
Observer Localization in Multiverse Theories

Marcus Hutter

RSISE @ ANU and SML @ NICTA

Canberra, ACT, 0200, Australia

marcus@hutter1.net www.hutter1.net

November 2010

Abstract

The progression of theories suggested for our world, from ego- to geo- to helio-centric models to universe and multiverse theories and beyond, shows one tendency: The size of the described worlds increases, with humans being expelled from their center to ever more remote and random locations. If pushed too far, a potential theory of everything (TOE) is actually more a theories of nothing (TON). Indeed such theories have already been developed. I show that including observer localization into such theories is necessary and sufficient to avoid this problem. I develop a quantitative recipe to identify TOEs and distinguish them from TONs and theories in-between. This precisely shows what the problem is with some recently suggested universal TOEs.

Contents

1	Introduction	2
2	Complete TOEs (CTOEs)	4
3	Complete TOE - Formalization	6
4	Universal TOE - Formalization	7
5	Conclusions	8
	References	9

Keywords

world models; observer localization; predictive power; Ockham's razor; universal theories; computability.

1 Introduction

A number of models have been suggested for our world. They range from generally accepted to increasingly speculative to apparently bogus. For the purpose of this work it doesn't matter where you personally draw the line. Many now generally accepted theories have once been regarded as insane, so using the scientific community or general public as a judge is problematic and can lead to endless discussions: for instance, the historic geo \leftrightarrow heliocentric battle; and the ongoing discussion of whether string theory is a theory of everything or more a theory of nothing. In a sense this paper is about a formal rational criterion to determine whether a model makes sense or not. In order to make the main point of this paper clear, below I will briefly traverse a number of models [Har00, BDH04, Hut09]. The presented bogus models help to make clear the necessity of observer localization and hence the relevance of this work.

Egocentric to Geocentric model. A young child believes it is the center of the world. Localization is trivial. It is always at “coordinate” (0,0,0). Later it learns that it is just one among a few billion other people and as little or much special as any other person thinks of themselves. In a sense we replace our egocentric coordinate system by one with origin (0,0,0) in the center of Earth. The move away from an egocentric world view has many social advantages, but dis-answers one question: Why am I this particular person and not any other?

Geocentric to Heliocentric model. While being expelled from the center of the world as an individual, in the geocentric model, at least the human race as a whole remains in the center of the world, with the remaining (dead?) universe revolving around *us*. The heliocentric model puts Sun at (0,0,0) and degrades Earth to planet number 3 out of 8. The astronomic advantages are clear, but dis-answers one question: Why this planet and not one of the others? Typically we are muzzled by semi-convincing anthropic arguments [Bos02, Smo04].

Heliocentric to cosmological model. The next coup of astronomers was to degrade our Sun to one star among billions of stars in our milky way, and our milky way to one galaxy out of billions of others, according to current textbooks. Again, it is generally accepted that the question why we are in this particular galaxy in this particular solar system is essentially unanswerable.

Multiverses. Many modern more speculative cosmological models (can be argued to) imply a multitude of essentially disconnected universes (in the conventional sense), often each with its own (quite different) characteristic: Examples are Wheeler's oscillating universe, Smolin's baby universe theory, Everett's many-worlds interpretation of quantum mechanics, and the different compactifications of string theory [Teg04]. They “explain” why a universe with our properties exist, since the multiverse includes universes with all kinds of properties, but they cannot *predict* these properties. A multiverse theory *plus* a theory predicting in which universe we happen to live would determine the value of the inter-universe variables for our

universe, and hence have much more predictive power. Again, anthropic arguments are sometimes evoked but are usually vague and unconvincing.

Universal TOE (UTOE). Taking the multiverse theory to the extreme, Schmidhuber [Sch00] postulates a universal multiverse, which consists of *every* computable universe. Clearly, if our universe is computable (and there is no proof of the opposite [Sch00]), the multiverse generated by UTOE contains and hence perfectly describes our own universe, so we have a Theory of Everything (TOE) already in our hands. Unfortunately it is of little use, since we can't use UTOE for prediction. If we knew our "position" in this multiverse, we would know in which (sub)universe we are. This is equivalent to knowing the program that generates *our* universe. This program may be close to any of the conventional cosmological models, which indeed have a lot of predictive power. Since locating ourselves in UTOE is equivalent and hence as hard as finding a conventional TOE of our universe, we have not gained much.

All-a-Carte models. Champernowne's normal number glues the natural numbers, for our purpose written in binary format, 1,10,11,100,101,110,111,1000,1001,... to one long string.

1101110010111011110001001...

Obviously it contains every finite substring by construction. The digits of many irrational numbers like $\sqrt{2}$, π , and e are conjectured to also contain every finite substring. If our space-time universe is finite, we can capture a snapshot of it in a truly gargantuan string u . Since Champernowne's number contains every finite string, it also contains u and hence perfectly describes our universe. Probably even $\sqrt{2}$ is a perfect TOE. Unfortunately, if and only if we can localize ourselves, we can actually use it for predictions. (For instance, if we knew we were in the center of universe 001011011 we could predict that we will 'see' 0010 when 'looking' to the left and 1011 when looking to the right.) Locating ourselves means to (at least) locate u in the multiverse. We know that u is the u 's number in Champernowne's sequence (interpreting u as a binary number), hence locating u is equivalent to specifying u . So a TOE based on normal numbers is only useful if accompanied by the gargantuan snapshot u of our universe. In light of this, such an "All-a-Carte" TOE (without knowing u) is rather a theory of nothing than a theory of everything.

Localization within our universe. The loss of predictive power when enlarging a universe to a multiverse model has nothing to do with multiverses per se. Indeed, the distinction between a universe and a multiverse is not absolute. For instance, Champernowne's number could also be interpreted as a single universe, rather than a multiverse. It could be regarded as an extreme form of the infinite fantasia land from the NeverEnding Story, where everything happens somewhere. Champernowne's number constitutes a perfect map of the All-a-Carte universe, but the map is useless unless you know where you are. Similarly but less extreme, cosmological inflation models produce a universe that is vastly larger than its visible part, and different regions may have different properties.

Predictive power. The exemplary discussion above has hopefully convinced the reader that we indeed lose something (some predictive power) when progressing to too large universe and multiverse models. Historically, the higher predictive power of the large-universe models (in which we are seemingly randomly placed) overshadowed the few extra questions they raised compared to the smaller ego/geo/heliocentric models. But the discussion of the (physical, universal, and all-a-carte) multiverse theories has shown that pushing this progression too far will at some point harm predictive power. We saw that this has to do with the increasing difficulty to localize the observer.

Contents. Classical models in physics are essentially differential equations describing the time-evolution of some aspects of the world. A Theory of Everything (TOE) models the whole universe or multiverse, which should include initial conditions. As argued above, it can be crucial to also localize the observer. I call a TOE with observer localization, a *Complete TOE* (CTOE). Section 2 gives an informal introduction to the necessary ingredients for CTOEs, and how to evaluate and compare them using a quantified instantiation of Ockham’s razor. Section 3 gives a formal definition of what accounts for a CTOE, introduces more realistic observers with limited perception ability, and formalizes the CTOE selection principle. The Universal TOE is a sanity critical point in the development of TOEs, and will be investigated in more detail in Section 4. Important extensions listed in Section 5 are detailed in [Hut09].

2 Complete TOEs (CTOEs)

A TOE by definition is a perfect model of the universe. It should allow to predict all phenomena. Most TOEs require a specification of some initial conditions, e.g. the state at the big bang, and how the state evolves in time (the equations of motion). In general, a TOE is a program that in principle can “simulate” the whole universe. An All-a-Carte universe perfectly satisfies this condition but apparently is rather a theory of nothing than a theory of everything. So meeting the simulation condition is not sufficient for qualifying as a Complete TOE. We have seen that (objective) TOEs can be completed by specifying the location of the observer. This allows us to make useful predictions from our (subjective) viewpoint. We call a TOE plus observer localization a Complete TOE. If we allow for stochastic (quantum) universes we also need to include the noise. If we consider (human) observers with limited perception ability we need to take that into account too. So

A complete TOE needs specification of

- (i) initial conditions
- (e) state evolution
- (l) localization of observer
- (n) random noise

(o) perception ability of observer

We deal with limited perception ability (o) in Section 3. Space prevents discussing stochastic theories (n); they are dealt with in [Hut09]. Next we need a way to compare TOEs.

Predictive power and elegance. Clearly we can never be sure whether a given TOE makes correct predictions in the future. After all we cannot rule out that the world suddenly changes tomorrow in a totally unexpected way. We have to compare theories based on their predictive success in the past. It is also clear that this is not enough: For every model we can construct an alternative model that behaves identically in the past but makes different predictions from, say, year 3000 on. Popper's falsifiability dogma is little helpful. Beyond postdictive success, the guiding principle in designing and selecting theories, especially in physics, is elegance (and mathematical consistency). The predictive power of the first heliocentric model was not superior to the geocentric one, but it was much simpler. In more profane terms, it has significantly less parameters that need to be specified.

Ockham's razor suitably interpreted tells us to choose the simpler among two or more otherwise equally good theories. For justifications of Ockham's razor, see [LV08]. Some even argue that by definition, science is about applying Ockham's razor, see [Hut05]. For a discussion in the context of theories in physics, see [GM94]. It is beyond the scope of this paper to repeat these considerations. In [Hut09] I prove that Ockham's razor is suitable for finding TOEs.

Complexity of a TOE. In order to apply Ockham's razor in a non-heuristic way, we need to quantify simplicity or complexity. Roughly, the complexity of a theory can be defined as the number of symbols one needs to write the theory down. More precisely, write down a program for the state evolution together with the initial conditions, and define the complexity of the theory as the size in bits of the file that contains the program. This quantification is consistent with our intuition, since an elegant theory will have a shorter program than an inelegant one, and extra parameters need extra space to code, resulting in longer programs.

Standard model versus string theory. To give an example, let us pretend that the standard model of particle physics + gravity (P) and string theory (S) both qualify as TOEs. P is a mixture of a few relatively elegant theories, but contains about 20 unexplained parameters that need to be specified (although some regularities can be explained [BH97]). String theory is truly elegant, but ensuring that it reduces to P needs sophisticated extra assumptions (e.g. the right compactification). It would require a major effort to quantify which theory is the simpler one in the sense defined above, but I think it would be worth the effort. It is a quantitative objective way to decide between theories that are (so far) predictively indistinguishable.

CTOE selection principle. It is trivial to write down a program for an All-a-Carte multiverse A. It is also not too hard to write a program for the universal

multiverse U, see Section 4. Lengthwise A easily wins over U, and U easily wins over P and S, but as discussed A and U have serious defects. Given all of the above, it now nearly suggests itself that we should include the description length of the observer location in our TOE evaluation measure. That is,

$$\text{among two CTOEs, select the one that has shorter overall length} \\ \text{Length}(i) + \text{Length}(e) + \text{Length}(l) \tag{1}$$

For an All-a-Carte multiverse, the last term contains the gargantuan string u , catapulting it from the shortest TOE to the longest CTOE.

TOE versus UTOE. Consider any (C)TOE and its program q , e.g. P or S. Since U runs all programs including q , specifying q means localizing (C)TOE q in U. So $U+q$ is a CTOE whose length is just some constant bits (the simulation part of U) more than that of (C)TOE q . So whatever (C)TOE physicists come up with, U is nearly as good as this theory. This essentially clarifies the paradoxical status of U. Naked, U is a theory of nothing, but in combination with another TOE it excels to a good CTOE, albeit slightly longer=worse than the latter.

Localization within our universe. So far we have only localized our universe in the multiverse, but not ourselves in the universe. Assume the about $10^{11} \times 10^{11}$ stars in our universe are somehow indexed. In order to localize our Sun we only need its index, which can be coded in about $\log_2(10^{11} \times 10^{11}) \approx 73$ bits. To localize earth among the 8 planets needs 3 bits. To localize yourself among 7 billion humans needs 33 bits. These localization penalties are tiny compared to the difference in predictive power (to be quantified later) of the various theories (ego/geo/helio/cosmo). This explains and justifies theories of large universes in which we occupy a random location.

3 Complete TOE - Formalization

Objective TOE. Since we essentially identify a TOEs with a program generating a universe, we need to fix some general purpose programming language on a general purpose computer. In theoretical computer science, the standard model is a so-called Universal Turing Machine (UTM) [LV08]. It takes a program coded as a finite binary string $q \in \{0, 1\}^*$, executes it and outputs a finite or infinite binary string $u \in \{0, 1\}^* \cup \{0, 1\}^\infty$. The details do not matter to us, since drawn conclusion are typically independent of them. In this section we only consider q with infinite output

$$\text{UTM}(q) = u_1^q u_2^q u_3^q \dots =: u_{1:\infty}^q$$

In our case, $u_{1:\infty}^q$ will be the universe (or multiverse) generated by TOE candidate q . So q incorporates items (i) and (e) of Section 2. Surely our universe doesn't look like a bit string, but can be coded as one as explained in [Hut09]. We have some

simple coding in mind, e.g. $u_{1:N}^q$ being the (fictitious) binary data file of a high-resolution 3D movie of the whole universe from big bang to big crunch, augmented by $u_{N+1:\infty}^q \equiv 0$ if the universe is finite. Again, the details do not matter.

Observational process and subjective complete TOE. As we have demonstrated it is also important to localize the observer. In order to avoid potential qualms with modeling human observers, consider as a surrogate a (normal not extra cosmic) video camera filming=observing parts of the world. The camera may be fixed on Earth or installed on an autonomous robot. It records part of the universe u denoted by $o = o_{1:\infty}$. (If the lifetime of the observer is finite, we append zeros to the finite observation $o_{1:N}$).

In a computable universe, the observational process within it, is obviously also computable, i.e. there exists a program $s \in \{0, 1\}^*$ that extracts observations o from universe u . Formally

$$\text{UTM}(s, u_{1:\infty}^q) = o_{1:\infty}^{sq} \quad (2)$$

where we have extended the definition of UTM to allow access to an extra infinite input stream $u_{1:\infty}^q$. So $o_{1:\infty}^{sq}$ is the sequence observed by subject s in universe $u_{1:\infty}^q$ generated by q . Program s contains all information about the location and orientation and perception abilities of the observer/camera, hence specifies not only item (l) but also item (o) of Section 2.

A Complete TOE (CTOE) consists of a specification of a (TOE,subject) pair (q, s) . Since it includes s it is a Subjective TOE.

CTOE selection principle. So far, s and q were fictitious subjects and universe programs. Let $o_{1:t}^{true}$ be the past observations of some concrete observer in our universe, e.g. your own personal experience of the world from birth till today. The future observations $o_{t+1:\infty}^{true}$ are of course unknown. By definition, $o_{1:t}$ contains *all* available experience of the observer, including e.g. outcomes of scientific experiments, school education, read books, etc.

The observation sequence $o_{1:\infty}^{sq}$ generated by a correct CTOE must be consistent with the true observations $o_{1:t}^{true}$. If $o_{1:t}^{sq}$ would differ from $o_{1:t}^{true}$ (in a single bit) the subject would have ‘experimental’ evidence that (q, s) is not a perfect CTOE. We can now formalize the CTOE selection principle as follows

$$\begin{aligned} &\text{Among a given set of perfect } (o_{1:t}^{sq} = o_{1:t}^{true}) \text{ CTOEs } \{(q, s)\} \\ &\text{select the one of smallest length } \text{Length}(q) + \text{Length}(s) \end{aligned} \quad (3)$$

Minimizing length is motivated by Ockham’s razor. Inclusion of s is necessary to avoid degenerate TOEs like U and A. The selected CTOE (q^*, s^*) can and should then be used for forecasting future observations via $\dots o_{t+1:\infty}^{forecast} = \text{UTM}(s^*, u_{1:\infty}^{q^*})$.

4 Universal TOE - Formalization

Definition of Universal TOE. The Universal TOE generates all computable universes. The generated multiverse can be depicted as an infinite matrix in which each

row corresponds to one universe.

q	UTM(q)
ϵ	$u_1^\epsilon u_2^\epsilon u_3^\epsilon u_4^\epsilon u_5^\epsilon \dots$
0	$u_1^0 u_2^0 u_3^0 u_4^0 \dots$
1	$u_1^1 u_2^1 u_3^1 \dots$
00	$u_1^{00} u_2^{00} \dots$
\vdots	$\vdots \quad \vdots \quad \vdots$

To fit this into our framework we need to define a single program \check{q} that generates a single string corresponding to this matrix. The standard way to linearize an infinite matrix is to dovetail in diagonal serpentes though the matrix:

$$\check{u}_{1:\infty} := u_1^\epsilon u_1^0 u_2^\epsilon u_3^\epsilon u_2^0 u_1^1 u_1^{00} u_2^1 u_3^0 u_4^\epsilon u_5^\epsilon u_4^0 u_3^1 u_2^{00} \dots$$

Formally, define a bijection $i = \langle q, k \rangle$ between a (program, location) pair (q, k) and the natural numbers $\mathbb{N} \ni i$, and define $\check{u}_i := u_k^q$. It is not hard to construct an explicit program \check{q} for UTM that computes $\check{u}_{1:\infty} = u_{1:\infty}^{\check{q}} = \text{UTM}(\check{q})$.

Remarks. Cutting the universes in bits and interweaving them into one string might appear messy, but is unproblematic for two reasons: First, the bijection $i = \langle q, k \rangle$ is very simple, so any particular universe string u^q can easily be recovered from \check{u} . Second, such an extraction will be included in the localization/ observational process s , i.e. s will contain a specification of the relevant universe q and which bits k are to be observed.

TOE versus UTOE. We can formalize the argument in the last section of simulating TOE by UTOE as follows: If (q, s) is a CTOE, then (\check{q}, \tilde{s}) based on UTOE \check{q} and observer $\tilde{s} := rqs$, where program r extracts u^q from \check{u} and then o^{sq} from u^q , is an equivalent but slightly larger CTOE, since $\text{UTM}(\tilde{s}, \check{u}) = o^{sq} = \text{UTM}(s, u^q)$ by definition of \tilde{s} and $\text{Length}(\check{q}) + \text{Length}(\tilde{s}) = \text{Length}(q) + \text{Length}(s) + O(1)$.

The best CTOE. Finally, one may define the best CTOE (of an observer with experience $o_{1:t}^{\text{true}}$) as

$$UCTOE := \arg \min_{q,s} \{ \text{Length}(q) + \text{Length}(s) : o_{1:t}^{sq} = o_{1:t}^{\text{true}} \}$$

where $o_{1:\infty}^{sq} = \text{UTM}(s, \text{UTM}(q))$. This may be regarded as a formalization of the holy grail in physics; of finding such a TOE.

5 Conclusions

Respectable researchers, including Nobel Laureates, have dismissed and embraced each single model of the world mentioned in the introduction, at different times in history and concurrently. (Excluding All-a-Carte TOEs which I haven't seen

discussed before.) As I have shown, Universal TOE is the sanity critical point. The most popular (pseudo) justifications of which theories are (in)sane have been references to the dogmatic Bible and Popper's limited falsifiability principle. This paper contained a more serious treatment of world model selection. I introduced and discussed the usefulness of a theory in terms of predictive power based on model *and* observer localization complexity. Extensions to more practical and realistic (partial, approximate, probabilistic) theories (rather than TOEs), more motivation and examples, and a proof that Ockham's razor is suitable for finding TOEs can be found in [Hut09].

References

- [BDH04] J. D. Barrow, P. C. W. Davies, and C. L. Harper, editors. *Science and Ultimate Reality*. Cambridge University Press, 2004.
- [BH97] A. Blumhofer and M. Hutter. Family structure from periodic solutions of an improved gap equation. *Nuclear Physics*, B484:80–96, 1997. Missing figures in B494 (1997) 485.
- [Bos02] N. Bostrom. *Anthropic Bias*. Routledge, 2002.
- [GM94] M. Gell-Mann. *The Quark and the Jaguar: Adventures in the Simple and the Complex*. W.H. Freeman & Company, 1994.
- [Har00] E. Harrison. *Cosmology: The Science of the Universe*. Cambridge University Press, 2nd edition, 2000.
- [Hut05] M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2005.
- [Hut09] M. Hutter. A complete theory of everything (will be subjective). 2009. arXiv:0912.5434.
- [LV08] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer, Berlin, 3rd edition, 2008.
- [Sch00] J. Schmidhuber. Algorithmic theories of everything. Report IDSIA-20-00, quant-ph/0011122, IDSIA, Manno (Lugano), Switzerland, 2000.
- [Smo04] L. Smolin. Scientific alternatives to the anthropic principle. Technical Report hep-th/0407213, arXiv, 2004.
- [Teg04] M. Tegmark. Parallel universes. In *Science and Ultimate Reality*, pages 459–491. Cambridge University Press, 2004.