Abstract

This paper relates the uncertainties involved in neural signal processing with the uncertainties in physical processes. It has been common practice to research the biological world of neural networks and the world of physical processes and phenomena independently of each other. As a matter of fact, biology and physics have developed into independent fields with only few interconnections. In this paper, however, I present a common information-theoretical model capable of explaining uncertainties in both worlds. In fact, I postulate that the uncertainties in the physical world follow from the uncertainties inherent in a biological neural system, and are basically given by the limitations of human perception. I measure entropy using a generalized form of Shannon’s entropy, which in different forms has established itself in physics and computer science. The foundation of my information-theoretical analysis of neural networks is the sigmoid function, which computer scientists typically use as threshold function to model natural neural networks. The sigmoid function, in combination with my generalized form of entropy, leads to a new theory for human neural information processing and perception. Based on this theory, I provide a neural analogy to the Heisenberg Uncertainty Principle that in its most common form states that it is not possible to simultaneously determine the position and momentum of a particle.

Keywords: Information Theory, Entropy, Sensor Fusion, Machine Learning, Neural Networks, Perception, Relativity, Heisenberg Uncertainty Principle.
1 Introduction

A well-known conundrum of our times is that the mere observation of a physical experiment can change its outcome. The classical example for this fact is the famous double slit experiment, which consists of letting particles, such as photons or electrons, diffract through two slits on a screen [13, 14]. This produces an interference pattern with light and dark regions corresponding to the locations where the particle waves interfere constructively and destructively. The interference pattern disappears if we try to determine which slit the particle passes through. In other words, the interference pattern disappears as soon as we know which way the particle passes through the slits. Physicists take this as evidence of the “wave-particle” duality, which holds that light and matter exhibit properties of both waves and particles. Our perception of the way of the particle leads to a collapse of the wave function, letting the photon or electron behave like a particle and not like a wave.

The baffling result of the double-slit experiment strongly suggests that any ultimate theory explaining these major physical phenomena will have to take the human observer into account. It is therefore very likely that we need to couple an appropriate model of perception with our existing physical models in order to explain these effects and come to a comprehensive understanding of the world we live in; as far as this is theoretically possible. The following brings a familiar neural activation function in connection with Heisenberg’s uncertainty principle and Einstein’s special theory of relativity. Information theory will serve as the bridge between the world of perception, as it is defined by neural activity, and the realm of the physical world with its uncertainty and relativity. Given the importance it has achieved in so many apparently very different fields such as physics and computer science, information theory seems in fact to be a good candidate for a unifying framework. In the form of entropy, information was introduced in the early 1860s by the German physicist Rudolf Clausius to mathematically account for the dissipation of energy in thermodynamic systems [9]. Later, in 1948, Shannon introduced entropy into computer science in order to calculate the uncertainty or randomness for probabilistic events [12, 10]. Since then, information theory has continued to play a crucial role in
both fields [9].

The paper is structured as follows: Section 2 introduces the information-theoretical basis and assumptions. Section 3 then makes the connection with the sigmoidal activation function often used in artificial neural networks. In Section 4, the assumptions made in Section 2 will manifest themselves as the golden ratio. Section 5 shortly explains the main assertions of Einstein’s special theory of relativity before it follows them logically from the proposed information-theoretical framework. Likewise, Section 6 first summarizes shortly the main statement of Heisenberg’s uncertainty principle and then shows how we can derive a similar principle from the framework proposed here. Finally, a conclusion with the main findings closes the paper.

2 Open World Assumptions

Dealing with uncertainty is a very important task in many disciplines of computer science, such as signal theory, pattern recognition, artificial intelligence, etc. In all these disciplines entropy has gained its proper place as a formalism for modeling uncertainty. According to Shannon, the entropy $H$ of a discrete random variable with $n$ possible outcomes is, apart from a multiplying factor, the sum of the expected information $K_i$ for each outcome $i$ [12]:

$$H = \sum_{i=1}^{n} K_i$$  \hfill (1)

Let us concentrate on the individual $K_i$ and disregard the subscript $i$ for the time being. Using the negative logarithm as a measure for information, and multiplying the measured information with the probability, gives us the expected information $K$:

$$K = -p \ln(p)$$  \hfill (2)

So far, this has been standard information-theoretical formalism. A closer look reveals that both appearances of $p$ in Eq. 2 play different roles, though obviously their nominal values are the same. Labeling the first $p$ as $p^E$ and resolving for the second $p$ leads to the following relationship for $p^E \neq 0$:

$$K = -p^E \ln(p)$$  \hfill (3)

$$\iff p = e^{-\frac{K}{p^E}}$$  \hfill (4)
We see that the observed probability \( p \) from which we measure information is the residual part of an exponential distribution; i.e. \( 1 - e^{-\frac{K}{pE}} \), with expectation value \( pE \). This result is not only interesting from a theoretical standpoint, it also motivates practical machine learning methods, as e.g. in [5].

We now come to an important assumption of this paper that will be the basis of all future derivations, namely the open world assumption. It states that we cannot perceive the entire world around us and that there is a world behind the realm of our perception or knowledge. Our sensory organs allow us to capture only a small part of the world that we perceive as closed and complete. As a consequence, we set everything that we observe into relation to this closed world we perceive or know, while the real world is open and can be much larger than our closed world. This also includes the complement of any event we observe; we can only set it into relation to the world we know and not to the entire world, which we only partly know in general. In fact, the complement of our world is also part of our world and does not overlap with the world we do not know. Furthermore, the observation of an event automatically entails that we, as the observer, are part of the event’s complement. For instance, if we look at something nice, we automatically consider ourselves ugly in whatever aspect we think is nice. Additionally, the open world assumption in the information-theoretical context introduced here, also asserts that, given two complementary events, we will always perceive the one with the smaller probability; i.e. the event with the higher information. We will always set its probability into relation to the higher probability of its complement, which represents the “known” world to us. For instance, when looking at the starry night sky, it will always be the brightest star that catches our eye first and not one of the many less bright stars.

According to the open world assumptions, the probability \( p \) used to measure information in Eq. 3 becomes the fraction \( \frac{1-p}{p} \); with \( 1 - p \leq p \) or equivalently \( 0 \leq \frac{1-p}{p} \leq 1 \), and thus \( p \in \left[ \frac{1}{2}, 1 \right] \). Inserting this fraction into Eq. 3 provides the following new formula for the expected information:

\[
K = -pE \ast \ln \left( \frac{1 - pE}{pE} \right)
\] (5)

The notation \( pE \) is used instead of \( p \) to indicate that \( p \) is also an expectation value. The
parameter \( p^E \) now denotes the true probability in the open world, while the fraction \( \frac{1 - p^E}{p^E} \) stands for the probability as we measure, or rather perceive, it in our closed world. Readers familiar with information-theory will recognize the similarity between the term on the right-hand side of Eq. 5 and the term used in the Kullback-Leibler divergence, which measures the difference between two probability distributions. However, instead of computing the difference between two independent distributions, we look here at the difference, or self-similarity, between an event and its complement.

3 Sigmoid Function

Naturally, one is interested in the open world’s true probability \( p^E \) and how one can compute it from the observed probability. Knowing the true probability \( p^E \) allows to compute the actual open world information according to Eq. 3. We can derive the true probability by resolving Eq. 5 for \( p^E \), which produces the following result:

\[
K = -p^E \ln \left( \frac{1 - p^E}{p^E} \right)
\]

\[
\iff p^E = \frac{1}{1 + e^{-\frac{K}{p^E}}}
\]

It shows that there is a direct relationship of \( p^E \) to a very common type of artificial neural networks. In particular, \( p^E \) defines the well-known sigmoid function that virtually all feedforward networks implement as output function for their neurons [1, 3]. The traditional explanation for the use of this particular function has always been its non-linearity and simple derivative. Non-linearity increases the expressiveness of a neural network, allowing decision boundaries in feature space that a simple linear network would not be able to model. On the other hand, a simple derivative facilitates backpropagation of errors during the training phase. While these are surely important points, the result of Eq. 6 shows that the sigmoid function has a deeper meaning, which is of an information-theoretical nature. It in fact computes the true probability \( p^E \) of the open world.

The sigmoid function is the result of a long evolution process in neural network theory that began with simple linear models and eventually lead to the non-liner models of today [7, 8, 11]. Nowadays, it has become the mainstream activation function for
networks of the feedforward type, providing high performance in various classification problems. Figure 1 depicts the sigmoid function for three different expectation values $p^E$, namely $1, \frac{1}{2},$ and $\frac{1}{3}$. As its name already suggests, the sigmoid function has an S-shape. It converges on 0 towards negative infinity, and it converges on 1 towards infinity. The expectation $p^E$ controls the steepness of the sigmoid function. For smaller values of $p^E$, the sigmoid function becomes steeper and approaches faster to either 0 or 1 on both ends. Independently of $p^E$, the sigmoid function is always 0.5 for $K = 0$.

4 Golden Ratio

In view of the fundamental difference between open and closed world, legitimate questions to ask are: Can we observe the true probability of the open world in our closed world and, if yes, when does the measured probability equal the true probability? The answer to these questions lies in the following relationship that the true probability $p^E$ has to meet in order for us to be able to measure it.

\[ p^E = \frac{1 - p^E}{p^E} \quad (T) \]

\[ \iff p^E \approx 0.618... \quad (\text{or} \quad p^E \approx -1.618...) \]

There are exactly two possible values satisfying Eq. 7, namely $\Phi \approx -1.618$ and $\varphi \approx 0.618$. These values define the well-known golden ratio, which some authors also call golden mean. Hence, according to the open world assumption, the golden ratio defines points...
The golden ratio is an irrational number, or rather two numbers, describing the proportion of two quantities [6, 4]. Expressed in words, two quantities are in the golden ratio to each other, if the whole is to the larger part as the larger part is to the smaller part; with the whole being the sum of both quantities. Figure 2 shows an example of a line divided into two segments that are in the golden ratio to each other. Historically, ancient mathematicians already began studying the golden ratio. It plays an important role in different fields like geometry, biology, physics, and others. Many artists and designers make use of it, either deliberately or unconsciously, because it seems that artwork based on the golden ratio has an esthetic appeal, featuring a kind of natural symmetry. Despite the fact that the golden mean is of paramount importance to so many fields, it is fair to say that we still do not have a full, or rather correct, understanding of its true meaning in science. The open world assumption proposed here offers a new information-theoretical explanation for the golden ratio.

5 Relativity

As shown in the last section, we can measure the true probabilities of the open world in points defined by the golden ratio. Let us now go one step further and use this result to make statements about how we perceive the open world in our limited closed world. According to Eq. 7, we can define the golden ratio by various equivalent relations. In particular, we can perform the following straightforward transformations of Eq. 7 to obtain equivalent definitions of the golden ratio:

\[ p^E = \frac{1 - p^E}{p^E} \]
\[\iffalse \implies p^{E^2} = 1 - p^E \tag{8}\] 
\[\iffalse \implies p^E = 1 - p^{E^2} \tag{9}\]

For instance, if we measure \(1 - p^E\) on the right-hand side of Eq. 8 as a probability in our closed world, then it follows from the open world assumption that the corresponding probability in the open world is in fact \(p^{E^2}\). We can enumerate all possible closed world observations and their open world equivalents by inserting the result of Eq. 9 into Eq. 5 and applying an algebraic transformation:

\[K = -p^E \ln(1 - p^{E^2}) \tag{10}\]
\[\iffalse \implies \frac{K}{2} = -p^E \ln(\sqrt{1 - p^{E^2}}) \tag{11}\]

Implementing the Pythagorean theorem, Eq. 11 illustrates that the probabilities of the closed world and open world form two different sides of a right-angled triangle with a hypotenuse of Length 1, respectively. Each possible measurement therefore corresponds to a point on the unit circle. The open world probability, which is the term outside the logarithm, and the closed world probability, which is the term inside the logarithm, are sine and cosine, respectively.

The exponential distribution has traditionally been used to model the time between statistical events that happen at a constant average rate, such as radioactive decay or the time until the next system failure. In this context, the expectation value of the exponential distribution means time, namely the expected time until the occurrence of the next event. This identity between expectation and time has a direct connection to Einstein’s theory of relativity [2]. The root expression in the logarithm of Eq. 11 becomes the well-known Lorentz factor, or rather the inverse Lorentz factor, when we replace \(p^E\) with \(\frac{v}{c}\); i.e., the velocity \(v\) of an object set in relation to the speed of light \(c\). Eq. 12 shows the result of this replacement.

\[\frac{K}{2} = -t \ln \left( \sqrt{1 - \left(\frac{v}{c}\right)^2} \right) \tag{12}\]

We see the open world expectation of Eq. 11 being replaced by the term \(t\), which stands for time, and the Lorentz factor appearing in the logarithmic expression. The Lorentz
factor plays a crucial part in Einstein’s special relativity. It describes how mass, length, and time change for a system whose velocity approaches the speed of light, which is the maximum possible velocity according to Einstein’s theory of relativity. With increasing speed, an observer will measure a shorter length, more mass, and a shorter lapse of time for a moving system (or inertial system to be precise). These effects become more pronounced as the moving object approaches the speed of light. Depending on the relative speed, the Lorentz factor describes the ratio of the quantity in the observer system to the quantity measured by the observer for the moving system. For instance, if $t$ is the time measured locally by the observer, then the corresponding time $t'$ measured for the moving system computes as follows:

$$t' = \sqrt{1 - \left(\frac{v}{c}\right)^2} \ast t \quad (13)$$

We can see that $t'$ converges to zero for increasing speed. The observer thus sees a dilation of time for systems approaching light speed. In particular, he measures no lapse of time for a system moving at the exact speed of light.

In the information-theoretical context of Eq. 12, the Lorentz factor in the logarithmic expression is the time measured in the closed world of the system’s observer, while $t$ is the time in the open world. In its role as the residual part of an exponential distribution, $p^E$ stands for a percentage of an observed quantity. The quantity can be any observable variable that an observer can realize up to its maximum value, which poses an upper limit, such as the speed of light. Another example is the error or recognition rate in a pattern recognition task, which an observer always measures in the range of 0% to 100%; and where a measurement of 0% indicates optimal performance, see e.g. [5].

6 Uncertainty

Starting out from the open world assumption, and assuming a sigmoidal activation function for neurons, the last section revealed an intrinsic connection between neural reasoning and a phenomenon of the physical world; i.e., Einstein’s special theory of relativity. This section shows that this is not the only relationship to the laws of physics. Given
the framework proposed here, we can observe a limitation in neural reasoning that is strongly reminiscent of Heisenberg’s uncertainty principle in quantum physics [13].

In the quantum mechanical world, the idea that we can locate objects exactly breaks down. Heisenberg’s uncertainty principle states that locating a particle in a small region of space makes the momentum of the particle uncertain; and conversely, that measuring the momentum of a particle precisely makes the position uncertain. For instance, let \( \Delta x \) be the uncertainty about the exact location of an electron, as measured for instance as the width of the slit in a double-slit type of experiment, and let \( \Delta p \) be the dispersion of its momentum. Then, Heisenberg’s uncertainty principle reads as follows: \( \frac{\hbar}{2} \leq \Delta x \times \Delta p \), where the symbol \( \hbar \) on the left-hand side of the inequality is Dirac’s constant, which is equal to Planck’s constant \( h \) divided by \( 2\pi \). This means that the combination of the error in position times the error in momentum is always greater than a positive constant. From this it follows that we cannot simultaneously find both the position and momentum of an electron, or an object in general, to arbitrary accuracy. The more precisely we determine the position of the electron, the less we will know about momentum; conversely, the more we know about the momentum, the less we will know about the position of the electron. This uncertainty of an electron’s position and momentum is established the moment it is observed, resulting in the measured values always being dispersed.

Due to its significance and appealing simplicity, the uncertainty principle is one of the most flamboyant formulas in physics. Moreover, when expressed in Planck units, the uncertainty principle looks even simpler. Planck units were introduced by the German physicist Max Planck to simplify existing physical laws and unify coexisting physical theories, such as general relativity and quantum physics, into a great unifying theory. Planck units are the result of normalizing several fundamental physical constants to 1, among which are light speed \( c \) and Dirac’s constant \( \hbar \), which then become natural units of measurement. In Planck units; i.e. with \( \hbar = 1 \), the uncertainty principle simplifies to:

\[
\frac{1}{2} \leq \Delta x \times \Delta p
\]

Interestingly enough, a small transformation of Eq. 11, which revealed the Pythagorean relationship between closed world probability and open world probability in Section 5,
leads to a similar inequation that allows us to infer comparable limitations for neural reasoning:

\[
\frac{K}{2} = -p^E \ln \left( \sqrt{1 - p^E^2} \right)
\]

\[
\implies \frac{1}{2} \leq \ln \left( \sqrt{1 - p^E^2} \right) \cdot \frac{1}{K} = \frac{I_{\text{closed}}}{I_{\text{open}}} \quad (15)
\]

Here, we see that the product of the information measured in our closed world and the inverse of the open world’s information \( K \) is always at least \( \frac{1}{2} \). However, we would like to see this product become 0, and in particular, we want \( I_{\text{closed}} \) to be equal to zero and \( I_{\text{open}} \) to reach infinity. The reason for this is twofold: First, we do not want our measurements to be flawed with uncertainty; hence \( I_{\text{closed}} = 0 \). Second, we want all measured probabilities in our closed world to be equal to the open world’s true probabilities; hence \( p^E = 1 \) and thus \( K = I_{\text{open}} = \infty \). Unfortunately, Eq. 15 tells us that this is not possible. We can either know the precise uncertainty in the open world, in which case we observe maximum uncertainty in our closed world; or, we can reduce the uncertainty in our closed world at the expense of measuring the open world’s uncertainty less accurately. Like we have seen for the uncertainty principle, it is only possible to minimize the ratio of the closed world information to the open world information up to the positive constant \( \frac{1}{2} \). Accordingly, we have to accept this limitation as we have to accept the uncertainty principle as an inherent limitation of our world.

7 Conclusion

Given the fact that we perceive the world through our sensory organs, which transform the physical input into neural signals to be processed by the nervous system, this paper suggested that prominent physical phenomena reflect in our neural signal processing. In fact, the underlying premise is that an appropriate neural model can explain all observable physical processes of the world we observe. The paper showed evidence that Einstein’s theory of relativity, in particular the dilation of time in special relativity, and a relationship reminiscent of Heisenberg’s uncertainty principle, which have been so far irreconcilable phenomenas on macroscopic and microscopic scale, manifest in the
information processing of nerve cells when assuming a sigmoidal activation function for nerve cells. Starting from an information-theoretical formalism not unlike the Kullback-Leibler divergence that distinguishes between the closed world of the observer and its surrounding open world, it became evident that the classical sigmoid function, which is the backbone of feedforward-type networks, computes the open world’s probability for any measurement performed in the closed world of the observer. A closer look revealed that two probabilities measured in the open world and closed world, respectively, coincide if they equal the golden ratio. Moreover, the definition of the golden ratio lead to the Lorentz factor, which is a corner stone in Einstein’s special theory of relativity and which introduces time dilation into the information-theoretical framework proposed, given that the expectation value describes the time that elapses between two events of a statistical process. Another important result that followed from the golden ratio is an inherent limitation on neural information processing for systems implementing the model introduced here. Similar to the Heisenberg uncertainty principle, we can only measure one of two quantities with infinite precision, while the other one will remain elusive. In particular, we cannot measure both the probability of an event in the observer’s closed world and its corresponding probability in the open world with infinite precision. We can only increase the precision of one of them at the expense of the precision of the other.

Continuing the train of thought from the beginning, where the double-slit experiment stood for a typical experiment whose outcome can be changed by mere observation, this paper should be understood as a contribution towards an unifying theory combining concepts from the physical word and ideas from cognitive science or machine learning. It seems very likely that only a combination of these hitherto completely unrelated fields can eventually lead to a comprehensive understanding of such bewildering physical phenomena as the outcome of the double-slit experiment.

References


