Check for updates

Probabilities and Shannon's Entropy in the Everett Many-Worlds Theory

Andreas Wichert * and Catarina Moreira

Department of Informatics, INESC-ID/Instituto Superior Técnico - University of Lisboa, Porto Salvo, Portugal

Following a controversial suggestion by David Deutsch that decision theory can solve the problem of probabilities in the Everett many-worlds we suggest that the probabilities are induced by Shannon's entropy that measures the uncertainty of events. We argue that a relational person prefers certainty to uncertainty due to fundamental biological principle of homeostasis.

Keywords: decision theory, Everett many-worlds, homeostasis, probability, Shannon's entropy

1. INTRODUCTION

The Everett many-worlds theory views reality as a many-branched tree in which every possible quantum outcome is realized [1–8]. The modern version of the wave-function collapse is based on decoherence and leads to the multiverse interpretation of quantum mechanics [9]. Every time a quantum experiment with different possible outcomes is performed, all outcomes are obtained. If a quantum experiment is preformed with two outcomes with quantum mechanical probability 1/100 for outcome A and 99/100 for outcome B, then both the world with outcome A and the world with outcome B will exist. A person should not expect any difference between the experience in a world A and B. The open question is the following one: Due to deterministic nature of the branching, why should a rational person care about the corresponding probabilities? Why not simply assume that they are equally probable due to deterministic nature of the branching [10]. How can we solve this problem without introducing additional structure into the many-worlds theory?

David Deutsch suggested that decision theory can solve this problem [11–13]. A person identifies the consequences of decision theory with things that happen to its individual future copies on particular branches. A person who does not care in receiving 1 on the first branch A and 99 on the second branch B labels them with a probability 1/2. A rational person that cares assigns the probability 1/100 for outcome A and 99/100 for outcome B. It should be noted that David Deutsch introduced a rational person into the explanation of the corresponding problem. The probability rule within unitary quantum mechanics is based on the rationality of the person. However since the branching is deterministic and no uncertainty is present, how can this rational person justify the application of decision-making?

In this work, we pretend to give an alternative and simpler explanation of Deutch's decision theoretic argument motivated by biological mechanisms and also based on Epstein's ideas that toward uncertainty, human beings tend to have aversion preferences. They prefer to choose an action that brings them a certain but lower utility instead of an action that is uncertain but can yield a higher utility [14]. We precede in two steps.

In the first step we propose to use Shannon's entropy as the expected utility in Deutsch's approach. Probabilities in Shannon's entropy function can be seen as frequencies; they can be measured only by performing an experiment many times and indicate us the past experience. Surprise is inversely related to probability. The larger the probability that we receive a certain

message, the less surprised we are. For example the message "Dog bites man" is quite common, has a high probability and usually we are not surprised. However, the message "Man bites dog" is unusual and has a low probability. The more we are surprised about the occurrence of an event the more difficult an explanation of such an event is. The surprise is defined in relation to previous events, in our example men and dogs.

In the second step we introduce the experience of identity derived from homeostasis as a fundamental biological principle. It is preformed subconsciously by our brain as a coherent explanation of events in a temporal window [15]. Events with higher surprise are more difficult to explain and require more energy. Before an event happens an explanation has to be initiated, so after the event happened it can be integrated in the present explanation in the temporal window. A rational person may not care about the attached weights during deterministic branching, but our brain machinery cares. This information is essential for the ability to give a continuous explanation of our "self" identity.

The paper is organized as follows:

- We review Deutsch's decision-theoretic argument.
- We propose Shannon's entropy as the expected utility function and surprisal as the utility function.
- We argue that probabilities that are the basis of surprise are essential for the ability to give a continuous explanation.

2. REVIEW OF DEUTSCH'S DECISION-THEORETIC ARGUMENT

Decision theory according to Savage is a theory designed for the analysis of rational decision-making under conditions of uncertainty [11]. A rational person faces a choice of acts as a function from the set of possible states to the set of consequences. There are some constraints on the acts of the rational person, as for example the preferences must be transitive. It can be then shown that there exists a probability measure p on states s and a utility function U on the set of consequences of an act A so that the expected utility is defined as

$$EU(A) := \sum_s p(s) \cdot U(A(s)). \tag{1}$$

It follows that a rational person prefers act A to act B if the expected utility of A is greater than that of B. The behavior corresponds to the maximization of the expected utility with respect to some probability measure [16].

2.1. Decision-Theoretic Argument

In the context of the many-worlds the rational person is able to describe each of its acts as a function from the set of possible future branches that will result from a given quantum measurement to the set of consequences [11]. Consequences are the things that happen to individual future copies of the person on particular branch. Act is a function from states to consequences, the preferences must be transitive. If a rational person prefers act A to act B, and prefers act B to act C, then the

same person must prefer act A to act C. This can be summarized by assigning a real number to each possible outcome in such a way that the preferences are transitive. The corresponding number is called the utility or value. The deterministic process of branching is identified as a chance setup for the rational person by a quantum game with a payoff function P associating a consequence with each eigenvalue of the observable $\hat{X}$. When the measurement is performed the state vector collapses into one or the other eigenstate of the observable being measured, a projection into a eigenstate is preformed. For observable $\hat{X}$ and the state $|y\rangle$ in which the expression for the probability reduces to $|\langle x|y\rangle|^2$ in which x is an eigenvector of $\hat{X}$. A quantum game is specified by the triple

$$(|y\rangle, \hat{X}, P).$$

It is assumed that the utilities of the possible payoffs have an additivity property [11]. The approach is based on two non-probabilistic axioms of decision theory. The principle of substitutability, constrains the values of sub-games. If any of the sub-games is replaced by a game of equal value then the value of the composite game is unchanged. The other axiom concerns two-player zero sum games. If there are two possible acts A and B with payoff c for A and $-c$ for B then playing A and B results in zero. The expected utility or value for playing A and B is

$$EU(A) + EU(B) = c - c = 0.$$

For any two acts A, B, the rational person prefers A to B if

$$EU(A) > EU(B).$$

The person acts if she regarded her multiple future branches as multiple possible futures. In classical decision theory two rational persons may be represented by different probability measures on the set of states. Not so in David Deutsch suggested approach that deals with many-worlds theory [10]. Expected value with an observable $\hat{X}$ is

$$EU(|y\rangle) = \langle y|\hat{X}|y\rangle \tag{2}$$

and since

$$\langle y|\hat{X}|y\rangle = \sum_i |\langle x_i|y\rangle|^2 \cdot x_i = \sum_i p_i \cdot x_i \tag{3}$$

representing the weighted mean over the eigenvalues of $\hat{X}$. In quantum physics $\langle y \rangle$ is called the expected value. A rational person that makes decision about outcomes of measurement believes that each possible eigenvalue x_i had the probability $|\langle x_i|y\rangle|^2 = p_i$ due to the process of maximizing the probabilistic expectation value of the payoff [11].

2.1.1. The Derivation of Probabilities for Real Amplitudes

We sketch the proof by David Elieser Deutsch, see Deutsch [11]. If $|y\rangle$ is an eigenstate $|x\rangle$ of $\hat{X}$ it follows that

$$EU(|y\rangle) = x$$

is equal to the eigenvalue x. By appealing twice to the additivity of utilities for eigenvectors $|x_i\rangle$ and adding a constant k we arrive at

$$EU\left(\sum_i \omega_i \cdot |x_i + k\rangle\right) = EU\left(\sum_i \omega_i \cdot |x_i\rangle\right) + k. \quad (4)$$

The rational person is indifferent in receiving the separate payoffs with utilities x_1 and x_2 or receiving a single payoff with utility $x_1 + x_2$. The expected utility of $|x_i + k\rangle$ has by additivity the same value as the expected utility of $|x_i\rangle$ followed by k. This is the central equation on which the proof of David Deutsch is based. The constant k on the left side corresponds to a combination of eigenstates. This is because it is required that for each branch the payoffs are present. The constant k on the right side corresponds to a combination of the corresponding eigenvalues.

According to additivity the left side of equation has the same expected utility as the superposition of possible branches and the payoffs represented by a corresponding combination of eigenvalues is on the right side. The other equation follows from the axiom concerns two-player zero-sum games

$$EU\left(\sum_i \omega_i \cdot |x_i\rangle\right) + EU\left(\sum_i \omega_i \cdot |-x_i\rangle\right) = 0. \quad (5)$$

If $|y\rangle$ is in a superposition

$$|y\rangle = \frac{1}{\sqrt{2}} \cdot |x_1\rangle + \frac{1}{\sqrt{2}} \cdot |x_2\rangle. \quad (6)$$

and with

$$k = -x_1 - x_2 \quad (7)$$

it follows

$$EU\left(\frac{1}{\sqrt{2}}\left(|x_1 - x_1 - x_2\rangle + |x_2 - x_1 - x_2\rangle\right)\right)$$
$$= EU\left(\frac{1}{\sqrt{2}}\left(|-x_2\rangle - |x_1\rangle\right)\right)$$
$$= -EU\left(\frac{1}{\sqrt{2}}\left(|x_1\rangle + |x_2\rangle\right)\right)$$
$$= EU\left(\frac{1}{\sqrt{2}}\left(|x_1\rangle + |x_2\rangle\right)\right) - x_1 - x_2 \quad (8)$$

and by the Equation the value of the expected utility is derived as

$$\frac{1}{2}(x_1 + x_2) = EU\left(\frac{1}{\sqrt{2}}\left(|x_1\rangle + |x_2\rangle\right)\right). \quad (9)$$

For a superposition of n eigenstates

$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n) = EU\left(\frac{1}{\sqrt{n}}\left(|x_1\rangle + |x_2\rangle + \cdots + |x_n\rangle\right)\right) \quad (10)$$

the proof is based on induction from the principle of substitutability and additivity. For $n = 2^m$ the proof follows

from substitutability by inserting a two equal amplitude game into remaining 2^{m-1} equal-amplitude outcomes. Otherwise the proof follows inductive hypothesis and additivity by replacing $n - 1$ by n, for details see Deutsch [11]. For unequal amplitudes

$$\frac{m}{n} \cdot x_1 + \frac{n-m}{n} \cdot x_2 = EU\left(\sqrt{\frac{m}{n}} \cdot |x_1\rangle + \sqrt{\frac{n-m}{n}} \cdot |x_2\rangle\right). \quad (11)$$

we introduce auxiliary system that can be in two states different states, either

$$\sqrt{\frac{1}{n}} \cdot \sum_{a=1}^{m} |z_a\rangle \quad (12)$$

or

$$\sqrt{\frac{1}{n-m}} \cdot \sum_{a=m+1}^{n} |z_a\rangle \quad (13)$$

with eigenstates $|z_a\rangle$ and eigenvalues z_a of the observable $\hat{Z}$ that are all distinct. Then the joint state is given by

$$\sqrt{\frac{1}{n}} \cdot \left(\sum_{a=1}^{m} |x_1\rangle|z_a\rangle + \sum_{a=m+1}^{n} |x_2\rangle|z_a\rangle\right). \quad (14)$$

When we measure the observable $\hat{Z}$ depending on the index a of the eigenvalues z_a we know that $\hat{X}$ is x_1 if $a < m + 1$ or x_2 otherwise. With additional properties of the eigenvalues

$$\sum_{a=1}^{m} z_a = \sum_{a=m+1}^{n} z_a = 0 \quad (15)$$

and all n values

$$x_1 + z_a, \ 1 \leq a \leq m \quad (16)$$

and

$$x_2 + z_a, \ m + 1 \leq a \leq n \quad (17)$$

are all distinct, the composite measurement with observable $\hat{X}$ and $\hat{Z}$ has the same value as the one with observable $\hat{X}$ alone. Because of the additivity the state is equivalent to the amplitude superposition

$$\sqrt{\frac{1}{n}} \cdot \left(\sum_{a=1}^{m} |x_1\rangle|z_a\rangle + \sum_{a=m+1}^{n} |x_2\rangle|z_a\rangle\right) \quad (18)$$

and it follows

$$\frac{1}{n} \cdot \left(\sum_{a=1}^{m} |x_1\rangle|z_a\rangle + \sum_{a=m+1}^{n} |x_2\rangle|z_a\rangle\right) = \frac{m}{n} \cdot x_1 + \frac{n-m}{n} \cdot x_2. \quad (19)$$

A rational person would choose the expected value of $|y\rangle$ with an observable $\hat{X}$

$$EU(|y\rangle) = \langle y|\hat{X}|y\rangle \quad (20)$$

and since

$$\langle y|\hat{X}|y\rangle = \sum_i \langle x_i|y\rangle|^2 \cdot x_i = \sum_i p_i \cdot x_i \qquad (21)$$

represents the weighted mean over the eigenvalues of $\hat{X}$, the rational person interprets p_i as probabilities. For complex amplitudes it is assumed that the unitary transformation

$$|x_a\rangle \rightarrow e^{i\cdot\theta_a}|x_a\rangle \qquad (22)$$

with a corresponding phase θ_a does not alter the payoff, the player is indifferent as to whether it occurs or not. The proof can be extended to irrational numbers, it is based on the idea that the state undergoes some unitary evolution before the measurement [11]. The unitary evolution leads to real amplitudes with eigenvalues that exceeds the original state. Each game played after the unitary transformation is as valuable as the original game, and the values of such games have a lower bound of the original game.

3. EXPECTED UTILITY AND ENTROPY

A fundamental property of rational persons is that they prefer certainty to uncertainty. Humans prefer to choose an action that brings them a certain, but lower utility instead of an action that is uncertain, but can bring a higher utility [14].

We measure the uncertainty by the entropy of the experiment. The experiment starts at t_0 and ends at t_1. At t_0, we have no information about the results of the experiment, and at t_1, we have all of the information, so that the entropy of the experiment is 0. We can describe an experiment by probabilities. For the outcome of the flip of a fair coin, the probability for a head or tail is 0.5, $p = (0.5, 0.5)$. A person A knows the outcome, but person B does not. Person B could ask A about the outcome of the experiment. If the question is of the most basic nature, then we could measure the minimal number of optimally required questions B must pose to know the result of the experiment. A most basic question corresponds to the smallest information unit that could correspond to a yes or no answer. For a fair coin, we pose just one question, for example, is it a tail? For a card game, to determine if a card is either red, clubs or spades, we have a different number of possible questions. If the card is red, then we need only one question. However, in the case in which the card is not red, we need another question to determine whether it is a spade or a club. The probability of being red is 0.5, of clubs 0.25 and spades 0.25, $p = (0.5, 0.25, 0.25)$. For clubs and spades, we need two questions. In the meantime, we must ask $1 \cdot 0.5 + 2 \cdot 0.25 + 2 \cdot 0.25$ questions, which would result in 1.5 questions. The entropy is represented by Shannon's entropy H for an experiment A

$$H(A) = -\sum_i p_i \log_2 p_i. \qquad (23)$$

It indicates the minimal number of optimal binary yes/no questions that a rational person must pose to know the result of an experiment [17, 18] . We can describe the process of measuring a state by an observable as measuring the entropy of the experiment. Before the measurement of a state $|x\rangle$ by an observable we are uncertain about the outcome. We measure the uncertainty by Shannon's entropy. After the measurement the state is in eigenstate, the entropy is zero, Shannon's entropy is defined for any observable and any probability distribution, according to Ballentin [19, p. 617] "It measures the maximum amount of information that may be gained by measuring that observable."

Assuming that a Hilbert space H_n can be represented as a collection of orthogonal subspaces

$$H_n = E_1 \oplus E_2 \oplus \ldots \oplus E_f \qquad (24)$$

with $f \leq n$. A state $|y\rangle$ can be represented with $|x_i\rangle \in E_i$ as

$$|y\rangle = \omega_1 \cdot |x_1\rangle + \omega_2 \cdot |x_2\rangle + \cdots + \omega_f \cdot |x_f\rangle.$$

For one dimensional subspaces $f = n$ and the value $|x_k\rangle$ is observed with a probability $\|\omega_k \cdot |x_k\rangle\|^2 = |\omega_k|^2$. Shannon's entropy is defined as

$$E(|y\rangle) = -\sum_{i=1}^n (|\omega_i|^2 \cdot \log_2 |\omega_i|^2) = -\sum_i p_i \log_2 p_i. \qquad (25)$$

3.1. Weighted Sum of Surprisals

We say that events that seldom happen, for example, the letter x in a message, have a higher surprise. Some letters are more frequent than others; an e is more frequent than an x. The larger the probability that we receive a character, the less surprised we are. Surprise is inversely related to probability.

$$s_i = \frac{1}{p_i}$$

The logarithm of surprise

$$I_i = -\log_2 \frac{1}{|\omega_i|^2} = 2 \cdot \log_2 |\omega_i| = -\log_{\sqrt{2}} |\omega_i| = \log_2 s_i \quad (26)$$

is the self-information or surprisal I_i. The Shannon's entropy H represents the weighted sum of surprisals.

$$H(A) = \sum_i p_i \cdot \log_2 s_i = \sum_i^n p_i \cdot I_i. \qquad (27)$$

It can be interpreted as an expected utility

$$EU(A) := \sum_i p_i \cdot U(A(i)) = \sum_i^n p_i \cdot I_i. \qquad (28)$$

with the utility function $U(A(i))$

$$U(A(i)) = I_i. \qquad (29)$$

For acts the expected utility is identified with the representation of the entropy of an action, represented by H

$$H(A) := EU(A).$$

That the rewards are determined by the negative information content is already known and is used for the utility representation problems in economics [20–22]. If there are two possible acts A and B with payoff c for A and $-c$ for B then playing A and B results in zero. The expected utility or value for playing A and B is

$$H(A) + H(B) = c - c = 0,$$

but for any two acts A, B, the rational person prefers A to B if

$$H(A) < H(B)$$

because the uncertainty is lower. The theoretic proof David Deutsch can be applied, for example

$$|y\rangle = \frac{1}{\sqrt{2}} \cdot |x_1\rangle + \frac{1}{\sqrt{2}} \cdot |x_2\rangle. \qquad (30)$$

and with

$$k = -x_1 - x_2 \qquad (31)$$

the right side of the equation with eigenvalues $x_1 = 1$ and $x_2 = 1$

$$k = -x_1 - x_2 = -2 \qquad (32)$$

or $U(A(i)) = I_i = \log_2 2 = 1$. Therefore

$$\frac{1}{2}(1+1) = \frac{1}{2}(x_1 + x_2) = EU\left(\frac{1}{\sqrt{2}}(|x_1\rangle + |x_2\rangle)\right) \qquad (33)$$

and

$$1 = 0.5 \cdot 1 + 0.5 \cdot 1. \qquad (34)$$

However we can as well recover the probabilities from the definition of Shannon's entropy. This strengthens idea even more.

3.2. Recovering Probabilities from Entropy

The entropy or the uncertainty is maximal in the case in which all probabilities are equal, which means that $p = (1/n, 1/n..., 1/n)$. In this case

$$H(F) = -\sum_i p_i \log_2 p_i = -\log_2 1/n = \log_2 n. \qquad (35)$$

the surprisal I_i is equal to the entropy

$$H(F) = \log_2 n = I_i. \qquad (36)$$

If $|y\rangle$ is in a superposition of n eigenstates

$$|y\rangle = \frac{1}{\sqrt{n}}(|x_1\rangle + |x_2\rangle + \cdots + |x_n\rangle) \qquad (37)$$

it follows that a surprisal I_i^* of each state represented by amplitude $\omega_i = \frac{1}{\sqrt{n}}$ is

$$I_i^* = \log_2 \sqrt{n} = \frac{1}{2} \cdot \log_2 n. \qquad (38)$$

We know that in the case of equal amplitudes $\log n$ yes/no questions have to be asked with $H(F) = I_i$ so the function $f(I^*)$ is equal to multiply by two

$$H(|y\rangle) = I_i = f(I_i^*) = 2 \cdot \log_2 \sqrt{n} = \log_2 n = -\log_2 \left(\frac{1}{\sqrt{n}}\right)^2 \qquad (39)$$

that is equivalent to

$$p_i = 2^{I_i} = \left(\frac{1}{\sqrt{n}}\right)^2. \qquad (40)$$

In the case of complex amplitudes with phase θ_a

$$\omega_i = e^{i \cdot \theta_a} \frac{1}{\sqrt{n}} \qquad (41)$$

$$I_i^* = \log_2 e^{i \cdot \theta_a} \sqrt{n} = \frac{i \cdot \theta_a}{\log 2} + \frac{1}{2} \cdot \log_2 n, \qquad (42)$$

the operations described by $f(I^*)$ would be, first get rid of the complex number $i \cdot \theta_a / \log 2$ by subtracting $-i \cdot \theta_a / \log 2$ and then multiply by 2. Or

$$p_i = 2^{I_i} = \left(\frac{1}{e^{-i \cdot \theta_a} e^{i \cdot \theta_a} \cdot \sqrt{n}}\right)^2 = \left(\frac{1}{\sqrt{n}}\right)^2. \qquad (43)$$

For unequal amplitudes for each surprisal I_i^* of each state we multiply it by two and recover from I_i the value $p_i = 2^{I_i}$. This operation leads to the minimal number of optimal binary yes/no questions. For example for

$$|y\rangle = \sqrt{\frac{m}{n}} \cdot |x_1\rangle + \sqrt{\frac{n-m}{n}} \cdot |x_2\rangle \qquad (44)$$

we get by operations described by $f(I^*)$

$$H(|y\rangle) = -\frac{m}{n} \cdot \log_2 \frac{m}{m} - \left(\frac{n-m}{n}\right) \cdot log_2 \left(\frac{n-m}{n}\right), \qquad (45)$$

the correct value of Shannon's entropy. We assume

$$p_i = \left(\frac{1}{|\omega_i|}\right)^2. \qquad (46)$$

3.3. Biological Principle of Energy Minimization

Identity is a concept that defines the properties of a rational person over time [23]. It is a unifying concept based on the biological principles of homeostasis [24, 25]. Organisms have to be kept stable to guarantee the maintenance of life, like for

example the regulation of body temperature. This principle was extended by Allostasis [26] for regulation of bodily functions over time. To preform this task efficient mechanisms for the prediction of future states are needed to anticipate future environmental constellations [27, 28]. This is done, because the homeostatic state may be violated by unexpected changes in the future. It means as well that every organism implies a kind of self identity over time [29]. This identity requires a time interval of finite duration within which sensory information is integrated. Different sensor information arrives at different time stamps. The fusion process has to be done over some time window. Similar problems are present during a sensor fusion task in a mobile robot. For example in visual and auditory perception in humans the transduction of the acoustic information is much shorter then the visual [30]. In it is suggested that in humans a temporal window with the duration of 3 s is created [15]. This window represents the psychological concept of "now" [29]. The consciousness concept of "now" represented by the temporal window is shifted backward in time of the consciousness itself, since a subconsciousness mechanism is required to preform the integration task.

3.3.1. Prediction of Events

One of the brain functions is to provide a casual consistent explanation of events to maintain self identity over time leading to the psychological concept of "now." Split brain research and stimulation or brain regions during awake operation suggest that the brain generates an explanation of effects that were not initiated by consciousness [31, 32]. Before an event happens an explanation has to be incited by the subconsciousness parts of the brain so that it is possible to integrate it into the temporal window of the self when the event happens. As well other organism functions need be put in alert due to some predicted possible events.

Events with higher surprise are more difficult to explain than events with low surprise values. An explanation has to be possible. When the surprise is too high, an explanation may be impossible and the identity of self could break. The idea is related to the general constructor theory of David Elieser Deutsch, see Deutsch [33]. The metabolic cost of neural information processing of explaining higher surprise events require higher energy levels then lower surprise events. Fechner's law states that there is a logarithmic relation between the stimulus and its intensity [34]. We assume as well that there is a logarithmic relation between the cost of initiation an explanation of an event and its surprise value s_i, that is $\log s_i$. Neuronal computation is energetically expensive [35]. Consequently, the brain's limited energy supply imposes constraints on its information processing capability. The costs should be fair divided into the explanation

of all predicted possible branches since the organism will be deterministic present in all of them. A possible solution is given by the Shannon's entropy. The corresponding value I_i is weighted in relation

$$p_i = 2^{I_i} \tag{47}$$

and the resulting costs are $p_i \cdot I_i$. The resulting costs of initiating n explanations for action A of n predicted branches before a split are

$$H(A) = \sum_i^n p_i \cdot \log_2 s_i = \sum_i^n p_i \cdot I_i \tag{48}$$

For the human (subconsciousness) brain it makes sense to choose A to B if

$$H(A) < H(B).$$

since it requires less energy.

4. CONCLUSION

Since the branching is deterministic and no uncertainty is present, how can a rational person justify the application of decision-making? Why not simply assume that they are equally probable due to deterministic nature of the branching. David's Deutsch probability rule within unitary quantum mechanics is based on rationality. Instead of rationality we introduced the biological principle of homeostasis. Before an event happens an explanation has to be be prepared. The costs of the explanation should be fair divided into the explanation of all predicted possible branches since the organism will be deterministic present in all of them. The costs are described by the negative expected utility represented by Shannon's entropy. The probabilities can be recovered from the Shannon's entropy.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

1. Everett H. "relative state" formulation of quantum mechanics. *Rev Mod Phys.* (1959) **29**:454–62. doi: 10.1103/RevModPhys.29.454
2. Wheeler J. Assessment of everett's "relative state." *Rev Mod Phys.* (1957) **29**:463–5. doi: 10.1103/RevModPhys.29.463
3. Dewitt BS, Graham N, (eds.). *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton, NJ: Princeton University Press (1973).
4. Deutsch D. *The Fabric of Reality.* London: Penguin Group (1997).
5. Deutsch D. The structure of the multiverse. *Proc R Soc A* (2002) **458**:2911–23. doi: 10.1098/rspa.2002.1015

6. Wallace D. Worlds in the everett interpretation. *Stud Hist Philos Mod Phys.* (2002) **33**:637–61. doi: 10.1016/S1355-2198(02)00032-1

7. Wallace D. Everett and structure. *Stud Hist Philos Mod Phys.* (2003a) **34**:87–105. doi: 10.1016/S1355-2198(02)00085-0

8. Byrne P. The many worlds of hugh everett. *Sci Am.* (2007) **297**:98–105. doi: 10.1038/scientificamerican1207-98

9. Bousso R, Susskind L. Multiverse interpretation of quantum mechanics. *Phys Rev D* (2012) **85**:045007. doi: 10.1103/PhysRevD.85.045007

10. Greaves H. Probability in the everett interpretation. *Philos Comp.* (2007) **1**:109–128.

11. Deutsch D. Quantum theory of probability and decisions. *Proc R Soc A* (1999) **455**:3129–97. doi: 10.1098/rspa.1999.0443

12. Wallace D. Everettian rationality: defending deutsch's approach to probability in the everett interpretation. *Stud Hist Philos Mod Phys.* (2003b) **24**:415–39. doi: 10.1016/S1355-2198(03)00036-4

13. Wallace D. Quantum probability from subjective likelihood: improving on deutsch's proof of the probability rule. *Stud Hist Philos Mod Phys.* (2007) **38**:311–32. doi: 10.1016/j.shpsb.2006.04.008

14. Epstein LG. A definition of uncertainty aversion. *Rev Econ Stud.* (1999) **66**:579–608. doi: 10.1111/1467-937X.00099

15. Pöppel E. Pre-semantically defined temporal windows for cognitive processing. *Philos Trans R Soc Lond B Biol Sci.* (2009) **364**:1887–1896. doi: 10.1098/rstb.2009.0015

16. Russell SJ, Norvig P. *Artificial Intelligemce: A Modern Approach, 2nd Edn.* Harlow: Prentice-Hall (2003).

17. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* (1948) **27**:379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

18. Topsoe F. *Informationstheorie*. Stuttgart: Teubner Sudienbucher (1974). doi: 10.1007/978-3-322-94886-1

19. Ballentine LE. *Quantum Mechanics A Modern Development, 2nd Edn.* Singapore: World Scientific (2012).

20. Candeal JC, De Miguel JR, Indura E, Mehta GB. Utility and entropy. *Econ Theory* (2001) **17**:233–8. doi: 10.1007/PL00004100

21. Ortega PA, Braun DA. A conversion between utility and information. In: *Proceedings of the Third Conference on Artificial General Intelligence*. Mountain View, CA: Springer; Heidelberg (2009). p. 115–120.

22. Robert FN, Jose VRR, Winkler RL. Duality between maximization of expected utility and minimization of relative entropy when probabilities are imprecise. In: *6th International Symposium on Imprecise Probability: Theories and Applications*. Durham (2009).

23. Pöppel E. Perceptual identity and personal self: neurobiological reflections. In: Fajkowska M, Eysenck MM, editors. *Personality From Biological, Cognitive,* *and Social Perspectives*. Clinton Corners, NY: Eliot Werner Public (2010). p. 75–82.

24. Bernard C. *An Introduction to the Study of Experimental Medicine*. New York, NY: Dover (1957).

25. Gross CG. Claude bernard and the constancy of the internal environment. *Neuroscientist* (1998) **4**:380–5. doi: 10.1177/107385849800400520

26. Sterling P. Principles of allostasis: optimal design, predictive regulation, pathophysiology and rational therapeutics. In: Schulkin J, editor. *Allostasis, Homeostasis, and the Costs of Adaptation*. Cambridge: University Press (2004). p. 17–64. doi: 10.1017/CBO9781316257081.004

27. von Holst E, Mittelstaedt H. Das reafferenzprinzip (wechselwirkungen zwischen zentralnervensystem und peripherie. *Naturwissenschaften* (1950) **37**:464–76. doi: 10.1007/BF00622503

28. Bao Y, Pöppel E, Liang W, Yang T. When is the right time? a little later! – delayed responses show better temporal control. *Proc Soc Behav Sci.* (2014) **126**:199–200. doi: 10.1016/j.sbspro.2014.02.370

29. Zhou B, Pöppel E, Bao Y. In the jungle of time: the concept of identity as a way out. *Front Psychol.* (2014) **5**:844. doi: 10.3389/fpsyg.2014.00844

30. Pöppel E, Schill K, von Steinbüchel N. Sensory integration within temporally neutral system states: a hypothesis. *Naturwissenschaftem* (1990) **77**:89–91. doi: 10.1007/BF01131783

31. Libet B. *Mind Time - The Temporal Factor in Consciousness*. Cambridge, MA; London: Harvard University Press (2004).

32. Coon D, Mitterer JO. *Introduction to Psychology Gateways to Mind and Behavior, 13th Edn*. Boston, MA: Wadsworth Publishing (2012).

33. Deutsch D. Constructor theory. *Synthese* (2013) **190**:4331–4359. doi: 10.1007/s11229-013-0279-z

34. Frisby JP, Stone JV. *Seeing, The Computational Approach to Biological Vision, 2nd Edn*. Cambridge, MA: MIT Press (2010).

35. Laughlin SB, de Ruyter van Steveninck RR, Anderson JC. The metabolic cost of neural information. *Nat Neurosci.* (1998) **1**:36–41. doi: 10.1038/236