on the nanoscale." *Reviews of Modern Physics* 81.1 (2009): 387.

Hartmann, Stephan, „The World as a Process: Simulations in the Natural and Social Science“, Theory and Decision Library, Dodrecht, (1996): 77-100.

Hartmann, Stephan. "Models and stones in hadron physics." *Models as mediators: Perspectives on natural and social science* 52 (1999): 326.

Heisenberg, Werner. "Nonlinear problems in physics." *Physics Today* 20 (1967): 27.

Hemmo, Meir, and Orly Shenker. "Prediction and retrodiction in Boltzmann's approach to classical statistical mechanics." (2007). http://philsci-archive.pitt.edu/3142/1/Hemmo_Shenker_on_Boltzmann_23Jan07.doc

Hemmo, Meir, and Orly R. Shenker. *The road to Maxwell's demon: conceptual foundations of statistical mechanics*. Cambridge University Press, (2012).

Hempel, Carl G. "Fundamentals of concept formation in empirical science, Vol. II. No. 7." (1952).*In Aspects of Scientific Explanation* (pp. 155-171). New York: Free Press/London: Collier Macmillan (1965).

Hoking, Stiven, *Kratka povest vremena*, Sfinga, (1988).

Humphreys, Paul. "Seven theses on thought experiments." *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grunbaum* (1993): 205-227.

Irvine, Andrew D. "Thought experiments in scientific reasoning." *Thought experiments in science and philosophy* In T. Horowitz & G. Massey (Eds.), *Thought experiments in science and philosophy* (1991): 149-165.

Jauch, Josef-Maria, and J. G. Baron. "Entropy, information and Szilard's paradox." *Maxwell's Demon: Entropy, information, computing* (1990): 160-172.

Kaneman, Danijel, *Misliti brzo i sporo*, Heliks, 2015

Kant, Immanuel. *Critique of pure reason*. Cambridge University Press, 1998.

Kožnjak, Boris, Who let the demon out? SHPS, 51, (2015): 42-52,

Krakauer, David C. "Darwinian demons, evolutionary complexity, and information maximization." *Chaos.* 21.3 (2011): 037110.

Kuhn, Thomas S. "Objectivity, Value judgement, and Theory Choice." *Thomas Kuhn (ed)* 76 (1973): 320-339.

Kuhn, Thomas S., "A function for thought experiments". In I. Hacking (Ed.), *Reprinted in Scientific Revolutions,* Oxford: Oxford University Press. (1981): 6-27.

Kutrovátz, Gábor. "Heat death in ancient and modern thermodynamics." OSID 8.4 (2001): 349-359.

Ladyman, James, et al. "The connection between logical and thermodynamic irreversibility." SHPMP 38.1 (2007): 58-79.

Ladyman, J., Presnell, S., & Short, A. J. 2008, "The use of the information-theoretic entropy in thermodynamics", SHPMP 39, 315–324.

Landauer, Rolf. "Irreversibility and heat generation in the computing process." *IBM journal of research and development* 5.3 (1961): 183-191.

Landauer, Rolf. "Computation and physics: Wheeler's meaning circuit?." *Found. Phys.* 16.6 (1986): 551-564.

Landsberg, Peter Theodore, and Eric J. Chaisson. "The enigma of time." *American Journal of Physics* 53.6 (1985): 601-602.

Laplace, Pierre. "La probabilité des causes parles événements." Œuvres Complétes de Laplace 8 (1774): 27-62.

Laplace, Pierre Simon. *Théorie analytique des probabilités*. Courcier, Paris, Couvier. (1820).

Layzer, David. 1976, “The Arrow of Time” Astrophysical Journal 206, pp. 559-569

Leff, Harvey. S. and Rex, Andrew. F, *Maxwell's Demon: Entropy, Information, Computing* Princeton, New Jersey: Princeton University Press. (1990).

Leff, Harvey S. and Rex, Andrew F. 2003, *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing* Philadelphia: Institute of Physics Publishing. (2003).

Leff, Harvey S. "Entropy, its language, and interpretation." *Found. Phys.* 37.12 (2007): 1744-1766.

Lockwood, Michael. *The labyrinth of time: introducing the universe*. Oxford University Press on Demand, (2005).

Mach, Ernst. "The science of mechanics: A critical & historical account of its development, (after the 9th German edition)." *Open Court. [JVB]* (1883).

Machta, John, Natural complexity, computational complexity and depth, *Chaos* 21, 037111 (2011).

Mahon, Basil, *The Man Who Changed Everything – the Life of James Clerk Maxwell*. Hoboken, NJ: Wiley. (2003).

Mainzer, Klaus. *Symmetries of nature*. Walter de Gruyter, (1996).

Maroney, Owen. "Information and entropy in quantum theory." quant-ph/0411172 (2004).

Maroney, Owen, "Information Processing and Thermodynamic Entropy", SEP (2009 ed.), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2009/entries/information-entropy/>.

Maruyama, Koji, Nori, Franco, and Vedral, Vlatko, Colloquium: The physics of Maxwell’s demon and information. *Reviews of Modern Physics*, *81*(1), (2009): p.1-23.

Maudlin, Tim, *The Metaphysics Within Physics* Oxford University Press, Oxford. (2007).

Maxwell, James Clerk, *Letter to P. G. Tait, 11 December 1867*, u C. G. Knott, Life and Scientific Work of Peter Guthrie Tait (Cambridge: Cambridge University Press, 1911), (1867): 213–215.

Maxwell, James Clerk, (1878), *Diffusion Encyclopedia Britannica*, 7,214, Reprinted in: Maxwell (1952): 625-646

Maxwell, James Clerk. "Essay for the Eranus Club on science and free will. Does the progress of physical science tend to give any advantage to the opinion of necessity (or determinism) over that of the contingency of events and the freedom of the will." *The scientific letters and papers of James Clerk Maxwell* 2 (1995): 814-823.

Maxwell, Nicholas. "Are probabilism and special relativity incompatible? " *Philosophy of Science* 52.1 (1985): 23-43.;

Maxwell, Nicholas. "Are probabilism and special relativity compatible? " *Philosophy of Science* 55.4 (1988): 640-645.

McAllister, Janet, "The evidential significance of thought experiments in science". SHPS, (1996), 27, pp. 233-250.

McAllister, Janet, "Thought experiments and the belief in phenomena". *Philosophy of Science, 71*, (2004): pp. 1164-1175**.**

Morgan, Mary S., and Margaret Morrison, eds. *Models as mediators: Perspectives on natural and social science*. Vol. 52. Cambridge University Press, (1999).

Nagel, Ernest, and David Hawkins. "The structure of science." *American Journal of Physics* 29.10 (1961): 716-716.

Norton, John A. "On thought experiments: Is there more to the argument?" *Philosophy of Science, 71,* (2004): pp. 1139-1151.

Norton, John D. "Thought experiments in Einstein's work." *Horowitz and Massey* 1991 (1991): 129-148.

Norton, John D. "Are thought experiments just what you thought?" *Canadian Journal of Philosophy* 26.3 (1996): 333-366.

Norton, John D. "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon." SHPMP 36.2 (2005): 375-411.

Norton, John D. "Waiting for Landauer." SHPMP 42.3 (2011): 184-198.

Pais, Abraham. *Subtle is the Lord: The Science and the Life of Albert Einstein: The Science and the Life of Albert Einstein*. Oxford University Press, USA, (1982).

Penrose, Oliver, *Foundations of Statistical Mechanics*, Oxford: Pergamon Press, (1970).

Penrose, Roger. *The Emperor's New Mind: concerning computers, brains and the laws of physics.* (1989).

Penrose, Roger. *The road to reality*, Alfred A. (2006).

Piechocinska, B. 2000, "Information erasure", Physical Review A, 61: 1–9.

Planck, Max, *Treatise on Thermodynamics*. Longmans; reprinted New York: Dover. (1926).

Popper, Karl, "Indeterminism in Quantum Physics and in Classical Physics, Part I" *The* BJPS **1**, (1950): 173-195.

Popper, Karl, *The logic of scientific discovery.* London: Hutchinson. (1959).

Popper, Karl R. *Autobiography of Karl Popper*. In P A Schlipp, editor, The philosophy of Karl Popper, Open Court (1974).

Price, Huw. *Time's arrow & Archimedes' point: new directions for the physics of time*. Oxford University Press, USA, (1997).

Price, Huw. "Cosmology, time's arrow, and that old double standard." In Savitt, Steven F. *Time's Arrows Today: Recent Physical and Philosophical Work on the Direction of Time* (1998): 66-96.

Raymond, Richard C. "The well-informed heat engine." *American Journal of Physics* 19.2 (1951): 109-112.

Reichenbach, Hans. *The direction of time*. Berkeley: University of California Press. 1956. (1991): 198.

Rex, Andrew and Larsen, Ross, "Entropy and Information for an Automated Maxwell's Demon," (1992). Workshop on Physics and Computation, pp. 93-101. URL =

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=615503&isnumber=13433

Rothstein, Jeffrey, "Information, Measurement, and Quantum Mechanics", *Science* 114: (1951): 171-175.

Schneider, Eric D., and James J. Kay. "Life as a manifestation of the second law of thermodynamics." Math. Comp. Mod. 19.6-8 (1994): 25-48.

Schrödinger, Erwin. "Conceptual models in physics and their philosophical value." *Science theory and man* (1957): 148.

Shenker, Orly R. "Logic and entropy." (2000). http://philsci-archive.pitt.edu/115/1/Shenker_Logic_and_Entropy.doc.

Shizume, Kousuke. "Heat generation required by information erasure." *Physical Review E* 52.4 (1995): 3495.

Sklar, Lawrence. "Time in experience and in theoretical description of the world." In Savitt, Steven, *Time's Arrows Today. Recent physical and philosophical work on the direction of time* (1998): 217-229.

Sklar, Lawrence. "The elusive object of desire: in pursuit of the kinetic equations and the second law." In Savitt, Steven F. *Time's arrows today: recent physical and philosophical work on the direction of time.* (1998): 191-217.

Skordos, Panayotis Augoustos. "Compressible dynamics, time reversibility, Maxwell's demon, and the second law." *Physical Review E* 48.2 (1993): 777.

Smoluchowski, Marian V. "Experimentally verifiable molecular phenomena that contradict ordinary thermodynamics." *Phys. Z* 8 (1912): 1069.

Sober, Eliot, *Filozofija biologije,* Plato, Beograd. (2006).

Steckline, Vincent. S. 1983, "Zermelo, Boltzmann, and the recurrence paradox," Am. J. Phys. 51, 894-897.

Strutt, RJ & Rayleigh, JW. *Life of John William Strutt Third Baron Rayleigh*. Arnold, Madison: University of Wisconsin Press. (1924).

Szilard, Leo, "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings", Zeit. Phys. 53: 840–856. (english translation in The Collected Works of Leo Szilard: Scientific Papers, B. T. Feld i G. W. Szilard (Cambridge, Massachusetts: MIT Press, 1972), (1929): pp. 103–129.

Uffink, Jos, "Bluff your way in the second law of thermodynamics", *Studies in the History and Philosophy of Modern Physics* 32, (2001): 305–394.

Uffink, Jos, "Boltzmann's Work in Statistical Physics", SEP (2017 Ed.), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/statphys-Boltzmann/>.

Von Neumann, John, *Mathematical Foundations of Quantum Mechanics* (Princeton, New Jersey: Princeton University Press). [1932] 1955.

Von Neumann, John. "Theory and organization of complicated automata." *Burks (1966)* (1949): 29-87.

Weinert, Friedel. "Einstein and the Representation of Reality." *Facta Philosophica,* 8 (1-2), (2006): 229-252.

Weinert, Friedel. "Theories, models and constraints." SHPS *Part A* 30.2 (1999): 303-333.

Weinert, Friedel, *The march of time*. New York: Springer. (2013).

Weinert, Friedel. *The Demons of Scienc*e. Springer. (2016).

Weinert, Friedel. "Theories, models and constraints." SHPS Part A 30.2 (1999): 303-333.

Zeh, Dieter. "The physical basis of the arrow of time." Springer-Verlag, New York, (1992).

Zhang, Kechen, and Kezhao Zhang. "Mechanical models of Maxwell's demon with noninvariant phase volume." *Physical Review A* 46.8 (1992): 4598.

Zurek, Wojciech H. "Reversibility and stability of information processing systems." PRL 53.4 (1984): 391.

Zurek, Wojciech H., "Maxwell's demon, Szilard's engine and quantum measurements", in *Frontiers of Nonequilibrium Statistical Physics*, G.T. Moore and M.O. Scully (eds.), New York: Plenum Press, (1986): pp. 151–161.

Biography

Jelena Dimitrijević was born in Split, Croatia in 1990. She enrolled at the Faculty of the Philosophy in Belgrade in 2009. She received B.E. degree in philosophy in 2013. with a grade point average of 9.12. She graduated on subject „Determinism in Einstein's General and Special Theory of Relativity." She enrolled at M.A. studies in philosophy at the same faculty the same year and received a Master's degree in philosophy next year. The grade point average was 10.00. The topic of the master's thesis was "Computer Simulations, Scientific Models and Experiments." She enrolled at a PhD program in philosophy at the same university in 2015.

Since 2014 she has been a journalist in a journal for literature and culture "Zvezdani Kolodvor" and from 2016 the deputy editor. Since 2016 she has been a member of the Center for the Study of Bioethics. She is also a member of the Serbian Unit of the UNESCO Chair in Bioethics. She has published scientific papers. She participated in a conference on "Genome Editing: Bioethical and Medical Perspectives" held in Belgrade in 2017 as a speaker. She was the coeditor of a collection of papers on Bioethics. She translated numerous papers and one book from English to Serbian.

In 2018, she joined the Department for philosophy as a Research Assistant, teaching on the subject science and rationality at the Faculty of Philosophy, University of Belgrade.

She is also poet and her poems have been published in numerous collections of poetry in Serbia. She was the coeditor of a collection of poems dedicated to Charles Baudelaire.

Образац 5.

Изјава о ауторству

Име и презиме аутора Јелена Димитријевић

Број индекса ОФ 14-8

Изјављујем

да је докторска дисертација под насловом

Philosophical Aspects of the Problem of Maxwell's Demon

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, __________________

____

Образац 6.

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора _______Јелена Димитријевић_______________________

Број индекса _________________ОФ 14-8________________________________

Студијски програм __________Философија________________________________

Наслов рада __ Philosophical Aspects of the Problem of Maxwell's Demon_____

Ментор ______др Слободан Перовић_________________________________

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањена у **Дигиталном репозиторијуму Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, ________________________

____

Образац 7.

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

Philosophical Aspects of the Problem of Maxwell's Demon

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)

2. Ауторство – некомерцијално (CC BY-NC)

3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)

4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)

5. Ауторство – без прерада (CC BY-ND)

6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, ________________________

____

1. **Ауторство**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.