The Relation Between Shannon Information and Semantic Information

MSc Thesis *(Afstudeerscriptie)*

written by

Nikki Weststeijn
(born September 26st, 1998 in Amsterdam, The Netherlands)

under the supervision of **Prof. dr. Sonja Smets** and **Dr. Soroush Rafiee Rad**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
July 7, 2022	Prof. dr. Sonja Smets
	Dr. Soroush Rafiee Rad
	Prof. dr. Yde Venema
	Dr. Aybüke Özgün
	Dr. Sebastian de Haro Ollé

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

Information is a concept that has become a point of focus in many disciplines other than information theory, but it is often unclear precisely which concept we are referring to. There are namely two main types of information. On the one hand, there is Shannon information, which is the most standard concept of information in information theory and which quantitatively measures the information contained in an information source. On the other hand, there is semantic information, which is more in line with the everyday understanding of the term "information" and which concerns information that is always *about* something. Whereas Shannon information is characterized as being quantitative, objective, physical and mind-independent, semantic information is characterized as being qualitative, subjective, mind-dependent and evidently semantic. It thus seems that there is a dichotomy between these two concepts of information. The main question of this thesis is to find out whether Shannon information is really that different from semantic information, i.e. to characterize the relation between Shannon information and semantic information. We will look at the concept of Shannon information in information theory and physics and clarify the relation between information and entropy. We will also look at quantitative accounts of semantic information by Dretske, Bar-Hillel and Carnap, and Floridi and at how we can use Situation Theory, Dynamic Epistemic Logic and Logic of Functional Dependence to model the flow of information.

Acknowledgements

First of all, I would like to thank those who have supervised this thesis. Thank you Soroush, for your feedback and advice throughout this process and the willingness to talk to me whenever I wanted to. Thank you Sonja for being willing to become the other official supervisor of my thesis so shortly before the deadline and for your insightful comments. I also would like to express my thanks to Johan van Benthem for starting this project with me. Even though things did not go as planned, I am grateful for all your insights and your help.

My thanks also go out to the other members of the Thesis Committee: Yde Venema, Aybüke Özgün and Sebastian de Haro Ollé. Thank you for reading my work and I am curious what you think.

Then I would like to thank Fred Muller and John Dougherty for kindly giving very useful feedback on the parts about physics.

I also want to thank Tanja for helping me both with this thesis process and personally. Talking to you has always made me feel better.

Finally, I want to thank Alex, my friend from Bristol, for continuing to talk to me about my projects and giving me feedback, even though we have never met in person. I will also thank my friends and family who have supported me.

Contents

Introduction

The Information Age

Although we might not often think about the concept of information, information affects people daily. It is said that we live in an Information Age: our ability to store information has grown exponentially, communication is easier than ever and the digitization of different services and media has had a profound impact on our lives. While we are coming to grips with how the new use of information has led to these practical changes, it is natural that we also turn our attention to the conceptual definition of information. What actually is information? Are there different kinds of information? How can we store information and transmit it? These are questions that will be dealt with in this thesis. And while the thesis will have little to do with the societal aspect of information or information studies, it nevertheless is part of the big project of trying to understand the world around us and the concepts that play an important role in it, of which information is a textbook example.

We could also say that, in a way, a large part of science is going through an Information Age. Information theory started out as a discipline after Shannon produced his Mathematical Theory of Communication (MTC) [87] in which he formulated a quantitative measure of information. This measure has proved to be extremely useful for calculating how we can store and transmit information. A focus on the concept of information has also led to a new perspective on several phenomena in philosophy, physics, cognitive science and biology.

In physics, it has become standard to refer to the entropy contained in a system as the "information" contained in that system, because of the similarity between the entropy measure and Shannon's measure of information. The entropy (or information) contained in a system can be thought of as the amount of randomness in a system. All physical systems are thought to have a certain entropy, i.e. information, and some physicists argue that information is the most fundamental building block of reality. Wheeler [94] for example defends an "It from Bit" philosophy according to which every particle derives its existence from information.

In cognitive science, an information-theoretic perspective on the human mind has become popular. This is the idea that the mind is constantly processing information and functions similarly to a computer. In biology, the fact that biological systems also carry information led to the idea that genes and hormones can send information to other parts of the body that work as instructions for development and other physiological processes. A focus on the concept of information has finally also had an impact on philosophy. On the one hand, the philosophy of information itself has become a new area of study with early accounts of information given by Bar-Hillel and Carnap [15] and a recent account given by Floridi [42]. On the other hand, a focus on information has led to new perspectives on meaning and knowledge.

The definition of information

It has become clear that in this scientific Information Age, information is a concept that is applied in a variety of area's. The precise definition of information that is used however differs quite a lot. There are at least three different conceptions of information that are being used across the disciplines.

First, there is the intuitive meaning that most people associate with the term "information", which is information understood as the knowledge that we obtain by reading a message. For example if we read the sentence "The milk is in the fridge", we get the information that we can find the milk in the fridge. This is the type of information that is referred to when we talk about "information" in the context of philosophy of language for example.

Second, there is the standard information-theoretic conception of information that follows from Shannon's MTC [87]. If we would ask an information theorist what the information is that is contained in the English sentence "The milk is in the fridge", they would take an entirely different approach. They would want to know what source emitted this message and how likely it was that this message would appear and then calculate a number in bits that captures the information of the sentence. Shannon's measure of information is the measure that is used in physics.

There is a third and different way of thinking about information, which is in terms of algorithmic complexity. When presented with the above sentence as a string of symbols, we can determine the Kolmogorov complexity of the string by calculating how long the shortest program would be that can produce it. This is taken to be the amount of algorithmic information of the sentence, also given in bits. While Shannon information and algorithmic information are different in some respects, it has been showed that their values also correspond in a certain way [68].

In biology and cognitive science, but also in philosophy of physics and philosophy of computation, both the everyday notion of information and the information-theoretic notions of information are used. It is in these area's that we notice the confusion that exists around the term "information". It is often not clear which interpretation an author has in mind. Are they talking about a quantity that we can ascribe to a source or about the meaning of a message? Or are they the same thing after all? The confusion between the information-theoretic concepts and everyday concept of information can also lead to further issues. In philosophy of physics and computation for example, the term "information" is often interpreted in the everyday sense, even when the subject is actually Shannon's notion of information. This can lead to people ascribing properties to Shannon information that only hold for the everyday concept of information, like the fact that it can tell us something about the content of information.

The question is how and to what extent the three different conceptions of information – the everyday notion, the Shannon measure, and the algorithmic measure– overlap. Are they merely different ways of approaching the same phenomenon or do they capture different types of information? There is reason to think that there is a difference between the type of information that we think of in an everyday sense and the type of information that information theorists deal with. Information theorists may use either Shannon information or algorithmic information, but I will focus on the difference between the everyday notion of information and Shannon information.

Shannon information vs. semantic information

Semanticity. The first immediate difference is that the everyday notion involves meaning whereas the Shannon measure does not. The everyday notion of information has to do with the knowledge that a message can give us in virtue of the meaning that we attribute to the message. Like in the previous example with the message about the milk, we associate a certain meaning with the words that together form a sentence and thereby obtain the knowledge that the milk is in the fridge. The everyday notion of information can therefore be referred to as semantic. The Shannon measure is not semantic in this respect.

Quantitativity. What also stands out is that Shannon information is a quantitative measure and that semantic information in principle seems to require a qualitative description rather than a quantitative description. For someone who is not very familiar with theories of information, it will seem strange to ask *how much* semantic information is contained in a sentence. A sentence like "the milk is in the fridge" for example, what amount of information would we associate with this?

Objectivity. Related to these last two differences is the fact that Shannon information is considered to be objective whereas the everyday notion of information generally is not. Concerning the semantic concept of information, we may ask what the information is that a sentence conveys to a specific person. The information that a sentence conveys to someone is highly dependent on this person's previous background knowledge. It is in this sense that semantic information seems *subjective*. Shannon information, on the other hand, is a mathematical quantity of which it is standardly assumed that this will be the same for different agents. It is therefore often considered objective.

Physicality. There is also another difference that can be noted. This is that Shannon information is often taken to be something physical and that the everyday notion is not. Here I distinguish information being physical itself from information requiring a physical carrier. Shannon information is used in physics and treated as a quantity that can be measured for all systems. It is therefore often considered to be a physical property itself. There is a famous paper written by physicist Landauer that is titled "Information is a physical entity" [63]. As I mentioned before, there are even physicists like Wheeler who argue that information is the most basic quantity and allows for the existence of everything that is physical [94]. Wheeler himself however does not explicitly argue that information itself is something physical. Whether it is indeed correct to say that physics tells us that information is physical is yet unclear and is something I will look into in this thesis. What we do already know, is that the everyday notion of information is not taken to be a physical phenomenon.

Mind-dependence. Next to that, when it comes to the ontological status of information, it could be assumed that the Shannon information contained in an object or an information source is something that is mind-independent. This can be understood both in the sense of *ontic* mind-independence and in the sense of *epistemic* mind-independence. It can be thought that an object or a source inherently has a certain quantity of information associated with it, independently of whether there is some cognitive agent that uses this information. Similarly, it can be thought that the Shannon information of a source can be *described* independently of any cognitive agent. It seems to be the case that physicists who believe in something like Wheeler's "It from Bit", must believe that Shannon information is ontologically mind-independent, in order to be able to believe that external reality is mind-independent.

Because of these differences, it can be wondered whether Shannon information quantifies something entirely different from semantic information. The idea that there are two types of information –semantic, qualitative, subjective and mind-dependent information on the one hand and physical, quantitative, objective and mind-independent information on the other hand– has been affirmed by philosophers like Olimpia Lombardi et al.[71] and Roman Krzanowski [59]. They have also discerned them as two types of information: semantic information and physical information. Lombardi characterizes semantic information as "strongly related with notions such as reference, meaning and representation: semantic information has intentionality — "aboutness", it is directed to other things" [71, p.1984]. According to the physical interpretation of information, information is a physical magnitude in the same way as energy is a physical magnitude. Krzanowski characterizes physical information as a concrete phenomenon that has no intrinsic meaning and that exists objectively [59, p.2].

While Krzanowski and Lombardi do not mention in it explicitly, it is in general thought that physical information is captured by the Shannon measure or the algorithmic measure and that semantic information is what we have in mind when we talk about information in an everyday sense. Physicist Christopher Timpson [89] has also contributed to the idea that semantic information and Shannon information are completely distinct; has argues that Shannon information has nothing to do with semantics.

This dichotomy between semantic information and Shannon information is, I think, the main problem when it comes to defining what information is. As has been pointed out by Adriaans and van Benthem [2] and by Floridi [40], finding a good definition of information can be considered one of the main open problems in the philosophy of information. A related problem is to find one unified theory of information. Here the dichotomy between Shannon information and semantic information is of course also a main part of the problem.

The main question

In this thesis, I will thus set out to answer the question whether Shannon information really describes a type of information that is different from semantic information. In other words, the main topic of this thesis is the relation between Shannon information and semantic information. An additional question that we will see pops up is whether both semantic information and

Shannon information can legitimately be considered "information" if they are so fundamentally different. To be able to answer these questions, I will attempt to define as carefully as possible what we mean when we talk about semantic information and what we mean when we talk about Shannon information. This will also be a valuable task in its own right, because the definition and interpretation of these types of information is also not clearly established.

In order to determine whether Shannon information and semantic information are really different types of information, I will consider if it is indeed the case that Shannon information is physical, objective and mind-independent and that semantic information is semantic, subjective and mind-dependent. This will be the main line of the chapters, but I will also aim to characterize Shannon information and semantic information in their own right.

My thesis will thus contribute to answering the main questions in the philosophy of information. This will allow for a better discussion of information in all of the fields in which information plays a role: linguistics, cognitive science, biology, physics, information theory and more. For physics and information theory it is clear that a clarification of the relation between Shannon information and semantic information is very much needed.

For linguistics, it can also be relevant to know more about semantic information and how concepts from Shannon's theory can be used to analyze natural languages. Information-theoretic tools have been used to analyze languages, but it has been doubted whether it can be applied or not, because Shannon information is considered not to be about semantics.

Knowing what semantic information is and how it relates to Shannon information will also contribute to attempts in cognitive science to formalize and quantify information processing in the brain [53]. In biology, a clarification of the relation between semantic information and Shannon's measure can show which concepts we could and should apply to analyze genetic information [88].

Clarification of concepts

Before we go into the thesis, I will first say something about the interpretation of some of the concepts we use; physicality, objectivity and mind-independence.

The interpretation of the term "physical" in the literature is debated and I will argue in the second chapter how we can best understand it. I will argue that it should be distinguished from the interpretation of "concrete", which is confused with physicality by philosophers like Krzanowski [59, p.3]. Concreteness is really about whether we can interact with an object or substance, whereas physicality can be about whether an object or quantity is described by physics or natural science. Lombardi et al. [71, p.2003] agree that there is a difference between the meaning of the terms "physical" and "concrete". She argues that an object can be both not concrete (i.e. abstract) and physical at the same time and that many observables in physics are of this kind. An example is gravity. We know that gravity has an observable effect on objects and it is an object of natural science, but it is not something we can interact with, it is not concrete. Lastly, as I noted before, I do not take "physical" to mean that something requires a physical carrier, but that it is actually physical itself.

Then we also see that the meaning of mind-dependence and objectivity are often confused with each other. I take it that mind-dependence is about the ontic or epistemic status of an object or phenemonon. Ontic and epistemic mind-dependence, however, often also go hand in hand, as we will see when it comes to information. I do not use the interpretation of contingent mind-dependence [25]. This is the idea that an object is also ontologically mind-dependent if it exists because cognitive agents created it, but it could have existed without them. An example of this is a house. Arguably, houses would not have existed if there were no cognitive agents, but they are not necessarily mind-dependent. If all cognitive agents would disappear, the houses could still exist. I can thus say that I use the interpretation of necessary mind-dependence. These different forms of mind-dependence are described in more detail by Botchkina [25].

Objectivity is specifically about the epistemic status of an object or phenomenon. Some philosophers like Kranozwski [59] use this term to refer to the ontic status of an object, but I argue that we should refrain from doing so. I argue that we should take "objectivity" as simply being about the fact that different agents can agree about something, whether this is an object or a property. When a physical quantity is objective, this should be taken to mean that we can come to a public agreement about the value of this quantity. This understanding of "objectivity" is also proposed by the Denbighs [34] and supported by Price [84]. It should also be noted that a quantity can thus be both mind-dependent and objective at the same time.

The structure of this thesis

We will begin the thesis with analyzing Shannon information. Shannon's theory of information is the first formal account of information and many accounts of semantic information are inspired by it. In the first chapter, we will go over Shannon's MTC and get familiar with the main concepts introduced by Shannon and some others that have been associated with Shannon's information theory. We will also look at the relation between Shannon information and algorithmic information theory and in what sense these measures correspond to each other. When we look at these information-theoretic concepts of information, we will also consider whether they are semantic or not.

In the second chapter, we will look at the interpretation of Shannon information and mostly at its interpretation in physics. This is because the role of Shannon information in physics is what has given rise to the idea that Shannon information is something physical. We will look at the relation between Shannon information and entropy and find out whether they are the same quantity or not. This will make it clear in what sense it is legitimate to refer to entropy as "information" and vice versa. We will also see why physicist Landauer argues that information is physical. We will at the end of the chapter judge whether Shannon information can be considered concrete, physical, mind-independent and objective.

In the third chapter we will come to semantic information. Two main philosophical accounts of semantic information are Dretske's account of environmental information [38] and Floridi's account of information as "truthful well-formed and meaningful data"[42]. These two accounts will already challenge the assumption that semantic information cannot be measured quantitatively, because both Dretske and Floridi propose quantitative measures of semantic information. When analyzing semantic information, we will first consider these two accounts and the measures they propose. We will also consider Bar-Hillel and Carnap's account [15], which was the first to propose a quantitative measure of information inspired by a concept from Shannon's information theory.

In the fourth chapter, we will then consider some logics that can be used to analyze semantic information. Logic provides a qualitative yet precise way to analyze semantic information and it will be interesting to see which aspects of information they capture. Originally, logic only related to information in the sense that proof theory provide the rules of valid inference. What information can we infer from this set of premises? Later, with Hintikka's Epistemic Logic [54], it became possible to model the information states of different agents. As pointed out by van Benthem and Martinez [22], Epistemic Logic treats *information as range*. The more states of the world an agent considers possible (i.e. are in their range), the less information this agent has. The so-called "Dynamic Turn" in logic that was amongst others initiated by works by van Benthem such as [21] has made it possible to explicitly model informational processes. With Dynamic Epistemic Logic as described in [91, 11], we can model how agents update their information based on newly received information.

A different perspective on information that can be modeled by logic is *information as correlation* [22]. This is focused on how physical systems can carry information about other physical systems, in virtue of the correlations between them. The first logic that was designed to model information in that way is Situation Theory [17]. A more recent logic that is also focused on correlations is the Logic of Functional Dependence by Baltag and van Benthem [10]. It will become clear that this logic can model both how agents obtain information and how physical systems can carry information about each other. This latter part of the thesis and in particular the Dynamic Turn in logic, also relates to a third problem that Floridi identified, namely to determine the dynamics of information, i.e. how the states of physical systems can carry information about other physical systems.

While considering these quantitative and qualitative accounts of information we will again check if and in what sense according to these accounts semantic information is indeed semantic, mind-dependent and subjective. We will also consider whether quantitative or qualitative accounts of semantic information are better suited at capturing semantic information.

In the final and fifth chapter, we will then make a final comparison between Shannon information and semantic information. We will here attempt to answer the question whether they characterize fundamentally different types of information and whether they can both legitimately be called "information".

Chapter 1

Shannon's theory of information

In this chapter we look at Shannon's theory of information. Claude E. Shannon is referred to as "the father of information theory", and even as "the father of the Information Age", because his information theory and other work on early versions of the chip and the computer have been extremely influential. His work showed how the digitization of information could make communication more efficient and this has been the foundation for the advanced technologies we use today.

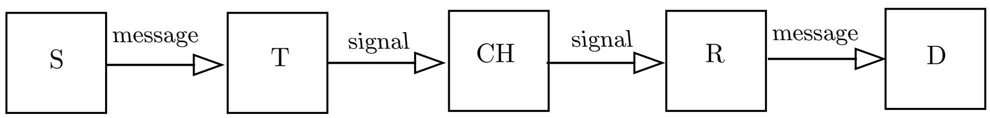
I will first explain some important elements of Shannon's theory that are required to understand in what context Shannon developed his measure of information. I will then consider the measure of information and how we can interpret it. There are two main possible interpretations of Shannon information: as a measure of uncertainty and as a measure of compressibility. Christopher Timpson [89], a philosopher of physics, argues that we should interpret Shannon information as a compressibility measure, not as a measure of uncertainty. Timpson thus holds that Shannon information does not have anything to do with notions like knowledge and uncertainty. I will go over Timpson's argument and consider whether it is possible that both interpretations will be used.

I will also consider some other concepts from Shannon's MTC, like individual Shannon information and mutual information, that will re-occur in this thesis and that can also be interpreted in relation to concepts like knowledge and uncertainty. In addition, I will weigh some of the arguments given by Timpson concerning these concepts. All of this will give us insight into the relation between the everyday semantic concept of information and Shannon information.

Finally, I will touch upon the relation between Shannon information and algorithmic information and briefly consider how algorithmic information relates to semantic information.

1.1 The Mathemathical Theory of Communication

The function that Shannon proposed as a measure information is part of his Mathemathical Theory of Communication (MTC) [87] which was brought out in 1948. This theory of communication was meant to make communication systems more efficient. At the time, the telegraph was the most used communication system. In the abstract, a communication system consists of an *information source* that sends a message, which is changed into a signal and then sent by a *transmitter* over a channel to a *receiver*. If it goes well, the received signal is interpreted as the original message at the *destination*. An important notion is the *capacity* of the channel, which is the maximal rate at which messages can be sent over the channel [87, p.3].



The focus in the MTC is on the *Noiseless Coding Theorem* that Shannon proves [87, p.16]. This theorem determines the relation between the capacity of the channel, how messages of the information source are coded into signals, and the information that we want to send. To make this relation clear, it is required to quantify the notion of information. Shannon does this by treating information as a property of the information source. He then uses the measure of information to determine what the required capacity of the channel should be when we use efficient coding. It is important to realize that this is the context in which the concept of Shannon information is developed. So how does Shannon propose to measure the information contained in an information source?

Shannon's measure of information

We should first look at how an information source can be characterized formally. An information source is treated as a stochastic process that produces one symbol after another with certain probabilities. Mathematically, a stochastic process is defined as a set of random variables that have the same probability space and that are indexed by some other set. In this case, the random variables are indexed by time so that the indexed random variables represent the sequential emitting of messages. If an information source transmits messages in a natural language like English, the values of the random variables can be letters of the English alphabet. There would thus be 26 different possible values, or 27 if we include the blank space. We know that in general in English texts, the letter 'a' occurs with a higher probability than the letter 'z'. We also know that it is likely that after 'th' the letter 'e' will follow. Each letter thus has both an independent probability of being produced by the information source and a probability of being produced after each other letter is produced.

Shannon uses these probabilities to calculate the amount of information produced by the information source. We can represent the information source with a single random variable X that takes different values sequentially. Then for every random variable X with n possible values and a probability $p(x_i)$ for each possible value x_i, the amount of information $H(X)$ produced by X is defined as follows:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{1.1}$$

Shannon uses the logarithm for practical purposes and he purposefully uses a logarithm with base 2 to get information in a unit of binary digits, i.e. *bits*. We can think of H as measuring how many bits on average are required to code a symbol coming out of the source. Shannon saw that calculating information in bits would go well together with the possibility of storing information in a device with two stable positions like an electric switch. This is the insight that led to the big technological advances in communication and information storage.

Shannon pointed out that qua form, the function H is very similar to Gibbs' function of entropy used in physics. He therefore sometimes refers to H simply as entropy, as has also become standard in the literature. In order to avoid confusion with Boltzmannian entropy I will however keep referring to H as "Shannon information" and only use the term "entropy" to refer to the specific form of the function 1.1. In the following chapter we look further into the relation between Shannon information and physics and it will become clearer how Shannon information and thermodynamic entropy are related.

1.2 A measure of compressibility or of uncertainty?

Two interpretations

Now that we have measure H, we can look at what it means to say that $H(X)$ is high for some X, i.e. to say that there is a lot of information in an information source. The interpretation that is close to the goal for which H was designed is that H is essentially a measure of *compressibility*. The idea is that if the information $H(X)$ of source X is low, this is because the probabilities of the signals are less spread out. Hence, some signals are considerably more likely than others. This means that clever coding can be used so that, for example, commonly used letters can be communicated with shorter signals in bits. The letter "e" for example can be coded with only one bit "1", while longer codes like "101" can be used for the letter "q". Since commonly used letters with shorter codes will be emitted more often, overall less bits are required to transmit messages. This means that the information of the source is more compressible. So if $H(X)$ is low, the information of X is more compressible. Vice versa, if $H(X)$ is high, the probabilities are more spread out and we cannot use shorter code in a clever way. Therefore, the information of the source is less compressible.

We must realize that Shannon designed his information measure in sight of the Noiseless Coding Theorem. This theorem states that if a source has information H (amount of bits required per symbol) and a channel has capacity C (amount of bits that can be transferred per second), then it is possible to code the symbols produced by the source in such a way that $\frac{C}{H} - \epsilon$ symbols per second can be transmitted where ϵ can be arbitrarily small. This bound cannot be exceeded. I will refer to this rate of symbol transmission (measured in symbols per second) with

"R^s". Thus, if H is higher and the capacity C remains the same, this means that the possible rate of transmitting symbols over the channel will be lower. Because if we need more bits to transfer symbols and we can only transfer so many bits per second, we can transfer less symbols per second.

Shannon argues that this theorem justifies interpreting H as the rate of generating information at the source. Because if we know that the overall rate of transmission becomes slower while the capacity stays the same, it seems that in some sense more information is coming out of the source. The type of "information" measured by H thus seems to be a very specific property that only makes sense tied to an information source and that is about how well we can code what is coming out of it. So it seems that Shannon information is not that similar to what most people have in mind when we think about information.

There is however another possible interpretation of H, one that is also pointed out by Shannon himself [87, p.11]. That is to think of H as a measure of uncertainty. It can be seen as a measure of uncertainty, because it measures how uncertain we are about which symbols the information source will emit. If $H(X)$ is high, the probabilities of each possible signal of the information source X are more spread out, and therefore we are more uncertain about the symbols X will emit. If $H(X)$ is very low and, for example, the probabilities of all but one signal are zero, we will know for sure that it will emit only that one signal. Hence, there will be minimal uncertainty.

Timpson's point of view

Timpson [89, p.27] holds that we should not interpret Shannon information as a measure of uncertainty. He acknowledges that we can interpret the function H as such, but claims that this is not the right interpretation of Shannon information. To make this argument, he first needs to argue that H as a measure of compressibility and H as a measure of uncertainty are two different concepts and then argue that the latter is not the concept that Shannon had in mind.

Timpson argues that H as a measure of compressibility and H as a measure of uncertainty are two different concepts, because as a measure of uncertainty, H is not necessarily unique, whereas as a measure of compressibility, it is unique. Shannon showed in the MTC that H is unique in satisfying the three axioms that he thinks a measure of information should satisfy. He describes the axioms as follows:

> 1. H should be continuous in the p_i.
> 2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n.
> 3. If a choice be broken down into two successive choices, the original H should be the weighted sum of the individual values of H. [87, p.10]

Timpson, however, notes that these three axioms are too stringent for a measure of uncertainty in general. Jos Uffink [90] has shown that the third axiom listed by Shannon does not necessarily have to be satisfied by a measure of uncertainty and that there are many other measures that have the same properties as H when it comes to measuring the uncertainty of a probability distribution. Hence, if H is interpreted as a measure of uncertainty, it is not unique.

Timpson notes that the only sense in which H can be the unique measure of information is if we require that the measure of information must measure compressibility. Uffink [90, p.71] makes it insightful that if we interpret H as a measure of compressibility, Shannon's third axiom does become a sensible requirement. I refer the reader to Uffink's work [90] for the technical details and this insight. It seems that it is this third axiom which leads to the Noiseless Coding Theorem. As a measure of compressibility, it is therefore reasonable to require the third axiom. Thereby, as a compressibility measure, H would be unique. Timpson concludes that since H is unique as a compressibility measure but not as an uncertainty measure, H as compressibility and H as uncertainty are two distinct concepts.

Two sides of the same coin

It is however not clear what Timpson has in mind when using the term "concept". Undoubtedly, H as a compressibility measure and H as an uncertainty measure are two distinct ways of looking at H. But to me it does not seem that this results in two entirely different "concepts". It seems to me that one function can only give rise to different concepts if it is also applied in a different way. In each application, the function is interpreted differently and could give rise to a different concept. However, both as a measure of compressibility and as a measure of uncertainty, the H function is applied to a random variable that ranges over possible outcomes of an information source. The two ways of looking at H fit this one application.

Concerning the fact that H is unique as a compressibility measure and not as an uncertainty measure: it can simply be the case that when we look at Shannon information as in the set of compressibility measures, it is the only one element in that set. And when we look at Shannon information as in the set of uncertainty measures, then it is not. That does not mean that the two perspectives on H lead to two entirely different concepts. There can be one concept: Shannon information, that can both be considered the unique measure of compressibility and one of the many measures of uncertainty.

After concluding that H as compressibility measure and as an uncertainty measure are two different concepts, Timpson asserts that we should interpret H as a measure of compressibility and not as a measure of uncertainty. His one and only argument as to why we should not interpret H as a measure of uncertainty seems to be that this was not the intended interpretation that Shannon himself focused on.

It can be granted that Shannon designed his concept of information in sight of the Noiseless Coding Theorem, which suggests that Shannon information should at least definitely be thought of as a compressibility measure. Shannon also said that his main justification for taking H to be the appropriate measure of information, was to be found in the Noiseless Coding Theorem which is about coding and therefore can be taken to depend on H as a measure of compressibility [87, p.16]. This, however, does not exclude the interpretation of H as a measure of uncertainty. As noted earlier, Shannon himself recognized this interpretation and even intended to call H an uncertainty measure [76].

I maintain that even though the interpretations are distinct, they are still both valid interpretations of Shannon information. We can also intuitively connect the two interpretations: what causes the uncertainty is the *unpredictability* of which symbol will be emitted. What makes it hard to compress is also the unpredictability, because if it is unpredictable which symbol will be emitted we cannot code them efficiently and therefore we cannot compress the symbols. The uncertainty and the compressibility interpretation can thus be considered two sides of the same coin, two properties of Shannon information.

In the next section, we will look at how Shannon information relates to our everyday understanding of information.

1.3 Our information about the information source

The everyday understanding of the concept of information has to with meaning. If we read the sentence "the milk is in the fridge" we say that the sentence contains the information that the milk is in the fridge. This is so, because the words in the sentence have a certain meaning for us. The information is always about something, some object or state of affairs. In this section we will see to what extent the link between this everyday, semantic understanding of information and Shannon information can legitimately be made.

Individual Shannon information

When Shannon's information theory is linked to our everyday understanding of what information is, it is often done by focusing on the uncertainty interpretation of H. It is also often mentioned that we can associate an amount of Shannon information with individual possible source outputs.[1] Some accounts of Shannon information treat this individual information value as part of Shannon's information theory, but originally it was not used or referred to by Shannon himself. It can, however, be derived from Shannon's theory and therefore we can take it into account when it comes to the interpretation of H. For every symbol x_i, the individual amount of information $I(x_i)$ that is generated at the source can be calculated as follows:

$$I(x_i) = -\log p(x_i) \tag{1.2}$$

It makes sense intuitively that this is the Shannon information for individual source outputs, because we can see how it relates to the general information H of the source. If $I(x_i)$ is the amount of information that is generated by each output x_i of source X and we want to get the overall amount of information $H(X)$ that is produced at the source, we simply need to compute the weighted average of the information $I(x_i)$ produced by all symbols. To do so we add the individual values for all symbols, multiplied by the probability that these symbols occur. And this is indeed what H looks like (see 1.1).

[1] One example is Dretske [38] who argues that the individual Shannon information can measure semantic information, just as Bar-Hillel and Carnap [15] have done. Isaac [3] also argues that individual Shannon information provides a basis to analyze the semantic content of messages.

Surprisal value

The individual amount of information per symbol is sometimes taken as an example of how Shannon information does agree with our intuitive understanding of the concept of information. If the probability that some symbol x will occur is low, this means that $I(x)$ will be higher. In this case we will also be more surprised if the source does emit this symbol. This is why $I(x)$ is often called the "surprisal" of symbol x [52, 89]. If we would only be interested in knowing what symbol a source will emit, a symbol with a higher surprisal value can be considered to give us more information. Since we expect symbols with a lower surprisal value, it will be less informative if this symbol appears. In this sense the function I can be considered to accurately measure information in a way that corresponds to our intuitive understanding of the term "information".

Timpson stresses [89, p.30] that it is only in this specific scenario, when we are exclusively interested in the symbols coming out of the source, that I can accurately measure our gain of information. If we were interested in measuring information in general, information that may be about any object whatsoever, I would not be a good measure of this. One reason for this is that Shannon information (whether we take H or I) does not have anything to do with what the symbols coming out of a source might mean to us. The only thing that I can tell us is how likely it is that some symbol will appear. And the only thing that H can tell us is how unpredictable the source is.

Shannon information and semantics

Shannon himself is also clear about the fact that his measure of information does not quantify the content of messages [87, p.1]. As is often quoted, he states in the introduction of the MTC that semantic matters are irrelevant to the problem that he is concerned with, the problem of communication:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. [87, p.1]

Here, Shannon also stresses the fact that it is about the set of messages from which an actual message is selected, i.e. that H is a property of the information source, not of a message taken on its own. The surprisal value I also depends on the information source, because it depends on the probability distribution that is connected with it. This is another reason why H and I cannot tell us anything about the amount of information that these symbols in general can convey, only about how much information they might convey if we are only interested in finding out the symbol output by the source. If we would want to use H and I to measure information in general, we would need a probability distribution over all possible messages that may exist.

In conclusion, there seems to be no connection between Shannon information and the content of messages or the meaning of symbols. In so far as we can, nevertheless, make a link between the Shannon information I generated at the source by an individual symbol and the information we gain from seeing that the source emits this symbol, this holds only if we confine ourselves to a small domain of possible information. Moreover, this link is not due to the fact that Shannon information tells us something about the content of messages, but because it tells us something about how likely a certain message is. So if the information I of a message is high, this is merely because we are surprised to learn that it is true. And we are surprised to learn that it is true not because of the content of the message, but because we knew beforehand that it was not likely to be true.

Surprisal value and uncertainty

Previously, I argued against Timpson that the interpretation of H as a measure of uncertainty can co-exist with the interpretation of H as a measure of compressibility and that they can be two sides of the same coin, of Shannon information. The interpretation of I as a surprisal value makes it more insightful how H measures uncertainty. Since $H(X)$ is the weighted average of the $I(x_i)$ for all possible output symbols x_i of a source X, we can think of $H(X)$ as the expected surprisal value of a symbol emitted by X. So the higher H, the higher the expected surprisal value, the more uncertain we are about the symbols H will emit. This connection to uncertainty is to be expected, because uncertainty about the source output and the information we could gain by learning the source output can be seen as two ways of approaching the same thing. This is why the interpretation of Shannon information as a measure of uncertainty is often used in explanations. It makes it more intuitive to think of Shannon information as a measure of information.

Mutual information

There is another concept in Shannon's information theory of which the interpretation has a link to everyday information. This is the value of *mutual information* [31, 89]. It tells us how much information the symbols arriving at the destination can give us about the symbols that were emitted at the source. In more general terms, it measures how much information we obtain about one random variable by measuring another.

If we call the source X and the destination Y the mutual information can be calculated as follows:

$$H(X:Y) = H(X) - H(X|Y) \tag{1.3}$$

In this equation, $H(X|Y)$ represents the *conditional entropy* of X given Y. This is understood as the average entropy of the conditional distributions of X given some value y_j of Y. The entropy for a conditional distribution of X given some specific value y_j of Y is as follows:

$$H(X|Y = y_j) = -\sum_i p(x_i|y_j) \log p(x_i|y_j) \tag{1.4}$$

From this we can calculate the conditional entropy by taking the weighted average of the entropies of the conditional distribution of X given y_j for all y_j:

$$H(X|Y) = \sum_j p(y_j)H(X|Y = y_j) = \sum_j p(y_j)(-\sum_i p(x_i|y_i)\log p(x_i|y_j)) = -\sum_{i,j} p(x_i, y_i)\log p(x_i|y_i) \tag{1.5}$$

Conditional entropy (1.5) thus measures how uncertain we are of x on average when we know y. Shannon also phrases it as follows: "It measures the average ambiguity of the received signal." [87, p.20] It should be noted that conditional entropy is technically not an entropy itself, because it is not of the same form as $H(X)$. It is rather an average of entropies.

Conditional entropy tells us something about the channel between a source X and destination Y. It can be the case that there is some degree of *noise* that affects the communication between X and Y. This causes that sometimes a signal is received at the destination that is not the same as the signal that was transmitted. If a channel is noiseless, the conditional entropy $H(X|Y)$ will be zero. In this case, if we receive a signal we can be sure that this is the same symbol that was transmitted at the source. The more noise in the channel, the higher $H(X|Y)$ becomes.

Mutual information in Shannon's theory

We can now get back to mutual information $H(X : Y)$ which is given above in 1.3. This is the entropy of the source X minus the conditional entropy of X given Y. Shannon himself did not give this quantity its own name, but the quantity occurs in the MTC [87]. The quantity that we call "mutual information" functions as the definition for the rate of transmission of information in the case of noisy channels. The rate R of transmission of information in a noisy communication system is lower than if there were no noise. That is because we cannot be sure that the symbol received is equal to the one that we sent. For noisy channels, Shannon therefore defines the rate R as follows:

$$R = H(X) - H(X|Y) \tag{1.6}$$

We see that R is what we now refer to as mutual information. We need to note that we are here talking about the rate R of how many bits per second can be transmitted, which is different from the previously used R^s which was measured in symbols per second. To get R, we simply multiply R^s with $H(X)$. Since the capacity of a noisy channel should be equal to the maximal transmission rate possible, the capacity C of a noisy channel is defined as the maximal mutual information:

$$C = Max(H(X) - H(X|Y)) = Max(H(X : Y)) \tag{1.7}$$

Here, we maximize the transmission rate of the channel for all possible information sources X.

Shannon's *Coding Theorem for Channels with Noise* states that if we have a channel with capacity C, then it is possible to encode the symbols coming out of a source X in such a way that the conditional entropy $H(X|Y)$ can get arbitrarily close to zero as long as the transmission rate R stays below C, i.e. $R < C$. This means that the error rate of the signals received can get arbitrarily close to zero as long as R stays below C. The reason that the conditional entropy can get arbitrarily small is that we can send information in a redundant way by sending messages multiple times. If the receiver then receives one message multiple times with a few altered due to noise, they can still infer the original message with almost complete certainty. So this is the context in which mutual information occurs in Shannon's theory. It defines the transmission rate and the maximum possible transmission rate (i.e. the maximal mutual information) is the capacity of the channel.

The interpretation of mutual information

Now we can consider the interpretation of mutual information. One way to see how we can interpret $H(X : Y)$ is as the average number of questions that are required to find out the value of X if we know the value of Y. This is because a bit can be considered to store the answer to a yes/no question and therefore $H(X)$ can be considered to give the number of such yes/no-questions that need to be asked in order to determine the value of X.

To see this, consider the fact that $H(X)$ is the average individual information contained in a signal emitted from the source. This is also the average amount of bits that are required to code a signal from the source. Suppose that $H(X) = 3$. This means that 3 bits are required to code a signal from the source. A bit can be thought of as storing the answer to a yes/no-question. So if we were able to ask 3 yes/no-questions, we would on average be able to find out the signal coming from the source.

We can then also think of $H(X|Y)$ as measuring the average number of questions required to find out the value of X if we know the value of Y. We then see that $H(X : Y)$, being equal to $H(X) - H(X|Y)$, measures how many fewer questions on average are required to know X if we know Y relative to when we do not know Y. We can therefore say that $H(X : Y)$ measures the reduction in uncertainty about X given Y. We can, in turn, interpret this as saying that $H(X : Y)$ measures how much information we can gain about X by getting to know the value of Y.

But is this information in the everyday sense or does $H(X : Y)$ measure the Shannon information contained in Y about X? As Timpson has pointed out, we must realize that $H(X : Y)$ does not measure how much *Shannon* information y gives about x. Shannon information cannot really be *about* anything. On top of that, it is also the case that it normally means that if the Shannon information of some object is high, our uncertainty about that object is also high. However, if $H(X : Y)$ is high, this means that Y decreases our uncertainty about X. We therefore cannot say that $H(X : Y)$ gives the amount of *Shannon information* that Y has about X, nor can we say that it gives the amount of *Shannon information* that we can gain about X by observing Y. It quantifies the information in the everyday sense that *we* can gain about the signal that is sent at X by observing Y.

We can also note, however, that the quantity of mutual information does not seem to tell us anything about the meaning of the information that we can get from one variable about another. It does not seem to dependent on the meaning of the different possible values that the variable can take.

Mutual information as a measure of dependency

There is also a different way of calculating mutual information which can give more insight into what it measures:

$$H(X : Y) = H(X) + H(Y) - H(X, Y) \tag{1.8}$$

Here, $H(X, Y)$ denotes the joint entropy of X and Y and is calculated as follows:

$$H(X, Y) = -\sum_i \sum_j p(x_i, y_j) \log(p(x_i, y_j)) \tag{1.9}$$

This way of calculating mutual information (as in 1.8) shows that mutual information can also be thought of as the difference between the entropy of the joint distribution and the entropy of the joint distribution of X and Y if they were independent. This is because $H(X) + H(Y)$ would be the entropy of the joint distribution if X and Y were independent and $H(X, Y)$ is the entropy of the actual joint distribution given the dependency between X and Y.

This means that mutual information quantifies the amount of dependency between X and Y. Since we can use this dependency to infer information about one variable from the other, it measures the amount of information we can derive about one variable from another. We will be able to use this later in Chapter 4 when we look at ways to quantify dependence within a logic of dependence.

1.4 The connection with algorithmic information theory

As I have mentioned in the introduction, although Shannon's measure of information is the most well-known and widely used measure of information, it is not the only one that is available. Another measure that is gaining popularity is that of *algorithmic information*. One important difference with Shannon information is that the algorithmic measure can be used to measure the information contained in individual objects without reference to any probability distribution or to any ensemble of which this object is a part. This is the main reason why algorithmic information is considered to be a more fundamental measure by some, for example by Cover and Thomas [31, p.3]. To clarify the relationship between Shannon information and other information-theoretic measures of information I will therefore also look into algorithmic information. I will consider some of the similarities and differences between Shannon information and algorithmic information and see how algorithmic information relates to the concept of everyday information.

Kolmogorov complexity

Algorithmic information theory is based on the idea of *Kolmogorov complexity*, which was amongst others proposed by Kolmogorov [58]. This is a measure that determines for an object x the length of the shortest program written in binary code that is required to produce x as its output. This is why it is referred to as "algorithmic information". We can also think of it as a measure that determines the length of the shortest description of the object x. If there is no regularity whatsoever in this object, then this means that it takes more code to describe the object or to design a program that produces the object. This means that in this case the complexity of the object is higher, and therefore the amount of information in object x is higher.

If the object that we want to calculate the complexity of is the string 1111111111, we can design a program that prints 10 times '1' and then halts. This will take less bits than required to print the string 1000101101 which also consists of 10 symbols but which shows more randomness. Thus the complexity of 1000101101 is higher than that of 1111111111, and therefore the amount of algorithmic information in the former string is also higher. And this can be determined independently of any probability distribution over a set of strings.

Comparing Shannon information and algorithmic information

For the purpose of this thesis, but also in general, it is interesting to know how much the values of Shannon information and algorithmic information differ from each other. For those who argue that information is an objective quantity that exists independently of mathematical construct, it would be significant if Shannon information and algorithmic information correspond. This can be taken to show that they are two ways of tracking the same mind-independent quantity. Li Ming and Paul Vitányi [68, p.275] write about this: "In mathematics the fact that quite different formalizations of concepts turn out to be equivalent is often interpreted as saying that the captured notion has an inherent relevance that transcends the realm of pure mathematical abstraction." They also suggest that if two different formalizations of the same concept that were not designed to correspond eventually do correspond, may be taken to show that this concept that was formalized is not merely something created by mathematics.

One may object that it can also be a coincidence that the two formalizations correspond, but then at least we know that Shannon information and algorithmic information can indeed both legitimately be referred to as two types of "information", or at least as two types of the same concept. If Shannon information and algorithmic information correspond, this would mean for this thesis that some of the conclusions drawn about Shannon information can also be applied to algorithmic information.

A surprising result

Li and Vitányi [68, 51] have proved that in fact there is a way in which Shannon information values and algorithmic information values correspond to each other. We first need to note something about Shannon information. It follows from Shannon's Noiseless Coding Theorem that for an information source X, the average length of the code words required for each output x will approach $H(X)$. This makes sense intuitively when we think of $H(X)$ as measuring the weighted average of the surprisal values $I(x_i)$. As explained in the sections before, this surprisal value $I(x)$ can be considered to represent how much information we gain from observing output

x. Thus, $H(X)$ can be considered to represent the average number of bits of information per output symbol x. It makes sense that this value is the amount of bits that is on average at least required to code this symbol. When *Shannon-Fano* code is used, which is the optimal coding that associates probable symbols with shorter codes and improbable symbols with longer codes, the average number of bits required for each symbol will get closer to $H(X)$ [51].

We can now look at how Shannon information and algorithmic information correspond. We need to consider an information source X with its corresponding probability distribution. We know that the average code word length required for source outputs approaches $H(X)$. We also know that for each possible source output x, we can find the complexity $K(x)$, i.e. the length of the shortest code required to describe x. We can then calculate the expected complexity (code word length) of a possible source output as follows:

$$\sum_i p(x_i)K(x_i) \tag{1.10}$$

Here we again take the sum of the complexities of the source outputs multiplied by the probability that this output occurs. Li and Vitányi [68] proved that the expected complexity of a source output differs from the Shannon information only within an additive constant c_P[2], given some restrictions on the probability distribution P of the source:

$$0 \leq \sum_i p(x_i)K(x_i) - H(X) \leq c_P \tag{1.11}$$

Here c_P is a constant that depends on P, the probability distribution of the source. Grünwald and Vitányi [51] and Cover and Thomas [31] argue that we can interpret this result as saying that given the restrictions on P, the expected Kolmogorov complexity and the Shannon information of a source are "close" to each other. Li and Vitányi [68, p.630] have also showed that the values of Shannon's mutual information and the expected algorithmic mutual information are close to each other as well.

The fact that algorithmic information and Shannon information will correspond in this way can seem suprising, because since K measures the information of individual objects and H only the information of objects in a collection with a probability distribution, one easily thinks that algorihmic information and Shannon information are completely different concepts. We see, however, that here we specifically look at the expected algorithmic complexity of the source outputs of some random variable. A correspondence between Shannon information and algorithmic information only holds if we already start with some random variable of which we know the Shannon information. We can also note that it make sense that $H(X)$ and $\sum_i p(x_i)K(x_i)$ correspond, because both $H(X)$ and $K(x)$ measure compressibility. Shannon information measures the compressibility of symbols coming out of a source and algorithmic information measures the compressibility of individual objects.

[2] The interested reader can find the proof for the result in 1.11 in Li and Vitányi's book [68, p.626]. It is too lengthy and too complex to reproduce here.

Similarities between Shannon information and algorithmic information

Whereas Shannon information derives the average compressibility from the *probability* of the outputs, algorithmic information derives compressibility from the *complexity* of the outputs. One could argue, however, that algorithmic information can also be used to derive compressibility from probabilities. This is because the Kolmogorov complexity of an object can be used to calculate the *universal probability* of this object. This probability is independent from any predetermined probability distribution and is simply based on how complex the object is. The idea is that we can think about how likely it is that an object x will be produced by a program assuming that this program is completely random. Since this program is a piece of code itself, we can think of it as being produced by randomly selecting either 0 or 1 sequentially. It has been shown by Vitányi [92] that roughly, the universal probability p_u of an object x is:

$$p_u(x) = 2^{-K(x)} \tag{1.12}$$

This probability is thus dependent on $K(x)$. This means that when we say that algorithmic information can be used to derive the average compressibility from probabilities, these probabilities are in turn dependent on algorithmic information.

We can now also express algorithmic information in terms of universal probability. We can rewrite 1.12 so that we get:

$$K(x) = -\log p_u(x) \tag{1.13}$$

This looks surprisingly similar to how we compute the individual Shannon information or surprisal value $I(x)$ (see 1.2). Both algorithmic information and individual Shannon information are calculated as the negative logarithm of a probability of an object. Both values also characterize the amount of information in object x, or the amount of information we can gain by observing the object. In general, $I(x)$ and $K(x)$ will, however, not have the same value for some object x, because the former is dependent on the probability distribution of some information source and the latter is not. There is no reason to assume that the universal probability of some message will be equal to the probability that this message will be emitted from a specific information source.

Two formalizations of the same concept?

All of this shows that while Shannon information and algorithmic information may seem very different, they in fact give similar results if we consider a scenario involving a source with different outputs. This would seem to allow us to say that Shannon information and algorithmic information are simply two formalizations of the same concept.

One may object to this that it is still the case that in scenarios when we just want to consider the information of an individual object we cannot calculate the Shannon information of this object. This suggests that they are not similar at all and therefore are two different concepts.

We can, however, reply to this that in the case of individual objects, we could choose to calculate the individual Shannon information by using the universal probability rather than a pre-determined probability distribution. This would then result in the same value as the algorithmic information of this object. In this light, Shannon information would then be a specific case of algorithmic information in which we do not consider the information of an object *an sich* but the information of an object in a specific context of an information source. This would mean that algorithmic information is indeed the more fundamental concept and that Shannon information is the same type of concept but in a specific scenario.

We can at the very least conclude that Shannon information and algorithmic information indeed can be thought of either two very similar concepts, or of one single concept applied in a more general and a more specified context.

Does information measure uncertainty?

We can also ask ourselves how the interpretation of algorithmic information may bear on the interpretation of Shannon information. It seems that algorithmic information can only be interpreted as a compressibility measure, and not as a measure of uncertainty. The fact that Shannon information can be interpreted as such seems to be inherent to the fact that it is calculated based on a random variable that causes uncertainty on our part about the value of this variable.

The fact that algorithmic information cannot be interpreted as a measure of uncertainty could be taken to show that Shannon information should also not be interpreted as measuring uncertainty. Lombardi et al. [71] argue that if one holds that Shannon information should be interpreted as a a measure of compressibility and not of uncertainty, it will be easier to assume that Shannon information and algorithmic information deal with the same concept. This would imply that if we would start from the idea that there is only one concept of information of which Shannon information and algorithmic information are merely two formalizations, we would have to conclude that Shannon information should not be interpreted as a measure of uncertainty.

We can, however, disagree with this line of thought. I argued in Section 1.2 that Shannon information can both be interpreted as a measure of uncertainty and of compressibility. It is indeed essential to the nature of Shannon information that it measures compressibility, but also that it measures uncertainty. If it did not measure uncertainty in the way it does, it is hard to see how it could measure compressibility at all. It measures compressibility by taking into account how spread out the probability distribution is. I referred to this as the unpredictability of the source. If the source is unpredictable, the compressibility of the source is lower. This unpredictability is also what leads to uncertainty on our part. It is therefore an essential part of Shannon information that it can measure uncertainty. If it did not, it is hard to see how it could come to agree with the expected algorithmic information as in result 1.11. I therefore argue that the fact that algorithmic information does not measure uncertainty should not be taken to show that Shannon information cannot be interpreted as such either.

Algorithmic information and everyday information

Although algorithmic information cannot be interpreted as a measure of uncertainty, we can consider whether it links in any other way to the everyday concept of semantic information. Does the Kolmogorov complexity of an object measure how much everyday information is contained in that object? On the face of it, it does make sense that a more complex object can contain more everyday information. If a sequence is very long, for example, it is also more complex and it can contain more information.

There is one main problem, however, if we really wish to use algorithmic information as a measure of everyday semantic information. The problem is that according to Kolmogorov's measure, a very random object is also very complex, whereas we do not consider random objects to contain a lot of everyday information. An example that Fouad Chedid [29] uses to illustrate this problem involves two strings x and y where x is a string of a 1000 characters from War and Peace by Tolstoy and y is a randomly generated string. String y will contain more algorithmic information because it does not have many regularities that allow a short description. String x will contain many regularities with words that are often repeated like "the" and "or" and letters that often occur together. This means that x will allow a shorter description and therefore contains *less* algorithmic information.

Chedid [29], however, points out that Kolmogorov himself provided a solution to this problem. It is possible to distinguish the random parts in a string from the regular parts. If we then want to measure the amount of everyday information in y we must look only at the regular pieces of string in y and measure the kolmogorov complexity of these parts. We know that string y is very random so almost all the algorithmic information in this string comes from the random parts. In string x, there are almost only regular parts. This means that string x will contain more information than y, which is what we want. A similar technique is used by Vitanyi [93] to define his concept of meaningful information.

One can still wonder, however, what happens if we then compare two strings that have the same length and in addition contain the same amount of regular parts. If we then calculate the complexity of the regular parts of these two strings, it could be that one string is more complex. But this complexity only tells us something about the string itself, not about what the information in the string is *about*. It seems that the only thing that the measure captures which is related to semantic information is the length of the regular parts, because this tells us something about how much data was actually intended to convey information. But it does not seem that it can tell us something about the content of the regular parts. It is thus still unclear if and in what sense algorithmic information can actually measure everyday semantic information.

1.5 Conclusion

We can now conclude this chapter on Shannon's information theory. We saw that Shannon information is a property of an information source. It is a mathematical function that measures the predictability of the outputs of the source.

We saw that Shannon information can both be used to measure the average compressibility of the source outputs and our uncertainty about the source outputs. Timpson argued that we should only interpret it as a measure of compressibility, but I argued that we can also interpret it as a measure of uncertainty.

We also considered the individual Shannon information of a message and the mutual information between a source and a destination. We saw that the individual Shannon information can measure our amount of surprise upon observing that message. In this sense it does measure something about everyday, semantic information but only confined to the context of the possible messages of the source. It does not quantify semantic information in general. The amount of individual Shannon information is only based on the probability that this message will occur and not on the meaning of this message.

Concerning the mutual information between two variables, we saw that it can be used to convey how much everyday information we can gain about one variable by observing the other variable. We also saw that we can interpret it as measuring the dependency between the two variables form which we can get information. We again noted that the amount of mutual information measured does not have anything to do with the meaning of the values of the variables. It thus quantifies everyday information only based on our uncertainty about the probability distribution over the values of this variable.

In the final section we also looked at the relation between Shannon information and algorithmic information. We saw that there are more similarities between algorithmic information and Shannon information than may seem at first sight. They both measure compressibility and the expected Kolmogorov complexity of the outputs of a source will be equal to the Shannon information of that source within some additive constant. We also noted that the Kolmogorov complexity of an object can be formulated in terms of the individual Shannon information of that object relative to a universal probability distribution.

We also noticed two differences with respect to the interpretation of Shannon information and algorithmic information. One is that Shannon information can be used as a measure of our uncertainty about the outputs of a source, whereas algorithmic information cannot. Another difference is that the fact that Kolmogorov complexity measure the compressibility of objects means that it could tell us something about how much everyday information is contained in this object. We however also noted that it remains to be seen how exactly it tells us anything about the the content of the information contained in these objects.

We have so far only considered the technical details of Shannon information and the relation with semantics. In the next chapter we will go more into the interpretation of Shannon information.

Chapter 2

The interpretation of Shannon information

In this chapter, we will look more into the interpretation of Shannon information. We will specifically look at whether Shannon information can indeed be characterized as objective, mind-independent and even as concrete and physical. We will start by considering whether Shannon's measure quantifies the flow of some sort of concrete substance. I will explain in line with Timpson [89] why this idea is not correct. Secondly, we look at whether Shannon information can be considered to be a physical quantity. For this purpose we will go into the interpretation of Shannon information in physics. We will finally also look at whether Shannon information can be considered objective and mind-independent.

2.1 Shannon information as a concrete substance

In the previous chapter, we characterized what Shannon information is. We saw that Shannon's measure is also used to quantify the amount of information per second that can be transferred over a channel. In this context, Shannon information could be interpreted as a concrete substance that flows from an information source to the destination. Lombardi and Timpson both point out that this interpretation exists [71, 89]. Krzanowski [59] also argues that the idea of concrete information exists in philosophical literature and refers to multiple authors who supposedly defend this idea. In this section, we will consider if it is possible that Shannon information is a concrete substance.

Pieces of information

I agree with Timpson [89], who argues that it would be wrong to interpret Shannon information as something concrete. He argues that information is abstract in nature and therefore cannot be concrete. His argument for why information is abstract is based on a distinction between two different types of Shannon information. The first type is the quantitative concept of Shannon information that we have examined in the previous chapter. The second type is that of a *piece* of Shannon information. Timpson argues that Shannon's theory also introduces a concept of what pieces of information are. Timpson defines a piece of Shannon information as "what is produced by an information$_t$ source that is required to be reproducible at the destination if the transmission is to be counted a success."[89, p.22]. Timpson uses the term "information$_t$" to refer to Shannon information.

This definition, however, cannot be found in Shannon's theory. Shannon never mentions anything about pieces of information in the MTC. This is also pointed out by Lombardi et al. [71, 70]. The definition simply seems to characterize a piece of information in the context of a communication channel which is perhaps inspired by the MTC. Nonetheless, it could indeed be that some people will think of Shannon's measure as being about pieces of information, which could make it more tempting to think of Shannon information as being concrete.

However, as Timpson's definition indicates, even such a piece of information cannot be concrete. The definition equates a piece of Shannon information with the *sequence type*, which is that which can be produced by a *sequence token* at the source and reproduced by another sequence token at the destination. A sequence type is abstract and therefore cannot be concrete.

Even if we would not follow this specific definition by Timpson, pieces of information in general, also in the context of semantic information, are not concrete. We will go into what pieces of semantic information are in the next chapter, but it is straightforward that one piece of information can be contained in or conveyed by different objects. Two utterances of a sentence can contain the same piece of information. Two different sentences, as in sentence types, can even contain the same piece of information. This suggests that a piece of information is something abstract and not concrete.

The quantity

Now that we have established that pieces of information are not concrete, we can look at what is generally understood to be Shannon information, which is simply the quantity as defined in the previous chapter. The quantity represents a property which can be ascribed to both an information source and to an individual message (with the surprisal value).

One reason why Shannon information could be considered concrete is because it is a quantity that is used in physics. Even though a physical system does not emit messages, we can still calculate the Shannon information contained in it. Instead of a probability distribution over different messages, we can take the probability distribution over the possible configurations that the physical system can be in. This comes down to the notion of entropy and we will look at the relation between Shannon information and entropy in depth in the next section.

We can thus ask the question "How much information is contained in region x?". Nevertheless, Shannon information is not something concrete. Timpson explains this by means of an analogy with energy, a physical quantity which also does not quantify a concrete substance.

> Or to take a more interesting physical example, the field in a certain spatio-temporal region might have a particular energy (the field has such-and-such an energy density

> associated with the various spatio-temporal points making up the region), but the property of having that energy isn't something located in the region: it is the object (part of the field)—the thing having the property—which is. To force the point home: to ask 'How much Shannon information is contained in region x?' is a very different kind of question from the question 'How much syrup is located in region x?' In the former case, one is asking what the value of energy possessed by the objects (if any) in the region x is; in the latter, one is asking about the spatio-temporal distribution of a physical stuff—something which genuinely has a location of its own. [89, p.19]

So in short: information is not like syrup. It is not some concrete substance that we can touch, that exists in space-time. How much information is contained in a certain region may partially depend on the concrete stuff that is present at that location, but the information itself is not concrete stuff. When we say that information "flows" from a source to a destination, we can only refer to the fact that information is contained at different locations in space-time, but the information itself does not move through space-time because it is not concrete.

Another reason why Shannon information may be called "concrete" is because the meaning of the term is confused with the meaning of the term "physical". Krzanowski [59] refers to authors who, according to him, defend the idea that information is concrete. However, in none of the articles to which he refers it is mentioned explicitly that information is concrete. Nor is it mentioned that Shannon information is something we could possibly interact with. What Krzanowski [59] may be referring to is the idea that information is the structure of physical objects. This is an idea which comes up in some of the articles that he refers to. But if information is the structure of objects, this does not immediately imply that information is concrete. Concrete objects have structure, but structure itself, taken separately from the matter, is not something that we could touch, and hence is not concrete. So either Krzanowski equates the meaning of the term "physical" with the term "concrete" or he does not represent these authors' views in an entirely correct way.

2.2 Information as a physical quantity

While Shannon information cannot be something concrete, it can still be something physical. An important reason why Shannon information can be considered physical is that it has started to play an important role in physics. There is a certain connection between thermodynamic entropy and Shannon information, but there is a lot of confusion about what this connection really is. The terms "information" and "entropy" are sometimes used interchangeably, but at other times are treated differently. I will attempt to uncover some of the confusion surrounding this connection. We will then look at a principle from physics called "Landauer's principle" which is also taken to show that information is physical.

Trivial or baffling

We can first look at what Timpson has to say about whether information is physical. He [89] has argued that it would be either baffling or "of little interest" to state that information is physical. This depends, he explains, on the interpretation of the term "information". If we refer to Shannon information, Timpson judges that the claim that information is physical is trivial, presumably because Shannon information is a quantity used in physics which automatically makes it physical. Shannon information is, however, not a quantity that was initially defined by a physical theory. It is therefore not trivial to show why Shannon information does indeed play an important role in physics, which is what I will seek to do in this section.

On the other hand, if we refer to the everyday concept of information, Timpson holds it would be baffling to say that information is physical "because the everyday concept of information is reliant on that of a person who might read or understand it, encode or decode it, and makes sense only within a framework of language and language users; yet it is by no means clear that such a setting may be reduced to purely physical terms."[89, p.2]

It is, however, not entirely clear what the criterion is for information to be considered physical. What does it mean that it should be reducible to physical terms?

What does it mean to say that information is physical?

The term "physical" can be interpreted in slightly different ways. In the Introduction I argued that it is best to separate the meanings of "physical" and "concrete". I hereby follow the interpretation suggested by Lombardi et al. [71]. A concrete object is something that has a location in space and time, something that we can in principle interact with. A physical object or property does not have to be concrete, but it should be described by physics, or any other natural science.

Many physical properties are observable properties, but not all of them; some physical properties are not observable. In practice, we do see that if properties discussed in physics are not observable, there is at least some significant link between that property and other observable properties. An observable property is one that can be observed with the human senses. Here I do not use the interpretation of unaided observability that philosopher van Fraassen [18] argues for, but for a more textbook understanding of observability in physics where we are allowed to use tools such as microscopes and minor inferences. In any case, we can say that if there is a significant link between a property and an observable physical property, this means that the former property can also be considered to be physical.

According to this interpretation of the term "physical", a physical object or property can also be abstract at the same time. We can look at the concept of energy as an example. Energy is an abstract property. It can have a particular physical manifestation, like kinetic energy that manifests in movement which we can observe, but energy itself is not some concrete stuff, it is not like syrup. It is, however, a physical property, because it is dependent on observable properties like temperature. If the kinetic energy of the particles of a gas is high, the temperature will be high and vice versa. Thus, energy is a described by physics and is a physical property.

So according to my use of the term “physical”, when we say that information is a physical property, we are essentially not saying more than that information is a property that is described by physics. In addition, if information is related in an important way to observable physical properties, we can say that it belongs to the domain of physics and can therefore be considered physical. In what follows we will check if this is the case.

2.2.1 Shannon information and entropy

We will begin by looking at the relation between Shannon information and entropy. Shannon himself already noted that his measure of information was very similar in form to the measure of entropy used in statistical mechanics which is known as Gibb’s entropy. To make matters more complicated, there is not just one way to calculate entropy in physics, but three. The first measure stems from the early days of thermodynamics, known as Clausius’ entropy. Then there are Gibb’s and Boltzmann’s measure of entropy which approach entropy by taking into account the possible microstates of a physical system. We will first look at Clausius’ measure.

Clausius’ entropy

The first time entropy was mentioned in physics, it had nothing to do yet with probabilities, and the formula to calculate entropy did not look anything like Shannon’s measure of information. The term “entropy” was proposed by Clausius in 1865 to refer to the quantity he used to formulate the *Second Law of Thermodynamics* [30]. One interpretation of this law is that heat always flows from hot systems to cold systems [23, p.126]. Another interpretation is that isolated systems are always more likely to go from lower entropy states to higher entropy states [23, p.141]. We can understand the connection between these two interpretation as follows: if we have two objects where object A is hot and object B is cool, we know that if A and B are in contact, heat will be transferred from A to B. Entropy was calculated as a difference in the entropy of a system that is caused by heat transfer. According to Clausius’ formula, the difference in entropy of a system can be calculated as follows:

$$\Delta S = \frac{\Delta Q}{T} \tag{2.1}$$

The quantity ΔQ is the amount of energy transferred (heat) to or from the system and its unit is Joule (J). T is the temperature of the system given in Kelvin K. The unit of entropy is therefore J/K. As defined here, ΔS is technically not a state variable, but if this is integrated on a path from one microstate to another, we do obtain a state variable $S = \int \frac{\Delta Q}{T}$.

If heat flows from object A to object B, it can be calculated with Clausius’ measure that the entropy lost from the hot object A will be lower than the entropy gained by cold object B. This means that overall, the entropy of the system consisting of the objects A and B will increase. This shows that if heat flows from hot to cold, the entropy of the system will increase.

We see directly from Clausius’ measure that entropy is related to temperature which is an observable quantity. A change in the temperature of a system will lead to a change in the entropy of that system. The way we understand entropy now, however, is that it is not only dependent on temperature but also on other macro-variables like volume and pressure. This is because we have started to calculate entropy in a different way which seems to have broadened our understanding of the concept.

Boltzmann's entropy

The two measures of entropy that are now widely used take a different approach than Clausius' measure of entropy. These are Boltzmann's and Gibbs' entropy that were proposed in 1877 [24] and in 1902 [46] respectively.

Boltzmann's work focused on the microscopic variables of systems like the positions of particles which gave rise to a new branch of physics: statistical mechanics. By applying probability theory to ensembles of microstates, he formulated a different way to calculate the entropy of a system. A *microstate* is the specific configuration of all the particles and energies in a system. It therefore takes into account the positions, velocities, mass and energy of the particles. A *macrostate* only takes into account the macroscopic variables of a system like temperature and volume. Boltzmann defined the entropy for a system in equilibrium as being dependent on the number of possible microstates that a system can be in relative to the macrostate that we can observe. Boltzmann's measure for entropy S is as follows:

$$S = k_B \ln W \tag{2.2}$$

The term W refers to the number of microstates of the system and ln is the natural logarithm. The constant k_B is the Boltzmann constant that ensures that S will have the same unit as Clausius' entropy, J/K. We can see that if the number of possible microstates that a system can be in is larger, the entropy of the system is higher.

To get an intuition for the relation between Boltzmann's measure and Clausius' measure, we can think of the kinetic energy of the particles in object A and B. We saw that if the heat from object A flows to object B so that they get the same temperature, the overall entropy will increase. If the temperature in object A and B is more spread out, this means that the kinetic energy of the particles is more spread out. This, in turn, means that there are more possible microstates for the system consisting of A and B to be in. And if the number W is higher, we also get that Boltzmann's entropy will be higher.

Gibbs' entropy

Gibbs' measure of entropy additionally considers the probability distribution over the possible microstates of the system. It thus takes into account how likely it is that a system is in a certain microstate. It is defined as:

$$-k_B \sum_i p(x_i) \ln p(x_i) \tag{2.3}$$

We see that this formula is very similar to Shannon's measure of information. In fact, the constant k_B might as well not be there, which would make it into the exact same formula aside from the choice of logarithm. Boltzmann included the constant in order to give the same units to his entropy as Clausius' entropy. Physicist Ben-Naim [19] has argued that Clausius' entropy could have been dimensionless if it was realized at the time that temperature can be expressed in units of energy. If Clausius' entropy had been dimensionless, there would be no need to add constant k_B and Boltzmann's entropy would also be dimensionless.

The only difference with Shannon information would then be the logarithm, but the choice of the base of the logarithm is conventional, both for Boltzmann's entropy and Shannon information. We could just as well choose the binary logarithm for the Boltzmann entropy or the natural logarithm for Shannon information. It just happens that it is useful to work with a binary logarithm for Shannon information and a natural logarithm for Boltzmann's entropy.

Gibbs' entropy and Boltzmann's entropy can be considered two different frameworks for calculating *micro-entropy*, i.e. entropy that takes into account the microstates of a system [36]. When a system is in a thermodynamic equilibrium, the Gibbs and Boltzmann entropies give the same result, but in non-equilibrium states they do not necessarily [48]. A system is in thermodynamic equilibrium when there is no flow of energy or matter and there are no macroscopic changes to the system. It is currently debated which micro-entropy gives the right values, but in practice the Gibbs entropy is taken to be the standard measure. Philosophers of physics Charlotte Werndl and Roman Frigg [44], for example, write that "GSM [Gibbsian statistical mechanics] is widely used and considered by many to be *the* formalism of statistical mechanics" and they cite statistical mechanics books that only explain the Gibbs measure, showing that it is taken to the the standard one. For the rest of this chapter, when I simply use the term "entropy" I will therefore refer to the Gibbs entropy.

The relation between Gibbs' entropy and Shannon information

When Shannon proposed his measure of information, he decided to call it an "entropy". About this decision, Shannon once said in an interview:

> My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty.' When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage.' [76, p.180].

It is a natural idea to call Shannon's measure "entropy", as it has the same form as Gibbs' formula. We have to say, however, that by following Van Neumann's advice, the confusion around the term "entropy" has gotten only greater. The terms "information" and "entropy" now seem to be used interchangeably in physics and in information theory and it is unclear if they are indeed exactly the same concept and have the same properties.

I subscribe to how Lombardi characterizes the relation between Shannon information and entropy. She writes: "Depending on how the probabilities involved in its definition are endowed with reference, the Shannon information can be interpreted as Boltzmann entropy or as Gibbs entropy". [70, p.228] Shannon's measure of information is an abstract mathematical function that can be used for different purposes. In each context of use, the function gets a different interpretation. We may use Shannon's measure to calculate the entropy of a physical system. In this case, the values of the random variable refer to the different possible microstates of the system. It is, however, still an application of the concept of Shannon information. Therefore, the entropy of a system can also be considered to measure our uncertainty about which microstate the system is in.

The idea that entropy measures uncertainty already existed before Shannon proposed his measure of information in 1948. Boltzmann himself already interpreted entropy as a degree of ignorance according to Jaynes [56]. As another example, physicist G. N. Lewis wrote in 1930 that "Gain in entropy means loss of information and nothing more." [67, p.573]. This suggests that indeed the concept of entropy can be legitimately thought of as a type of information and that it is not merely a coincidence that they have the same measure.

Physicist Christoph Adami [1] also expresses the view that statistical thermodynamics can be considered an application of information theory. In both cases, we quantify the uncertainty about a measurement outcome. The difference between Shannon information and entropy comes down to the level of generality. Do we specifically look at the microstates of a physical system, or do we consider any set of states of an information source that is possibly abstract? We can consider an information source that emits symbols and calculate the Shannon information of this source by taking account of the probability distribution over which symbol the source emits. We can, however, also look at the microstate of the physical system which is the information source.

As an example, which is inspired by Adami, we can think of a die with six sides. We can calculate the Shannon information of this die based on how likely it is that the die will land on each side. We can also, however, calculate the entropy of this die by looking at the probability distribution over all the possible microstates of the die. It all depends on what we take to be the level of precision, what we consider to be the different states over which we have a probability distribution.

In his book *A Farewell to Entropy*, physicist Ben-Naim [19] agrees with the idea that Shannon information is the more general concept and argues that we can simply drop the use of the term "entropy", even in thermodynamics [19]. I think that while entropy can be considered an application of Shannon information, we can still call it "entropy" whenever we are considering the microstates of a physical system. On the other hand, I maintain it is best not to refer to all Shannon information as "entropy", because this obscures the fact that entropy is only one application of Shannon information. These terminological matters aside, the important take-away here is that entropy can be considered an application of Shannon information.

Entropy is a physical property

We can now go back to the question of whether Shannon information is a physical property. One line of thought could be that if entropy is a physical property, then information is a physical property too.

Entropy is indeed considered a physical property, because, as is clear from Clausius' measure, entropy can be derived from observable quantities like temperature and heat. Although there is no measuring instrument that can directly measure entropy on its own, like we can use a thermometer to measure temperature or a scale to measure mass, we can still say that it is an indirectly measurable property because we can derive it from other measurable physical properties. In any case, it is clear that the value of entropy is dependent on observable quantities and vice versa. So it is a physical property.

According to Gibbs' measure of entropy, we calculate entropy based on the number of possible microstates and their probability distribution. Since this probability distribution is based on our knowledge about the system, this measure might at first sight not seem to quantify a physical property. However, the Gibbs (and Boltzmann) entropy can roughly be considered to measure the same quantity as Clausius' thermodynamic entropy, because in most cases they depend on variables in the same way as Clausius' entropy. I say only in most cases, because when it comes to a change in volume, Clausius' thermodynamic entropy and micro-entropies may give different results.[1] In any case, the micro-entropies also depend on observable properties and vice versa. So we can also say of the micro-entropies that they define physical properties.

Hence, in case we calculate the Shannon information of an object by looking at the probability distribution over different possible microstates, we can say that Shannon information is a physical property. So in so far as Shannon information is used to calculate entropy, it is physical.

In the next section we will see why Shannon information in general can also be considered to be physical.

2.2.2 Landauer's principle

A reason to think that Shannon information is physical is *Landauer's principle*. It is named after physicist Rolf Landauer who explicitly argued that information is physical, amongst others in his articles titled "Information is a physical entity" [63] and "The physical nature of information" [64].

Landauer's principle roughly states that if information is erased, heat must be emitted. In a more precise and general formulation which is used by James Ladyman et al [62], the principle states that a logically irreversible transformation is always accompanied by a thermodynamically irreversible process. In a thermodynamically irreversible process, entropy increases, and in many cases this is in the form of heat emission to the environment. A recent proof for Landauer's principle is also provided by Ladyman et al. [62]

[1] Dieks argues that Clausius' entropy is extensive while Boltzmann's and Gibb's entropy are non-extensive. We can see this if we think of a box filled with gas with a partition in the middle. At first the gas is only on one side of the partition. If we remove the partition, the gas will have doubled in volume. According to the Clausius entropy the total entropy of the gas will be double the entropy of each of the two halves. According to Gibbs' entropy, however, the entropy will become more because each particle will have twice as much space to move through so there are many new trajectories. See Dieks' article for more details on the relation between Clausius' entropy and micro entropies [36]

Erasing information

Landauer's principle is quite puzzling, especially if we look at the first statement. How can it be that the erasure of information can cause a system to emit heat? To understand this, we must first look at what is meant with the "erasure of information". What is meant is an operation in which a system loses information that it previously contained. We can think of a computer that performs a RESET operation. This logical operation can take as input either 0 or 1 and always outputs 0. Hence, in order to reset a bit, the computer does not need to actually consider the value of the bit. It just changes the value of the bit to 0, no matter what the original value.

To see in what sense information is erased, we can zoom in on the physical system which represents the bit that the computer will perform RESET on. Since the computer does not check the value of the bit at the start at time t_1, we say that the computational state of the bit is either 0 or 1 with 0.5 probability each. If we compute the Shannon information or *computational entropy* of this bit, we get $H = 0.5 \times \log(0.5) + 0.5 \times \log(0.5) = 1$ bit. When the computer performs RESET, it sets the value of the bit to 0. So at time t_2, after the operation, there is only one possible value of the bit. In this case, $H = 1 \times \log(1) = 0$. There is no Shannon information contained in the bit anymore, and no uncertainty on our part about the value of the bit. In this sense, we erase the information in the computational device. We as external observers, however, actually gain information about the value of the bit.

Liouville's theorem

Now why is it the case that such erasure of information goes together with an increase of entropy? This can be explained in terms of the number of microstates of the computational device.

One type of proof that has been given for Landauer's principle is based on *Liouville's theorem*, which states that the overall number of microstates must remain constant in a system. See [43, p.17] or [75] for derivations of Landauer's principle from Liouville's theorem.

Liouville's theorem might seem in opposition to the Second Law, which states that the entropy of an isolated system is more likely to increase. It would seem that if according to Liouville's theorem the overall number of microstates in a system remains constant, the entropy could never increase. This is only an apparent opposition however, because the thermodynamic entropy is calculated from the perspective of the observers. As a system evolves over time, it is impossible for us observers to track the exact consequence of certain physical processes. Thus, we become more uncertain about which microstate a system is in. This means that even though the overall amount of microstates remains constant, for us there will be more possible microstates consistent with what we can measure about the system. Hence, entropy can spontaneously increase. If we would always know the exact dynamics of the isolated system, however, the entropy would remain the same because the overall number of possible microstates remains the same.

If we now think of the state of the bit at time t_1, we know that it is either in computational state 0 or state 1. Each computational state, which is represented by a macrostate, will be consistent with a number of particular microstates, let's say with 2 microstates for this example.

At t_2, after the RESET operation, we know that the bit is in state 0. From Liouville's theorem, we can derive that the number of microstates corresponding to the state 0 at t_2 must be equal to the sum of the number of microstates corresponding to the possible states at t_1, which is 4. This is because the overall number of microstates must stay the same. This means that the thermodynamic entropy of the physical system increases.

When we say that the thermodynamic entropy increases, we have in mind the entropy of the physical system when the computational state is given. Frank [43] also refers to this as the *non-computational entropy.* If the computational state at t_1 is given, then there are only 2 possible microstates consistent with it. This means that the entropy of the computational device actually goes from 2 microstates to 4 microstates. This means that the relevant entropy will increase. This increase in entropy can be in any form. The volume can get bigger, the density, or the temperature may go up.

In Figure 2.1. below, we can see a representation of what happens when a physical object's Shannon information decreases. We see that it starts in state C_I in which there are two possible computational states, c_{I1} and c_{I2}, that each correspond to two microstates. We see that the entropy that is based on the overall number of microstates, referred to as S_I, remains the same as we go the the next sate C_F, which follows from Liouville's theorem. The symbol k is the unit used for entropy here, an alternative for J/K. We also see that at C_I, the Shannon information H_I is 1 bit and at C_F, the Shannon information H_F is 0 bit. As a consequence of this, we see that the non-computational entropy represented by S_nc increases by 1 bit.

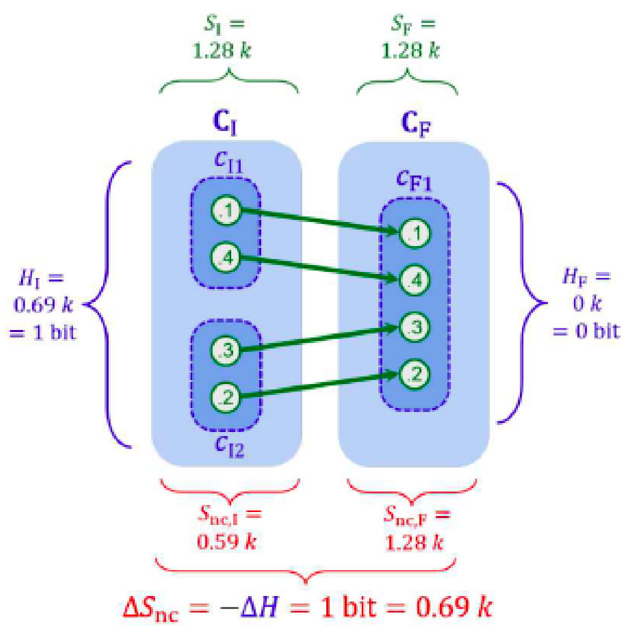


Figure 2.1: Entropy ejection from the computational state [43]

Work is required to erase information

It is puzzling that resetting a bit can cause macroscopic variables of a system, like temperature, to change. There is, however, a way to look at this increase of entropy a bit differently.

We can take a box with a gas inside consisting of only one molecule and a partition in the middle as in Figure 2.2, to represent one bit. We assume that the partition can be inserted and removed in the middle of the box without any friction. If this partition is inserted and the molecule is to the right side of the partition, we say that the bit has value 0. If the molecule is to the left of the partition, the bit has value 1.

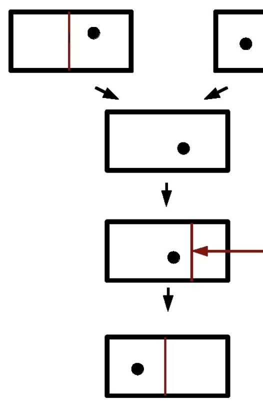


Figure 2.2: The erasure of one bit of information. [75]

So suppose the box is either in state 0 or 1, but we don't know which. We can then first remove the partition, frictionlessly, so that the molecule can move freely. After that, we push the molecule to the right side of the box with a piston and then insert the partition again. This ensures that the box will end in state 0. So the information contained in the box has been erased.

Pushing the molecule to the right side of the box with the piston will actually require work during which heat is emitted. It requires work, because the piston needs to push the molecule forward. The work causes energy to be emitted in the form of heat, which causes the increase in entropy. A different way to phrase Landauer's principle is therefore that the erasure of information requires work [75]. And it turns out that this work causes heat to be emitted.

It is important to realize that it is crucial in this example how much we know about the location of the molecule. If we were able to observe the position of the molecule, heat would not be emitted if we changed the state to 0. In case the molecule was already on the right side of the partition, we would not have to do anything. In case the molecule was on the left side of the partition, we could simply remove the partition and wait for the molecule to move to the right and then quickly insert back the partition. This means that if we could observe the position of the molecule, we could have set the box to state 0 without doing any work. In this case, no information would be erased, because the initial state is known by the agent who controls the device. The agent stores the information about the location of the particle in their memory which means that the information does not get lost. In this case we thus see that the entropy of the system does not increase.

Landauer on information

We now see why Landauer argued that information is a physical. It seems that a change in the Shannon information contained in a system goes together with a change in measurable properties like heat. And in this case, it does not seem to be a situation in which we would refer to the Shannon information as "entropy", because the calculation of the Shannon information is not based on a probability distribution over microstates of a physical system. It is, however, based on a distribution over two computational states of the physical system.

2.2.3 Is Shannon information physical?

We can now judge whether it is indeed correct to say that Shannon information is physical. It seems that we can say that Shannon information is a physical property, but we must slightly nuance this claim.

Shannon information is neutral

Lombardi et al. [71] also argue that information can be considered a physical property. They also point out, however, that Shannon information is essentially a formal concept. It is a mathematical theory that can be given any interpretation, depending on what is taken to be the information source and what are taken to be possible values. Lombardi et al. [71] therefore adopt a pluralist view of information and argue that Shannon information can in some contexts be considered physical, while it does not have to be considered physical in others.

I agree that Shannon information is indeed technically only a formal concept. It can be applied to any domain, and not just to physical objects. As a tool, it can also be applied in linguistics to quantify various properties of natural languages. In this case, we are not necessarily applying it to a physical system or information source. Thus, it would be more correct to say that Shannon information is physical only if it is applied to a physical object or physical information source. This means that in the context of physics and engineering, Shannon information is physical.

Epistemic relevance

Javier Anta [6] agrees with Lombardi et al.'s formalist view on technical information concepts according to which these concepts are neutral with respect to their content or interpretation. He disagrees, however, with Lombardi et al. concerning the acceptance of the idea that information can be physical. Anta argues that whether a concept can be considered physically significant should depend on whether the concept is epistemically relevant to some physical theory. He then argues that the concept of information is not epistemically relevant for any physical theory, although it is pragmatically relevant.

A concept is *pragmatically relevant* if it is only useful for obtaining knowledge. If "its content contributes cognitively to knowledge elucidation" [6, p.9], a concept is *epistemically relevant.* Anta illustrates this as follows:

> For example, while the concept of 'meter' will be pragmatically relevant for understanding how an apple falls to the ground (one could always use another metric system based on 'feet'), the concepts of 'gravity' and 'acceleration' will be somehow epistemically relevant for this very task, since they make explicit certain key factors that give rise to such a phenomenon. [6, p.9]

To see if information is physically significant, Anta thus considers whether the concept of Shannon information is epistemically relevant for thermal physics. He concludes that it is not. He argues that Shannon information cannot be considered equivalent to Gibbs entropy, nor can it be considered equivalent to Clausius entropy. He argues that the fact that Gibbs entropy has a different logarithm and a different unit because of the Boltzmann constant k, is an essential difference. He also argues that since Gibbs and Clausius' entropy are slightly different measures, it is even less correct to say that Shannon information is equivalent to Clausius' entropy. Apparently he judges this equivalence to be essential for the epistemic relevance of Shannon information.

As I argued in a previous section of this chapter, even though Gibbs entropy and Clausius entropy do not depend on all variables in the same way, the Gibbs entropy itself does have a crucial and epistemically relevant role in thermal physics because it still depends on other physical variables and vice versa. Thus, it is only necessary to see in what sense Shannon information and Gibbs entropy correspond. It can hardly be argued that the constant and logarithm are so important, because as I argued previously, these are merely conventional choices. As Ben-Naim has argued [19], we may as well work with Shannon's measure instead of Gibbs's measure, i.e. omit the constant and use the binary logarithm.

Thus, the fact that Gibbs entropy is epistemically relevant amounts to the fact that Shannon information is epistemically relevant for physics. And we can be sure that Gibbs entropy is epistemically relevant because the whole field of thermodynamics is dependent on this concept and without it we could not understand it. Anta also makes no remark about Landauer's principle, which would be an important aspect to consider when determining the physical significance of Shannon information.

Shannon information is physical

All in all, we can conclude that Shannon information can be considered a physical property. This does not mean that it is physical in general, but that when Shannon information is applied to a physical source or object, it can be considered a physical property of that object.

2.3 Information as an objective property

Besides the fact that Shannon information is considered a physical property, it is often also taken to be objective. It is thereby contrasted with semantic information which is considered to be subjective. There is, however, one context in which the supposedly subjective aspects of Shannon information are highlighted. Some physicists think that the fact that Shannon information is subjective could imply that entropy is also subjective. Atkins [8], for example, has argued for this reason that entropy and Shannon information should not be compared with each other:

> I have deliberately omitted reference to the relation between information theory and entropy. There is the danger, it seems to me, of giving the impression that entropy requires the existence of some cognizant entity capable of possessing 'information' or of being to some degree 'ignorant.' It is then only a small step to the presumption that entropy is all in the mind, and hence is an aspect of the observer. I have no time for this kind of muddleheadedness and intend to keep such metaphysical accretions at bay. For this reason I omit any discussion of the analogies between information theory and thermodynamics. [8]

Other physicists like Jaynes [57, 56] and Lewis [67] argue that entropy is subjective in its own right. However, since we know that entropy is a type of Shannon information, it can still be that some of the arguments concerning the subjectivity of entropy might apply to Shannon information. In this section we will look at some of these arguments and judge whether Shannon information can be considered objective or not.

Semantic information

The main reason that Shannon information is often considered to be objective when contrasted with semantic information, is that it is has less to do with the knowledge of specific agents. Semantic information is taken to be subjective because it is thought to be more dependent on how individual agents understand the information and on their own background knowledge. The idea is that this is not the case for Shannon information. The fact that Shannon information is a mathematically defined quantity and does not have anything to do with the meaning of symbols also causes people to associate it with objectivity. The question is whether this is justified.

Information and entropy as subjective in physics

We can first look at possible arguments as to why Shannon information would be subjective. Although not aimed directly at Shannon information, some physicist like Jaynes [57], Lewis[67] and Callen [26] have argued that entropy is subjective. Since entropy can be considered a form of Shannon information, we can see if the reasons for taking entropy as subjective would apply to Shannon information too.

One physicist who argued that entropy is a subjective concept is G.N. Lewis [67]. He also equated entropy with loss of information, as we saw earlier. The sense in which Lewis considers entropy to be subjective seems to be that entropy is dependent on the available knowledge that observers have about the system. Gibbs and Boltzmann entropy, the micro-entropies, are calculated from the number of possible microstates of a system that are consistent with what we can observe about that system. These entropies are thus dependent on what we know about the system. If we know less about the system, the number of possible microstates is bigger and thus the entropy is higher. Lewis has described entropy as follows:

> Gain in entropy means loss of information and nothing more. It is a subjective concept, but we can express it in its least subjective form, as follows. If, on a page, we read the the description of a physicochemical system, together with certain data which help to specify the system, the entropy of the system is determined by these specifications. If any of the essential data are erased, the entropy becomes greater: if any essential data are added, the entropy becomes less. [67, p. 573]

Thus Lewis argues that the entropy of a physical system is subjective, because it is dependent on the data we have about the system. When Lewis argues that entropy means a loss of information, he however cannot be referring to Shannon information. Because gain in entropy means a gain in Shannon information. So he must be referring to information in the more everyday sense. The information that *we* have about the system.

Shannon information as subjective

We can, however, ask ourselves whether the same argument applies to Shannon information. Is Shannon information dependent on the knowledge we have about the object or source in question? There is a sense in which it is, because Shannon information is dependent on the probability distribution that we associate with the object under consideration. This probability distribution, in turn, can be considered to depend on what we know about the source or object.

In any case, just like with entropy, we cannot say that an object has an absolute Shannon information. We cannot really talk about *the* information of an object or information source, because the Shannon information of an object or source is also dependent on the probability distribution. We could be in a situation in which we ask "What is the Shannon information of object x?" where it is not even specified yet what signals we are interested in. Then there would not be one absolute answer. Lombardi agrees [71] that Shannon information is relative in this sense and gives an example:

> For instance, a roulette wheel can be described as a source with 37 states when we are interested in a single number, or as a source with 3 states when we are interested in color: although the physical system is the same, in informational terms the two sources are completely different. [71, p.1998]

So Shannon information is not an absolute property of an object or source. But the question is whether this makes Shannon information a subjective quantity.

Objectivity

To answer this question it is important to be clear on the interpretation of the term "objectivity". In the introduction I have argued that I think it is best to use a weaker interpretation than it is sometimes given. Some of the physicists who argued that entropy is subjective, for example, can have intended it to mean that entropy is not an absolute and mind-independent property of a physical system, which is the stronger interpretation of objectivity. Jaynes first argued that entropy is subjective [57] but in a later paper [56] referred to entropy as "anthropomorphic" rather than as subjective:

> From this we see that entropy is an anthropomorphic concept, not only in the well-known statistical sense that it measures the extent of human ignorance as to the microstate. Even at the purely phenomenological level, entropy is an anthropomorphic concept. For it is a property, not of the physical system, but of the particular experiments you or I choose to perform on it. [56, p. 398]

Jaynes argues that entropy can be considered an anthropomorphic property because it is dependent on what we can do with a system, on the relation between the system and the observer. The term "anthropomorphic" can thus be understood to be similar in meaning to the term "mind-dependent". Both point out a dependency relation with a cognitive agent. But if some quantity has a dependency relation with a cognitive agent, this does not necessarily amount to that quantity not being objective.

Denbigh [33] also pointed out that there are two ways in which objectivity can be understood.

> (i) A broad, or weak, meaning which refers to statements which can be publicly agreed (philosophers sometimes refer to this as intersubjectivity) and (ii) a more restricted, or stronger, meaning, much used in science, which refers to statements about things which can be said to exist, or about events which can be said to occur, quite independently of man's thoughts and perceptions, or of his presence in the world. [33, p.109]

I have thus argued that we should understand objectivity in the first sense, so that a physical quanitity will be objective if it can be publicly agreed upon what the value of this quantity is. The second sense I take to characterize mind-independence. Denbigh also argues in his article that entropy can be objective in the first sense, while it may be mind-dependent or anthropomorphic in some ways.

Shannon information is objective

Now that we have this understanding of objectivity in mind, we can answer the question whether Shannon information is objective. I argue that it is, because it is a quantity of which the value can be publicly agreed upon.

In most cases, it is clearly determined what outcomes of a source we are interested in before we determine the Shannon information of a source. And in case that this is not so, different agents can still discuss what outcomes they are interested in and then agree on a value. The value of the Shannon information of the source will also be dependent on the probability distribution over the different outcomes. While the ontological status of probability is also a subject of much debate, it can be argued that we only calculate the Shannon information of sources for which the probability distribution is one that can be publicly agreed upon. This would make probabilities also objective in the weak sense that I use it. For example, it could be objectively determined what the probability distribution over different outputs of a source is by observing the frequency with which these outputs are emitted.

An illustration of why the individual knowledge of agents does not matter for Shannon information, can be given as follows. We can think of a situation in which we want to calculate the amount of Shannon information of a source that emits sentences of English. It can be the case that one person observes this source and sees that it has emitted the symbols 'H', 'e', 'l', 'l' in a sequence. Since it is likely that this source will next emit an 'o' to form the common word of greeting, this person will have more knowledge about what symbol the source is going to emit next. However, this type of knowledge is not considered relevant in general when we want to calculate the Shannon information of a source. We normally assume that the probability distribution is set and that this cannot differ for individual agents. This is of course just one example, but it illustrates that when calculating Shannon information we implicitly assume that the probability distribution is publicly agreed upon.

Shannon information is mind-dependent

We can say a little bit more about the sense in which Shannon information can be considered to be mind-dependent. Even though the rough interpretation of mind-dependence is quite clear, namely that of being in some way dependent on the mind, the precise meaning is harder to specify. Mind-dependence can be both about the ontological and epistemic status of an object or a property, as described by Ekaterina Botchkina [25]. Ontological mind-dependence means that an entity cannot exist without some relation to a cognitive agent, whereas epistemic mind-dependence means that an entity cannot be described or determined without reference to a cognitive agent.

Entropy could also be considered mind-dependent to some extent. Huw Price [84] writes about entropy:

> ...that entropy might be argued to be subjective in the sense that secondary qualities are subjective. This is not in the sense that they exist only when and where human or other observers exist, rather in the sense that their nature and distinctive character is in some way tied to the responses or capacities of suitably equipped such observers. [84, p.115]

Where Price uses the term "subjective", I would thus use the term "mind-dependent". This interpretation of entropy also corresponds to Jaynes' idea that entropy is anthropocentric. It is also argued for by Ladyman [60]. He argues that the entropy of a system is dependent on our knowledge of the system which determines what we can do with the system. Myrvold [79] has similarly referred to entropy as "means-relative".

I argue that Shannon information can be thought of in a similar way, since it is dependent on a probability distribution. This probability distribution may be determined beforehand, but it is still dependent on some relation between a cognitive agent and the system in question. At least, this is the position that I take on probability and that philosophers like Lewis [65], Mellor [77] and Price [84] also take. Price writes, for example: "chance cannot be understood in isolation from cognitive creatures who must deliberate in contexts of incomplete information." [84, p.116] There are accounts of objective probability out there, namely frequentism and propensity theory, but both accounts have problems which means that it remains hard to think of probability as something that exists independently of cognitive agents.

According to *frequentism* [85, 78], the probability of an outcome is determined by the relative frequency of this outcome in a sequence of trials. It seems though, that this also makes probability dependent on cognitive agents because it is us who observe the trials and then calculate the relative frequency. If it is argued that probability is not the relative frequency in an actual observed sequence of trials, but in a hypothetical and infinite sequence of trials, then probability again becomes something abstract. It can be maintained that it is abstract yet mind-independent, but this is hard to justify. The other option is to go for a *propensity theory* [83]. This account postulates probabilities as mind-independent properties of events.

According to this account, objects or circumstances can have an inherent propensity to cause events with specific probabilities. A die, for example, would have a propensity to land on each side with a probability $\frac{1}{6}$, independently of any observer. This account however has quite some problems, as has been argued for by Gillies. [47] One of the issues is that it is hard to grasp what it is exactly that possesses a propensity. In the case of a die, there are many other circumstances that determine the outcome, like the surface of a table and the wind. So it seems the propensity should be a property of the entire universe. Another problem is that if the propensity is a property of the entire universe, it is likely that it would be much more determined what the outcome of the die would be, which would make the probability different than $\frac{1}{6}$.

An extensive discussion of the discussion about probability lies beyond the scope of this thesis. I have pointed out, however, that I see problems with defining probability as mind-independent both within the frequentist account and the propensity theory of probability. I thus argue that probability can be objective, but that it can only be defined as dependent on cognitive agents and can only exist because of cognitive agents. This thus implies that Shannon information would both be epistemically and ontologically mind-dependent.

2.4 Conclusion

We are now in a position to consider the full picture of Shannon information. In the previous chapter we saw that Shannon information is a quantity that can be attributed to information sources. In this chapter it became clear how it can be attributed to objects in general. The probability distribution can range over any set of states, not only over possible signals that can be emitted by an information source. We now also understand how entropy and Shannon information are related. Entropy can be taken to be an application of Shannon information to thermodynamics. In this case the probability distribution is always taken to range over the possible microstates of a physical system.

We have seen why Shannon information cannot be taken to quantify a concrete substance. Information is always a type and never a token, it is always carried by something concrete but cannot be concrete itself. When information "flows" from one place to another, this is not an actual flow of material stuff, i.e. information is not like syrup. What flows is the abstract type.

I then argued why Shannon information can be considered to be a physical property, even if we use Anta's criterion of epistemic relevance. I have explained in line with Lombardi why this does have to be restricted to applications of Shannon information to physical objects or information sources. We can say that Shannon information is a physical property in this case, because its value is relevant for physics. The Shannon information of a physical object is related to observable properties like heat and volume. This follows from Landauer's principle which states that the erasure of information always goes together with an increase of entropy, which is often in the form of heat emission.

Finally, I have argued that Shannon information is an objective property. This is so because it can be publicly agreed upon what the value of the Shannon information of an object is. I have also explained that Shannon information is not an absolute property, but a relative property. This is because it can vary what we choose to be the different outcomes of the probability distribution. I have also argued that Shannon information can be considered both epistemically and ontologically mind-dependent if we assume that probability is mind-dependent in the same way.

So we now have a clearer picture of the ontological and epistemic status of Shannon information. We also understand the role of Shannon information in physics. We are ready to go on to the next chapter in which we will look at the other main concept of information: semantic information.

Chapter 3

Semantic information

After having looked at the main technical concept of information, Shannon information, we will now consider semantic information. As I explained in the Introduction, semantic information is the type of information that is closer to the more everyday interpretation of the term. An English sentence like "The milk is in the fridge" contains the semantic information that the milk is in the fridge. We see that the semantic information of the sentence depends on the meaning we associate with the words in the sentence. We can also note that, as Lombardi puts it, semantic information has a certain "aboutness", or intentionality [71, p.1984]. The semantic information of the English sentence given above, for example, is information *about* the milk and *about* the fridge. Aboutness is not limited to sentences in natural language. If we associate a certain meaning with the symbols @, # and *, any string of these symbols can contain semantic information. Any type of signal can be attributed a meaning in virtue of which it can convey semantic information.

Semantic information is generally considered to be subjective and mind-dependent. It is also associated with needing a qualitative description, in contrast with Shannon information which allows a quantitative measure. In this chapter and the following chapter, we will look at whether these ideas about semantic information are correct. We will also consider two other themes that concern the nature of (semantic) information. One is that of the truthfulness of semantic information and the other is about what the main function is of information: world description or communication. We will see that accounts of semantic information can disagree when it comes to these properties of information. This will in turn also make the relation between semantic information and Shannon information more precise.

When we look at accounts that analyze semantic information, we immediately see that there are both qualitative and quantitative accounts. This already shows that it is possible to measure something about semantic information in quantitative form. Most of the measures designed to measure semantic information are in some way based on probabilities, just like Shannon information. We will have to see, however, to what extent such measures satisfactorily describe semantic information.

Three quantitative accounts of information

In this chapter, I will first focus on accounts of information that propose a quantitative measure and later go into the qualitative or logic-based accounts. I will consider three different accounts that all propose their own measure of information. We will first look at Bar-Hillel and Carnap's account which stems from 1953 [15], a few years after Shannon's MTC was brought out. We will then look at Floridi's account of information which is the most recent account of semantic information [42]. Finally we will look at Dretske's account [38] which is from the 80's and which has a slightly different focus than the other two.

Linguistic information and environmental information

We will see that Bar-Hillel and Carnap's account [15] and Floridi's account [42] focus on *linguistic information*, i.e. information contained in sentences from natural language. Dretske [38], however, focuses on the concept of *environmental information*, although he did not use this term yet. Other philosophers like Floridi use this term to characterize Drestke's concept of information [42, p.30]. Environmental information is the type of information that is contained in natural signs. An example that is often used to illustrate what environmental information is, is that involving smoke and fire: if we see smoke, this can give us the information that there is a fire somewhere. The smoke can thus be understood as a signal that carries information and it does so in virtue of the fact that we know that there is a certain correlation between smoke and fire.

While environmental information may at first sight not be considered to be semantic information, it can in fact be considered semantic information as we will see in this chapter. Nevertheless, it is a very different perspective on semantic information than the other two accounts have. We will also see that Dretske explicitly construes information as something that is objective and mind-independent. Floridi's and Bar-Hillel and Carnap's account lean more towards the idea that information is subjective and mind-dependent.

Truthfulness

When it comes to the truthfulness of semantic information, both Dretske and Floridi argue that semantic information should be truthful. They hold that misinformation is actually pseudo-information, i.e. not really information. Bar-Hillel and Carnap do not yet take this stance and formulate a measure of information that can apply both to true and false statements.[1]

[1]It seems that all accounts treated here assume something similar to a correspondence theory of truth (the idea that a belief or a proposition is true if it corresponds to a fact in the world). Dretske and Bar-Hillel and Carnap do not explicitly mention it and Floridi proposes an information-based Correctness Theory of Truth [42] which can be understood to be more similar to a correspondence theory than to a coherence theory of truth. It could be that when a different theory of truth is adopted, the idea that information should be truthful can be evaluated differently.

World description or communication

When it comes to quantitatively measuring how much information is contained in a signal, there are different aspects of information that we can focus on. We can either focus on how surprised we are to learn a piece of information, or on how much a piece of information says about the world. These aspects of information reflect two main possible functions of information: *communication* and *world description*. These two perspectives on information were pointed out by van Benthem in the opening lecture of the Congress of Logic, Methodology and Philosophy of Science in Helsinki in 2015. We will see that Bar-Hillel and Carnap and Dretske focus more on communication and that Floridi focuses on world description.

3.1 Bar-Hillel and Carnap: semantic information

Bar-Hillel and Carnap were the first to come up with a formal theory of semantic information [15, 28]. They noticed that Shannon's MTC [87], which, as we saw, puts forth a formal procedure to quantitatively measure information, was not suited to measure the semantic information of a sentence of natural language. In line with Carnap's bigger project to apply methods involving Inductive Logic in the empirical sciences [27], they felt that it would be useful to have a quantitative measure of semantic information. One of their goals was, for example, to be able to use this measure to see how they could maximize the information obtained from an experiment [7].

Bar-Hillel and Carnap therefore propose a quantitative measure to capture the semantic information contained in a sentence. The key idea of their approach is that the amount of semantic information in a sentence depends on the likelihood that the sentence is true. If it is very likely that a sentence is true, the amount of information contained in that sentence is low. If it is not likely that the sentence is true, the amount of information contained in the sentence is high. They restrict their attention to sentences of a formal language with predicates and names for individuals and with the standard Boolean connectives [28].

The content of a sentence

Bar-Hillel and Carnap first consider the concept of the *content* of a sentence as a means to qualitatively describe the information conveyed by a sentence [15, p. 149]. They define the content of a sentence as the set of *content-elements* that follow logically from that sentence. A content-element is a disjunction of *atomic sentences*. An atomic sentence is of the form "Pa" where the property that "P" stands for is attributed to the individual with the name "a". For all atomic sentences in the language, a content-element contains either this sentence or its negation. It is essentially a negation of a state-description. A *state-description* is a conjunction of atomic sentences that gives a complete description of all objects in the domain and that individuates a possible world.

Bar-Hillel and Carnap hold that this concept of the content of a sentence is a good description of the information of a sentence. A tautological sentence will not have any information, because it will have no content-element as a consequence. A contradictory sentence will in some sense contain all information because it will imply all content-elements. It is also the case that if a sentence s implies another s', s will contain all information that s' contains, because the content of s' will be contained in the content of s.

Contradictions containing maximal information

It may not seem obvious that a contradictory sentence should contain maximal information. Bar-Hillel and Carnap [28] argue, however, that if a sentence is contradictory, this can be because it contains too much information.

> It might perhaps, at first, seem strange that a self-contradictory sentence, hence one which no ideal receiver would accept, is regarded as carrying with it the most inclusive information. It should, however, be emphasized that semantic information is here not meant as implying truth. A false sentence which happens to say much is thereby highly informative in our sense. Whether the information it carries is true or false, scientifically valuable or not, and so forth, does not concern us. A self-contradictory sentence asserts too much; it is too informative to be true. [28, p.7-8]

Two quantitative measures of semantic information

Next to this qualitative definition of information, Bar-Hillel and Carnap also propose a quantitative measure of information. They first consider the following measure that is also called '*cont*', which is somewhat confusing as it really is a different concept from the qualitative notion of the content of a sentence.

$$cont(s) = 1 - p(s) \tag{3.1}$$

Here $p(s)$ is the probability that sentence s is true.

They also consider a different measure $inf(s)$ that has slightly different properties[2]. It is defined as follows:

$$inf(s) = \log_2 \frac{1}{1 - cont(s)} \tag{3.2}$$

This reduces to:

$$inf(s) = \log_2 \frac{1}{p(s)} = -\log_2 p(s) \tag{3.3}$$

[2]The *cont* function is additive only when two sentences are *content independent*. The inf function, on the other hand, is additive when two sentences are inductively dependent. Two sentences s and s' are content independent when they do not have any overlapping factual consequences, in the sense that there will be content-elements that are both a part of s's and s''s content. Two sentences are *inductively independent* if the the truth of one sentence does not make it more likely that the other sentence is true and vice versa. Bar-Hillel and Carnap point out that we might want that the information content of two sentences should be additive if two sentences are inductively independent, not content independent.

We can note that the range of the inf function is between 0 and ∞, rather than between 0 and 1 like the $cont$ function.

Logical probability

The probability $p(s)$ that is referred to in the measures, is a type of logical probability. Carnap, and later also Carnap and Bar-Hillel together [27, 28], wrote about what such a probability distribution over sentences should look like. There are certain constraints on the distribution like that tautologies should have probability one and that the sum of the probabilities of all the state-descriptions should sum to one. Because of the constraints on what such a logical probability function should look like, the probabilities of all sentences will be determined once the probabilities of the state-descriptions are determined.

It can be decided freely what we take to be the probabilities of the state-descriptions. We could choose to give them all the same probability if we have no contextual knowledge at all. However, we could also decide to assign them probabilities depending on the background knowledge that some agent or a group of agents have. In any case, if we choose a distribution that satisfies the constraints chosen by Carnap, we will get that a sentence that is less complete than a state-description will have a higher probability. This will mean that $cont(s)$ will be lower. This, in turn, corresponds to the idea that sentences that claim less about the world should have less information. We also get the similar result as with the content of a sentence that a contradiction will have probability zero and thus maximal information and that if one sentence s implies another s' it will be the case that the probability of s' will be higher than s and therefore the amount of information contained in s will be higher than in s'.
Here $p(s)$ is the logical probability that sentence s is true.

Communication and world description

It seems that Bar-Hillel and Carnap's measures reflect the world description perspective to some extent, but mostly the communication perspective.

Both of the measures that they propose are such that the more unlikely it is that a sentence is true, the higher the information contained in this sentence. They even go as far as to say that false sentences can convey a lot of information if they are very surprising, with contradictions containing maximal information. This fits the idea that the main function of information is communication: the more surprised we are to learn some proposition, the more information it contains. We also see that on Bar-Hillel and Carnap's account, the semantic information contained in a sentence can be different for different agents, depending on the probability they associate with that sentence being true. This, in turn, depends on their background knowledge.

On the other hand, it could also be argued that the measures do incorporate the idea that the main function of information is world description. That is because of the way they define the probability function over the sentences in the language. A complete state-description will get lower probability than a less complete description of the world. This means that sentences that are more complete in describing the world will get a higher information value.

However, it can be argued that the communication perspective is reflected more strongly in their measures. If we would have two sentences that are equally complete in describing the world, but one of them is more likely, this one would immediately get a lower information value. In addition, it can be that a sentence that is more complete than another can get a lower information value because it is deemed to be more likely to be true. It can finally also be argued that if the main function of semantic information would be to give the most accurate description of the world, it would be unacceptable that false sentences can contain a lot of semantic information. That is because they do not describe anything about the actual state of the world at all.

3.1.1 Evaluation of Bar-Hillel and Carnap's measures

Whether Bar-Hillel and Carnap's measures can be accepted as good measures of semantic information seems to depend on what we consider to be the most important function of information. We will thus evaluate the measures from both perspectives.

From the perspective of world description

If the main function of information is taken to be world description, then Bar-Hillel and Carnap's account is arguably not very satisfactory. D'Alfonso [32] and Anta [7] have both argued that Bar-Hillel and Carnap's account does not do justice to our intuitions about information. D'Alfonso writes about contradictions getting maximal information content: "not only is $A \wedge \neg A$ false, but it does not at all do a good job of describing how things presumably are or could be. It does not discriminate and selectively narrow down on a potential state of affairs (unless a contradiction does actually occur!)." [32, p.66] On this basis, D'Alfonso turns down Bar-Hillel and Carnap's account. D'Alfonso thus seems to assume that one of the main functions of information is world description.

Even if we would require that *cont* and inf can only be applied to truthful statements, we would arguably still get information values that do not perfectly agree with the idea that the main function is to describe the world. It could be, for example, that some agent considers it to be very unlikely that it is true that the milk is in the fridge. They consider it more likely that it is true that the milk and the eggs are not in the fridge, that the salad and tomatoes *are* in the fridge and that there is a pizza in the oven. Even though the latter statement describes more about the world, it would, according to Bar-Hillel and Carnap's measures, contain less information for that agent.

From the perspective of communication

What about if we think of communication being the main function of information, would we then be satisfied with Bar-Hillel and Carnap's account? I would argue that even in this case, it does not make sense that contradictions would get maximal information. One agent may, in some sense, get a lot of information out of some false statement made by another agent. However if an agent would plainly state a contradiction, it does not seem to be that this agent could get any information out of that statement. If one agrees with the idea that the bigger the surprise, the higher the information, regardless of whether that information is truthful, then Bar-Hillel and Carnap's measures could be satisfactory. Alternatively we could require that *cont* and inf can only be applied to truthful statements and we would get an acceptable measure.

3.2 Floridi: truthful semantic information

Luciano Floridi [42] has also proposed an account of semantic information which he has presented in full detail in his book *The philosophy of information* in 2011. He argues that the fact that contradictions come out as having maximal information on Bar-Hillel and Carnap's account does not agree with our intuitions about what information is.

Information should be truthful

Floridi argues that it should be a necessary requirement for information that it is truthful. He argues that even though the concept of false information may exist, what is meant in this case is a proposition which does not qualify as information. Thus, Floridi argues that the term "false" in the phrase "false information" is used attributively and not predicatively. It tells us that the piece of information is actually pseudo-information, not that there is a piece of information that is also false.

Floridi suggests that when we are tempted to think that there can be cases of false but genuine information, this is because we think about the hypothetical scenario in which it could have qualified as genuine information. [42, p.95]

We can, for example, think of a situation in which Alice would tell Bob that it is 5:30 because she has just checked her watch which said 5:30. In reality it is 6:30. However, the Daylight Saving Time (DST) just started the night before and Alice forgot to set her clock one hour forward. Therefore, the information that it is 5:30 is false. But can we not still say that it is genuine information? Floridi would argue that we should not. We could be tempted to think it is information, because we can think of the hypothetical scenario in which it is actually 5:30. In fact, it is very easy to imagine this scenario in this example because if DST had not been started, it would still be 5:30 when Alice told Bob the time.

Floridi, however, argues that even though a statement could be true in a different scenario, it cannot be considered information if it is not actually true. It must be considered pseudo-information. Floridi adds that he does not deny that people can often be misinformed in the sense that Bob is misinformed by Alice. However, he argues that if Alice misinformed Bob, this strictly means that Alice did not inform Bob, i.e. did not communicate to him a piece of genuine semantic information.

A theory of strongly semantic information

Floridi thus proposes an account of truthful semantic information in which contradictions have no degree of informativeness, nor do tautologies. Floridi makes a distinction between propositions that are informative and propositions that qualify as genuine semantic information. Contingently false statements can be informative, but cannot be considered actual semantic information. He calls his account a Theory of Strongly Semantic Information (TSSI) and contrasts this with Bar-Hillel and Carnap's account which he takes to be a Theory of Weakly Semantic Information (TWSI).

The degree of discrepancy of a sentence

The idea of Floridi's account is to look, rather than at the likelihood that a sentence is true, at how precisely a sentence describes the actual state of the world. This is what Floridi calls the "discrepancy" of a sentence [42, p.120]. Two sentences can both be true, but have a higher or lower degree of discrepancy with respect to the actual state of the world. Floridi gives an example of the sentences "there is someone in the library" and "there are 9 or 10 people in the library". In reality, there are 9 people in the library. In this case, Floridi argues that the latter sentence is closer to the actual state of the world and therefore has a lower degree of discrepancy with respect to the actual world than the former.

Floridi defines a function f that assigns a value of discrepancy (ϑ) to all possible sentences. This value can be between -1 and 1, so either positive or negative. A contradiction will have a discrepancy value of -1 and a tautology a discrepancy value of 1. A tautology has a maximal discrepancy value because it is supported by every possible world and therefore is most distant from a characterization of the actual state of the world. The sentence that gives the most complete description of the world will have discrepancy value 0. Floridi concedes that this function f that assigns discrepancy values will be difficult to define in real-life and that we will not always have complete knowledge of the actual state of the world. This means that the discrepancy values of a sentence will often be based on the evidence we have that this sentence confirms to the state of the world.

The informativeness of a sentence

The informativeness ι of a sentence σ is then calculated based on the discrepancy as follows:

$$\iota(\sigma) = 1 - \vartheta(\sigma)^2 \tag{3.4}$$

In Figure 3.1 below we can see how the informativeness and the discrepancy of a sentence are related to each other. The x-axis represents the informativeness ι and the y-axis represents the degree of discrepancy ϑ. The sentences with a high discrepancy have low informativeness and sentences with a low discrepancy have a high informativeness. We also see that both false and true sentences can have a positive value of informativeness. The more the discrepancy goes below 0, the more inaccurate a sentence is and the more the discrepancy goes above 0, the more vacuous a sentence is.

Since Floridi requires that information should be truthful, the informativeness of a sentence is not sufficient to determine the semantic information of a sentence. A sentence also needs to be truthful to be considered semantic information. The semantic information of a sentence ι^* is

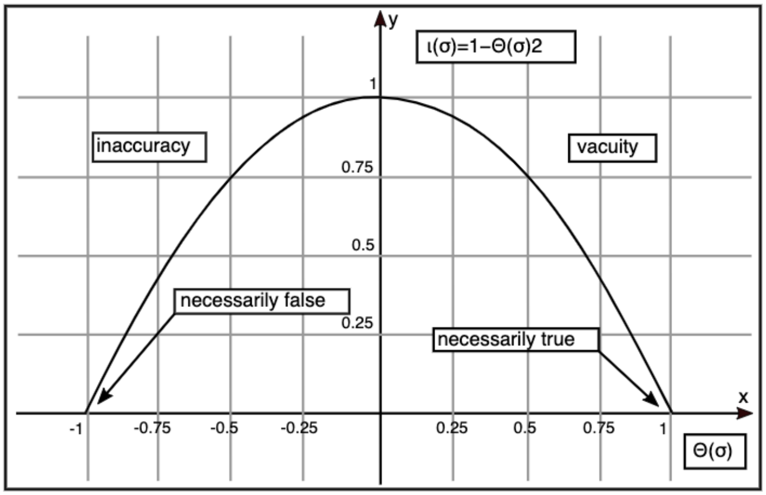


Figure 3.1: Floridi's account of information [41]

calculated as follows where α denotes the maximal amount of information that a sentence can have and β stands for the amount of vacuous information in a sentence:

$$\iota^*(\sigma) = \alpha - \beta \tag{3.5}$$

The maximal possible information that a sentence can have α is set to be equal to the surface of the area under the curve from $\vartheta = 0$ until $\vartheta = 1$. So we get that $\alpha = \int_1^0 1d(x) = \frac{2}{3}$. To calculate the amount of vacuous information in a sentence σ, we take the surface of the area under the curve from $\vartheta = 0$ until $\vartheta = \vartheta(\sigma)$. This is therefore dependent on the discrepancy of σ. The higher the degree of discrepancy in a sentence, the more vacuous the information and so the lower the amount of semantic information.

In Figure 3.2 we see that when we have a sentence σ with discrepancy $\vartheta = 0.25$, β is equal to the surface of the black area. The whole surface on the right side under the curve is α. So that means that surface of the blank area, $\alpha - \beta$, is equal to the semantic information $\iota^*(\sigma)$ contained in σ.

Information is well-formed meaningful data

Next to his measure $\iota^*(\sigma)$, Floridi also has a theory about what can be considered semantic information. He argues for a data-based approach for information according to which semantic information is well-formed and meaningful data, which as we saw also needs to be truthful.

But what are data? Floridi comes to the conclusion that a *datum* is a lack of uniformity. He calls this the *diaphoric interpretation* of data, because *diaphora* means "difference" in Greek. If there is a difference between one and another thing, there is a datum.

And what does it mean that the data must be well-formed? Floridi argues that data must respect "the rules of the chosen system, code or language being analyzed" [42, p.84]. He writes:

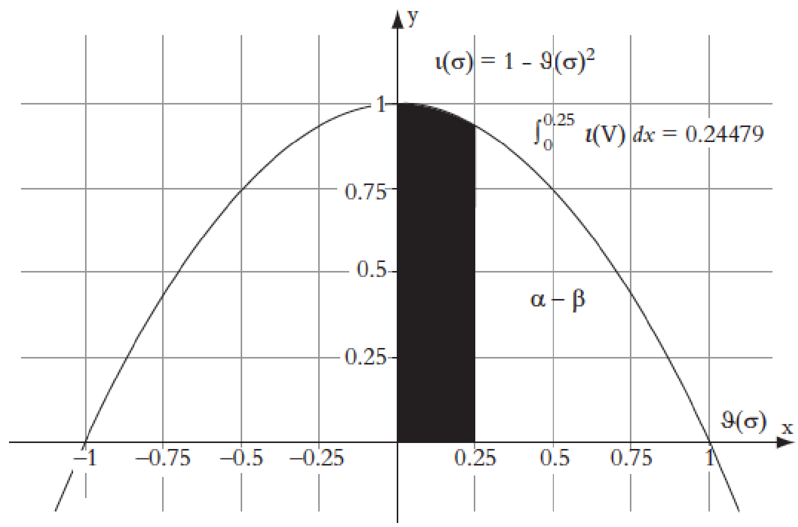


Figure 3.2: The amount of semantic information in a sentence with $\vartheta = 0.25$ [42]

"Syntax here must be understood broadly (not just linguistically), as what determines the form, construction, composition, or structuring of something. Engineers, film directors, painters, chess, and gardeners speak of syntax in this broad sense." [42, p.84] So even if there is no explicitly defined syntax as a language has, there still is assumed to be a set of rules associated with a signal.

Concerning the interpretation of "meaningful", Floridi also writes that this means "that the data must comply with the meanings of the chosen system, code, or language in question" [42, p.84]

Floridi also refers to the *level of abstraction* of an information-processing agent. A level of abstraction is a type of framework in terms of which an agent can interpret data. It is also related to Kant's concepts that shape our experiences. The level of abstraction, but also the goal of the agent is relevant for how information is interpreted. Depending on the purpose of an agent in a context and the level of abstraction, data can be interpreted as answers to certain queries. Floridi writes about this:

> The point made was that data are never accessed and elaborated (by an information agent) independently of a level of abstraction. Understood as relational entities, data are constraining affordances: they allow or invite certain constructs (they are affordances for the information agent that can take advantage of them) and resist or impede some others (they are constraints for the same agent), depending on the interaction with, and the nature of, the information agent that processes them. With an analogy, a red light flashing repetitively and the car engine not starting allow one to construct the information that (a) the battery is flat. But it makes it more difficult to construct the information that (b) there is a short circuit affecting

> the proper functioning of the low-battery indicator, where the engine fails to start because there is no petrol in the tank, a fact not reported by the relevant indicator which is affected by the same short circuit. Still as constraining affordances, data are exploitable as input of adequate queries, at a given level of abstraction, that correctly semanticize them to produce information as output, for a particular purpose. In short, semantic information can also be described erotetically as *data + queries*.

Here Floridi gives an example that shows how certain data can lead to the construction of information by an information agent. Depending on the purpose and the level of abstraction of the agent, the data can be interpreted as input to queries of the agent which can then be used to construct information. Here the purpose of the agent is to get to their destination and therefore to get their car running again. In combination with the driver's level of abstraction it will interpret the flashing light as an answer to their query of why the engine is not starting. The flashing red light makes it likely that the battery is flat and therefore the driver will extract this information from the situation.

The mind-dependency of semantic information

Just as is the case for Shannon information, it is debated whether semantic information is something that exists independently of cognitive agents, i.e. if it is mind-independent. It may seem straightforward that semantic information cannot exist mind-independently because it requires meaning and meaning is generally considered to be something that only exists in the minds of cognitive agents. Nonetheless, some philosophers have argued that semantic information is less dependent on cognitive agents than is often assumed.

Floridi, for example, has argued that data *can* have a semantics independently of an informee. Here a distinction is made between an *informer* and an *informee*. So Floridi argues that there can be semantic information even if there is no informee that understands it. This is the thesis he calls the *genetic neutrality of information*. Floridi gives an example involving the Rosetta Stone which contains texts in both Ancient Egyptian and Ancient Greek. He argues that even though at some point there was no one who understood hieroglyphs, the text in Ancient Egyptian was considered information. Floridi writes about this: "The identification of an interface between Greek and Egyptian did not affect the semantics of the hieroglyphics, but only its accessibility." [42, p.91]

Floridi thus argues that the hieroglyphs had meaning even when there was no one who understood their meaning at some point. He also claims that: "Meaning is not (at least not only) in the mind of the user" [42, p.91]. He does, however, leave open the possibility that meaning depends in some way on the minds of users, albeit counterfactually, in the sense that for data to have meaning it must be *possible* that someone can understand it.

This becomes clear also from the fact that Floridi distinguishes his position from Fred Dretske's view on the matter [38]. Dretske argues for the more realist idea that information can have a semantics even without an informer. Thus, Floridi does require that there should have been an informer that created the information, i.e. someone that did understand the meaning of the information at some point.

World description and truthfulness

It is clear that Floridi's measure fits best with the idea that the function of information is world description. His measure is such that sentences that describe the most about the world, have the most semantic information. This goes hand in hand with his requirement that genuine semantic information must be truthful.

3.2.1 Evaluation of Floridi's account

Measuring truthlikeness

Again, whether we consider Floridi's measure to be satisfactory depends on our view on the function of information. If we take it to be world description, it can be considered quite satisfactory. One possible aspect in which it is lacking it that it is not very clear how the degree of discrepancy of a sentence will be determined. To determine this we need to know the actual state of the world and then quantify this, but this may not be very evident. It is also argued by D'Alfonso [32] that Floridi's measure lacks the technical detail that is required for it to actually be used. D'Alfonso proposes, however, that we can use other accounts of truthlikeness like that of Oddie [81] and Niiniluoto [80] that can provide a more precise way to measure the truthlikeness of a sentence.

From the perspective of communication

If, however, we do not take the main function of information to describe the world, it can be argued that Floridi's account is too narrow. Lundgren [74], for example, argues that it is too strict to require that all information must be truthful. He argues that with respect to the role of information in communication, it is not practical to require that all information must be truthful. There are situations in which we take something to be information previously to knowing for sure whether it is truthful or not. Lundgren gives the example of a weather forecast which tells us that it will rain tomorrow [74, p.2893]. When we listen to this forecast, we consider the prediction to convey semantic information. However, it can be that in fact it will not rain the day after, which would imply that the information turned out to be false. Nevertheless, it can be considered information.

Lundgren argues that Floridi's account is actually a theory about how *informative* a sentence is, which he distinguishes from being information. He argues that truthful semantic information should be considered a subconcept of semantic information, a type of information which is also informative. He thus uses the term "informativity" in a different way than Floridi, who argued that also false statements could be informative. Lundgren seems to take it to mean that it is information that can potentially lead to knowledge.

Lundgren also argues that if it is required that information is truthful, like Floridi's measure requires, it will be impossible to ascribe an amount of information to liar sentences like "This sentence is not semantic information." If we say that this sentence is truthful, this implies that the sentence contains semantic information. We then have a contradiction with what is said by the sentence however. If we say that the sentence is not truthful, this implies according to Floridi's account that the sentence is does not contain semantic information. However, if the sentence is false then this would imply that the sentence *is* semantic information. Thus, on Floridi's account the sentence is semantic information if and only if it is not semantic information, which is paradoxical. We will next consider an account of semantic information that also requires semantic information to be truthful.

3.3 Dretske: environmental information

Dretske presented his account of information a while earlier in his book *Knowledge and the Flow of Information* that was brought out in 1981. It is more radical in its attempt to naturalize semantic information than Floridi's account is, but it did inspire Floridi's account in certain ways.

Dretske focuses specifically on a type of information which Floridi has called *environmental information* and what is also known as *natural information*. Floridi describes this as a certain perspective on information according to which information is understood as "patterns of physical signals, which are neither true or false" [42, p.30]. An example given by Floridi is that the rings inside the trunk of a tree can give us information about the age of the tree. In these cases, there is no informer who intentionally creates information but it just "exists" in the environment.

This environmental information is the type of information that Dretske focuses on, but he does not give his perspective on information a special name and just refers to his account as "A Theory of Semantic Information" [38, p.63]. It seems that Dretske wants to argue that all information is a type of environmental information: it must always be truthful and Dretske argues that it can be considered an objective commodity. Dretske writes that when information is clearly distinguished from meaning:

> ...one is free to think about information (though not meaning) as an objective commodity, something whose generation, transmission, and reception do not require or in any way presuppose interpretive processes. [38, p.vii]

Nonetheless, Dretske does not deny the semantic character of information. So it seems that it is indeed Dretske's ambition to show that information can have a semantics even without an informer or informee, as Floridi points out. The question is whether Drestke succeeds in providing an account that shows that this is indeed possible.

Dretske argues that we can use insights from Shannon's MTC to construe such an account. He writes that with Shannon's MTC and his own adaptions of Shannon's measure to semantic information:

> ...we have all the ingredients necessary for understanding the nature and function of our cognitive attitudes, all that is necessary for understanding how purely physical

systems could occupy states having a content (meaning) characteristic of knowledge and belief" [38, p.xi].

Here we see clearly Dretske's naturalistic project. Drestke's theory of information is also a part of his broader project to give a naturalistic account of knowledge that is information-based. Knowledges is identified with belief that is produced by information. And the idea is that if information can be reduced to physical phenomena, then knowledge can too.

Dretske's measure of information

Dretske realizes that Shannon's theory is not concerned with semantics. However, Dretske believes that the theory may still serve as a means to characterize semantic information.

Dretske [38, p.13] uses the measure for the individual Shannon information, i.e. the surprisal value, of a message as a starting point for a measure of semantic information. This surprisal value will be based on the probability distribution over the possible outcomes of a situation. Dretske uses the example of a situation in which a group of people need to choose a leader by randomly taking out a piece of paper with one of their names on it. When there are 8 people and 8 pieces of paper, and we assume there is equal probability for each piece of paper to be taken out, then the individual surprisal value for the event that Alice's paper will be taken out (e_a) is:

$$I(e_a) = -\log(\frac{1}{8}) = 3 \tag{3.6}$$

Dretske [38, p.51] then goes on to define a way to quantify the amount of information that can be contained in a signal about another object or event. In the previous example we can think about the piece of paper being handed over to an outsider of the group who must read it out aloud. This person who receives the paper reading out the name 'Alice' is the event r_a and the amount information in r_a about e_a is:

$$I_{e_a}(r_a) = -\log p(e_a) - E(r_a) \tag{3.7}$$

Dretske uses $E(r_a)$ to refer to the amount of uncertainty (also known as conditional entropy or equivocation) about whether e_a occurred when we know that r_a occurred, a quantity that has not been individually defined by Shannon. In the MTC it occurs only as the conditional entropy between two variables (as in equation 1.5 from Chapter 1). Dretske [38, p.24] defines the quantity he needs as follows[3]:

$$E(e_a) = -\sum_i p(e_i|r_a)\log p(e_i|r_a) \tag{3.8}$$

[3] Lombardi [69, p.29] has pointed out that the equation in 3.8 proposed by Dretske is an incorrect way to measure the individual uncertainty in one signal at the destination about one specific signal at the source, which is what Dretske needs. The definition Dretske gives is the same as the definition for uncertainty in one signal at the destination about the source in general (as in equation 1.4 from Chapter 1). Lombardi also shows that this leads to an inconsistency between Dretske's equations. She argues that it should be replaced with the following definition: $E_{e_a}(r_a) = -\log p(e_a|r_a)$

We see that the lower the uncertainty about whether e_a occurred after observing r_a , the higher the information contained in r_a about e_a. While this is a way to calculate the amount of information contained in one signal about another object or event, it does not yet define what information is.

The Xerox principle

Next to this quantitative measure, Dretske also formulates a definition of semantic information. He argues that any theory of semantic information should first and foremost satisfy the *Xerox principle*:

> Xerox principle: If A carries the information that B, and B carries the information that C, then A carries the information that C.

It is a simple principle that is crucial to the idea of environmental information. We can consider the example of when we are on the phone with our friend. The microphone of the cell phone of our friend records that the friend says "yes". The cell phone will change this into a digital signal that is transmitted to our own cell phone. There it is changed into an analog sound wave again which comes out of their speaker as "yes".

Here we see that the vibrations of the friend's microphone carries the information that the friend says yes. The digital signal carries the information about the vibrations of the friend's microphone. The speaker's vibrations of our phone carry information about the digital signal. Applying the Xerox principle twice, we get that the speaker's vibrations carry the information that our friend says "yes".

Three conditions for semantic information

Dretske also proposes three other conditions for a definition of semantic information. The first is that if a signal carries the information that some object s is F, then the following should hold:

> (A) The signal carries as much information about s as would be generated by s's being F.

In the context of the example of the selection of a leader, we can understand this condition as saying that if the event of reading out the name on the piece of paper can carry information about the paper that was selected, then it needs to carry the same amount of bits as the selection of the paper. In this example this is so, because there are also 8 possibilities of names that can be read aloud. If the person who must read out the paper is biased and will always read out Alice's name even though it is not the name on the paper, the event will not carry information about the selection of the paper. This condition also amounts to requiring that the individual equivocation or conditional entropy $E_{e_a}(r_a)$ is zero. If this equivocation is not zero, then the amount of information that the signal carries about s will be smaller than the amount that is generated by the fact that s is F.

The second condition is that:

(B) s is F.

This is simply the requirement that the information should be truthful.

Then there is the third condition:

(C) The quantity of information the signal carries about s is (or includes) that quantity generated by s's being F (and not, say, by s's being G).

As Dretske concedes himself, this requirement is formulated in a rather vague manner. Dretske argues, however, that it is about that aspect of semantic information which (A) and (B) together cannot capture and that (C) requires that the information that the signal carries is really about the fact that s is F.

A definition of informational content

Dretske [38, p.65] then proposes a definition of *informational content* that satisfies these three conditions:

A signal r carries the information that s is F = The conditional probability of s's being F, given r (and k), is 1 (but, given k alone, less than 1).

Here the k stands for what the receiver of the information knows about the possibilities at the source, i.e. whether s can be G or F. This definition captures the idea that something can be considered information when it makes a probabilistic difference, when the conditional probability that F is s is increased to 1.

We indeed see that this definition satisfies the three conditions. It satisfies (A), because if the conditional probability of s being F given r is 1, then we must be certain that s is F. This means that the equivocation will be zero. According to Dretske's measure of information 4.4, this means that $I_{e_a}(r_a) = I(e_a)$. Thus, the amount of information carried by the signal about s being F will be the same as the amount of information generated by s's being F.

It satisfies (B), because if the conditional probability of s's being F given r is 1, then this implies that s is F. Dretske writes about this in a footnote: "A conditional probability of 1 between r and s is a way of describing a lawful (exceptionless) dependence between events of this sort, and it is for this reason that I say (in the text) that if the conditional probability of s's being F (given r) is 1, then s is F." So we see that since Dretske assumes that the conditional probability of 1 reflects the actual lawful dependence between the signal and the source, he can infer that s is indeed F.

When it comes to condition (C), Dretske does not give a clear explanation as to how this is satisfied by his definition. We can note, however, that when the conditional probability of s's being F is increased by receiving the signal r, then this means that the information in signal r is really generated by s's being F, that there is some relation between the signal and s's being F. However, the fact that the conditional probability is increased by observing r is, it seems to me, caused by the fact that we assume that such a relation exists. So Dretske's definition of informational content does not really explain how it can be ensured that there is a dependence between r and s. But possibly this is also not the goal of the definition.

Background knowledge

We see here that Dretske takes into account the background knowledge k of the receiver when defining semantic information. Whether a signal carries information depends on whether the receiver can extract this information from the signal. Suppose that a receiver knows that s must either be G or F and then receives the information that s is not G. Because of their background knowledge they get the information that s is F. If the receiver did not know that s must either be G or F then they would not have received this information and thus with respect to them the signal would not carry that information.

This suggest that Dretske's information *cannot* have a semantics independently of an informee. Dretske himself, however, holds [38, p.87] that it can. He argues that knowledge can, in turn, be defined in terms of information theory and that this does not require reference to any interpretative process. According to Dretske's definition of knowledge, if it is known as in the previous example that s is F or G, then the receiver must have previously received the information that s is F or G. This is, again, simply the consequence of lawful relations between the source and the receiver according to Dretske. If, to explain whether they got the information that s is F or G, we need to look again at the background knowledge of the receiver, we will look at the information they previously received and so on, until we reach a situation in which the receiver has no background knowledge at all. Dretske calls this the recursive character of the definition of knowledge. This is essential to the idea that information can have a semantics completely independently of any interpretative process.

In Dretske's definition of informational content, we also see that the conditional probability of s's being F cannot be 1 given k alone. It can only be 1 once the signal r is received. This means that it cannot be a necessary fact that s is F or that it is already known by the information-processing agent that s is F. This means that according to Dretske, a signal can only carry information about some object if this is information is not already known by the receiver.

Communication and truthfulness

We see that Dretske's measure of information is similar in form to Bar-Hillel and Carnap's inf measure, except that Dretske also takes into account the correlation between a signal and what the information is about, which is represented by the equivocation. We subtract the equivocation, i.e. the uncertainty in the signal about the source, from the surprisal value. However, Dretske also requires that for some signal to carry genuine semantic information, the equivocation must be zero. That means that the measure ends up being the same as Bar-Hillel and Carnap's. A difference is, however, that Dretske requires that semantic information must be truthful.

The fact that Dretske uses the surprisal value of a signal to measure semantic information still indicates that his account is more in line with a communication perspective on information. And it can also be in line with a communication perspective to require that information must be truthful.

3.3.1 Evaluation of Dretske's account

As a theory of information that has communication as the main function, it can be argued that Dretske's account is quite satisfactory. There has, however, been some critique on Dretske's measure. First and foremost, the critique given by Lundgren against accounts that require all semantic information to be truthful, is also applicable and directed against Dretske's requirement that information must be truthful. Then there are also two other points of critique that I will discuss here.

Timpson on Dretske's measure

Timpson [89], for example, has argued that Dretske's measure is not suited as a measure of the information that a signal can carry about another object or event. Timpson argues that the surprisal value of the event r_a and the uncertainty in r_a about e_a are quantities that are too independent of each other. He points out that it can be the case that while the uncertainty in r_a about e_a can be very high, the information that r_a carries about e_a can still be arbitrarily large if the initial surprisal value of e_a is very large. This is wrong according to Timpson, because if the uncertainty in r_a about e_a is high, the information in r_a about e_a cannot be high. Timpson concludes that $I_{e_a}(r_a)$ is unacceptable as a measure.

I would argue that Timpson's argument can be nuanced slightly. There is still a sense in which it is correct that $I_{e_a}(r_a)$ can be high if $I(e_a)$ is very high even though the uncertainty in r_a about e_a is also big. If the initial information in e_a is very high, it does not matter how uncertain the correlation is, but any suggestion that e_a might be the case could still convey some information.

It does feel strange, however, to say that r_a then carries a lot of information about e_a. Dretske, however, has imposed condition (A) which ensures that the quantity of information in a signal about a source must be equal to the amount of information generated by the event at the source. According to Dretske's definition of informational content, if there is any uncertainty in r_a about e_a, r_a cannot contain any information about e_a. This is because the conditional probability that e_a is the case must be 1. Thus, we see that Dretske would not really count those situations described by Timpson as information.

From the perspective of world description

If the main function of information is considered to be world description, then it can be seen that the measure proposed by Dretske will not capture the right aspect of information. Just as with Bar-Hillel and Carnap's account, we will get that two statements that give a similarly complete description of the world can carry different amounts of information because one is more unlikely than the other.

Dretske, in his book, gives an example of the information content in a weather report to argue in favor of his measure of information. He writes:

> It is a bit like listening to the weather forecast during the monsoon season (probability of rain on any given day = .9) and listening to the forecast during the normal

> season (when, say, the probability of rain = .5). In both cases some information is conveyed by a forecast (whether it be "rain" or "sunny") but the forecasts during the monsoon period will, on the average, be less informative. Everyone will be expecting rain anyway, and most of the time they will be right without the help of the official forecast. During the monsoon season the official forecast eliminates less "uncertainty" (on the average) and therefore embodies less information.

This argument, however, assumes one specific perspective on information, namely that its main function is in communication. If we think of information being some more or less complete description of the world, then this example can tell us the opposite. From a world description perspective, it seems quite strange that the statement "it will rain tommorrow" can carry different amounts of information depending on people's expectations about the weather. Both during the monsoon season and during the dry season, the statement gives an equally complete description of the actual state of the world.

The recursive definition of knowledge

There has also been some critique which does not concern Dretske's quantitative measure of information, but his argument that information does not require any interpretative processes. Dretske realizes that the semantic information is dependent on background knowledge, but argues that this background knowledge of receivers can be defined recursively and solely based on previously received information. This previously received information would then not require any interpretative process either. This argument is, however, not extremely convincing.

Dretske argues that all background knowledge can be traced back to previously received information until an initial state is reached in which the cognitive agent does not have any background knowledge. As Alston [4] and Van den Herik [53] note, however, it is unclear how a cognitive agent, as a *tabula rasa*, can realize that some signal carries information. It seems that we always need some awareness about the correlation between the signal and the source.

Alston [4] also points out that it is additionally yet unclear how the rest of all knowledge and information-processing may follow from this initial state. In order to believe in Dretske's account it seems that we need a more in depth explanation of how this can happen.

Awareness of correlations

Barwise [16] also stresses the fact that some awareness of the correlation between the signal and the source is required for an agent to perceive something as information. When Dretske refers to the background knowledge k that is taken into account in the definiton of informational content, he does not refer to this awareness. He only refers to the knowledge about the possible states of the source. Barwise argues that this is a problem:

> Dretske emphasizes that it is only due to nomic relations between types of situations that one can carry information about another. But when he defines the basic notion, he relativizes only to what the receiver knows about alternative possibilities. More

> important is what the receiver knows about the nomic relations. While information is out there, it informs only those attuned to the relations that allow its flow. This is true in general, but obvious where the relations are conventional, as with language. If you don't understand English, the doctor's uttering "You have the flu" will not inform you. Dretske gives a recursive account of knowledge of contingent fact - but only in terms of knowledge of the nomic relations that allow information flow. [16, p.65]

We must agree with what Barwise writes here. Dretske's recursive account of knowledge will only be successful if we can assume knowledge of the correlations that allow information flow.

3.4 Conclusion

In this chapter, we have seen three accounts of semantic information. We have seen how, with respect to information with as its main function communication, Dretske's account and Bar-Hillel and Carnap's account are best-suited. The difference between these two accounts is that Bar-Hillel and Carnap also allow information to be false, whereas Dretske does not. From the perspective of world description, Floridi's account would be best-suited.

It can be argued, however, that measuring information from the communication perspective aligns better with our general intuitions about semantic information. It can be agreed upon, for example, that the semantic information contained in a sentence can differ for agents with different background information.

In this concluding section we will first look at how Dretske's account and Bar-Hillel and Carnap's account relate to Shannon's theory of information, as their measures are inspired by Shannon's surprisal value. We will also consider to what extent the quantitative measures we have seen in this chapter satisfactorily capture semantic information. Finally, we will look whether semantic information is mind-dependent and subjective.

3.4.1 Comparison with Shannon's theory

We saw that Dretske's measure and Bar-Hillel and Carnap's inf measure are in form very similar to Shannon's measure of individual information, i.e. the surprisal value $I(x)$. This quantity is also inversely dependent on the likelihood that this message x will appear. The more unlikely a message, the higher the Shannon information in that message and the bigger our surprise if it does appear.

One thing that distinguishes Shannon's measure of individual information from Bar-Hillel and Carnap's measure of semantic information is the domain that the information can be about. In the case of Shannon information, the probability distribution is over the possible signals that the source will emit. We can only calculate the individual Shannon information of one of these signals relative to that source.

In the case of Bar-Hillel and Carnap's measure of semantic information, we can calculate the amount of information contained in every possible sentence of the language. The logical probability function that is used is defined for all sentences of the language and the sum of the probabilities of all state-descriptions sums to one. This means that the domain of this measure of semantic information is much bigger and more general.

Dretske's measure of information is more similar to Shannon's theory, because the probability distribution used to calculate the surprisal value ranges only over the possibilities at the source about which we get information. This means that Dretske's measure of information is more confined to a specific domain than Bar-Hillel and Carnap's measure.

One thing in which both measures of semantic information differ from Shannon's theory is that in Shannon's theory it is seems to be assumed that the probabilities are known and determined beforehand. They could have been determined based on the frequency of the outputs of the source, but the final probability associated with an output of a source does not have much to do with the meaning of that output. In the context of the theories of semantic information, the probabilities will be determined based on the knowledge of the agents. In Dretske's theory, the agents will consider different possible outcomes of some event and decide which outcome is more likely based on other knowledge. In Bar-Hillel and Carnap's framework, the state-descriptions can also get different probabilities assigned to it based on the knowledge of the agent. Here it seems that the meaning of the possible outcomes of the event is taken into account.

3.4.2 Quantitativity of semantic information

All three accounts have proposed quantitative measures to measure semantic information. We can ask ourselves if we think that these measures sufficiently capture what semantic information is about.

It can be argued that if we think about what semantic information is, a number does not sufficiently describe this. We can think again about the sentence "The milk is in the fridge". Following Dretske's account, if we want to consider the semantic information contained in this sentence we must look at the possibilities at the source. Suppose that there are 3 possibilities: there is no milk in the house, the milk is in the cellar and the milk is in the fridge. Suppose that there is a 0.5 probability of there being no milk, a 0.25 probability of the milk being in the fridge and similarly for the cellar. According to Dretske's measure the semantic information contained in the sentence would be $-\log(0.25) \approx 1.39$, assuming that the equivocation is 0. This number, 1.39, however seems to capture only one aspect of the semantic information contained in the sentence. It only tells us something about how surprised we will be to obtain this information. Bar-Hillel and Carnap's measure would give a result that is also mainly determined by how likely we think it is that the milk is in the fridge.

If we would use Floridi's account, we would obtain some other number x that would tell us only something about how much about the world the sentence describes. It would not give an answer to the question "what is the information contained in this sentence?". If we compare this with Shannon information, we see that if we ask "what is the individual Shannon information contained in this message?", a number will in fact be a satisfactory answer.

We can thus say that the quantitative measures we have seen so far can only capture one aspect of information: how much it describes of the world or how much new information it will give to someone. To answer the question "what is the information contained in this sentence?" it seems that a qualitative characterization is necessary. In the following chapter, we will see how logic can help with such a characterization.

3.4.3 Mind-dependence

In sight of our main question, to see whether semantic information and Shannon information are two different types of information, we will now consider whether semantic information is something that is indeed inherently subjective and mind-dependent. We will first consider the mind-dependency of semantic information.

We saw that Floridi argued that information can have a semantics independently of an informee, but not independently of an informer. This meant that Floridi holds that data can be considered to contain semantic information if we see that it is in principle possible to interpret it, like with the Rosetta stone. However, there does not need to be an informee at that exact time and place who can understand the information. This therefore seems to amount to the idea that semantic information cannot be completely mind-independent.

We saw that Dretske argued that semantic information does not require any interpretative process. Floridi therefore argues that according to Dretske, information can have a semantics without any informer or informee. It is, however, not clear whether Dretske would put it like that. Thus it is unclear if Dretske would argue that semantic information can be completely mind-independent. Because even though semantic information may not require an interpretative process, it does require some cognitive agent who memorizes previously received information and then uses this to obtain the new piece of information. This Dretske might agree with.

Nevertheless, the claim made by Dretske that semantic information does not require any interpretative process does not seem well-argued for. Dretske argued that the background knowledge of an agent that semantic information is dependent on can be materialistically defined in terms of previously received information. We already pointed out, however, that this recursive definition of knowledge needs more refinement because it is not clear how one may ever obtain information in the beginning, if there would be no previously received information.

Going back to the question of whether semantic information is mind-dependent, it seems that both epistemically and ontologically, semantic information cannot be mind-independent. It seems that it should at least be possible that there is some agent who will understand some signal to be semantic information for this signal to actually be semantic information. It is because a cognitive agent has certain goals and knows about certain correlations or conventionally assigned meanings that it can subtract information from signs. It is in the mind of the cognitive agent that data becomes semantic information. Of course, we can talk as if semantic information "resides" in an object independently of any agent, but that is because we all agree on the fact that it contains this information for each and everyone of us as cognitive agents. It is also remarkable that one sign can carry so many different pieces of semantic information for each agent.

In sight of environmental information, it may seem more acceptable that information could be mind-independent. The relation between smoke and fire or the relation between a phone ringing and someone calling just seem to be “out there” because of the natural laws and the way that a cell phone is designed. However, the case of environmental information is similar to linguistic information. While it may be the case that lawful relations exist between physical objects or phenomena, it still requires a cognitive agent to realize that this relation exists and thereby subtract information.

This all suggests that a sign or data can only carry semantic information when it can be interpreted by an agent. Barwise says that “information is out there, but it informs only those attuned to the relations that allow its flow”[17]. I would even argue that only relations are out there and that information is there for those attuned to the relations that allow its flow.

3.4.4 Subjectivity

The next thing to consider is whether semantic information is subjective or objective.

Bar-Hillel and Carnap have not said much about whether semantic information is objective or subjective. However, the fact that their measure of information is dependent on the probability an agent associates with a sentence, implies that they do consider semantic information to be something that can differ between agents. This suggests that semantic information is something subjective.

Floridi has also argued that the level of abstraction (which includes their background knowledge) of an agent and the goals of this agent all determine what semantic information they can extract from a situation. This can obviously differ from different agents and this will thus differ. When we look specifically at Floridi’s measure of information, it can seem that it is designed to be objective. The degree of discrepancy of a sentence does not have much to do with different agents. However, it first needs to be determined what propositions actually fall under the semantic information that is subtracted from a signal. Once that is done, it is then calculated how much these propositions say about the world.

Dretske argues that information is something objective. This, however seems strange if we look at his definition of informational content. According to this definition, whether a signal carries information is dependent on the background knowledge of agents. This would imply that there are at least some cases in which different agents assign different semantic information to the same signal.

According to the interpretation of objectivity handled in this thesis, information is objective if it can be publicly agreed upon that it is there. If we consider a sentence like “The milk is in the fridge”, for example, we could thus say that it objectively contains the information that the milk is in the fridge. Everyone who understands English will get this information. This might be the same sense of objectivity that Dretske had in mind.

There will also be cases, however, where one person can extract additional information from a signal that another person cannot. As an example, we can consider again the sentence “The milk is in the fridge”. Suppose that someone knows that the milkman always comes by to bring milk. Then this sentence will for this person also carry the information that the milkman came by that day. For someone who doesn’t know about the milkman, this sentence will not carry this information. This, Dretske would certainly agree with. It thus seems that in principle, semantic information is subjective but that there are cases in which this can become objective, in the sense of intersubjectivity.

Chapter 4

Logic-based approaches to semantic information

In this chapter we will look at some logic-based approaches to semantic information. These approaches provide a more precise way to analyze semantic information in a qualitative manner. They can also tell us more about the dynamics of information, i.e. how it can flow from one object to another or from a signal to an agent.

As pointed out in the Introduction, logic is traditionally only associated with information in the sense that a logical system determines which inferences can be made from which premises. These proof-theoretical rules can thus be considered to determine what information an agent can logically infer from some sentence given their background knowledge. More recently, logic is also employed to model the epistemic states of agents specifically. With the "Dynamic Turn" in logic, which is amongst other initiated by works like [21], we are now also using logic to study informational processes like communication, belief revision, and measurement processes.

Van Benthem and Martinez [22] have made a distinction between logical accounts of information that treat *information as range* and those that treat *information as correlation*. We will first look at information as range as it is treated in Epistemic Logic and Dynamic Epistemic Logic. In these logics, the idea is that the more states of this world the agent considers possible, the less information the agent has. This will become more clear during the next section.

We will then look at information as correlation as treated in Situation Theory and Logic of Functional Dependence. These accounts focus on the fact that states of affairs in the world carry information about other states of affairs. This perspective on information is highly influenced by Dretske's account of environmental information. Barwise and Perry [17] write, for example, that one of their main points they wanted to convey with their book on Situation Theory is the following:

> That the ecological approach, looking at people as living things among other living things in an environment full of information, can shed light on the special case of language. [17, p.274]

One of the main initiators of this ecological approach is Dretske, as Barwise and Perry recognize [17, p.274]. Dretske was the first to focus on the correlations that exist in the world between physical objects and events and use that in order to explicate concepts like knowledge, information and meaning.

We will consider what aspects of semantic information these logics formalize. We will also formulate a way to define the concept of the semantic information contained in a sentence. We will then compare this qualitative definition of semantic information with the quantitative measures of information we considered in the previous chapter. At the end we will also take into account these logic-based approaches to information into our evaluation of semantic information as semantic and mind-dependent.

4.1 Information as range

We will first look at approaches that consider information as range. The first logic that we can use to model information as range is Epistemic Logic. Jaakko Hintikka [54] formulated the first idea for such a logic. The idea is that we can use Epistemic Logic to represent the knowledge and beliefs of agents based on a possible world semantics.

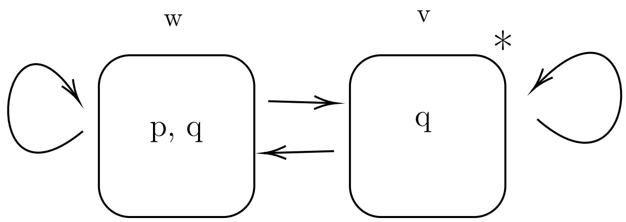
4.1.1 Epistemic Logic

The language of Epistemic Logic contains the modal operators $\mathcal{K}_i\phi$ for every agent i we want to consider. This operator means that agent i knows that ϕ is the case. The language also contains propositional variables p, q, and so on, and it contains the Boolean operators negation $\neg$ and disjunction $\vee$.

One possible way to give a semantics to Epistemic Logic is by using Kripke models. Kripke models $\mathcal{M}$ are models of the form $\mathcal{M} = \{W, R, \| \bullet \|\}$ where W is a set of possible worlds, R is a set of relations between these worlds and $\| \bullet \|$ is a valuation function that determines which atomic propositions are true in which possible worlds.

We can illustrate how this model works based on an example. In the figure below we see a model $\mathcal{M}$ with two worlds w and v. These worlds represent possible states that the actual world is in. In this example we only look at the epistemic states of one agent. The arrows drawn represent which state of the world the agent considers possible. World v represents the state that the world is actually in, so it is marked with an asterisk. The idea is that $\mathcal{K}_i\phi$ is true in a world w when in all states that this agent considers possible it is the case that ϕ:

$$\mathcal{M}, w \models \mathcal{K}_i\phi \text{ iff for all } v \in W \text{ such that } wRv \text{ it is the case that } M, v \models \phi$$



We can now check which sentences are true in this model. We see that in world v, which is the actual world, it is true that $\mathcal{K}q$, because in all worlds that this agent considers possible (world w and world v) it is the case that q. We also see that $\neg\mathcal{K}\neg p$ which can be interpreted as saying that this agent considers it possible that p is true. This holds because there is a world in which p is true which this agent considers possible (namely world w).

From knowledge to information

Although the operator $\mathcal{K}_i\phi$ is traditionally interpreted as "agent i knows tha ϕ", it can easily be considered to stand for the information an agent has received. In light of considering how Epistemic Logic can be a theory of information, Van Benthem and Martinez [22] propose to interpret $\mathcal{K}_i\phi$ as saying "to the best of agent i's information it is the case that ϕ".

Dretske argued that knowledge is caused by previously received information. This would mean that if an agent knows ϕ, they would have been informed that ϕ. In this light we could also interpret $\mathcal{K}_i\phi$ as "agent i is informed that ϕ". In this way Epistemic Logic can tell us something about the information that different agents have received.

By concatenating knowledge operators we can also represent what information agents have about the information that others have, i.e. "$\mathcal{K}_a\mathcal{K}_b\phi$" would mean that agent a knows that agent b knows that ϕ. In informational terms: a is informed that agent b is informed that ϕ. Another possibility is to add the operator $Ck\phi$ which means that it is common knowledge that ϕ. One interpretation of common knowledge is that it is required that everyone knows that ϕ but also that everyone knows that everyone knows that ϕ, and so on until infinity. This interpretation is due to Schelling [86] and Lewis [66].

In informational terms, we may think of this as saying that everyone is informed that ϕ and that everyone is informed that everyone is informed that ϕ, etc.

In the next section, we can show how this can be taken one step further.

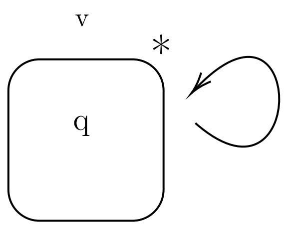
4.1.2 Dynamic Epistemic Logic

With what is called Dynamic Epistemic Logic, or "DEL" [21, 12, 37], we can model how information updates change the epistemic attitudes of the agents.

We can consider what happens, for example, if some information is publicly announced. This means that all agents become aware of this information. A truthful public announcement is what can lead to common knowledge because everyone will know that everyone has heard the information. A public announcement can change which states of the world the agents consider possible. Public announcements were first included in the Logic of Public Announcements by Plaza [82] and also by Gerbrandy and Groeneveld [45].

Plaza [82] proposed to represent a public announcement as a change in the model and Baltag and Moss [11] follow Plaza in doing so. If it is truthfully and publicly announced that ϕ, this means that the model must be updated so that all the worlds in which ϕ is false are deleted. The arrows to these worlds must also be deleted. This idea of considering how the information states of agents are updated can also be found in types of Dynamic Semantics such as what is used in Dynamic Predicate Logic by Groenendijk and Stokhof [49].

In the previous example, we could consider a public announcement that $\neg p$. This would mean that the world w would be deleted and the updated model would look as follows:



A symbol in the language of Epistemic Logic is included that represents the action of a public announcement: $\phi!$. We then also include a dynamic operator in the language: $[\phi!]\psi$. The intended reading of this operator is that in the new model, after the update with ϕ, ψ will be true. The exact change of the model and the truth conditions of this operator can be found in Baltag and Moss' paper [11].

In our example, however, it can be intuitively seen that $[\neg p!]\mathcal{K}\neg p$ would be true. After the update with $\neg p$, it is true that the agent knows that p is false. This is so because in all state of the world the agent considers possible (i.e. only world v) it is the case that p is false.

Baltag and Moss [11] also show how we can model private announcements, which represent situations in which, in a group of agents, only one agent or a group of agents is informed about something.

The dynamics of information

We thus see that we can use Dynamic Epistemic Logic to represent how newly received information changes an agent's epistemic attitude. In this way, it can capture something about the flow of information.

We can also use DEL to make more precise how one public announcement can lead to more knowledge for some agents than for others. Suppose that it is publicly announced that "the milk is in the fridge". For an agent who knows that the only way to get milk is if the milkman comes by, this sentence will lead to the knowledge that the milkman came by. This can all be modeled with DEL, for example by having p represent the proposition that the milkman came by and q that the milk is in the fridge. Some agent could know that $p \to q$. If it is publicly announced that p, this would mean that all the worlds in which p are false would be deleted from the model. Thus, the agent would end up knowing that q.

We can formalize this concept of the information that an agent obtains from a public announcement. We can do this by looking at the set of statements that an agent a knows after the update and did not know before the update. This would then be relative to a model M and update model M' and a world w in that model. If we update with a sentence ϕ, the semantic information contained in ϕ for agent a can look as follows:

$$inf_a^w(\phi) = \{\psi : M, w \nvDash \mathcal{K}_a\psi \wedge M', w \models \mathcal{K}_a\psi\}$$

Now, a sentence ϕ will contain as semantic information all and only those sentences ψ that an agent will learn from ϕ. It can be argued, however, that even if an agent already knows that ϕ is true, a public announcement with ϕ still contains the semantic information that ϕ for them. This is to require, for example, that if someone tells me that the milk is in the fridge and I already know that this is the case, their message still contains the semantic information that the milk is in the fridge.

Dretske implicitly argued against this requirement with his definition of informational content [38, p.65]. He argued that previously to receiving some signal, an agent cannot know for sure that ϕ for that signal to carry the information that ϕ. I personally think, however, that an exception should be made for the direct semantic information contained in a sentence, i.e. the meaning of the sentence itself. It also seems to hold for some cases of environmental information. If I know that smoke implies that there is a fire, seeing smoke will convey the semantic information to me that there is a fire, even if I already know that there is a fire. Thus, we could also use the following definition:

$$inf_a^w(\phi) = \{\psi : \psi = \phi \vee (M, w \nvDash \mathcal{K}_a\psi \wedge M', w \models \mathcal{K}_a\psi)\}$$

Here we express that a sentence ϕ contains the semantic information that ϕ and those sentences ψ that an agent learns from receiving ϕ.

With DEL, we can also formalize some of the correlations that agents are aware of and in terms of which they can obtain information. The knowledge that if the milk is in the fridge, the milkman must have come by, is an example of the awareness of such a correlation. Another is the knowledge that smoke implies fire. We can thus look at conditional knowledge of the form $\mathcal{K}(p \to q)$ that will be used to infer new information from some signal.

Weaker than knowledge

We have now looked at (Dynamic) Epistemic Logic with knowledge operators. We may also want to model, however, cases of incomplete information which may result in belief rather than knowledge. We can then use the operators $B_i\phi$ for each agents which means that "agent i believes that ϕ". These operators will have the same semantics as $\mathcal{K}_i\phi$ in the sense that $B_i\phi$ is also true if in all worlds that agent i considers possible it is the case that ϕ is true. However, now, whether an agent considers a world possible is based on their belief, i.e. on whether this possible world is consistent with their beliefs, rather than on whether it is consistent with their knowledge.

A different option is to use so-called *plausibility models* [14, 50]. In these models, we interpret the relation between worlds as indistinguishability rather than as epistemic possibility and we add a relation which expresses whether worlds are considered more or less plausible. This approach can be found in an article by Baltag and Smets [14].

There are many additional aspects related to public and private announcements that we can model with DEL that we have not discussed here, such as the relation between truthful and false announcements, the reliability of the source, etc. These factors could also play a role in how the information states of the agents will be updated. In general there are also more informational processes that we can model with DEL that we have not discussed here. One example of this is measurement processes in physics [9, 13].

Information as range or information as correlation

We have seen in this section that (Dynamic) Epistemic Logic indeed treats *information as range.* The more worlds are in an agent's range of possibilities, the less an agent knows, i.e. the less information an agent has. If an agent receives more information, the range becomes smaller. In the next section, we will look at a different logical approach to information which treats *information as correlation.*

4.2 Information as correlation

In this section we will look at logics that focus on the correlations in the world that allow the flow of information. Hence why it treats information as correlation. This perspective can be traced back to Dretske's view on information. When we see smoke, we can subtract the information that there is a fire because we know that there is a correlation between fire and smoke.

4.2.1 Situation Theory

Situation Theory is the first logic that inherited some ideas from Dretske and formalized them into a theory of information. The foundational work on Situation Theory is by Barwise and Perry [17]. Devlin [35] has later also written about Situation Theory with an explicit focus on the concept of information. Situation Theory is in principle a type of semantics that uses a complex language to formulate propositions, infons, situations, situation types and constraints. Here I will not go into the language of Situation Theory but aim to explain the main concepts of the theory and how they can be used to formalize how agents obtain information from the environment.

Situations: Situations are parts of reality. They determine whether a certain state of affairs is actually the case or not. We can think of the situation that it is true that there is smoke on top of the mountain. The *state of affairs* is that there is smoke on top of the mountain and the situation determines that this is indeed the case. This situation is then in turn of the abstract *situation type* "SMOKE". All concrete situations in which there is smoke are of this type.

Propositions: Propositions are what can be true or false. There are propositions that express that there *exists* a situation of a certain type and propositions that express that a specific situation is of a certain *type*. The truth of these propositions is thus dependent on whether this is indeed the case.

Infons: can be considered to be abstractions from states of affairs. For example, if there is smoke on top of the mountain, we can abstract from this and say "there is smoke somewhere". This is an infon. A situation can be considered to support an infon, in the sense that if there is a situation in which there is smoke, then the infon "there is smoke somewhere" is supported by this situation.

Constraints: A constraint is what holds for agents in virtue of which they can derive information from a certain situation. For example, the constraint that if there is smoke, there is fire, holds for agents. This constraint thus links the type of situations in which there is smoke with the type of situations in which there is a fire. A constraint is itself is also an infon.

Situation Theory then makes it precise how the fact that a situation is of a certain type in combination with a certain constraint makes it that one situation can carry information about another. For example, because the situation that there is smoke on top of the mountain is of the situation type "SMOKE" and because there is the constraint that if there is smoke there is fire (SMOKE $\rightarrow$ FIRE), we can infer that there is a situation in which there is a fire. This situation, in turn, supports the infon "there is a fire somewhere". We can then derive the proposition that it is true that there is a fire somewhere.

As van Benthem and Martinez [22, p.32] note, the *informational content* of the situation that there is smoke can thus be considered to be the proposition that there is a situation which is of the type that there is a fire. The informational content of a situation is that proposition that an agent can derive from the situation. The informational content of a situation can thus differ for different agents.

Constraints and recognizing situations

We see that just as in Dretske's theory, it is represented clearly in the theory that we can only subtract information from a situation because we are aware of certain constraints that exist in the world. In Situation Theory the focus is, however, on the fact that agents think these constraints hold, while in Dretske's theory the focus lay on correlations that are taken to exist independently from us.

As van Benthem and Martinez [22] note, Situation Theory has also made it clear that the flow of information is dependent on the fact that we recognize a concrete situation as being of a certain type. We recognize the smoke on top of the mountain as a situation in which there is smoke and we are aware of a certain constraint in virtue of which we infer that there is a fire.

4.2.2 Logic of Functional Dependence

Here we will consider a more recent logic-based approach which focuses on the correlations that exist in the world in virtue of which we can infer information from one situation about another situation. We will look at the Logic of Functional Dependence (LFD) proposed by Baltag and van Benthem [10]. As the name suggests, it models the dependence between variables. This dependence can be interpreted both in an *ontic* way, i.e. as existing correlations in the world, or in an *epistemic* way, i.e. as constraints in virtue of which we can learn one thing from another thing. We can, as will become clear, also use LFD to model the epistemic attitudes of agents as we were able to do with Epistemic Logic.

Local dependence and global dependence

There are two types of dependence between variables that can be distinguished. One is *global dependence* and the other is *local dependence* [10]. If one variable y depends on another variable x in general, this is global dependence. An example of this is when we take monthly salary as one variable x and expenditures on food as another variable y. We can then say that there is a global dependence of y on x, because people who have a higher salary will generally spend more on food. Global dependence is what is normally considered by other logics of dependence, but LFD focuses on the notion of local dependence.

Local dependence is when a variable y depends on a variable x when x has a specific value. This is more suited to be applied to specific states of affairs that are dependent on each other. For example, we can think of a situation in which someone just ate blueberries and has blue spots around their mouth. In this case, the fact that their mouth is blue determines that they just ate blueberries. Put differently, whether they just ate blueberries is dependent on the fact that their mouth is blue. Or in epistemic terms: from the fact that their mouth is blue we can infer that they ate blueberries. However, if their mouth is not blue, this does not tell us for sure that they did not eat blueberries, because they could have wiped their mouth clean. Hence, whether they just ate blueberries is not dependent on the fact that their mouth is clean. This means that if the color of their mouth is variable x and whether they ate blueberries is variable y, we can say that y is only locally dependent on x. It is locally dependent on x on the specific assignment w such that $w(x) = blue$.

The modal version of LFD

LFD adds atomic sentences that can express local dependence to a generalization of the semantics of First Order Logic known as "CRS", a logic of generalized assignment models [5]. I will here focus specifically on the modal perspective on LFD, which provides an alternative semantics. That means that we will not look into generalized assignment models.

In the semantics for this modal version of LFD, we treat the possible assignments of values to variables as possible worlds. The relations between the possible worlds are then used to express dependencies between variables.

Baltag and van Benthem [10] define so-called *relational models*. Relational models are of the form $\mathcal{M} = \{W, \sim, \| \bullet \|\}$, where W is a set of worlds, $\sim$ is a map $V \to \mathcal{P}(W \times W)$ that assigns to each variable $x \in V$ an equivalence relation $\sim_x$. It also assigns relations $\sim_X$ to sets of variables $X \subseteq V$. $\| \bullet \|$ is a valuation which associates to each atomic formula Px a set of worlds in W. $\| \bullet \|$ must satisfy the requirement that if $x \sim_X v$ and $\| Px_1, ..., x_n \| = w$ for some $x_1, ... x_n \in X$, then it must also be the case that $\| Px_1, ..., x_n \| = v$.

The equivalence relation can be interpreted as follows: if $w \sim_x v$, this means that assignment v assigns the same value to x as assignment w. Similarly for $w \sim_X v$: this means that assignment v assigns the same values to the variables in X as assignment w.

Baltag and van Benthem then define $D_X y$ as saying that y locally depends on values of X if and only if all assignments that assign the same values to X assign the same value to y. In a model $\mathcal{M}$:

$$w \models D_X y \text{ iff for all worlds } v \in W \text{ such that } w \sim_X v \text{ it is the case that } w \sim_y v$$

The sentence $D_X y$ can be interpreted as that y locally depends on X, or that X locally determines y. We can now express global dependence of y on X as well by means of a dependence on the empty set: $D_\emptyset D_X y$

Baltag and van Benthem also define a dependence modality $\mathbb{D}_X \phi$ that has a similar semantics as the $K\phi$ modality we saw earlier in Epistemic Logic.

$$w \models \mathbb{D}_X \phi \text{ iff for all worlds } v \in W \text{ such that } w \sim_X v \text{ it is the case that } v \models \phi$$

We can interpret $\mathbb{D}_X y$ as saying that X locally determines the truth of ϕ. This clause requires that in all worlds (assignments) that assign the same values to X as assignment w, it must be the case that ϕ is true.

An informational interpretation of dependency

Baltag and van Benthem already point out [10, p.14] that we can give LFD an informational interpretation. According to this interpretation the dependencies between variables represent constraints in virtue of which agents can obtain information.

We can thus think of the possible worlds as epistemic possibilities, just as in Epistemic Logic. These epistemic possibilities are assignments of values to variables that specify what the world could be like. One of the assignments specifies the actual state of the world.

As van Benthem and Martinez note [22], in a logic of dependence like LFD, we can think of the different variables as representing something similar to situations as in situation theory. The value of this variable then specifies what this situation is like. We can alternatively think of the variables as representing objects or subsystems.

When can then also interpret $D_X y$ specifically with respect to some agent. It would then represent that the agent has conditional knowledge. This agent can come to know the value of y, if it knows the specific values of X. Analogously, we can interpret $\mathbb{D}_X \phi$ as saying that if the agent knows the value of X, it will know that ϕ is true.

An epistemic reading

So far we considered how we can interpret the dependencies as relative to an agent, i.e. as constraints. A different possibility also noted by Baltag and van Benthem [10] is to take the variables to stand for agents themselves.

In this case, $\mathbb{D}_x\phi$ will mean that agent x knows that ϕ. For groups of agents X, we get that $\mathbb{D}_X\phi$ expresses that together, the agents in X will know that ϕ. This is a form of knowledge which is known as *distributed knowledge* [39]. Furthermore, a sentence like D_xy would mean that if agent x considers a world possible, then agent y would also consider this world possible. This implies that agent x will know at least as much as agent y. The case for groups of agents is similar.

Just as we did in epistemic logic, we can also interpret the knowledge of agents in terms of information received. If $\mathbb{D}_x\phi$, then agent x has the information that ϕ. If D_xy, then agent x has at least all the information that agent y has.

With this epistemic reading, LFD can thus express similar things as Epistemic Logics.

A mixed reading

It is also possible to handle a mixed reading of LFD. Some variables can represent agents while other variables represent situations (objects, subsystems). In this way, we can use LFD both to represent the epistemic attitudes of agents and to represent the dependencies that exist in the world.

It may seem strange that the same operator can be interpreted in such different ways, but if we think about it from an informational perspective it makes sense.

We said that $\mathbb{D}_x\phi$ can be interpreted as: agent x *has* the information that ϕ. This can be taken to imply that the agent x *carries* the information that ϕ. Israel and Perry [55] have argued that an agent having certain information is equal to that agent carrying that information such that this has a certain effect, for example in their behavior. This idea is very much in line with Dretske's idea that knowledge is caused by previously received information and that all semantic information is a form of environmental information, simply caused by a chain of observations and correlations.

We also said that $\mathbb{D}_x\phi$ can mean that a situation x determines that ϕ. If a situation x determines that ϕ, this means that situation x carries the information that ϕ.

Hence, we see that whether the variables are interpreted as agents or as situations, $\mathbb{D}_x\phi$ can always be taken to mean that x carries the information that ϕ. The same goes for D_xy. If a variable y is dependent on variable x, this means that variable x carries information about variable y. If agent x has the information that agent y has, then agent x carries the information that agent y carries.

Knowing this, we can also get combinations between agents and situations in one sentence. If x is an agent and y is a situation, then D_xy means that agent x knows about situation y. Or in informational terms, agent x carries information about situation x.

Information as range and as correlation

We thus see that that LFD can treat both information as range and information as correlation. This combination between information as range and information as range is represented also in the Interpreted Systems by Fagin et al. [39], Epistemic Constraint Logic as described by van Benthem and Martinez [22] and General Epistemic Logic by Baltag and Smets [9].

Dynamic LFD

Just as has been done with Epistemic Logic, we can add dynamic operators to LFD. Baltag and van Benthem [10] show how we can add two types of dynamic operators: $[\phi]\psi$ and $[X]\phi$.

The first operator $[\phi]\psi$ represents the action similar to the public announcement operator we saw in DEL. If it is publicly announced that ϕ, the model will change so that all worlds (assignments) in which ϕ are false will be deleted. We then get that $w \models [\phi]\psi$ holds in a model if, in the updated model, it is the case that $w \models \psi$.

The other operator $[X]\phi$ represents the pubic announcement of the values of the variables in X. If it is publicly announced what the values of X are, the updated model will only contain the assignments that assign these values to X. Then $w \models [X]\phi$ holds in a model if it is the case that $w \models \phi$ in the updated model.

What can we say about information?

LFD will allow us to express even more about information and its dynamics then DEL.

If we also use an epistemic reading of the dependence operators, we can, just as with DEL, represent the knowledge that agents have, or in other words the information that agents have. In addition, we can, just as with DEL, formalize the concept of the information contained in a sentence ϕ for an agent, here denoted with "x" because in LFD x can refer both to an agent and to a variable. Relative to a model M and a world w and an updated model M' after the public announcement that ϕ, we define:

$$inf_x^w(\phi) = \{\psi : M, w \not\models \mathbb{D}_x\psi \wedge M', w \models \mathbb{D}_x\psi\}$$

Additionally, we can consider what conditional knowledge an agent has by looking at what knowledge of the form $\mathbb{D}_x(p \to q)$ agent x has. Or, if we do not use an epistemic reading of the dependence operators, we can just look at the dependencies that exist in the model. This can then still be interpreted from the perspective of an agent who is aware of these dependencies.

What we can also use LFD for, is to model the correlations that exist in the world or in a specific situation. We can also model how public announcements can lead to the awareness of new correlations. It can be the case, for example, that an agent, for example a small child, does not know that the red sign on the door knob of a toilet door indicates that the toilet is occupied. If we would draw a dependence model from the perspective of this child, then there would be no dependency between the color of the door knob x and the vacancy of the toilet y. Suppose that their parent tells them that a red color indicates that the toilet is occupied. This can be represented by a public announcement that $p \to q$ where p represents that the door knob is red and q represents that the toilet is occupied. After this public announcement, the dependence model from the perspective of the child would contain a dependency between the toilet door being red, e.g. $x = red$ and the vacancy of the toilet y.

What is also interesting to note is that in LFD, we see represented clearly Dretske's Xerox principle. We namely get that the dependence relation is transitive. If y depends on x ($D_x y$) and z depends on y ($D_y z$) then also ($D_x z$). If we use an informational interpretation of the dependence operators, then we get that if x carries information about y and y carries information about z, then x carries information about y. This is exactly Dretske's principle.

Quantitatively measuring dependence

So far, LFD only includes ways to formalize that two variables are locally dependent, globally dependent or completely independent from each other. However, it is in principle possible to quantify *how* dependent one variable is on another.

One way to measure the dependence between two variables is with Shannon's measure of mutual information $H(X : Y)$. Note that here we again use the notation commonly used in information theory, so a capital X representing a variable and small x representing one specific value of that variable:

$$H(X : Y) = H(X) - H(X|Y) \tag{4.1}$$

We saw that this measure captures the dependence between two variables X and Y. It is calculated by subtracting the conditional entropy from the entropy of X. In Chapter 1, we saw that we can interpret this measure of mutual information also as quantifying how much information we could gain about variable X by observing Y and vice versa. We also saw that this concerns information in the everyday, semantic sense, not Shannon information.

We should note, however, that mutual information is symmetric, which means that $H(X : Y) = H(Y : X)$. This means that it tells us the mutual dependence of the variables on each other, not specifically the dependence of one variable on the other, which might differ in some cases. It would also quantify global dependence and not local dependence.

If we want to quantify local dependence specifically, we could look at the conditional entropy of X given a specific value y_i of Y:

$$H(X|Y = y_j) = -\sum_i p(x_i|y_j) \log p(x_i|y_j) \tag{4.2}$$

This measure can be interpreted as quantifying the uncertainty about X given the value y_i of Y. This on its own would not quantify dependency, but something like the following could quantify local dependency:

$$H(X : Y = y_i) = H(X) - H(X|Y = y_j) \tag{4.3}$$

Similarly to mutual information, this measure quantifies the difference between our uncertainty about the value of X and our uncertainty about X given a specific value of Y. This could thus be considered to measure the local dependency of X on Y for some specific value y_i.

The next question is how we would incorporate this into a logical system like LFD. We can imagine a dependence operator which also indicates the amount of dependency h, such as $D_x^h y$. This would then be interpreted as: the dependency of variable y on variable x is of amount h.

A full specification of what such an extension of LFD would like does not belong to the scope of this thesis, but it is interesting to note how quantitative concepts from Shannon's theory could be incorporated into a logic like LFD.

Interpreting the measure of local dependence

We can now ask ourselves what the measure in 4.3 represents, what it quantifies.

It quantifies local dependency, but from an informational perspective it also quantifies how much information we can get about one variable by knowing a specific value of another. It quantifies how much information the fact that $Y = y_i$ can give us about X. We can note the similarity with Dretske's measure of information here. He argued that we can measure the amount of information in a signal r_a about the some event e_a as follows:

$$I_{e_a}(r_a) = -\log p(e_a) - E(r_a) \tag{4.4}$$

Here, $-\log p(e_a)$ is the individual Shannon information in e_a and $E(r_a)$ represents how certain we can be that e occurred when we get the signal r_a. This is thus a different type of conditional entropy where we consider two specific values of the variables. It quantifies how certain we can be that $X = x_i$ given that $Y = y_i$. With 4.3, we quantify how certain we can be about X in general given that $Y = y_i$, but the idea is similar.

We could thus argue that just as Dretske's measure, the measure of local dependency, or of "local information" can be used as a proper measure of semantic information. If, just as in the case of Dretske, we apply it to a specific situation or event and we determine the probabilities of the different outcomes based on how likely we think they are, it seems that it can tell us something about the semantic information we can obtain about situation X by knowing what situation Y is like. The value of local dependence, however, does seem independent of what the variable represents, i.e. whether it just represents which drink is in the fridge or that it represents the whole state of the world. Therefore, from a world description perspective, it may not be considered a satisfactory measure of semantic information. Thus it seems that the value of local dependence or local information also measure information from the communication perspective.

Statistical dependency

Dretske argued that in order for one value of variable Y to really carry information about a value of variable X, there must be full dependency. With the measure of local dependency proposed above, maximal local dependency would come down to the fact that $H(X : Y = y_i) = H(X)$. In these cases, there would be no uncertainty about X anymore given $Y = y_i$.

If one holds that for one variable to carry information about another there must be full dependency, then from an informational perspective, it would not add much to add quantifications of dependency. In that case we would always require full dependency with $H(X : Y = y_i) = H(X)$. However, it can be also be argued that in reality, some signals can give us probabilistic information, in the sense that it can make it more likely that some event occurred. For this purpose, the quantifications of dependency could be insightful.

We would thus make a distinction between *deterministic* and *statistical* dependence, where deterministic dependence allows one to infer with certainty the value of variable X given value y_i of Y and where statistical dependence allows one to make a more educated guess about the value of X.

4.3 Conclusion

In this chapter we looked at logical systems that can be used to analyze information and how it flows. We have seen how DEL can be used to model information as range where the information an agent is dependent on which possible states of the world the agent considers possible. The less states the agent considers possible, the more information the agent actually has about the world. We have also seen that with LFD we can model the dependencies that exist in the world. We showed how we can use a mixed reading of the dependence operators to represent both the correlations existing in the world and the information that agents have about the world and about these correlations.

Concerning the dynamics of information, we saw that we can use both DEL and LFD to model the changes in the epistemic attitudes of agents caused by public announcements. We saw that after a public announcement, agents will gain more information based on the inferences they can make by combining their background knowledge and the newly received information. We also saw that newly received information can lead to new knowledge about dependencies in the world which can be represented in an LFD model.

Finally, we made two suggestions of ways to improve the analysis of information within the logics we discussed. We formally defined the semantic information contained in a sentence ϕ relative to an agent within DEL and LFD. We also pointed out that we can use an adapted version of Shannon's concept of mutual information in order to quantify statistical local dependencies. We saw that this measure of local dependence can capture how much semantic information we can get about one variable by knowing the value of another variable. We argued that it measures semantic information in a way that is more in line with the communication perspective than with the world description perspective on information.

4.3.1 Qualitative definitions of information

In the previous chapter, we studied three proposals for quantitative measures of semantic information. I argued that these measures all capture a different aspect of information and that they cannot fully answer the question of "what is the information contained in this signal"? When it comes to semantic information, this is arguably the most important question.

In this chapter we considered three qualitative ways to analyze information. In Situation

Theory, we saw how situations can carry information about other situations in virtue of the constraints that agents are attuned to. We also saw that the informational content of a situation can be considered to be those propositions that an agent can derive from that situation.

For DEL and LFD, I showed that we can formalize the information contained in a sentence for an agent. We did this by comparing the knowledge of the agent prior to a public announcement with ϕ and after the announcement. Thus, the semantic information contained in some signal for some agent is the set of all propositions that the agent has come to know by observing that signal.

Both these definitions of informational content are qualitative yet formal. They do answer to the question "what is the information contained in this signal?". Arguably, these qualitative definitions of information are thus more essential to semantic information than quantitative measures.

4.3.2 Mind-dependence and subjectivity

We can now consider how the logic-based approaches discussed in this chapter bear on the question whether semantic information is mind-dependent and subjective.

Subjectivity

We can first approach the question of whether semantic information is subjective. In DEL and LFD, we see clearly that if there is a public announcement, some agents can be able to extract more information from this announcement than others. According to the definition of semantic information I proposed in this chapter, the semantic information contained in a signal thus differs for different agents. This is dependent on their background knowledge.

In Situation Theory it is also made explicit that different agents can extract different information from the same situation depending on which constraints they are attuned to. This all confirms the idea that semantic information is subjective.

Mind-dependence: information as range

What about whether semantic information is mind-dependent? In the beginning of this chapter we distinguished two perspectives on information that we see in logic-based approaches to information. One was information as range and the other was information as correlation. These two perspectives fit with different philosophical ideas about the ontological status of information.

In principle the idea of information as range already assumes that information is mind-dependent. The information an agent has is dependent on which states of this world the agent considers possible. Information is only considered from the perspective of the information that an agent has.

The definition of the information contained in a sentence ϕ that I proposed is also inherently dependent on a cognitive agent. Thus, in Dynamic Epistemic Logic the focus in general seems to lie on semantic information as being mind-independent.

Mind-dependence: information as correlation

In this chapter we, however, also saw approaches that treat information as correlation. This perspective is more connected to the possibility that information is mind-independent.

In Situation Theory, we do see a lot of Dretske's ideas reflected. Dretske leaned more towards the idea that information can be mind-independent as we saw in the previous chapter. However, Dretske's idea that semantic information does not require an interpretative process is not repeated by Barwise and Perry nor by Devlin.

Devlin even explicitly argues that information is something that is abstract and can have no physical existence whatsoever [35, p.48]. This does not immediately imply that information is considered mind-dependent, but it can be assumed that if it is abstract, it either exists only in virtue of cognitive agents or it exists in some Platonic realm. In any case, Devlin argues that it has the same ontological status as mathematical objects like numbers.

We also see that Situation Theory places the focus more on agents that are aware of constraints and are thereby able to extract information from a situation. Dretske was focused more on the idea of correlations that exist independently of us and could thereby produce information.

What about LFD? This logic is not explicit about the ontological status of semantic information. It also allows different readings of the operators, treating both information as range and information as correlation.

If one thinks that information is mind-independent, LFD could be a logic that allows to anlyze such information. I showed that we can read the dependence operators $D_x y$ such that variable x *carries* information about variable y. This interpretation leaves open whether this variable carries this information independently of any cognitive agent or not.

We can think again of the example of eating blueberries. From the fact that someone's mouth is blue we can infer that they just ate blueberries. We can thus say that their blue mouth carries information about whether they ate blueberries. One could argue that this is so independently of whether some cognitive agent is actually able to use this information. It can also be argued, however, that independently of any agents, the only thing that is "out there" in the world is the dependency. There is a dependency between someone's mouth being blue and them having eaten blueberries. It is only because we are attuned to this dependency, this correlation, that we say that one thing carries information about the other. From this perspective, saying that one thing "carries" information about another thing, is merely a manner of speaking that leaves out the fact that it only carries this information with respect to us.

We can thus say that LFD is neutral on the ontological status of information and can be used for different purposes.

Chapter 5

The relation between semantic information and Shannon information

This will be the final and concluding chapter of this thesis. I will here revisit the main question and formulate an answer based on what we have seen in the previous chapters.

My main question was whether Shannon information measures a type of information that is different from semantic information. As noted in the introduction, there seems to be a dichotomy between semantic information and Shannon information where Shannon information is related to a type of physical information that is not semantic. Shannon information is depicted as objective, quantitative and mind-independent and semantic information as subjective, qualitative, mind-dependent. I will here directly address the question of to what extent this depiction is correct.

5.1 Shannon information

What is "Shannon information"?

In the first chapter, we saw that Shannon information is a concept that originates from Shannon's MTC. Shannon proposed a measure that he first considered to call "uncertainty" or "information", but eventually chose to call "entropy" based on Von Neumann's advice. This measure is what led to the concept of Shannon information. We saw that mathematically, Shannon's measure can be applied to a random variable and measures its predictability. If it is predictable and the Shannon information is low, then we have little uncertainty about the random variable. High Shannon information can thus be associated with high uncertainty on our part. We argued, against Timpson [89], in support of this uncertainty interpretation.

In information theory, the random variable represents an information source. The values that this variable can take represent the possible outputs of this information source. Hence, in this context, Shannon's measure can also be considered to measure the compressibility of the messages from the information source. If the random variable is predictable, this means that we can use shorter codes for regular symbols and thereby use less code to send the messages over a channel to the destination.

In Chapter 2, we saw that Shannon's measure can also be applied in a different context, namely in physics. We saw that the same measure can be used to measure entropy. The values of the random variable then represent the possible microstates of a physical system that are consistent with the macrostate that we can observe.

In Chapter 3 and 4, we also saw that concepts from Shannon's theory can in fact be applied to measure some aspects of semantic information, even though it seemed that this was not inherently possible with Shannon's measure in Chapter 1.

Distinguishing the formal object from the application

What we can take away from this, and what Lombardi [71] also argued for, is that we can and must separate the mathematical measure of Shannon information, i.e. the function described in 1.1, from the interpretation of this measure. If we equate the term "Shannon information" with the mathematical measure, then we must say that Shannon information in itself is just an abstract formal concept. Thus, it does not make sense to check whether Shannon information *tout court* is concrete, physical, semantic, etc. We must consider a specific application of Shannon measure to a certain domain and check whether in this context it quantifies something that satisfies these properties.

We will now first reconsider our conclusions concerning the nature of Shannon information in the context of physics and communication engineering.

Concreteness

I argued similarly to Timpson [89], that nor in the context of communication engineering, nor in the context of physics, Shannon information quantifies something concrete.

In communication engineering, the idea existed that Shannon information quantifies a concrete substance that flows from the source to the destination over the communication channel. We, however, argued that the content of the messages emitted from the source, i.e. the pieces of information, can never be concrete. Information is that which can be contained in a concrete representation of a symbol. This concrete instance is the token and the information is the type. Pieces of information are thus always abstract.

In physics, the quantity of Shannon information was considered concrete because it is used to measure the entropy of a system. Since it is a sensible question in physics to ask "How much Shannon information is contained in region x?", it can seem as if Shannon information is something that is concretely present in space-time, that it is something that we can in principle interact with. This is, however, not the case because entropy is, although it may be a physical property, not something concrete.

Physicality

We next argued that Shannon information *can* be considered to quantify a physical property of an object or system, both in the context of physics and in the context of communication engineering. I argued that a property can be considered to be physical if it is described by physics or natural science. Thus, if we see that some property is observable or is related to observable properties, we can take this property to be physical.

I first argued that entropy can be considered an application of the measure of Shannon information. It is the same measure as the one used in information theory but it is interpreted in a different way. The probability distribution now ranges over the possible microstates that a system can be in relative to the macrostate. I thus argued that if Shannon information is used as entropy, i.e. if the probability distribution ranges over possible microstates of a physical system, it is indeed physical. This is the case because it is related to properties that are observable like temperature.

I also argued, however, that Shannon information when it is not applied to microstates but to computational states, can also be considered physical. We saw that Landauer's principle states that if information is erased, heat is emitted. This was explained as follows: We can take a physical object which can be in either of two computational states: 0 or 1. We do not know what computational state this object is in, so the amount of Shannon information of this object is 1. If we now reset this physical object to state 0, without checking which state it started in, we see that the entropy in the physical object must increase. This was the case because in the original state, for each computational state there are two possible microstates consistent with it. When we reset the object to computational state 0, there is only one computational state left. We saw that Liouville's theorem says that the overall possible number of microstates must remain constant, which means that since there are $2 + 2 = 4$ possible microstates in the original state, there will be also be 4 possible microstates consistent with the computational state 0 in the final state. This implies that the non-computational, thermodynamic entropy of the physical object will increase. This, in turn, will have an observable effect in terms of an increase in temperature or volume for example. This means that Shannon information can be considered to be a physical property if it is applied to measure the information contained in some physical object.

Mind-dependence and objectivity

In Chapter 2 I also argued that Shannon information can be considered objective but is still mind-dependent. This applies both to Shannon information applied in communication engineering and in physics.

I argued that it is an objective quantity, because it can be publicly agreed what the amount of Shannon information is in some object. I did note that Shannon information is not absolute, but relative. If we take some object and we ask "what is the Shannon information contained in this object?" there are several possible correct answers, depending on what we understand to be the possible states of the object, or the possible symbols of the information source. This and the probability distribution over these states determines the amount of Shannon information. However, in the context of communication engineering it is assumed that the random variable is determined beforehand. It is not taken to be part of the question "what is the amount of Shannon information in this object?" to find out what probability distribution we are actually considering. In the context of physics, when the probability distribution ranges over possible microstates consistent with the macrostate, it is known that we will consider the microstates of that object. Thus, it can be objectively agreed upon what the entropy of a system or object is.

I did argue, on the other hand, that Shannon information is *not* mind-independent. I explained why entropy, i.e. Shannon information in the context of physics, can be considered mind-dependent, because it depends on our knowledge of a physical system. I also argued that Shannon information in the context of communication engineering can be considered mind-dependent, because probability itself is a mind-dependent notion. If there would be no cognitive agent to observe the possible states or outputs of an information source, the Shannon information could not be determined.

5.2 Semantic information

What is semantic information?

In Chapter 3 and 4 we looked at the concept of semantic information and considered both quantitative and qualitative accounts. We saw that semantic information is information that has "aboutness", it is always about something. Signs that carry semantic information have some informational content. We saw that both linguistic information and environmental information can have informational content and therefore can be considered semantic information.

In all accounts, both quantitative and qualitative, we noted that one signal can contain different semantic information for different agents. Dretske pointed out that the information contained in a signal depends on the background knowledge we have about the possibilities at the source about which we get information. Floridi also argued that the level of abstraction and the goals of the agent determines what semantic information is contained in a signal. In Chapter 4 we saw that the qualitative definition of semantic information which I proposed in the language of Dynamic Epistemic Logic also assumes that the semantic information contained in a sentence can be different for different agents depending on their background knowledge. We also saw that Situation Theory puts a focus on the fact that agents need to be attuned to constraints that allow them to infer information from one situation about another situation. Thus, different agents will get different semantic information out of the same situation or signal.

Truthfulness

We saw that while some philosophers argue that semantic information must always be truthful, others hold that it does not. Bar-Hillel and Carnap [15] held that semantic information can also be false and even contradictory, whereas Floridi [42] and Dretske [38] argue that information must be truthful.

Quantitative measures only capture one aspect of semantic information

In Chapter 3, we considered three different ways to quantitatively measure semantic information. Bar-Hillel and Carnap and Dretske both propose to measure semantic information in terms of the surprisal value of the signal. If the probability of this signal is low, the information contained in the signal is high. We noted that in doing so, Bar-Hillel and Carnap and Dretske capture one specific aspect of semantic information. They measure how surprised we are to obtain some piece of information and are thus in line with the perspective that sees communication as the main function of information.

Floridi proposed to measure semantic information in terms of how much a signal tells us about the world. The amount of information contained in a signal is dependent on the degree of discrepancy of that sentence. Thus, this measure is more in line with the perspective that sees world description as the main function of information.

In Chapter 4, we saw that the qualitative definition of semantic information that can be formulated both in the framework of Dynamic Epistemic Logic and the Logic of Functional Dependence, captures semantic information in a more all-round and complete way. It formulates an answer to the question: "what is the semantic information contained in this signal?" in terms of a set of sentences that an agent will know after receiving the signal and which they did not know before. The quantitative measures that we saw in Chapter 3 can only measure one aspect of semantic information, whereas this set can capture more of the essence of semantic information.

Subjective

Since we saw in Chapter 3 and Chapter 4 that the semantic information contained in a signal can differ for different agents, I argued that semantic information is in principle subjective. However, it can be that agents will publicly agree upon their background knowledge which will mean that they will also publicly agree about what semantic information is contained in some signal. This means that in some case we can speak of the "objective semantic information" contained in some signal relative to some group of agents with the same background knowledge, or relative to a pre-determined set of assumptions.

Some signals, for example, carry some pieces of information for everyone, like smoke carries the information that there is fire, because basically every cognitive agent will be attuned to the correlation between smoke and fire. In this case we can say that smoke *objectively* contains the semantic information that there is fire. We also pointed out that we could say that sentences will objectively contain their own meaning as semantic information, because everyone who understands that language will get that information. In Chapter 4 we showed how we can incorporate the idea that a sentence always contains its own meaning as semantic information in the definition of semantic information.

We can compare this with what I argued with respect to Shannon information. I argued that Shannon information in the context of physics and communication engineering is objective, because in those contexts, it is assumed that the probability distribution is known beforehand. We also saw, however, that Shannon's surprisal value can be used to measure one aspect of semantic information, as was done by Dretske and Bar-Hillel and Carnap. I argued that in this context, this measure of individual Shannon information is subjective. Now that we have realized that we cannot consider Shannon information *tout court*, but must consider Shannon information in a specific application, this is not a contradictory result. Applied to semantics, Shannon's measure is in principle subjective and applied to communication engineering and physics we can say that it is objective.

When Shannon information is applied in the context of semantic information, it is standardly assumed that the probability distribution based on which the surprisal value will be calculated can in practice differ for different agents. In addition, the surprisal value will be calculated of all the informational content in some signal. Whether an agent recognizes something to be part of the informational content of a signal also depends on their background knowledge as pointed out before, which creates an additional layer of subjectivity.

Mind-dependent

With respect to the mind-dependency of semantic information, we learned that there are different possible perspectives. Dretske argued that information can have a semantics independently of any interpretative process. Dretske puts the focus on the idea that there are objective correlations out there in the world and argues that information is simply the consequence of a sequence of causal relations between objects. Floridi argued that information can have a semantics independently of an interpreter, if it is in principle possible to interpret the information.

We also saw in Chapter 4 that whereas Dynamic Epistemic Logic seems to fit better with the idea that information is mind-dependent, the Logic of Functional Dependence can be used for both perspectives. In Situation Theory, we saw that while it inherits some ideas from Dretske, it relays the focus on the agent and the constraints that the agent is attuned to.

I have argued at the end of Chapter 3 that I do not think that it is tenable to argue that semantic information is mind-dependent. Simply the fact that the semantic information contained in an object can differ for different agents seems to show that it is mind-dependent. I also argued that Dretske's recursive definition of knowledge is not clear enough to explain how information can have a semantics independently of any interpretative process.

5.3 Some final considerations

5.3.1 Shannon information and semantics

In Chapter 1, I argued that Shannon information does not have anything to do with the meaning of symbols. Nevertheless, in Chapter 3 we saw that concepts from Shannon's information theory are applied to measure semantic information. We saw that Bar-Hillel and Carnap and Dretske both use a measure of semantic information that is inspired by Shannon's individual surprisal value. We saw that if we think of the main function of information to be communication, these measures can be satisfactory in determining the amount of semantic information contained in some signal. At the end of Chapter 4, we also saw that mutual information can be used to measure how much semantic information we get about one variable by observing a specific value of the other variable. We also argued, however, that this measure of local dependence captures only semantic information from the communication perspective.

What about the function H itself? We have seen that the surprisal value and mutual information can be used to measure semantic information, but we have not explicitly considered the Shannon information measure itself. In Chapter 1, however, we saw that H can be considered to measure the average surprisal value of an output from an information source. We also argued that the uncertainty interpretation is a valid interpretation of H: the higher $H(X)$, the more uncertain we are about the value of X and thus the more information we can gain by learning its value.

This suggests that we could use H to measure the average semantic information contained in an announcemnt about the value of a variable, for example if we let this variable represent which drink is in the fridge. There can be a probability distribution ranging over milk, beer and coke and then we can measure the average semantic information contained in an announcement about which drink is actually in the fridge. The same thing would apply here, however, namely that it will only measure the semantic information content from the communication perspective.

So it seems that concepts from Shannon's theory *are* suitable to measure semantic information, but only one aspect of it. As the role of communication is an important aspect of semantic information, we are free to say that concepts of Shannon's theory can be used to measure semantic information.

Does this result contradict what I argued for in Chapter 1? If we phrase it carefully, it does not. It again has to do with the fact that we need to consider the context of application of Shannon information. In the context of communication engineering, Shannon information does not tell us anything about the meaning of the outputs of an information source. However, in the context of semantic information, we can see that Shannon information can tell us something about semantic information.

This is because, in the context of measuring semantic information, the surprisal value of the information is not just based on a probability that is determined solely based on frequency as in communication engineering, but that is determined based on the meaning of the information. We associate a certain probability with the information content of a signal based on how likely we think it is that it is true. Hence, here is where the difference lies. The same goes for the measure of local dependence: if we use it in a context where we do take into account the meaning of the different values of the variables, we can use it to measure semantic information.

Hence why we can say that in the context of semantics, concepts of Shannon's information theory can convey something about semantics.

5.3.2 Semantic information and physics

There is one thing we have not had a chance to consider so far, which is the relation between semantic information and physics. We saw that Shannon information can be used to measure aspects of semantic information, which suggests that Shannon information and semantic information are not so strictly divided from each other as we originally thought. It turns out that some physicists also talk about information as playing an important role in physics, and they seem to be referring to something other than entropy or merely Shannon information. In order to understand the relation between semantic information and Shannon information we will thus in this section attempt to clarify the place of semantic information in physics.

Informational ontologies

As already referred to in the introduction, there are philosophers and physicists who argue that information can be considered the most basic entity of the universe. We will therefore shortly consider some of these ontologies to see whether they are actually talking about semantic information.

Wheeler famously argues for "It from Bit" [94]. He describes this view as follows:

> Otherwise put, every **it** — every particle, every field of force, even the spacetime continuum itself — derives its function, its meaning, its very existence entirely — even if in some contexts indirectly — from the apparatus-elicited answers to yes or no questions, binary choices [52], **bits**. It from bit symbolizes the idea that every item of the physical world has at bottom — at a very deep bottom, in most instances — an immaterial source and explanation; that what we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and this is a **participatory universe**. [94, p. 310] (Emphasis made by Wheeler)

Wheeler thus argues that information is the most basic entity of the universe from which all physical things derive their existence. He, however, also suggests that information is immaterial. He understands information as answers to yes-no questions that are posed in experiments and stored in bits. So it seems that Wheeler does, in some sense, have a semantic, subjective type of information in mind. Long also argues that Wheeler has in mind a "subjective participatory-constructivist cosmology" [73, p.653].

Floridi has also argued for a type of informational ontology which he calls "Informational Structural Realism" (ISR) [42, 41]. He also argues that information is the basic structure of the universe. He writes:

> A significant consequence of ISR is that, as far as we can tell, the ultimate nature of reality is informational, that is, it makes sense to adopt LoAs [levels of abstraction] that commit our theories to a view of reality as mind-independent and constituted by structural objects that are neither substantial nor material (they might well be, but we have no need to suppose them to be so) but informational. [41, p.340]

Floridi here must be referring to semantic information, because he explains in his book that he defends a semantic view of information [42]. Floridi [42, p.345] argues that this type of structural realism will provide a midway between Epistemic Structural Realism and Ontic Structural Realism [61]. The epistemic view is that we can only get knowledge about the relations between objects, the structure. The ontic view is that there are no objects, only structure. Beni [20] has argued, however, that Floridi's metaphysics is more similar to Epistemic Structural Realism. In the end, as Long [73] argues, Floridi's form of structural realism is not that different from the view proposed by Wheeler.

Bruce Long [72] also argues for an informational ontology. He defends a type of physicalist Ontic Structural Realism. Whereas Floridi's account is not physicalist, Long's account is. Long also argues explicitly that information is semantic but also that information is physical. He argues that "information only exists physically" [72, p.20]. He means by this "that it necessarily existentially depends upon and reduces to physical spatiotemporal structures and structures that reduce to such structures". [72, p.20] He thus argues for a picture of information similar to Dretske, where physical structure and physical causal relations create semantic information. He thus also argues that semantic information is mind-independent.

Is semantic information physical?

Long argues that semantic information is physical, but to what extent can this be legitimately argued for? I would argue that semantic information cannot be physical and the reason is similar as for why I argued that semantic information is not mind-independent.

I argued that we should understand "physical" to mean that it is described by physics and thus that it must be related to observable quantities. Long argues that it means "that it necessarily existentially depends upon and reduces to physical spatiotemporal structures and structures that reduce to such structures" [72, p.20].

We can agree that semantic information does depend on physical structures, because they are required to carry the information. This was also one of the reasons why Landauer argued that information is physical [63]. One argument was Landauer's principle, but another argument he makes is that all information requires physical representation. We can take it that Landauer meant this in the sense that all information requires some concrete carrier. Floridi also argued that "most people agree that there is no information without (data) representation" [42, p. 42]. It indeed seems sensible that all information requires some sort of data representation.

Floridi distinguishes three types of data, of which the most basic version is called *data de re*, which he understands as the lack of uniformity in the world out there. We can thus understand data as being reducible to basic relations between objects, as physical structure. Indeed, also semantic information requires some type of data that can represent it. There is, however, a difference between semantic information requiring some sort of physical or concrete carrier and semantic information itself being physical.

Long argues that semantic information can be reduced to physical spatiotemporal structures. For the same reasons as why I argued that semantic information cannot be mind-independent, I also argue that it cannot be reduced to physical structures. Different agents can get different semantic information from the same signal because the agents are required to be attuned to correlations that relate the signal to its content in order to get any information. It seems that those who argue that semantic information is physical are actually arguing that data are physical and that causal relations are physical. However, semantic information is what is recognized by a cognitive agent that is attuned to these causal relations.

I thus agree with what Timpson [89] said about the physicality of semantic information, which he refers to as the "everyday concept of information". He argued that it would be baffling to say that the everyday concept of information is physical "because the everyday concept of information is reliant on that of a person who might read or understand it, encode or decode it, and makes sense only within a framework of language and language users; yet it is by no means clear that such a setting may be reduced to purely physical terms."[89, p.2]

How does this relate to the fact that Shannon information is physical?

I argued in this thesis that Shannon information *can* be considered physical, because it relates to observable quantities as described by Landauer's principle. This principle stated that if information is erased, heat is emitted.

I would argue that this does not concern the erasure of semantic information, however. The scenario we considered was the erasure of 1 bit of Shannon information from a physical object. This erasure was performed by setting the object to one computational state without regarding its initial state. However, the informational content of the bit of information that was erased was not in any way considered. Different agents could have recognized completely different semantic information in the bit. It was uninterpreted information that was erased. We could also refer to this simply as "data". One bit of data was erased and therefore the thermodynamic entropy of the physical system increased.

This brings us to the idea that we can refer to Shannon's measure as measuring data, not information. We already saw that Shannon's measure measures compressibility, just as algorithmic information does. It measures how many bits on average are required to code a message coming out of an information source. This fits with the idea that Shannon's measure measures "data". Both Floridi [42] and van den Herik [53] have also argued that we should keep the term "information" as referring only to semantic information and that information-theoretic measures should be considered to measure data, not information. Van den Herik proposes the following: "If the aim of a certain theory or field is not to talk about the semantic properties of data, the usage of 'information' can almost certainly be replaced with 'data'." [53, p.12]

In relation to what we said earlier, that we must always consider the context in which we apply Shannon information, we could say that in the context of communication engineering, i.e. when the values of a variable represent different possible computational states, Shannon information measures data. When the values represent different possible microstates, Shannon information measures entropy. And when the values represent different interpreted epistemic possibilities, Shannon information, can be used to measure an aspect of semantic information, namely the amount of surprise contained in a symbol.

5.4 Conclusion

After some of the latter new considerations concerning Shannon information and semantic information, we will do one final recap of what we have found out in this thesis.

My main question was whether Shannon information describes a type of information that is different from semantic information. We originally started with the assumption that there is a dichotomy between Shannon information and semantic information: Shannon information is physical, objective, quantitative, mind-independent and non-semantic and semantic information is subjective, qualitative, mind-dependent and non-physical.

I argued that Shannon information is a mathematical measure with different possible applications. I suggested that we can think of Shannon information in communication engineering as measuring data. In physics, Shannon information is used to measure entropy. In both cases, the measure indeed quantifies a physical property. I also argued that it is an objective measure and evidently it is quantitative. I argued that the measure is nevertheless mind-dependent, contrarily to our assumption. We also saw that, contrarily to the initial characterization, concepts of Shannon's information theory can be used to measure one aspect of semantic information.

I argued that semantic information is indeed subjective and mind-dependent. I also showed that there are both quantitative and qualitative ways to analyze information, but argued that the quantitative measures only capture one aspect whereas a qualitative definition captures the essence of semantic information. We saw that semantic information does play a role in physics, but I argued that it does not have to be considered "physical".

Concerning the question whether Shannon information and semantic information are different types of information we can thus say the following: if we contrast semantic information with Shannon information as applied in physics or communication theory, we can say that they are really different concepts than semantic information. However, we can also use Shannon information or related concepts to measure semantic information, which means that in that case they would be closely aligned, though not the same, because Shannon information can only measure one aspect of semantic information.

When it comes to terminology, one could say that semantic information is the type of information that we think of when we think of the term "information". Shannon information, on the other hand, is an uninterpreted mathematical measure that can be used for different purposes. It can be used to measure data and entropy, and it can also be used to quantify aspects of semantic information.

Bibliography

[1] C. Adami. *The Physics of Information.* 2004.

[2] P. Adriaans and J. van Benthem. "Introduction: Information is what Information Does". In: *Philosophy of Information.* Vol. 8. Elsevier, 2008, pp. 3–26. ISBN: 9780444517265. DOI: 10.1016/B978-0-444-51726-5.50006-6.

[3] M. C. I. Alistair. "The Semantics Latent in Shannon Information". In: *British Journal for the Philosophy of Science* 70.1 (2019), pp. 103–125. DOI: 10.1093/bjps/axx029.

[4] W. P. Alston. "Dretske on Knowledge". In: *Behavioral and Brain Sciences* 6.1 (1983), pp. 63–64. DOI: 10.1017/s0140525x00014643.

[5] H. Andréka, I. Németi, and J. van Benthem. "Modal Languages and Bounded Fragments of Predicate Logic". In: *Journal of Philosophical Logic* 27.3 (1998), pp. 217–274. ISSN: 00223611, 15730433.

[6] J. Anta. "Can Information Concepts Have Physical Content?" In: *Perspectives on Science* 1.1 (2022), pp. 1–21.

[7] J. Anta. "Make Information in Science Meaningful Again". In: *Logos and Episteme: An International Journal of Epistemology* 3 (2021), pp. 263–286. DOI: 10.5840/logos-episteme202112321.

[8] P. W. Atkins. *The Second Law.* Scientific American Library, 1994. ISBN: 9780716750055.

[9] A. Baltag and S. Smets. "Correlated Knowledge: An Epistemic-Logic View on Quantum Entanglement". In: *International Journal of Theoretical Physics* 49.12 (2010), pp. 3005–3021. ISSN: 00207748. DOI: 10.1007/s10773-010-0411-5.

[10] A. Baltag and J. van Benthem. "A Simple Logic of Functional Dependence". In: *Journal of Philosophical Logic* 50.5 (2021), pp. 939–1005. ISSN: 15730433. DOI: 10.1007/s10992-020-09588-z.

[11] A. Baltag and L. S. Moss. "Logics for Epistemic Programs". In: *Synthese* 139.2 (2004), pp. 165–224. ISSN: 1573-0964. DOI: 10.1023/B:SYNT.0000024912.56773.5e.

[12] A. Baltag and B. Renne. "Dynamic Epistemic Logic". In: *The Stanford Encyclopedia of Philosophy.* Ed. by E. N. Zalta. Metaphysics Research Lab, Stanford University, 2016.

[13] A. Baltag and S. Smets. "Correlated Information: A Logic for Multi-Partite Quantum Systems". In: *Electronic Notes in Theoretical Computer Science* 270.2 (2011), pp. 3–14. ISSN: 1571-0661. DOI: `https://doi.org/10.1016/j.entcs.2011.01.020`.

[14] A. Baltag and S. Smets. "Dynamic belief revision over multi-agent plausibility models". In: *Proceedings of LOFT* (2006).

[15] Y. Bar-Hillel and R. Carnap. "Semantic Information". In: *British Journal for the Philosophy of Science* 4.14 (1953), pp. 147–157. DOI: `10.1093/bjps/IV.14.147`.

[16] J. Barwise. "Information and Semantics". In: *Behavioral and Brain Sciences* 6.1 (1983), pp. 65–65. DOI: `10.1017/s0140525x00014679`.

[17] J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, 1983.

[18] C. V. F. Bas. *The Scientific Image*. Oxford, England: Oxford University Press, 1980.

[19] A. Ben-Naim. *A Farewell to Entropy: Statistical Thermodynamics Based on Information : S*. World Scientific, 2008. ISBN: 9789812707062.

[20] M. D. Beni. "Epistemic Informational Structural Realism". In: *Minds and Machines* 26.4 (2016), pp. 323–339. ISSN: 1572-8641. DOI: `10.1007/s11023-016-9403-4`.

[21] J. van Benthem. *Logical dynamics of information and interaction*. Cambridge: Cambridge University Press, 2011. ISBN: 9780521765794.

[22] J. van Benthem and M. Martinez. "The Stories of Logic and Information". In: *Philosophy of Information* November (2008), pp. 217–280. DOI: `10.1016/B978-0-444-51726-5.50012-1`.

[23] S. J. Blundell and K. M. Blundell. *Concepts in Thermal Physics*. 2nd ed. Vol. 9780199562091. Oxford: Oxford University Press, 2010, pp. 1–512. ISBN: 9780191718236. DOI: `10.1093/acprof:oso/9780199562091.001.0001`.

[24] L. Boltzmann. *Wissenschaftliche Abhandlungen*. Ed. by F. Hasenöhrl. Vol. 1. Cambridge Library Collection - Physical Sciences. Cambridge University Press, [1909] 2012. DOI: `10.1017/CBO9781139381420`.

[25] E. Botchkina. "Issues in Objectivity and Mind-dependence". PhD thesis. Massachusetts Institute of Technology, 2010.

[26] H. B. Callen. *Thermodynamics and an introduction to thermostatistics; 2nd ed.* New York: Wiley, 1985.

[27] R. Carnap. *Logical Foundations of Probability*. Chicago, IL, USA: Chicago University of Chicago Press, 1950.

[28] R. Carnap and Y. Bar-Hillel. "An Outline of a Theory of Semantic Information". In: *Journal of Symbolic Logic* 19.3 (1954), pp. 230–232. DOI: `10.2307/2268645`.

[29] F. B. Chedid. "Kolmogorov Complexity and Information Content". In: (2017), pp. 1–12.

[30] R. Clausius. *The Mechanical Theory of Heat*. John van Voorst, 1865.

[31] T. M. Cover. *Elements of information theory*. 2nd ed. A Wiley-Interscience publication. Hoboken, N.J: Wiley-Interscience, 2006. ISBN: 0471241954.

[32] S. D'alfonso. "On quantifying semantic information". In: *Information (Switzerland)* 2.1 (2011), pp. 61–101. ISSN: 20782489. DOI: 10.3390/info2010061.

[33] K. Denbigh. "How subjective is entropy". In: *Chemistry in Britain* 17 (1982), pp. 168–184.

[34] K. Denbigh and J. Denbigh. *Entropy in relation to incomplete knowledge*. Cambridge University Press, 1985. ISBN: 0521256771.

[35] K. J. Devlin. *Logic and information*. Cambridge University Press, 1991. ISBN: 0521410304. DOI: 10.1016/0160-9327(92)90011-d.

[36] D. Dieks. "Is There a Unique Physical Entropy? Micro Versus Macro". In: *New Challenges to Philosophy of Science*. Ed. by H. Andersen et al. Springer Verlag, 2012, pp. 23–34.

[37] H. van Ditmarsch. *Dynamic epistemic logic*. Synthese library, vol. 337. Dordrecht: Springer, 2007. ISBN: 9781402058387.

[38] F. I. Dretske. *Knowledge and the Flow of Information*. MIT Press, 1981.

[39] R. Fagin. *Reasoning about knowledge*. Cambridge, Massachusetts: MIT Press, 2004. ISBN: 0-262-25609-6.

[40] L. Floridi. "Open Problems in the Philosophy of Information". In: *Metaphilosophy* 35.4 (2004), pp. 554–582. DOI: 10.1111/j.1467-9973.2004.00336.x.

[41] L. Floridi. "Semantic Conceptions of Information". In: *Stanford Encyclopedia of Philosophy*. 2008.

[42] L. Floridi. *The Philosophy of Information*. Oxford University Press, 2011.

[43] M. P. Frank. "Physical Foundations of Landauer's Principle". In: *Reversible Computation*. Ed. by J. Kari and I. Ulidowski. Cham: Springer International Publishing, 2018, pp. 3–33. ISBN: 978-3-319-99498-7.

[44] R. Frigg and C. Werndl. "Can Somebody Please Say What Gibbsian Statistical Mechanics Says?" In: *British Journal for the Philosophy of Science* (2018), pp. 1–27. DOI: 10.1093/bjps/axy057.

[45] J. Gerbrandy and W. Groeneveld. "Reasoning about Information Change". In: *Journal of Logic, Language and Information* 6.2 (1997), pp. 147–169. ISSN: 1572-9583. DOI: 10.1023/A:1008222603071.

[46] J. W. Gibbs. *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. Cambridge Library Collection - Mathematics. Cambridge University Press, [1902] 2010. DOI: 10.1017/CBO9780511686948.

[47] D. Gillies. "Varieties of Propensity". In: *British Journal for the Philosophy of Science* 51.4 (2000), pp. 807–835. DOI: 10.1093/bjps/51.4.807.

[48] S. Goldstein et al. "Gibbs and Boltzmann Entropy in Classical and Quantum Mechanics". In: *arXiv: Statistical Mechanics* (2019).

[49] J. Groenendijk and M. Stokhof. "Dynamic predicate logic". In: *Linguistics and Philosophy* 14.1 (1991), pp. 39–100. ISSN: 01650157. DOI: 10.1007/BF00628304.

[50] A. Grove. "Two Modellings for Theory Change". In: *Journal of Philosophical Logic* 17.2 (1988), pp. 157–170. ISSN: 00223611, 15730433.

[51] P. Grünwald and P. Vitányi. "Kolmogorov Complexity and Information Theory". In: *Journal of Logic, Language and Information* 12 (2003), pp. 497–529.

[52] J. Hale. "Information-theoretical Complexity Metrics". In: *Language and Linguistics Compass* 10.9 (2016), pp. 397–412. ISSN: 1749818X. DOI: 10.1111/lnc3.12196.

[53] J. van den Herik. "Is Information Out There?" In: *Erasmus Student Journal of Philosophy* 5.2 (2013), pp. 20–33.

[54] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca: Cornell University Press, 1962.

[55] D. J. Israel and J. Perry. "What is Information?" In: *Information, Language and Cognition*. Ed. by P. P. Hanson. University of British Columbia Press, 1990.

[56] E. T. Jaynes. "Gibbs vs Boltzmann Entropies". In: *American Journal of Physics* 33.5 (1965), pp. 391–398. ISSN: 0002-9505. DOI: 10.1119/1.1971557.

[57] E. T. Jaynes. "Information Theory and Statistical Mechanics". In: *Phys. Rev.* 106 (1957), pp. 620–630. DOI: 10.1103/PhysRev.106.620.

[58] A. N. Kolmogorov. "Three approaches to the quantitative definition of information". In: *International Journal of Computer Mathematics* 2.1-4 (1968), pp. 157–168. DOI: 10.1080/00207166808803030.

[59] R. Krzanowski. "Why Can Information Not Be Defined as Being Purely Epistemic?" In: *Philosophical Problems in Science* 68 (2020), pp. 37–62.

[60] J. Ladyman. "Intension in the Physics of Computation: Lessons from the Debate about Landauer's Principle". In: *Physical Perspectives on Computation, Computational Perspectives on Physics*. Landauer. Cambridge University Press, 2018, pp. 219–239. DOI: 10.1017/9781316759745.011.

[61] J. Ladyman. "What is Structural Realism?" In: *Studies in History and Philosophy of Science Part A* 29.3 (1998), pp. 409–424. DOI: 10.1016/s0039-3681(98)80129-5.

[62] J. Ladyman et al. "The connection between logical and thermodynamic irreversibility". In: *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics* 38.1 (2007), pp. 58–79. ISSN: 13552198. DOI: 10.1016/j.shpsb.2006.03.007.

[63] R. Landauer. "Information is a physical entity". In: *Physica A: Statistical Mechanics and its Applications* 263.1 (1999). Proceedings of the 20th IUPAP International Conference on Statistical Physics, pp. 63–67. ISSN: 0378-4371. DOI: https://doi.org/10.1016/S0378-4371(98)00513-5.

[64] R. Landauer. “The physical nature of information”. In: *Physics Letters, Section A: General, Atomic and Solid State Physics* 217.4-5 (1996), pp. 188–193. ISSN: 03759601. DOI: 10.1016/0375-9601(96)00453-7.

[65] D. K. Lewis. “A Subjectivist?s Guide to Objective Chance”. In: *Studies in Inductive Logic and Probability, Volume II*. Ed. by R. C. Jeffrey. Berkeley: University of California Press, 1980, pp. 263–293.

[66] D. K. Lewis. *Convention: a Philosophical Study*. Vol. 11. 2. Cambridge, Mass: Harvard University Press, 1970, pp. 14–15. ISBN: 0674170253. DOI: 10.1111/j.1468-0149.1970.tb00042.x.

[67] G. N. Lewis. “The Symmetry of Time in Physics”. In: *Science* 71.1849 (1930), pp. 569–577. ISSN: 00368075, 10959203.

[68] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Texts in Computer Science. Cham: Springer International Publishing AG, 2019. ISBN: 3030112977.

[69] O. Lombardi. “Dretske, Shannon’s Theory and the Interpretation of Information”. In: *Synthese* 144.1 (2005), pp. 23–39. DOI: 10.1007/s11229-005-9127-0.

[70] O. Lombardi, S. Fortin, and C. López. “Deflating the deflationary view of information”. In: *European Journal for Philosophy of Science* 6.2 (2016), pp. 209–230. ISSN: 18794920. DOI: 10.1007/s13194-015-0128-7.

[71] O. Lombardi, F. Holik, and L. Vanni. “What is Shannon Information?” In: *Synthese* 193.7 (2016), pp. 1983–2012. DOI: 10.1007/s11229-015-0824-z.

[72] B. Long. “A Scientific Metaphysical Naturalisation of Information”. PhD thesis. University of Sydney, 2018.

[73] B. Long. “ISR is Still a Digital Ontology”. In: *Erkenntnis* 85.3 (2020), pp. 649–664. ISSN: 15728420. DOI: 10.1007/s10670-018-0041-5.

[74] B. Lundgren. “Does semantic information need to be truthful?” In: *Synthese* 196.7 (2019), pp. 2885–2906. ISSN: 15730964. DOI: 10.1007/s11229-017-1587-5.

[75] O. Maroney. “Information Processing and Thermodynamic Entropy”. In: *Stanford Encyclopedia of Philosophy* (2009), Stanford Encyclopaedia of Philos.

[76] M. T. E. McIrvine. “Energy and Information”. In: *Scientific American* 224 (1971), pp. 179–186.

[77] D. H. Mellor. *The Matter of Chance*. Cambridge University Press, 1971.

[78] R. V. Mises and H. Geiringer. “Probability, Statistics and Truth”. In: *Philosophy of Science* 26.4 (1959), pp. 387–388.

[79] W. C. Myrvold. “Statistical mechanics and thermodynamics: A Maxwellian view”. In: *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics* 42.4 (2011), pp. 237–243. ISSN: 13552198. DOI: 10.1016/j.shpsb.2011.07.001.

[80] I. Niiniluoto. *Truthlikeness*. Dordrecht: Reidel, 1987.

[81] G. Oddie. *Likeness to Truth*. Dordrecht and Boston: Reidel, 1986.

[82] J. Plaza. "Logics of Public Communications". In: *Synthese (Dordrecht)* 158.2 (2007), pp. 165–179. ISSN: 0039-7857.

[83] K. Popper. *A World of Propensities*. Thoemmes, 1990.

[84] H. Price. "Reviewed Works: Entropy in Relation to Incomplete Knowledge by K. G. Denbigh and J. S. Denbigh : The Physical Basis of the Direction of Time by H. D. Zeh". In: *British Journal for the Philosophy of Science* 42.1 (1991).

[85] H. Reichenbach. *The Theory of Probability*. Berkeley: University of California Press, 1949.

[86] T. C. Schelling. *the Strategy of Conflict*. Vol. 9. 3. Cambridge, MA: Harvard University Press, 1969, pp. 3–8. ISBN: 0674840305. DOI: `10.1002/j.2326-1951.1969.tb02529.x`.

[87] C. E. Shannon. "A Mathematical Theory of Communication". In: *SIGMOBILE Mob. Comput. Commun. Rev.* 5.1 (2001), pp. 3–55. ISSN: 1559-1662. DOI: `10.1145/584091.584093`.

[88] J. M. Smith. "The Concept of Information in Biology". In: *Philosophy of Science* 67.2 (2000), pp. 177–194. DOI: `10.1086/392768`.

[89] C. G. Timpson. *Quantum Information Theory and the Foundations of Quantum Mechanics*. Oxford University Press, 2013.

[90] J. B. M. Uffink. *Measures of uncertainty and the uncertainty principle*. 1990.

[91] J. Van Benthem, J. Van Eijck, and B. Kooi. "Logics of communication and change". In: *Information and Computation* 204.11 (2006), pp. 1620–1662. ISSN: 10902651. DOI: `10.1016/j.ic.2006.04.006`.

[92] P. M. B. Vitanyi. "Conditional Kolmogorov complexity and universal probability". In: *Theoretical computer science* 501 (2013), pp. 93–100. ISSN: 0304-3975.

[93] P. M. B. Vitányi. "Meaningful Information". In: *CoRR* cs.CC/0111053 (2001).

[94] J. A. Wheeler. "Information, Physics, Quantum: The Search for Links". In: *Proceedings III International Symposium on Foundations of Quantum Mechanics*. 1989, pp. 354–358.