Marcus Hutter - A Complete Theory of Everything (will be subjective) (2010)

Tom Rochette <tom.rochette@coreteks.org>

December 21, 2025 — 77e1b28a

0.1 Context

0.2 Learned in this study

0.3 Things to explore

1 Overview

2 Notes

2.1 1 Introduction

- Theory is meant as "any model which can explain/describe/predict/compress our observations, whatever the form of the model"
- One can show that the more one can compress, the better one can predict, and vice versa
- ToE: Theory of Everything
- CToE: Complete Theory of Everything, includes observer localization
- It will be shown that the observer is indispensable for finding or developing any (useful) ToE

2.2 4 Completes 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)
- A complete ToE needs specification of
  - (i) initial conditions
  - (e) state evolution
  - (l) localization of observer
  - (n) random noise
  - (o) perception ability of observer
- Among two CToEs, select the one that has shorter overall length
  - Length(i) + Length(e) + Length(l)

2.3 5 Complete ToE - Formalization

- Universal Turing Machine (UTM)
- $UTM(q) = u_1^q u_2^q u_3^q ... =: u_{1:\infty}^q$
- $u_{1:\infty}^q$ is the (fictitious) binary data file of a high-resolution 3D movie of the whole universe from the big bang to big crunch, augmented by $u_{N+1:\infty}^q \equiv 0$ if the universe is finite
- A camera records part of the universe u denoted by $o = o_{1:\infty}$

- The only philosophical pressuposition made 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
- $UTM(q, u^q_{1:\infty}) = o^{sq}_{q:\infty}$
  - $u^q_{1:\infty}$ an infinite input stream
  - $o^{sq}_{q:\infty}$ the sequence observed by subject s in universe $u^q_{1:\infty}$ 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 4
- $o^{true}_{1:t}$ the past observations of some concrete observer in our universe (e.g. your own personal experience of the world from birth till today)
- $o^{true}_{t+1:\infty}$ are unknown
- The observation sequence $o^{sq}_{q:\infty}$ generated by a correct CToE must be consistent with the true observation $o^{true}_{1:t}$
- If $o^{sq}_{q:\infty}$ would differ from $o^{true}_{1:t}$ (in a single bit) the subject would have "experimental" evidence that (q, s) is not a perfect CToE
- Among a given set of perfect ($o^{sq}_{q:\infty} = o^{true}_{1:t}$) CToE $\{(q, s)\}$, select the one of smallest length Length(q) + Length(s)

2.4 6 Universal ToE - Formalization

- 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 (and the column represent the observation at time t)
- The standard way to linearize an infinite matrix is to dovetail in diagonal serpentines through the matrix

$$\breve{u}_{1:\infty} := u^{\epsilon}_1 u^0_1 u^{\epsilon}_2 u^{\epsilon}_3 u^0_2 u^1_1 u^{00}_1 u^1_2 u^0_3 u^{\epsilon}_4 u^{\epsilon}_5 u^0_4 u^1_3 u^{00}_2 \ldots$$

- Define a bijection $i = \langle q, k\rangle$ between a (program, location) pair (q, k) and the natural number $i \in \mathbb{N}$, and define $\breve{u}_i := u^q_k$. We can then construct an explicit program $\breve{q}$ for UTM that computes $\breve{u}_{1:\infty} = u^q_{1:\infty} = UTM(\breve{q})$
- One may define the best CToE (of an observer with experience $o^{true}_{1:t}$) as

$$UCTOE := \arg\min_{q,s} \left\{ Length(q) + Length(s) : o^{sq}_{1:t} = o^{true}_{1:t} \right\}$$

- where $o^{sq}_{1:\infty} = UTM(s, UTM(q))$

2.5 7 Extensions

- Partial theories: Let $o^{true}_{1:t}$ be the complete observation, and (q, s) be some theory explaning only some observations but not all. The other bits in $o^{qs}_{1:t}$ are undefined. We can augment q with a (huge) table b of all bits for which $o^{qs}_i = o^{true}_i$. Together, (q, b, s) allows to reconstruct $o^{true}_{1:t}$ exactly. Hence, for two different theories, the one with smaller length should be selected

$$Length(q) + Length(b) + Length(s)$$

- Some proponents of pluralism and some opponents of reductionism argue that we need multiple theories on multiple scales for different (overlapping) application domains. They argue that a ToE is not desirable and/or not possible.
- Consider two Theories (T1 and T2) with (proclaimed) applications domain A1 and A2, respectively
  - If predictions of T1 and T2 coincide on their intersection $A1 \cap A2$ (or if A1 and A2 are disjoint), we can trivially "unify" T1 and T2 to one theory T by taking their union. Of course, this does not result in any simplification, i.e. if $Length(T) = Length(T1) + Length(T2)$, we gain nothing. But since nearly all modern theories have some common basis, e.g. use natural or real numbers, a formal unification of the generating programs nearly always leads to $Length(q) < Length(q_1) + Length(q_2)$

- The interesting case is when T1 and T2 lead to different forecasts on $A1 \cap A2$. For instance, particle versus wave theory with the atomic world at their intersection, unified by quantum theory. Then we need a reconciliation of T1 and T2, that is, a single theory T for $A1 \cup A2$. Ockham's razor tells us to choose a simple (elegant) unification. This rules out naive/ugly/complex solutions like developing a third theory for $A1 \cap A2$ or attributing parts of $A1 \cap A2$ to T1 or T2 as one sees fit, or averaging the prediction of T1 and T2. Of course T must be consistent with the observations
- One problem with pluralism: which principle should one use in a concrete situation?

3 8 Justification of Ockham's Razor

- Ockham's razor could be regarded as correct if among all considered theories, the one selected by Ockham's razor is the one that most likely leads to correct predictions
- $Q_L := \{q : \text{Length}(q) \leq L \text{ and } \text{UTM}(q) = u_{1:t}^{true}*\}$
  - $*$ is any continuation of $u_{1:t}^{true}$
- $|Q_L| \approx 2^{L-l}$
  - Q_L is the set of all consistent universes (which is non-empty for large L)
  - L is a given length limit
  - l is the length of the shortest description of these consistent universes
- We are most likely in a universe that is (equivalent to) the simplest universe consistent with our past observations

4 See also

5 References

- Hutter, Marcus. "A complete theory of everything (will be subjective)." Algorithms 3.4 (2010): 329-350.
- http://arxiv.org/abs/0912.5434