The Subjective Computable Universe

Marcus Hutter

Research School of Computer Science
Australian National University
Canberra, ACT, 0200, Australia
&
Department of Computer Science
ETH Zürich, Switzerland

10 October 2011

Abstract

Nearly all theories developed for our world are computational. The fundamental theories in physics can be used to emulate on a computer ever more aspects of our universe. This and the ubiquity of computers and virtual realities has increased the acceptance of the computational paradigm. A computable theory of everything seems to have come within reach. Given the historic progression of theories from ego- to geo- to helio-centric models to universe and multiverse theories, the next natural step was to postulate a multiverse composed of *all* computable universes. Unfortunately, rather than being a theory of everything, the result is more a theory of nothing, which actually plagues all too-large universe models in which observers occupy random or remote locations. The problem can be solved by incorporating the subjective observer process into the theory. While the computational paradigm exposes a fundamental problem of large-universe theories, it also provides its solution.

Contents

Keywords

world models; observer localization; computability; predictive power; Ockham's razor; universal theories; inductive reasoning; simplicity and complexity.

1 Introduction

The idea of a mechanical universe is quite old and has been expressed by many scholars, including Leibniz and Newton [Dol11], Zuse [Zus69], Schmidhuber [Sch97], Wolfram [Wol02], and many others. With computers and virtual realities increasingly pervading our everyday life, the computable universe metaphor seems to have become ever more accepted. The question of *which* computer program governs our universe, though, remains. Actually it is “just” the age-old quest for a theory of everything (ToE) in a new guise. The new emphasis on computability leads to novel insights and fundamentally new possibilities, but also exhibits new problems.

For instance, consider the simple “All-a-Carte” program that enumerates all natural numbers in binary: 1 10 11 100 101 110 111 1000 ... Assume our space-time universe is finite and somehow coded into a gargantuan but finite bit string. This string will appear somewhere in the above enumeration, hence formally, this simple All-a-Carte program is a theory of everything, but on closer inspection it actually is more a theory of nothing [Sta06]. While this particular theory might simply be dismissed as nonsense, serious proposals of modern multiverse theories like Wheeler’s oscillating universe, Smolin’s baby universe theory, Everett’s many-worlds interpretation of quantum theory, the different compactifications of string theory, and inflationary theories with variable fundamental constants [Teg04] start to get contaminated by similar philosophical problems. The above All-a-Carte model just elucidates the problem in naked and intensified form with all the complex physics stripped off.

The focus of this article is to exploit the opportunities the computable universe paradigm offers, but at the same time avoid its pitfalls.

Section 2 starts with the philosophical prerequisites: how theories or models explain or describe observations, the relation between data compression and prediction, how to avoid difficult epistemological questions about knowledge via the bit-string ontology, how Ockham's razor and information theory solve the induction problem, and the role of observers and their localization in theories of everything. In order to make the main point of this article clear, Section 3 first traverses a number of models that have been suggested for our world, from generally accepted to increasingly speculative and questionable theories, and discusses their relative merits, in particular their predictive power (precision and coverage). We will see that localizing the observer, which is usually not regarded as an issue, can be very important. Section 4 gives an informal introduction to the necessary ingredients for Complete ToEs (CToEs), and how to evaluate and compare them using a quantified instantiation of Ockham's razor. Section 5 gives a slightly more formal definition of what accounts for a CToE, introduces more realistic observers with limited perception ability, formalizes the CToE selection principle, and discusses extensions to more realistic limited theories (rather than ToEs). Section 6 summarizes and discusses the assumptions underlying the CToE selection principle, and Section 7 concludes.

An extended version of this article with technical details has been published in [Hut10].

2 Philosophical Background

This article describes an *information-theoretic* and *computational* approach for addressing the *philosophical* problem of judging theories (of everything) in *physics*. The philosophical prerequisites are introduced in this section. In order to keep it generally accessible, I've tried to minimize jargon, and focus on the core problems and their solution.

Theories/models. By *theory* I mean any *model* which can explain $\approx$ describe $\approx$ predict $\approx$ compress [Hut06] our observations, whatever the form of the model. Scientists often say that their model *explains* some phenomenon. What is usually meant is that the model *describes* (the relevant aspects of) the observations more compactly than the raw data. The model is then regarded as capturing a law (of nature), which is believed to hold true also for unseen/future data.

Induction. This process of inferring general conclusions from example instances is called *inductive reasoning*. For instance, observing 1000 black ravens but no white one supports but cannot prove the hypothesis that all ravens are black. In general, induction is used to find properties or rules or models of past observations. The ultimate purpose of the induced models is to use them for making predictions, e.g. that the next observed raven will also be black. Arguably inductive reasoning is even more important than deductive reasoning in science and everyday life: for scientific discovery, in machine learning [Hut11], for forecasting in economics, as a philosophical discipline, in common-sense decision making, and last but not least to find theories of everything. Historically, some famous, but apparently misguided philosophers [Sto82, Gar01], including Popper and Miller, even disputed the existence, necessity or validity of inductive reasoning. Meanwhile it is well-known how minimum encoding length principles [Wal05, Grü07], rooted in (algorithmic) information theory [Hut07], quantify Ockham's razor principle, and led to a solid pragmatic foundation of inductive reasoning [RH11]. Essentially, one can show that the more one can *compress*, the better one can *predict*, and vice versa.

Theory=model=compressed data. A deterministic theory/model allows from initial conditions to determine an observation sequence, which could be coded as a bit string. For instance, Newton mechanics maps initial planet positions+velocities into a time-series of planet positions. So a deterministic model with initial conditions is "just" a compact representation of an infinite observation string. A stochastic model is "just" a probability distribution over observation strings.

(Compete) theories (of everything). 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 I will argue, it can be crucial to also localize the observer, i.e. to augment the ToE with a model of the properties of the observer, even for non-quantum-mechanical phenomena. I call a ToE with observer localization, a *Complete ToE* (CToE).

The role of observers in previous theories. That the observer itself is important in describing our world is well-known. Most prominently in quantum mechanics, the observer plays an active role in 'collapsing the wave function'. This is a specific and relatively well-defined role of the observer for a particular theory, which is *not* my concern. I will show that (even the localization of) the observer is indispensable for *finding* or developing *any* (useful) ToE. Often, the anthropic principle is invoked for this purpose (our universe is as it is because otherwise we would not exist). Unfortunately its current use is rather vague and limited, if not outright unscientific [Smo04]. It is possible to give a precise and formal account of observers by explicitly separating the observer's subjective experience from the objectively existing universe or multiverse, which besides other things, as pointed out in [Sta06], shows that we also need to localize the observer within our universe (not only which universe the observer is in).

Epistemology. To facilitate this, I will assume that the observers' experience of the world consists of a single temporal binary sequence which gets longer with time. This is definitely true if the observer is a robot equipped with sensors like a video camera whose signal is converted to a digital data stream, fed into a digital computer and stored in a binary file of increasing length. In humans, the signal transmitted by the optic and other sensory nerves could play the role of the digital data stream. Of course (most) human observers do not possess photographic memory. We can deal with this limitation in various ways: digitally record and make accessible upon request the nerve signals from birth till now, or allow for uncertain or partially remembered observations. Classical philosophical theories of knowledge [Alc06] (e.g. as justified true belief) operate on a much higher conceptual level and therefore require stronger (and hence more disputable) philosophical presuppositions. In my minimalist "spartan" information-theoretic epistemology, a bit-string is the only observation, and all higher ontologies are constructed from it and are pure "imagination".

3 Predictive Power & Observer Localization

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 article is about a formal rational criterion to determine whether a model makes sense or not. In order to make the main point of this article clear, below I will briefly traverse a number of models [Har00, BDH04, Hut10]. The presented bogus models help to make clear the necessity of observer localization and hence the relevance of this article.

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 themself. 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 of 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 their 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 (U). 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 (U) contains and hence perfectly describes our own universe, so we have a theory of everything already in our hands. Unfortunately it is of little use, since we can't use (U) 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 (U) is equivalent and hence as hard as finding a conventional ToE of our universe, we have not gained much.

All-a-Carte models (A). In the introduction I have pushed the idea even further: 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.

$$11011100101110111100010 01...$$

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 Never-Ending 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.

4 Complete ToE Selection Principle

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 subjective or 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 will ignore noise and perception ability in the following and resume to these issues in Section 5. Next we need a way to compare ToEs.

Predictive power and elegance. Whatever the intermediary guiding principles for designing theories/models (elegance, symmetries, tractability, consistency), the ultimate judge is predictive success. Unfortunately 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 the latter 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 2020 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 fewer parameters that need to be tuned.

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 article to repeat these considerations. One can show that simpler theories more likely lead to correct predictions, and therefore Ockham's razor is suitable for finding ToEs [Hut10].

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 known as algorithmic information or Kolmogorov complexity [LV08] and 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 [Wal05, Grü07]. From now on I identify theories with programs and write Length(q) for the length=complexity of program=theory q in bits.

Example: standard model+gravity (P) versus string theory (S). To keep the discussion simple, let us pretend that standard model of particle physics plus gravity (P) and string theory (S) each qualify as ToEs. (P) is a mixture of a few relatively elegant theories, but contains about 20 parameters that need to be specified. String theory is truly elegant, but ensuring that it reduces to the standard model needs sophisticated extra assumptions (e.g. the right compactification).

(P) can be written down in one line, plus we have to give 20+ constants, so lets say one page. The meaning (the axioms) of all symbols and operators require another page. Then we need the basics, natural, real, complex numbers, sets (ZFC), etc., which is another page. That makes 3 pages for a complete description in first-order logic. There are a lot of subtleties though: (a) The axioms are likely mathematically inconsistent, (b) it's not immediately clear how the axioms lead to a program simulating our universe, (c) the theory does not predict the outcome of random events, and (d) some other problems. So to transform the description into an e.g. C-program simulating our universe, needs a couple of pages more, but I would estimate around 10 pages overall suffices, which is about 20'000 symbols=bytes. Of course this program will be (i) a very inefficient simulation and (ii) a very naive coding of (P). I conjecture that the *shortest* program for (P) on a universal Turing machine is much shorter, maybe even only one tenth of this. The numbers are only a quick rule-of-thumb guess. If we start from string theory (S), we need about the same length. S is *much* more elegant, but we need to code the compactification to describe our universe, which effectively amounts to the same. Note that everything else in the world (all other physics, chemistry, etc,) is emergent.

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) [Sch00, Hut10]. Lengthwise (A) easily wins over (U), and (U) easily wins over (P) and (S), but as discussed, (A) and (U) have serious defects. On the other hand, these theories can only be used for predictions after extra specifications: Roughly, for (A) this amounts to tabling the whole universe, (U) requires defining a ToE in the conventional sense, (P) needs 20 or so parameters and (S) a compactification scheme. Hence localization-wise (P) and (S) easily win over (U), and (U) easily wins over (A). Given this trade-off, it has been suggested in [Sta06, Hut10], to include the description length of the observer location in our ToE evaluation measure. That is,

among two CToEs, select the one that has shorter overall length

$$\text{Length}(i) + \text{Length}(e) + \text{Length}(l)$$

For an All-a-Carte multiverse, the last term contains the gargantuan string u, catapulting it from the shortest ToE to the longest CToE, hence (A) will not minimize the sum.

ToE versus (U). Consider any ToE and its program q, e.g. (P) or (S). Since (U) runs all programs including q, specifying q means localizing ToE q in (U). So (U)+q is a CToE whose length is just some constant number of bits (the simulation part of (U)) more than that of ToE q. So whatever 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 q, it excels to a good CToE, albeit slightly longer=worse than q.

Localization within our universe. So far we have only localized our universe in the multiverse, but not ourselves in the universe. To localize our Sun, we could e.g. sort (and index) stars by their creation date, which the model (i)+(e) provides. Most stars last for 1-10 billion years (say an average of 5 billion years). The universe is 14 billion years old, so most stars may be 3rd generation (Sun definitely is), so the total number of stars that have ever existed should very roughly be 3 times the current number of stars of about $10^{11} \times 10^{11}$. Probably "3" is very crude, but this doesn't really matter for sake of the argument. In order to localize our Sun we only need its index, which can be coded in about $\log_2(3 \times 10^{11} \times 10^{11}) \doteq 75$ bits. Similarly we can sort and index planets and observers. To localize Earth among the 8 planets needs 3 bits. To localize yourself among 7 billion humans needs 33 bits. Alternatively one could simply specify the (x, y, z, t) coordinate of the observer, which requires more but still only very few bits. These localization penalties (l) 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.

5 Formalization & Extensions

This section formalizes the CToE selection principle and what accounts for a CToE. Universal Turing machines are used to formalize the notion of programs as models for generating our universe and our observations. I also introduce more realistic observers with limited perception ability.

Objective ToE. Since we essentially identify a ToE 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 binary program q, executes it and outputs a binary string u:

$$\text{UTM}(q) = u$$

. The details do not matter to us, since drawn conclusions are typically independent of them. In our case, u will be interpreted as a binary representation of the space-time universe (or multiverse) generated by ToE candidate q. So q incorporates items (i) and (e) of Section 4. Surely our universe doesn't look like a bit string, but can be coded as one as explained below and in more detail in [Hut10]. We have some simple coding in mind, e.g. u being the (fictitious) binary data file of a high-resolution 3D movie of the whole universe from big bang to big crunch. Again, the details do not matter.

Observational process and subjective complete ToE. As I 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 (conventional 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 ω.

I only consider *direct* observations like with a camera. Electrons or atomic decays or quasars are not directly observed, but with some (classical) instrument. It is the indicator or camera image of the instrument that is observed (which physicists then usually interpret). This setup avoids having to deal with any form of informal correspondence between theory and real world, or with subtleties of the quantum-mechanical measurement process. The only philosophical presupposition I make is that it is possible to determine uncontroversially whether two finite binary strings (on paper or file) are the same or differ in some bits.

In a computable universe, the observational process within it, is obviously also computable, i.e. there exists a program s that extracts observations ω from universe u. Formally

$$\text{UTM}(s, u) = \omega$$

where the UTM runs program s on input u to produce observation ω. So ω is the observation by subject s in universe u generated by program 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 4.

> *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 subject and universe programs. Let o be the past "true" observations of some concrete observer in our universe, e.g. your own personal experience of the world from birth till today. The future observations are of course unknown. By definition, o contains *all* available experience of the observer, including e.g. outcomes of scientific experiments, school education, read books, etc.

The proposal $\omega = \omega(q, s)$ generated by a correct CToE must be consistent with the true observations o in the sense that ω starts with o, denoted by $o... = \omega$. If ω would differ from o (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

> *Among a given set of perfect CToEs* $\{(q, s)\}$
> *select the one of smallest Length*(q) + *Length*(s).

The best CToE. Finally, one may define the best CToE (of an observer with experience o) as

$$(q^*, s^*)[o] \;:=\; \arg\min_{q,s}\{\text{Length}(q) + \text{Length}(s) : o... = \text{UTM}(s, \text{UTM}(q))\} \qquad (1)$$

This may be regarded as a formalization of the holy grail in physics; of finding such a ToE.

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 $o... = \text{UTM}(s^*, \text{UTM}(q^*)$ will (by construction) output observation o followed by future observations "$...''$ taken as prediction.

Extensions. The CToE selection principle is applicable to perfect, deterministic, discrete, and complete models q of our universe. None of the existing sane world models is of this kind. But the principle can easily be extended to more realistic, partial, approximate, probabilistic, and/or parametric models for finite, infinite and even continuous universes [Hut10].

Most existing theories only partially model some aspects of our world. Any theory that only predicts parts of our complete observation o, can be augmented, in the simplest case by tabulating all unpredicted bits. The complexity of this table then has to be added to that of q and s in (1). Similarly, for theories that are only approximately correct, one can table or code their errors and include them in (1). Some theories like quantum mechanics make only probabilistic predictions. A theory that predicts universe u with probability $Q(u)$ and observation o in universe u with probability $S(o|u)$, induces a probability distribution $P(o) = \sum_u S(o|u)Q(u)$ over observations. The observed noise can then be coded in $|\log_2 P(o)|$ bits [Wal05] to be added to (1) as indicated in item (n) of Section 4. Many theories in physics also depend on real-valued parameters. They need to be specified to some minimally sufficient accuracy [Grü07] and their code length added to (1). Theories of infinite or continuous spaces like 3+1 dimensional Minkowski space can be discretized to arbitrary precision. Such discretization is always possible, since all spaces occurring in physical theories are separable. An even more fundamental solution to construct countable models is to use a result by Loewenheim and Skolem [Sch00]. A final note on pluralistic approaches favoring multiple theories on multiple scales for different (overlapping) application domains: While currently in fashion and convenient in practice, they have fundamental consistency problems and cannot serve jointly as theories of everything, unless reconciliated in a reductionist fashion.

See [Hut10] for examples and details.

6 Discussion of the Assumptions

I will now discuss the assumptions which led to the CToE selection principle; more precisely, under which assumptions the principle will result in good models for our universe. I have argued in this article that the assumptions are sufficient for constructing sensible theories of everything.

> *(i) Bit-string ontology:* The observers' raw experience of the world can be cast into a single temporal binary sequence o. All other physical and epistemological concepts are derived.

What exactly knowledge is and how humans acquire it is philosophically still controversial. Discussions revolve around justification, truth, and belief, which are themselves subtle concepts. These problems are avoided by operating on the much lower ontological "data" level, namely that of an observer capable of perceiving an ordered stream of bits, and nothing else. For a robot or a human in a cyber-world we can take the binarized and linearized data stream from all sensors. For a human in the real world we can approximately digitally record all raw sensory input. All higher-level interpretations of this bit-string (leading to what is traditionally called 'knowledge') is theory-laden, where suitable theories are induced via CToE selection.

> *(ii) Realism:* There exists an objective world independent of any particular observer in it.

Although solipsists claim that the world and other minds do not exist outside and independent of their own mind, and idealists place ideas and spiritual experience at the center of existence, pragmatically, realism is the least controversial assumption. It can (much better than the alternatives) explain many features of our experience e.g. evolution and society. In any case, the CToE selection principle is powerful enough to determine whether the a-priori assumption of an objective world is warranted ($\text{Length}(q^*) \gg 0$) or not ($\text{Length}(q^*) \approx 0$). Indeed, it can determine precisely which and how much of our experience should be ascribed to an objective world (namely q^*) and what is subjective (namely s^*).

> *(iii) Computable universe:* The world is computable, i.e. there exists an algorithm (a finite binary string) which, when executed, outputs the entire space-time universe.

The idea of a mechanical universe is old and has been expressed by many researchers [Zus69, Sch97, Wol02]. The more we understand our world, the more plausible it becomes. As indicated in Sections 3 and 4, physical theories describe (aspects of) our universe with increasing precision and scope. All those theories are computable in the sense that they can simulate the physical phenomena on a computer, although quantum randomness and aspects of string theory complicate this picture. Given this trend, it is natural to conjecture that the total world u is computable. Note that this assumption implicitly assumes (i.e. implies) that temporally stable binary strings exist, which connects it with assumption (i).

> *(iv) Computable observer process:* The observer is a computable process within the objective world.

If/since the universe is computable, then an observer who is part of it, is obviously computable too, hence the observation bit-string o should be a computable function of the universe u. This is not at odds with free will [Hut05, Sec.8.6.3]. The important point is to acknowledge that the observer process is important, even if we are/were only interested in objective theories or aspects of the world. Note that observer localization is neither based on the controversial anthropic principle, nor has it anything to do with the quantum-mechanical observation process, although there may be some deeper yet to be explored connections.

> *(v) Ockham's razor principle:* Choose the simplest theory consistent with the observations.

Ockham's razor principle has so far been invaluable for understanding our world. The assumption is that the models selected by it will continue to lead to most-likely-correct predictions. Minimum encoding length principles, rooted in (algorithmic) information theory, quantify Ockham's razor principle, and have led to a quantitative, pragmatic and universal foundation of inductive reasoning [RH11]. Indeed, it seems to be a necessary and sufficient founding principle of science itself, in contrast e.g. to the popular but insufficient falsifiability principle. Until other necessary and sufficient principles are found, it is prudent to accept Ockham's razor as the foundation of science. A-priori justifications of Ockham's razor are possible too, but of course they must rest themselves on (other) assumptions. For instance, if one assumes that a-priori all universe and observer programs (q, s) are equally likely, then one can show that Ockham's razor 'works' [Hut10, Sec.8]. Hence one could replace (v) by this so-called universal self-sampling assumption (not to be confused with informal anthropic arguments or the no free lunch myth [LH11]).

7 Conclusions

The computational paradigm exposed a fundamental problem of large-universe theories, which could be overcome by taking serious the role of the observer. I discussed a quantitative method of world model selection by analyzing the usefulness of a theory in terms of predictive power based on model *and* observer localization complexity. In particular I have shown the following:

- Unlike falsificationism, quantified versions of Ockham's razor can serve as the foundation of science.
- A theory that perfectly describes our universe or multiverse, rather than being a Theory of Everything (ToE), might also be a theory of nothing.
- A predictively meaningful theory can be obtained if the theory is augmented by the localization of the observer.
- A truly Complete Theory of Everything (CToE) (q, s) consists of a conventional (objective) ToE q plus a (subjective) observer process s.
- The bit-string ontology, realism, computability, subjectivism, and Ockham's razor quantified in terms of code-length minimization enable a scientifically meaningful and systematic quest for a theory of everything.

- More precisely, the CToE Selection Principle allows a rigorous and quantitative comparison of CToEs and can even be used to select the "best" CToE (q^*, s^*).
- As a side result, this allows to separate objective knowledge q from subjective knowledge s.
- One might even argue that if q^* is non-trivial, this is evidence for the existence of an objective reality.
- Another side result is that there is no hard distinction between a universe and a multiverse; the difference is qualitative and semantic.

References

[Alc06] N. Alchin. *Theory of Knowledge.* John Murray Press, 2nd edition, 2006.

[BDH04] J. D. Barrow, P. C. W. Davies, and C. L. Harper, editors. *Science and Ultimate Reality.* Cambridge University Press, 2004.

[Bos02] N. Bostrom. *Anthropic Bias.* Routledge, 2002.

[Dol11] E. Dolnick. *The Clockwork Universe: Isaac Newton, the Royal Society, and the Birth of the Modern World.* Harper, 2011.

[Gar01] M. Gardner. A skeptical look at Karl Popper. *Skeptical Inquirer*, 25(4):13–14,72, 2001.

[GM94] M. Gell-Mann. *The Quark and the Jaguar: Adventures in the Simple and the Complex.* W.H. Freeman & Company, 1994.

[Grü07] P. D. Grünwald. *The Minimum Description Length Principle.* The MIT Press, Cambridge, 2007.

[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.

[Hut06] M. Hutter. Human knowledge compression prize, 2006. open ended, http://prize.hutter1.net/.

[Hut07] M. Hutter. Algorithmic information theory: a brief non-technical guide to the field. *Scholarpedia*, 2(3):2519, 2007.

[Hut10] M. Hutter. A complete theory of everything (will be subjective). *Algorithms*, 3(4):329–350, 2010.

[Hut11] M. Hutter. Universal learning theory. In C. Sammut and G. Webb, editors, *Encyclopedia of Machine Learning*, pages 1001–1008. Springer, 2011.

[LH11] T. Lattimore and M. Hutter. No free lunch versus Occam's razor in supervised learning. In *Proc. Solomonoff 85th Memorial Conference*, *LNAI*, Melbourne, Australia, 2011. Springer.

[LV08] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications.* Springer, Berlin, 3rd edition, 2008.

[RH11] S. Rathmanner and M. Hutter. A philosophical treatise of universal induction. *Entropy*, 13(6):1076–1136, 2011.

[Sch97] J. Schmidhuber. A computer scientist's view of life, the universe, and everything. In *Foundations of Computer Science: Potential - Theory - Cognition*, volume 1337 of *LNCS*, pages 201–208. Springer, Berlin, 1997.

[Sch00] J. Schmidhuber. Algorithmic theories of everything. Report IDSIA-20-00, arXiv:quant-ph/0011122, IDSIA, Manno (Lugano), Switzerland, 2000.

[Smo04] L. Smolin. Scientific alternatives to the anthropic principle. Technical Report hep-th/0407213, arXiv, 2004.

[Sta06] R. Standish. *Theory of Nothing.* BookSurge Publishing, 2006.

[Sto82] D. C. Stove. *Popper and After: Four Modern Irrationalists.* Pergamon Pres, 1982.

[Teg04] M. Tegmark. Parallel universes. In *Science and Ultimate Reality*, pages 459–491. Cambridge University Press, 2004.

[Wal05] C. S. Wallace. *Statistical and Inductive Inference by Minimum Message Length.* Springer, Berlin, 2005.

[Wol02] S. Wolfram. *A New Kind of Science.* Wolfram Media, 2002.

[Zus69] K. Zuse. *Rechnender Raum.* Friedrich Vieweg & Sohn, Braunschweig, 1969. English translation: *Calculating Space,* MIT Technical Translation AZT-70-164-GEMIT, Massachusetts Institute of Technology (Proj. MAC), Cambridge, Mass. 02139, Feb. 1970.