

# Journal of Theoretics

---

## Why Ockham's Razor?

Russell K. Standish  
High Performance Computing Support Unit  
University of New South Wales  
Sydney, 2052  
Australia

[R.Standish@unsw.edu.au](mailto:R.Standish@unsw.edu.au)  
<http://parallel.hpc.unsw.edu.au/rks>

**Abstract:** In this paper, I show why in an ensemble theory of the universe, we should be inhabiting one of the elements of that ensemble with least information content that satisfies the anthropic principle. This explains the effectiveness of aesthetic principles such as Ockham's razor in predicting usefulness of scientific theories. I also show, with a couple of reasonable assumptions about the phenomenon of consciousness, that quantum mechanics is the most general linear theory satisfying the anthropic principle.

## Introduction

Wigner[12] once remarked on "the unreasonable effectiveness of mathematics", encapsulating in one phrase the mystery of why the scientific enterprise is so successful. There is an aesthetic principle at large, whereby scientific theories are chosen according to their beauty, or simplicity. These then must be tested by experiment -- the surprising thing is that the aesthetic quality of a theory is often a good predictor of that theory's explanatory and predictive power. This situation is summed up by William de Ockham "Entities should not be multiplied unnecessarily" known as Ockham's Razor.

We start our search into an explanation of this mystery with the *anthropic principle*[1]. This is normally cast into either a weak form (that physical reality must be consistent with our existence as conscious, self-aware entities) or a strong form (that physical reality is the way it is *because* of our existence as conscious, self-aware entities). The anthropic principle is remarkable in that it generates significant constraints on the form of the universe[1,9]. The two main explanations for this are the *Divine Creator explanation* (the universe was created deliberately by God to have properties sufficient to support intelligent life), or the *Ensemble explanation*[9] (that there is a set, or ensemble, of different universes, differing in details such as physical parameters, constants and even laws, however, we are only aware of such universes that are consistent with our existence). In the Ensemble explanation, the strong and weak formulations of the anthropic principle are equivalent.

Tegmark introduces an ensemble theory based on the idea that every self-consistent mathematical structure be accorded the ontological status of *physical existence*. He then goes on to categorize mathematical structures that have been discovered thus far (by humans), and argues that this set should be largely universal, in that all self-aware entities should be able to uncover at least the most basic of these mathematical structures, and that it is unlikely we have overlooked any equally basic mathematical structures.

An alternative ensemble approach is that of Schmidhuber's[8] -- the "Great Programmer". This states that all possible halting programs of a universal Turing machine have physical existence. Some of these programs' outputs will contain self-aware substructures -- these are the programs deemed interesting by the anthropic principle. Note that there is no need for the UTM to actually exist, nor is there any need to specify which UTM is to be used -- a program that is meaningful on  $UTM_1$  can be executed on  $UTM_2$  by prepending it with another program that describes  $UTM_1$  in terms of  $UTM_2$ 's instructions, then executing the individual program. Since the set of programs (finite length bitstrings) is isomorphic to the set of whole numbers  $\mathbb{N}$ , an enumeration of  $\mathbb{N}$  is sufficient to generate the ensemble that contains our universe. The information content of this complete set is precisely zero, as no bits are specified. This has been called the "zero information principle".

In this paper, we adopt the Schmidhuber ensemble as containing all possible descriptions of all possible universes, whilst remaining agnostic on the issue of whether this is all there is.<sup>1</sup> Each self-consistent mathematical structure (member of the Tegmark ensemble) is completely described by a finite set of symbols, and a countable set of axioms encoded in those symbols, and a set of rules (logic) describing how one mathematical statement may be converted into another.<sup>2</sup> These axioms may be encoded as a bitstring, and the rules encoded as a program of a UTM that enumerates all possible theorems derived from the axioms, so each member of the Tegmark ensemble may be mapped onto a Schmidhuber one.<sup>3</sup> The Tegmark ensemble must be contained within the Schmidhuber one.

An alternative connection between the two ensembles is that the Schmidhuber ensemble is a self-consistent mathematical structure, and is therefore an element of the Tegmark one. However, all this implies is that one element of the ensemble may in fact generate the complete ensemble again, a point made by Schmidhuber in that the "Great Programmer" exists many times, over and over in a recursive manner within his ensemble. This is now clearly true also of the Tegmark ensemble.

# Universal Prior

The natural measure induced on the ensemble of bitstrings is the uniform one, i.e. no bitstring is favoured over any other. This leads to a problem in that longer strings are far more numerous than shorter strings, so we would conclude that we should expect to see an infinitely complex universe.

However, we should recognise that under a UTM, some strings encode for identical programs as other strings, so one should equivalence class the strings. In particular, strings where the bits after some bit number  $n$  are "don't care" bits, are in fact equivalence classes of all strings that share the first  $n$  bits in common. One can see that the size of the equivalence class drops off exponentially with the amount of information encoded by the string. Under a UTM, the amount of information is not necessarily equal to the length of the string, as some of the bits may be redundant. The sum

$$P_U(s) = \sum_{p:U \text{ computes } s \text{ from } p \text{ and halts}} 2^{-|p|}, \tag{1}$$

where  $|p|$  means the length of  $p$ , gives the size of the equivalence class of all halting programs generating the same output  $s$  under the UTM  $U$ . This measure distribution is known as a *universal prior*, or alternatively a Solomonoff-Levin distribution<sup>[6]</sup>. We assume the *self-sampling assumption*<sup>[5,2]</sup>, essentially that we expect to find ourselves in one of the universes with greatest measure, subject to the constraints of the anthropic principle. This implies we should find ourselves in one of the simplest possible universes capable of supporting self-aware substructures (SASes). This is the origin of physical law -- why we live in a mathematical, as opposed to a magical universe. This is why aesthetic principles, and Ockham's razor in particular are so successful at predicting good scientific theories. This might also be called the "minimum information principle".

There is the issue of what UTM  $U$  should be chosen. Schmidhuber sweeps this issue under the carpet stating that the universal priors differ only by a constant factor due to the compiler theorem, along the lines of

$$P_V(s) \geq P_{UV} P_U(s)$$

where  $P_{UV}$  is the universal prior of the compiler that interprets  $U$ 's instruction set in terms of  $V$ 's. The inequality is there because there are possibly native  $V$ -code programs that compute  $s$  as well. Inverting the symmetric relationship yields:

$$P_{UV} P_U(s) \leq P_V(s) \leq (P_{VU})^{-1} P_U(s)$$

The trouble with this argument, is that it allows for the possibility that:

$$P_V(s_1) \ll P_V(s_2), \text{ but } P_U(s_1) \gg P_U(s_2)$$

So our expectation of whether we're in universe  $s_1$  or  $s_2$  depends on whether we chose  $V$  or  $U$  for the interpreting UTM.

There may well be some way of resolving this problem that leads to an absolute measure over all bitstrings. However, it turns out that an absolute measure is not required to explain features we observe. A SAS is an information processing entity, and may well be capable of

universal computation (certainly *homo sapiens* seems capable of universal computation). Therefore, the only interpreter (UTM) that is relevant to the measure that determines which universe a SAS appears in is the SAS itself. We should expect to find ourselves in a universe with one of the simplest underlying structures, according to our own information processing abilities. This does not preclude the fact that other more complex universes (by our own perspective) may be the simplest such universe according to the self-aware inhabitants of that universe. This is the bootstrap principle writ large.

## The White Rabbit Paradox

An important criticism leveled at ensemble theories is what John Leslie calls *the failure of induction*[\[4, §4.69\]](#). If all possible universes exist, then what is to say that our orderly, well-behaved universe won't suddenly start to behave in a disordered fashion, such that most inductive predictions would fail in them. This problem has also been called the White Rabbit paradox[\[7\]](#), presumably in a literary reference to Lewis Carroll.

This sort of issue is addressed by consideration of measure. We should not worry about the universe running off the rails, provided it is extremely unlikely to do so. Note that Leslie uses the term *range* to mean what we mean by *measure*. At first consideration, it would appear that there are vastly more ways for a universe to act strangely, than for it to stay on the straight and narrow, hence the paradox.

However, things are not what they seem. Consider an observer looking at the world around it. Up until the time in question, the world behaves according to the dictates of a small number of equations, hence its description is a fairly short bitstring of length  $n$ . Next suppose an irreducibly bizarre event happens. Now, let's be quite clear about this. We're not talking about some minute, barely observable phenomenon -- eg an electron being somewhere it shouldn't -- and we're not talking about a phenomenon that might be described by adding new physical laws, as in the explanation of the precession of Mercury by General Relativity. We're talking about undeniable, macroscopic violations of physical law, for instance the coalescing of air molecules to form a fire breathing dragon. Such an event will have a large description,  $m$  bits, that will resist compression.

Consider the expanded space of all bitstrings of length  $n+m$ , sharing a common  $n$ -length prefix encoding the laws of physics that describe the world up until the bizarre event. The observer is a finite state machine in general, so there are a finite variety of these events that can be recognised by the observer. In general, the  $m$ -bit strings will be perceived as random noise by the observer, with a comparative minority being recognised as vaguely like something (as in Rorschach plots, or shapes in clouds), and a vastly rarer number having the convincing fidelity necessary to sustain a belief that the miracle in fact happened.

Thus the initial presumption that law breaking events will outnumber the law abiding ones is shown to be false. On the contrary, they will be extremely rare in comparison.

## Quantum Mechanics

In the previous sections, I demonstrate that members of the Tegmark ensemble are the most compressible, and have highest measure amongst all members of the Schmidhuber ensemble. In this section, I ask the question of what is the most general (i.e. minimum information content) description of an ensemble containing self-aware substructures.

There are a number of assumptions that need to be stated up front. The first relates to the nature of consciousness, as referred to by the Anthropic Principle. We have already stated that the conscious entity must be performing some kind of information processing, so as to interpret the universe. Human beings are capable of universal computation and perhaps all forms of consciousness must be capable of universal computation.

The ability to compute requires a time dimension in which to compute. The only mathematical structures in the Tegmark ensemble capable of being observed from within must have a time dimension in which that observation is interpreted. Denote the state of an ensemble by  $\psi$ .

The most general form of evolution of this state is given by:

$$\frac{d\psi}{dt} = \mathcal{H}(\psi, t) \tag{2}$$

Some people may think that discreteness of the world's description (ie of the Schmidhuber bitstring) must imply a corresponding

discreteness in the dimensions of the world. This is not true. Between any two points on a continuum, there are an infinite number of points that can be described by a finite string -- the set of rational numbers being an obvious, but by no means exhaustive example. Continuous systems may be made to operate in a discrete way, electronic logic circuits being an obvious example. Therefore, the assumption of discreteness of time is actually a specialisation (thus of lower measure according to the Universal Prior) relative to it being continuous.

The conscious observer is responsible, under the Anthropic Principle, for converting the potential into actual, for creating the observed information from the zero information of the ensemble. This can be modeled by a partitioning for each observable  $A : \psi \rightarrow \{\psi_a, \mu_a\}$ , where  $a$  indexes the allowable range of potential observable values corresponding to  $A$ , and  $\mu_a$  is the measure associated with  $\psi_a$  ( $\sum_a \mu_a = 1$ ). The  $\psi_a$  will also, in turn, be solutions to equation (2).

Secondly, we assume that the generally accepted axioms of set theory and probability theory hold. Whilst the properties of sets are well known, we outline here the Kolmogorov probability axioms[6]:

(A1)

If  $A$  and  $B$  are events, then so is the *intersection*  $A \cap B$ , the *union*  $A \cup B$  and the *difference*  $A - B$ .

(A2)

The *sample space*  $S$  is an event, called the *certain event*, and the *empty set*  $\emptyset$  is an event, called the *impossible event*.

(A3)

To each event  $E$ ,  $P(E) \in [0, 1]$  denotes the *probability* of that event.

(A4)

$P(S)=1$ .

(A5)

If  $A \cap B = \emptyset$ , then  $P(A \cup B) = P(A) + P(B)$ .

(A6)

For a decreasing sequence

$$A_1 \supset A_2 \supset \dots \supset A_n \dots$$

of events with  $\bigcap_n A_n = \emptyset$ , we have  $\lim_{n \rightarrow \infty} P(A_n) = 0$ .

Consider now the projection operator  $\mathcal{P}_{\{a\}} : V \rightarrow V$ , acting on a state  $\psi \in V$ ,  $V$  being an all universes ensemble, to produce  $\psi_a = \mathcal{P}_{\{a\}}\psi$ , where  $a \in S$  is an outcome of an observation. We have not at this stage assumed that  $\mathcal{P}_{\{a\}}$  is linear. Define addition for two distinct outcomes  $a$  and  $b$  as follows:

$$\mathcal{P}_{\{a\}} + \mathcal{P}_{\{b\}} = \mathcal{P}_{\{a,b\}} \tag{3}$$

from which it follows that

$$\mathcal{P}_{A \subset S} = \sum_{a \in A} \mathcal{P}_{\{a\}} \tag{4}$$

$$\mathcal{P}_{A \cup B} = \mathcal{P}_A + \mathcal{P}_B - \mathcal{P}_{A \cap B} \tag{5}$$

$$\mathcal{P}_{A \cap B} = \mathcal{P}_A \mathcal{P}_B = \mathcal{P}_B \mathcal{P}_A \tag{6}$$

These results extend to continuous sets by replacing the discrete sums by integration over the sets with uniform measure. Here, as elsewhere, we use  $\Sigma$  to denote sum or integral respectively as the index variable  $a$  is discrete or continuous.

Let the state  $\psi \in V \equiv \{\mathcal{P}_A\psi | A \subset S\}$  be a "reference state", corresponding to the certain event. It encodes information about the whole ensemble. Denote the probability of a set of outcomes  $A \subset S$  by  $P_\psi(\mathcal{P}_A\psi)$ . Clearly

$$P_\psi(\mathcal{P}_S\psi) = P_\psi(\psi) = 1$$

by virtue of (A4). Also, by virtue of equation (5),

$$P_\psi((\mathcal{P}_A + \mathcal{P}_B)\psi) = P_\psi(\mathcal{P}_A\psi) + P_\psi(\mathcal{P}_B\psi) \quad (7)$$

Consider the possibility that  $A$  and  $B$  can be identical. Equation (7) may be written:

$$P_\psi((a\mathcal{P}_A + b\mathcal{P}_B)\psi) = aP_\psi(\mathcal{P}_A\psi) + bP_\psi(\mathcal{P}_B\psi), \forall a, b \in \mathbb{N} \quad (8)$$

Thus, the set  $V$  naturally extends by means of the addition operator defined by equation (3) to include all linear combinations of observed states, at minimum over the natural numbers. If  $A \cap B \neq \emptyset$ , then  $P_\psi((\mathcal{P}_A + \mathcal{P}_B)\psi)$  may exceed unity, so clearly  $(\mathcal{P}_A + \mathcal{P}_B)\psi$  is not necessarily a possible observed outcome. How should we interpret these new nonphysical states? The answer lies in considering more than one observer. The expression  $P_\psi((a\mathcal{P}_A + b\mathcal{P}_B)\psi)$  must be the measure associated with  $a$  observers seeing outcome  $A$  and  $b$  observers seeing outcome  $B$ . Since in general in the multiverse, the number of distinct observers is uncountably infinite, the coefficients may be drawn from a measure distribution, instead of the natural numbers. The most general measure distributions are complex, therefore the coefficients, in general are complex[3]. We can comprehend easily what a positive measure means, but what about complex measures? What does it mean to have an observer with measure -1? It turns out that these non-positive measures correspond to observers who chose to examine observables that do not commute with our current observable  $A$ . For example if  $A$  were the observation of an electron's spin along the  $z$  axis, then the states  $|+\rangle + |-\rangle$  and  $|+\rangle - |-\rangle$  give identical outcomes as far as  $A$  is concerned. However, for another observer choosing to observe the spin along the  $x$  axis, the two states have opposite outcomes. This is the most general way of partitioning the Multiverse amongst observers, and we expect to observe the most general mathematical structures compatible with our existence.

The probability function  $P$  can be used to define an inner product as follows. Our reference state  $\psi$  can be expressed as a sum over the projected states  $\psi = \sum_{a \in S} \mathcal{P}_{\{a\}}\psi \equiv \sum_{a \in S} \psi_a$ . Let  $V^* = \mathcal{L}(\psi_a)$  be the linear span of this basis set. Then,  $\forall \phi, \xi \in V^*$ , such that  $\phi = \sum_{a \in S} \phi_a \psi_a$  and  $\xi = \sum_{a \in S} \xi_a \psi_a$ , the inner product  $\langle \phi, \xi \rangle$  is defined by

$$\langle \phi, \xi \rangle = \sum_{a \in S} \phi_a^* \xi_a P_\psi(\psi_a) \quad (9)$$

It is straightforward to show that this definition has the usual properties of an inner product, and that  $\psi$  is normalized ( $\langle \psi, \psi \rangle = 1$ ). The measures  $\mu_a$  are given by

$$\mu_a = P_\psi(\psi_a) = \langle \psi_a, \psi_a \rangle$$

$$\begin{aligned}
&= \langle \psi, \mathcal{P}_a \psi \rangle \\
&= |\langle \psi, \hat{\psi}_a \rangle|^2
\end{aligned} \tag{10}$$

where  $\hat{\psi}_a = \psi_a / \sqrt{P_\psi(\psi_a)}$  is normalised.

Until now, we haven't used axiom (A6). Consider a sequence of sets of outcomes  $A_0 \supset A_1 \dots$ , and denote by  $A \subset A_n \forall n$  the unique maximal subset (possibly empty), such that  $\bar{A} \cap_n A_n = \emptyset$ . Then the difference  $\mathcal{P}_{A_i} - \mathcal{P}_A$  is well defined, and so

$$\begin{aligned}
\langle (\mathcal{P}_{A_i} - \mathcal{P}_A)\psi, (\mathcal{P}_{A_i} - \mathcal{P}_A)\psi \rangle &= P_\psi((\mathcal{P}_{A_i} - \mathcal{P}_A)\psi) \\
&= P_\psi((\mathcal{P}_{A_i} + \mathcal{P}_{\bar{A}} - \mathcal{P}_S)\psi) \\
&= P_\psi(\mathcal{P}_{A_i \cap \bar{A}}).
\end{aligned} \tag{11}$$

By axiom (A6),

$$\lim_{n \rightarrow \infty} \langle (\mathcal{P}_{A_i} - \mathcal{P}_A)\psi, (\mathcal{P}_{A_i} - \mathcal{P}_A)\psi \rangle = 0, \tag{12}$$

so  $\mathcal{P}_{A_i}\psi$  is a Cauchy sequence that converges to  $\mathcal{P}_A\psi \in V$ . Hence  $V$  is complete under the inner product (9). It follows that  $V^*$  is complete also, and is therefore a *Hilbert* space.

Finally, axiom (A3) constrains the form of the evolution operator  $\mathcal{H}$ . Since we suppose that  $\psi_a$  is also a solution of eq 2 (ie that the act of observation does not change the physics of the system),  $\mathcal{H}$  must be linear. The certain event must have probability of 1 at all times, so

$$\begin{aligned}
0 &= \frac{dP_{\psi(t)}(\psi(t))}{dt} \\
&= d/dt \langle \psi, \psi \rangle \\
&= \langle \psi, \mathcal{H}\psi \rangle + \langle \mathcal{H}\psi, \psi \rangle \\
\mathcal{H}^\dagger &= -\mathcal{H},
\end{aligned} \tag{13}$$

i.e.  $\mathcal{H}$  is  $i$  times a Hermitian operator.

Weinberg[10,11] experimented with a possible non-linear generalisation of quantum mechanics, however found great difficulty in producing a theory that satisfied causality. This is probably due to the nonlinear terms mixing up the partitioning  $\{\psi_a, \mu_a\}$  over time. It is usually supposed that causality[9], at least to a certain level of approximation, is a requirement for a self-aware substructure to exist. It is therefore interesting, that relatively mild assumptions about the nature of SASes, as well as the usual interpretations of probability and measure theory lead to a linear theory with the properties we know of as quantum mechanics. Thus we have a reversal of the usual ontological status between Quantum Mechanics and the *Many Worlds Interpretation*.

## Bibliography

1

J. D. Barrow and F. J. Tipler.  
*The Anthropic Cosmological Principle.*

Clarendon, Oxford, 1986.

B. Carter.

The anthropic principle and its implications for biological evolution.  
*Phil. Trans. Roy. Soc. Lond.*, A310:347-363, 1983.

D. L. Cohn.

*Measure Theory*.

Birkhäuser, Boston, 1980.

J. Leslie.

*Universes*.

Routledge, New York, 1989.

J. Leslie.

*The End of the World*.

Routledge, London, 1996.

M. Li and P. Vitányi.

*An Introduction to Kolmogorov Complexity and its Applications*.

Springer, New York, 2nd edition, 1997.

B. Marchal.

Conscience et mécanisme.

Technical Report TR/IRIDIA/95, Brussels University, 1995.

J. Schmidhuber.

A computer scientist's view of life, the universe and everything.

In C. Freska, M. Jantzen, and R. Valk, editors, *Foundations of Computer Science: Potential-Theory-Cognition*, volume 1337 of *Lecture Notes in Computer Science*, pages 201-208. Springer, Berlin, 1997.

M. Tegmark.

Is "the theory of everything" merely the ultimate ensemble theory.

*Annals of Physics*, 270:1-51, 1998.

S. Weinberg.

Testing quantum mechanics.

*Annals of Physics*, 194:336-386, 1989.

S. Weinberg.

*Dreams of a Final Theory*.

Pantheon, New York, 1992.

E. P. Wigner.

*Symmetries and Reflections*.

MIT Press, Cambridge, 1967.

## Acknowledgments

I would like to thank the following people from the "Everything" email discussion list for many varied and illuminating discussions on this and related topics: Wei Dai, Hal Finney, Gilles Henri, James Higgs, George Levy, Alastair Malcolm, Christopher Maloney, Jaques Mallah, Bruno Marchal and Jürgen Schmidhuber.

In particular, the solution presented here to the White Rabbit paradox was developed during an email exchange between myself and Alastair Malcolm during July 1999, archived on the everything list (<http://www.escribe.com/science/theory>). Alastair's version of this solution may be found on his web site at <http://www.physica.freemove.co.uk/p101.htm>.