In this work, I present a formal construction of the axiomless position. This construction ultimately leads to the theory of everything in physics (ToE). The purpose of this paper is to first derive the ToE from the axiomless position, then to derive sufficient physics from it so as to convince the skeptical reader of the validity of the methodology.

Part I is the axiomless derivation of the ToE. From this derivation I obtain, in part II, a master equation formulated as a Gibb’s ensemble and relating the algorithmic notions of program-observables to that of entropy.

Part III is the derivation of physical laws so as to provide evidence that it is the ToE. I recover, in an axiomless manner, the exact mathematical formulation of the major theories of physics; including statistical mechanics, quantum mechanics (QM), special and general relativity (GR). These equations are derived entirely from pure reason with no appeal to physical observations. Deriving both QM and GR as necessary consequences and limiting cases of this equation is suggestive of unification.

Contents

1 Philosophy 3
   1.1 Primary motivation 3
   1.2 Explanatory gap 3
   1.3 No uniqueness theorem 4
   1.4 State of physics 4
   1.5 Summary of the argument 4

I An axiomless derivation of the Theory of Everything 6
   1.6 Axioms 6
   1.7 Language 7
   1.8 Arithmetic 14
   1.9 Translation 17
   1.10 Multiplication 18
   1.11 Set theory 21
   1.12 Representation of any theory of knowledge 22

2 The Theory of Everything 24
   2.1 Elimination of physically impossible theories 24
   2.2 The universal theory of knowledge 25
1 Philosophy

What is the axiomless position?

Plato\(^1\) recognized that most of the disagreements in philosophy are ultimately linked to the choice of assumptions made by the parties involved. He believed that by grinding away at those assumptions, one could recover a kind of universal truth. He believed that this universal truth, comprised of whatever survives the grinding process, could ultimately be used to build a logical framework in a manner that is **entirely irrefutable**. Furthermore, any argument that would be constructed exclusively from this logical framework will inherit its irrefutable character.

The goal of obtaining such a framework was revisited by René Descartes in 1641. He used a universal doubt method to obtain the 'cogito ergo sum' or 'I think, therefore I am', claiming the existence of the thinking self to be an absolute truth and suggested that it should be used as the foundation of philosophy.

The mathematician David Hilbert, in 1900, posed 23 problems to the mathematical community. Hilbert’s second problem is of interest here. He challenged mathematicians to establish a system of axioms to explain all of science and mathematics and then to prove that this system is consistent. In this spirit, Bertrand Russell published Principia Mathematica in 1910, proving that logic was the same as mathematics.

However, in 1930, Kurt Gödel would publish a proof on the incompleteness of arithmetic. His theorem proved that any formal system strong enough to construct self-referential sentences will necessarily have true, but unprovable statements. Hence it would be incomplete. This suggests that the axiomatic approach is highly problematic.

In this work, the axiomless position refers to the construction of a theory of logic in such a manner as to not contain any axioms assumed to be true. All foundational statements of the theory will be undeniably proven. Hence, as it will be devoid of assumptions, the irrefutable character argued by Plato will apply to it.

1.1 Primary motivation

A theory of everything (ToE) constructed exclusively from the axiomless position will inherit its irrefutable character. This is the primary motivation for this work.

1.2 Explanatory gap

An axiomatic ToE will necessarily have a foundational gap in its ability to explain the universe. Indeed, no axiomatic theory can explain

why its axioms are true over other axioms. The proposed ToE, as it will be derived from the axiomless method, will avoid this explanatory gap.

**Remark 1.1.** The falsifiability argument is a notable non-constructive exception consisting of claiming that these axioms best reproduce the scientific observations to date. But if such an argument is used to justify its axioms, then the theory cannot provably be the ToE and it would forever be an potentially intermediary theory.

### 1.3 No uniqueness theorem

An axiomatic ToE cannot have a global uniqueness theorem. Any uniqueness theorem it may contain will ultimately rest upon the validity of its axioms, and could be false under different axioms. Only an axiomless ToE can fully answer the question; why this theory and not another?

### 1.4 State of physics

The formal theories of physics are built upon axioms which are not proven from reason but are instead justified by a series of experiments or observations. For example, why is the speed of light a constant? Because of the failure of the Michelson–Morley experiment. And so on. Since our observational capabilities are limited by our technology, the consequence of building physics upon a series of observations is that we end up with multiple logically independent theories which don’t quite fit together.

### 1.5 Summary of the argument

This paper is best summarized by the following two results.

The first result is a theory of logic constructed in such a manner as to inherit the irrefutable character argued by Plato. Such theory is not constructed from axioms, but from primitive theorems. We will call this theory sentence theory.

The second result is a method to construct a ToE from this theory. The result is an equation irrefutably proven to be the ToE in the sense given by Plato.

As this equation is derived from pure reason yet makes physical claims, it will be best to immediately summarize, as a sketch, the key insight allowing such derivation from pure reason. First we consider all statements of the type \( k \vdash t \), where \( k \) is a set of axioms and \( t \) is a theorem,

\[
k \vdash t
\]  
(1.2)
Then, on the premises that

1. for all $k$ and $t$ where $k \vdash t$ holds, a mathematician can, in principle, print out an effective proof of such, and

2. said mathematician resides in the universe described by the ToE,

then it follows that for any possible ToE, the following statement must hold.

$$\forall k \forall t [(k \vdash t) \rightarrow \text{ToE} \vdash (k \vdash t)]$$

Hence, the ToE must be a universal theory of knowledge \footnote{Defined in theorem 2.1}. In plain english, this sentence means that for all $k$ and $t$, if $k$ proves $t$ then the ToE must prove that $k$ proves $t$.

In this paper, I prove that, as unintuitive as it might seem, for the statement 1.3 to hold, the laws of physics must also hold. The result obtained is constructive of the laws of physics and allows us to derive them from pure reason. It does not appear possible to derive laws that do not occur in nature, hence the construction is not over-predictive.

We will limit ourselves to showing how the restriction imposed by 1.3 implies both general relativity and quantum mechanics and show how its unification resolves singularities in black holes. Deriving dark energy and the standard model will not be done in this paper and will be investigated by the author at a later time.
Part I

An axiomless derivation of the Theory of Everything

1.6 Axioms

In this work, we will use the following definition for axioms.

**Definition 1.4 (Axiom).** An axiom is an unprovable sentence of a language that can be true or false within a formal logic system, but is considered to be true without proof.

We emphasize the underlined elements of the definition;

1. An axiom must be an unprovable sentence. This is required otherwise we could call any theorem an axiom which would negate the distinction.

2. An axiom must be a sentence that can be true or false. This prevents tautologies, necessary truths and contradictions from being axioms. For example, tautologies are considered to be theorems because they are provably always true.

3. The choice of a formal logic system must be established before we can formulate a correct axiom for it. For example, the sentences of first-order logic are not compatible with propositional logic hence writing axioms as correct sentences comes after the choice of a logic system.

In this work, we will use first order logic to write axioms and statements. However, to prove these statements we will not use the full facilities of first order logic. Instead, we will only make use of a maximally reduced logic system.

To prove first order logic sentences, we will only accept a direct construction of the object the existence of which is claimed by the sentence. As a trivial example, we could prove the existence of the symbol 1, represented in first order logic by $\exists x [\text{isSymbol}(x) \land (x = 1)]$ by writing 1 as the proof. Sentences proven that way are said to be primitive theorems.

**Definition 1.5 (Primitive Theorem).** A primitive theorem is a sentence of first order logic that is provable by direct construction of the object the existence of which it claims. Theorems proven exclusively from other primitive theorems are also primitive theorems.
Remark 1.6 (Primitive theorems are not axioms). *An axiom is an unprovable sentence of a language, whereas primitive theorems are provable with a proof by construction.*

Definition 1.7 (Axiomless theory). *A theory is axiomless if it is constructed exclusively from primitive theorems.*

The flexibility of this proof method is significantly diminished from the more expressive formal logic systems that are commonly in use. To reduce assumptions to a minimum, we have gotten rid of most principles of logic such as the law of excluded middle, etc. We only preserve the proof by construction as anything less would be too weak to prove anything.

In a very real sense the pen and paper we used to write formal logic statements becomes a source of undeniable evidence for some statements about the existence of language, symbols and their properties. As we are limited to only constructing sentences or symbols to offer as proof, it is natural to use language as a starting point to build upon.

1.7 Language

It is quite difficult, and most likely impossible, to even imagine a concept that cannot be formulated in language. As such, language is a very general concept - perhaps even the most general of all concepts.

The limits of my language mean the limits of my world [...] Whereof one cannot speak, thereof one must be silent.

–Ludwig Wittgenstein

So how will we define language? The first potential concern is that we are unavoidably using language to define language. This might seem circular, but it is a consequence of how fundamental the concept is. Furthermore, this will allow us to prove the existence of language axiomlessly, so for us it will actually be an advantage.

A real problem however is that of the infinite regression of the definitions. Suppose we define a language by its symbols. Then how do we then define symbols? Do we say that symbols are unique identifiers? If so, then what is an identifier - is it a shape in one’s mind? If so, what is a mind, or a shape? This goes on forever. To break the cycle we introduce a precision cutoff in our definitions and we instead assume that the reader knows intuitively what is talked about.

Formally speaking, these cutoffs are *primitive notions.*

Definition 1.8 (Primitive notion). *A primitive notion is a term that we use but that we do not define. The term should be understood by a mixture*
of examples, intuition and by the theorems and definitions that result from its usage.

For example, Euclidean geometry under Hilbert’s axioms has six recognized primitive notions; point, line, plane, congruence, betweeness, and incidence. Set theory has two; set and element of.

To define language, we will introduce two primitive notions.

**Primitive Notion 1.9** (Symbol). A symbol is a unique distinguishable identifier. It can be a shape, a sound, a sign, etc. By definition, there are only two predicates for a symbol and both are primitive predicates. They are,

\[
\begin{align*}
\text{isSymbol}(x) & \quad \text{(true if } x \text{ is a symbol)} \\
\text{areDistinguishable}(x, y) & \quad \text{(true if } x \text{ and } y \text{ are different symbols)}
\end{align*}
\]

**Primitive Notion 1.10** (Sentence). A sentence consists of taking multiple symbols and joining them together in a single group or unit. The order of occurrence of the symbols in the sentence matters and repetitions are allowed. The primitive predicate related to the sentence is

\[
\text{isSentence}(p) \quad \text{(true if } p \text{ is a sentence)}
\]

**Remark 1.11.** For completeness, single symbols as well as the empty sentence are also considered to be sentences. The empty sentence will be denoted with $\epsilon$.

**Remark 1.12.** Unlike symbols, sentences can have many more predicates. For example, containsSymbol($x, p$) could be one, isLength($p, n$) could be another, etc. We will not concern ourselves with these at this point. If we need any of these later on, we will define them from within the theory itself.

We will now pose definitions relying on these primitive notions.

**Definition 1.13** (Alphabetical sentence). An alphabetical sentence $S_\alpha$ is a sentence of finite length with no repetitions of symbols.

**Definition 1.14** (Language). A language $L$ is defined by a specific alphabetical sentence $S_\alpha$ such that;

1. The order of occurrence of the symbols in $S_\alpha$ is the alphabetical order of $L$.
2. If a sentence contain symbols not present in $S_\alpha$, then it is not a sentence of $L$.

As examples, the following languages (on the left) are defined by their $S_\alpha$ (on the right).
nullary := \varepsilon \quad \text{(1.15)}
unary := 1 \quad \text{(1.16)}
binary := 01 \quad \text{(1.17)}
ternary := 012 \quad \text{(1.18)}
decenary := 0123456789 \quad \text{(1.19)}
roman alphabet := abcdefghijklmnopqrstuvwxyz \quad \text{(1.20)}
roman numbers := IVXG \quad \text{(1.21)}

Remark 1.22 (Notation). When writing the sentences of a language, we adopt the following notational conventions to eliminate ambiguity. For example, ambiguity occurs if we write 10; is it the decimal number ten, or the binary number two.

• If we list all possible sentences of a language from shortest to longest and from alphabetical first to alphabetical last, we suffix the sentences with 1 for unary, 2 for binary, 3 for ternary, 4 for quaternary, etc. For example, binary would be enumerated as \varepsilon_2, 0_2, 1_2, 00_2, 01_2, 10_2, etc.

• If we list all possible sentences of a language according to positional notation, we suffix the sentences with u for unary, b for binary, t for ternary, q for quaternary, V for quintary, VI for sextary, etc. For example, positional binary would be enumerated as \varepsilon_b, 0_b, 1_b, 10_b, 11_b, etc.

• We note that for unary, both enumeration methods are identical hence the suffix u can be used interchangeably with 1. However, to my eye 1_1 reads a bit more confusing than 1_u, therefore we will pick u for unary in this work.

• By convention, we will name positional decenary to be decimal and no suffix is used for its sentences. The sentences of decimal are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc.

Remark 1.23. Positional notation skips some sentences from its enumeration, notably those with leading zeros. More on that later in the section on multiplication.

If we were to construct a conventional axiomatic theory of knowledge, we would pose our first axioms at this stage. A favorite is usually the axiom of existence taking the paraphrased form of: "We pose that a simple language exists, such as the binary language". Then using this axiom and possibly other axioms we would derive theorems, such as the existence of other languages, etc. For example, in a certain formulation of set theory, the axiom of the empty set takes the following form:
Remark 1.24 (Axiom of the empty set in set theory). In some formulation of set theory, an axiom is introduced to obtain a first set.

\[ \exists x \forall y (\neg (y \in x)) \]

In other formulations of set theory, the axiom of subsets, or the axiom of an infinite set takes the place of the generator of the first set.

Here however, we are not planning to describe sets or numbers but language. Since language is required to describe any formal system of logic, its existence is guaranteed (within logic). In the theory of knowledge that we will introduce, all initial statements of the theory will be provable as primitive theorems. Hence no axioms will be used in the making of this theory.

This is a unique property of using language to describe language. It cannot be used with any other abstract notions such as with sets. The sets of set theory although describable via language are nonetheless too abstract for language to provide a proof of their existence by construction thereof. Hence, set theory must be derived from axioms taken as true.

Definition 1.25 (Laws of sentence theory).

\[ \exists x [\text{isSymbol}(x)] \]  
(Existence of a symbol)

\[ \exists x \exists y [\text{isSymbol}(x) \land \text{isSymbol}(y) \land \text{areDistinguishable}(x, y)] \]  
(Existence of another symbol)

\[ \exists p [\text{isSentence}(p)] \]  
(Existence of a sentence)

\[ \forall x [\text{isSymbol}(x) \rightarrow \exists c [\text{isSentence}(c) \land (c x = x)]] \]  
(Existence of the empty sentence)

\[ \forall p [\text{isSentence}(p) \rightarrow \exists x [\text{isSymbol}(x) \land (x := p)]] \]  
(Unrestricted definitions)

Equality of symbols

\[ \forall x [\text{isSymbol}(x) \rightarrow (x = x)] \]  
(Reflexivity)

\[ \forall x \forall y [\text{isSymbol}(x) \land \text{isSymbol}(y) \rightarrow (x = y \rightarrow y = x)] \]  
(Symmetry)

\[ \forall x \forall y \forall z [\text{isSymbol}(x) \land \text{isSymbol}(y) \land \text{isSymbol}(z) \rightarrow (((x = z) \land (x = y)) \rightarrow y = z)] \]  
(Transitivity)

Equality of sentences

\[ \forall p [\text{isSentence}(p) \rightarrow (p = p)] \]  
(Reflexivity)

Rule to construct well-formed sentences

\[ \forall x \forall p [\text{isSymbol}(x) \land \text{isSentence}(p) \rightarrow \text{isSentence}(px)] \]  
(Concatenation)

Remark 1.26. First order logic is used only to write down the above laws in a clear and unambiguous manner. We do not use the full facilities of first-order logic to prove theorems.
Remark 1.27. The only proof method that we accept is the simplest proof method possible: direct construction of the object the existence of which we want to prove.

Remark 1.28. The law of unrestricted definitions is to be interpreted in a similar manner as the deduction rule of first order logic. Once specific definitions are posed, the sentences that are provable as a result of those definitions are only so within the definitions posed. We note the following special case.

Definition 1.29 (Theorem). If a theorem \( t \) of an axiomatic theory of knowledge \( k \) is proven within sentence theory but necessitate an appeal to the law of unrestricted definitions to do so, then \( \exists k [k \vdash t] \) is a primitive theorem of sentence theory, but \( t \) is not.

In what follows, we will prove the laws of sentence theory as primitive theorems.

Primitive Theorem 1.30 (Existence of a symbol).

\[ \exists x [\text{isSymbol}(x)] \]

Proof. We offer a proof by construction.

1

\[ \square \]

Primitive Theorem 1.31 (Existence of another symbol).

\[ \exists x \exists y [\text{isSymbol}(x) \land \text{isSymbol}(y) \land \text{areDistinguishable}(x, y)] \]

Proof. We offer a proof by construction.

0

\[ \square \]

Primitive Theorem 1.32 (Existence of a sentence).

\[ \exists p [\text{isSentence}(p)] \]

Proof. We offer a proof by construction.

1

\[ \square \]

Primitive Theorem 1.33 (Equality of symbols).

\[ \forall x [\text{isSymbol}(x) \rightarrow (x = x)] \] \hspace{1cm} (1.34)
\[ \forall x \forall y [\text{isSymbol}(x) \land \text{isSymbol}(y) \rightarrow (x = y \rightarrow y = x)] \] \hspace{1cm} (1.35)
\[ \forall x \forall y \forall z [\text{isSymbol}(x) \land \text{isSymbol}(y) \land \text{isSymbol}(z) \rightarrow (((x = z) \land (x = y)) \rightarrow y = z)] \] \hspace{1cm} (1.36)
Proof. Equality is definable from second order logic by the following principles of Leibniz:

\[
\forall x \forall y [x = y \rightarrow \forall P (Px \leftrightarrow Py)] \quad \text{(Indiscernibility of identicals)}
\]

\[
\forall x \forall y [\forall P (Px \leftrightarrow Py) \rightarrow x = y] \quad \text{(Identity of indiscernibles)}
\]

From the primitive notion 1.9 of the symbol - a unique distinct identifier, there exists two and only two predicates available to describe a symbol. The first one is \(\text{isSymbol}(x)\) and it is true if and only if \(x\) is a symbol. The second is \(\text{areDistinguishable}(x, y)\) and it is true if and only if \(x\) is distinct from \(y\).

For any symbol and from these two predicates, it is clear that the Leibniz principles are respected when we use the same symbol twice. As reflexivity, symmetry and transitivity of the equality are derivable from the Leibniz principles, we have proven the theorem.

Remark 1.37. The fact that \(\text{areDistinguishable}(x, y)\) is a “soft” converse of the notion of equality does not invalidate the proof nor does it render it circular. The predicate \(\text{areDistinguishable}(x, y)\) has the same value \(\forall x \forall y (x = y)\) therefore it does respect the Leibniz principles and the proof is valid.

Primitive Theorem 1.38 (Existence of the empty sentence).

\[
\forall x [\text{isSymbol}(x) \rightarrow \exists e [\text{isSentence}(e) \land (ex = x)]]
\]

(1.39)

Proof. Since \(e\) is defined as a blank statement it follows that replacing it with its definition yields \(ex \leftrightarrow x\), hence \(ex = x\).

Primitive Theorem 1.40 (Unrestricted definitions).

\[
\forall p [\text{isSentence}(p) \rightarrow \exists x [\text{isSymbol}(x) \land (x := p)]]
\]

Proof. We offer a proof by construction. Suppose a language \(L\) with \(n\) symbols \(s_1, s_2, \ldots, s_n\), where \(n\) is finite. We can construct every possible sentence of the language from shortest to longest in alphabetical order as shown in table 1.

To each of the sentences herein constructed, we associate a unique symbol. For example, we can associate symbols to sentences as such:

\[
\begin{align*}
\triangle & := \epsilon \\
\square & := s_1 \\
\circ & := s_2 \\
\vdots & \\
\end{align*}
\]

(1.41) (1.42) (1.43)

For each newly constructed sentence we associate a symbol different than the previous one. Without loss of generality, this can be done by

Table 1: We construct all possible sentences of \(L\) as shown here.
using the regular polygons starting from the triangle and each time adding an additional side to the shape. We never run out of symbols and we associate every sentences to a unique symbol.

**Primitive Theorem 1.44** (Reflexive equality of sentences).

\[ \forall p [\text{isSentence}(p) \rightarrow (p = p)] \]

**Proof.**

\[
\begin{align*}
    x &= x & \quad & \text{(Reflexive equality of symbols)} \\
    x &:= p & \quad & \text{(Unrestricted definitions)} \\
    (x = x) &\iff (p = p) & \quad & \text{(Replacement of } x \text{ by its definition)} \\
    p &= p & \quad & \text{ }(x = x \text{ is always true)}
\end{align*}
\]

We will now prove more advanced theorems from the basic laws.

**Primitive Theorem 1.45.** There exist the unary language

**Proof.** As a proof by construction, take the language generated by the following alphabetical sentence

\[ 1 \]

**Definition 1.46** (Unary language). *The unary language is defined by the alphabetical sentence of one symbol: 1. Some examples of the sentences of unary are: ε, 1_u, 11_u, 111_u, \ldots. The subscript u (for unary) is optional but it is added in this work to avoid confusion with other usages of the symbol 1.*

Other languages can be defined (and proved) in a similar manner. For example, binary.

**Primitive Theorem 1.47.** There exists the binary language

**Proof.** As a proof by construction, take the language generated by the following alphabetical sentence

\[ 01 \]

**Definition 1.48** (Binary language). *The binary language is defined by the alphabetical sentence of two symbols: 01. Some examples of the sentences of binary are: ε_2, 0_2, 1_2, 00_2, 01_2, 10_2, 11_2, 000_2, \ldots. The subscript 2 (for binary) is optional but it is added in this work to avoid confusion with other usages of the symbols 0 and 1.*
Primitive Theorem 1.49. There exists the natural numbers

Proof. The natural numbers are defined by recursion against the starting element, such that the successor of any natural number is also a natural number. To recover a model using the unary language, we

1. associate the empty sentence with the starting element,
2. associate each successively longer sentence with the next element.

\[ \square := 1_u \land \square := 1 \implies 1 = 1_u, \]
\[ \bigcirc := 11_u \land \bigcirc := 2 \implies 2 = 11_u, \]
\[ \bigcirc := 111_u \land \bigcirc := 3 \implies 3 = 111_u, \]
\[ \vdots \]

To translate between different notations of natural numbers such as from unary to decimal we can pose the following definitions

\[ \triangleq := e_u \land \triangleq := 0 \implies 0 = e_u \quad (1.50) \]

Definition 1.51 (Alphabetical position of a sentence). We repeat theorem 1.49 for other languages. To every sentence of a language we associate a natural number in the following way:

1. The number 0 is the empty sentence \( \epsilon \).
2. The number 1 is the alphabetical first symbol.
3. If there are more symbols, then the next number is associated to the alphabetical second symbol. Otherwise, it is associated to the sentence with two symbols.
4. And so on.

Each natural number is associated to a corresponding sentence from shortest to longest and from alphabetical first to alphabetical last.

Table 2 shows the alphabetical enumeration of select languages along with their alphabetical position indicated on the left in decimal.

1.8 Arithmetic

Definition 1.52 (Disjoin). A disjoin is made by inserting a period "." between two symbols of a sentence.
To avoid confusion with the usage of the period in normal text, sentences that are disjoined will be enclosed in brackets when used in normal text. Example: $[111.11_u]$.  

**Definition 1.53** (Join). *A join is made by removing a period "." between two symbols of a sentence.*  

Since the period symbol "." is not a symbol of the language of the sentences that are being split, the presence or the absence of the symbol does not affect the equality. For example, $[111.111_u]$ is concatenated to $[1111111_u]$, and $[111.111_u] = [1111111_u]$. The "." symbol should be seen in a manner similar to a comment in computer code. Although it does not have any impact on the laws of unrestricted language, it will nonetheless significantly improve the readability of the coming proofs.  

**Primitive Theorem 1.54.** *Joining any two sentences produces another sentence.*  

*Proof.* The proof will be offered for the unary language, but it is trivial to generalize it to other languages. The unary suffix $u$ will be omitted.  

We now list the possible joins of the unary language.

\[
\begin{bmatrix}
\epsilon.\epsilon = \epsilon & \epsilon.1 = 1 & \epsilon.11 = 11 & \ldots \\
1.\epsilon = 1 & 1.1 = 11 & 1.11 = 111 & \ldots \\
11.\epsilon = 11 & 11.1 = 111 & 11.11 = 1111 & \ldots \\
& \vdots & \vdots & \vdots \\
\end{bmatrix}
\]

The period "." behaves like comments in code hence are erasable from the equalities. Let us see an example of when they are removed.
111 = 11.1\epsilon \\
111 = 111\epsilon \quad \text{(removing .)} \\
111 = 111 \quad \text{(removing } \epsilon \text{)}

Disjoining a sentence with the period, or by inserting \( \epsilon \), has no impact on the equality. \(\square\)

**Primitive Theorem 1.55.** The joins over unary are commutative.

*Proof.* It is easy to see in the table for the proof of theorem 1.54, each element of the table has a corresponding commuted and equal join. For example, both \( [1_u.\epsilon_u] = [1_u] \) and \( [\epsilon.1_u] = [1_u] \) are present. This is the case for all terms. Furthermore, the diagonal terms are equal to their commutation. \(\square\)

**Primitive Theorem 1.56.** The joins over the unary is addition, where the symbol + is replaced by the disjoin symbol “.” (period).

*Proof.* The Peano axioms of addition are;

\[
\begin{align*}
a + 0 &= a \\
a + S(b) &= S(a + b)
\end{align*}
\]

(1.57) (1.58)

**Lemma 1.59.**

\[ a + 0 = a \]

*Proof.* Without loss of generality, we pose \( a = 1 \).

\[
\begin{align*}
1 + 0 &= 1_u + \epsilon_u \\
&= 1_u.\epsilon_u \\
&= 1\epsilon_u \\
&= 1_u \\
&= 1 \\
\text{(decimal to unary)} \\
\text{(definition of addition)} \\
\text{(join)} \\
\text{(elimination of } \epsilon \text{)} \\
\text{(unary to decimal)}
\end{align*}
\]

\(\square\)

**Lemma 1.60.**

\[ a + S(b) = S(a + b) \]

*Proof.* Without loss of generality, we pose \( a = 1 \) and \( b = 2 \) and \( S(b) = 3 \).

\[
\begin{align*}
1 + 3 &= 1_u + 111_u \\
&= 1.111_u \\
&= 1111_u \\
&= 4 = S(1 + 3) \\
\text{(decimal to unary)} \\
\text{(definition of addition)} \\
\text{(join)} \\
\text{(unary to decimal)}
\end{align*}
\]

\(\square\)
1.9 **Translation**

**Definition 1.61** (Translation sentence). A translation sentence is a binary sentence that can be used to arbitrarily associate each sentence of a language to a matching sentence in another language.

**Primitive Theorem 1.62.** There exists a translation sentence.

*Proof.* Take any binary sentence $S_2$. This is how $S_2$ can be interpreted as a translation sentence from $L_1$ to $L_2$.

We disjoin the sentence after any occurrence of the symbol 0. Each segment is unitary encoded. For example, suppose the $S_2$ is $10110111100...$. The disjoined segments are $\{10, 110, 11110, 0,...\}$. Each segment can be made to correspond to a natural number such that:

\[
\begin{align*}
0 &:= 0 \\
1 &:= 10 \\
2 &:= 110 \\
3 &:= 1110 \\
4 &:= 11110 \\
&\vdots
\end{align*}
\]

We then interpret the translation sentence as follows: the first unitary segment of $S_2$ associates the first sentence of $L_1$ to the $n^{th}$ sentence of $L_2$, where $n$ is the natural number associated with the unitary segment, and so on.

Using this construction, we can arbitrarily translate any sentence of $L_1$ to $L_2$.

**Definition 1.64** (Encoding). An encoding is a bijective translation sentence to and from the same language $L$.

**Definition 1.65** (Unitary encoding). Unitary encoding maps every sentence of binary to the following sentences of binary

\[
\{0_2, 10_2, 110_2, 1110_2, 11110_2, \ldots \}
\]

The translation sentence from binary to unitary encoding is

\[
10111101111111110... \tag{1.66}
\]

**Theorem 1.67.** All languages can be encoded in unitary.

*Proof.* Using an appropriate translation sentence, first translate any language $L$ to binary. Then using the unitary translation sentence (1.66), translate a second time from the binary language to the unitary encoding.
1.10 Multiplication

We can prove the existence of a model \( t \) for an axiomatic theory \( k \) within sentence theory if we can produce such a model using the law of unrestricted definitions. This will now be done for multiplication. We then obtain a primitive theorem of the form \( \exists k ((k \in L) \land (k \vdash t)) \) which has a similar interpretation as the deduction rule in first order logic.

For the purposes of multiplication, the alphabetical enumeration (shown in Table 2) is of no help here. We must introduce a restricted enumeration (shown in Table 3) which removes certain sentences from the list. We will call this enumeration the positional enumeration. It should not come as a surprise that some sentences must be removed from the enumeration as we introduce multiplication because, as is well-known, any theory of arithmetic with multiplication must be incomplete (Gödel).

There are two main culprits responsible for the need for this restricted enumeration.

1. The first one is the multiplication by 1, or in binary: 0\(_b\). To multiply 1 as a join, we obtain 0\(_b\) \times 1\(_b\) = 0.1\(_b\) = 01\(_b\). But this must also be equal to 1\(_b\) because 1 is the identity of multiplication e.g. (01\(_b\) = 1\(_b\) under multiplication). As a workaround we define 1 as \( \epsilon \) instead of as 0 in Table 3 for languages of 2 or more symbols.

2. The other culprit is that the concatenation associated to a multiplication is similar to the bit-shift operation of binary numbers producing a multiplication by 2. For the bit-shift operation to be connected to a multiplication by 2, we require that the natural numbers must be associated to binary sentences under the rule that 001\(_b\) = 01\(_b\) = 1\(_b\), etc.

Definition 1.68 (Multiplication). We define the multiplication of two natural numbers as a join of two sentences of a language enumerated according to table 3, such that:

- The multiplication of any \( n \in \mathbb{N}^+ \) by 2 is the join of \( n \) in binary with 1\(_b\).
- The multiplication of any \( n \in \mathbb{N}^+ \) by 3 is the join of \( n \) in ternary with 2\(_t\).
- The multiplication of any \( n \in \mathbb{N}^+ \) by 4 is the join of \( n \) in quaternary with 3\(_q\).
- The multiplication of any \( n \in \mathbb{N}^+ \) by 5 is the join of \( n \) in quinary with 4\(_v\).
- etc.
Positional enumeration

<table>
<thead>
<tr>
<th></th>
<th>nullary</th>
<th>unary := 1</th>
<th>p-binary := 01</th>
<th>p-ternary := 012</th>
<th>p-decenary* := 0123456789</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>e</td>
<td>e</td>
<td>≠</td>
<td>≠</td>
<td>≠</td>
</tr>
<tr>
<td>1</td>
<td>e</td>
<td>1</td>
<td>e</td>
<td>e</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>e</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>e</td>
<td>111</td>
<td>10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>e</td>
<td>1111</td>
<td>11</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>e</td>
<td>11111</td>
<td>100</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>e</td>
<td>111111</td>
<td>101</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>e</td>
<td>1111111</td>
<td>110</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>e</td>
<td>11111111</td>
<td>111</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>e</td>
<td>111111111</td>
<td>1000</td>
<td>22</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>e</td>
<td>1111111111</td>
<td>1001</td>
<td>100</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3: A natural number is associated to each sentence of select languages. Since this is the positional enumeration, each sentence with leading zeros is removed from the enumeration.

Already we recover all possible multiplication operations for all \( n > 1 \). Binary, ternary, quinary, septenary, etc. are the prime languages. We are left with the need to define the multiplication by and of 0 and multiplication by 1 as special cases.

We start with multiplication by and of 0.

- The multiplication of any \( n \in \mathbb{N} \) by 0 is the join of \( n \) in nullary with \( e \).
- The multiplication of 0 by any \( n \in \mathbb{N} \) is the join of \( e \) with \( n \) in nullary.

Now we define the multiplication of 1. The unary language is already defined for arithmetic. We therefore cannot use it for the multiplication by 1. Instead we define multiplication by 1 as such

- The multiplication of any \( n \in \mathbb{N}^+ \) by 1 is the join of 1 in \( n \)-ary with the alphabetical last symbol of \( n \)-ary.

All other joins are undefined under multiplication.

Using these operations, we can prove their existence by writing a sentence of a language that has the required properties. For example, the sentence \([10.2_t] = [102_t]\) is the multiplication of 4 by 3.

\[
4 \times 3 = 4 \times 3 \\
= 10_t \times 2_t \quad \text{(conversion from decimal to ternary)} \\
= 10.2_t \quad \text{(definition of multiplication)} \\
= 102_t \quad \text{(join)} \\
= 12 \quad \text{(conversion from ternary to decimal)}
\]

Commutation is respected
3 × 4 = 3 × 4
= \(10_b \times 11_b\) (conversion from decimal to binary)
= 10.11_b (definition of multiplication)
= 1011_b (join)
3 × 4 = 12 (conversion from binary to decimal)

**Theorem 1.69.** The join over Table 3 as defined in 1.68 is the arithmetic multiplication.

**Proof.** The Peano axioms of multiplication are;

\[
a \times 0 = 0
\]
\[
a \times S(b) = a + (a \times b)
\]

**Lemma 1.70.**

\[a \times 0 = 0\]

**Proof.** Without loss of generality, we pose \(a = 1\).

\[a \times 0 = 1 \times 0\]
\[= \epsilon_n \times \epsilon_n\] (conversion from decimal to nullary)
\[= \epsilon \epsilon_n\] (definition of multiplication)
\[= \epsilon \epsilon_n\] (join)
\[= \epsilon_n\] (elimination of redundant \(\epsilon\))
\[= 0\] (conversion from nullary to decimal)

**Lemma 1.71.**

\[a \times S(b) = a + (a \times b)\]

**Proof.** Without loss of generality, we pose \(a = 1, b = 2\) and \(S(b) = 3\).

\[a \times S(b) = 1 \times 3\]
\[= \epsilon \times 2_t\] (conversion decimal to ternary)
\[= \epsilon.2_t\] (definition of multiplication by 3)
\[= \epsilon 2_t\] (join)
\[= 2_t\] (elimination of \(\epsilon\))
\[1 \times 3 = 3\] (conversion from ternary to decimal)

It is a theorem that, in the context of an appropriate set of unrestricted definitions, multiplication is proven to exist.
1.11  **Set theory**

We can repeat the process used for multiplication but now for set theory. We first find a context for unrestricted definitions which produces a model for set theory. We will only offer a sketch here.

**Sketch 1.72. Set theory**

**Lemma 1.73. Equality**

\[ \forall x \forall y[\forall z(z \in x \leftrightarrow z \in y) \implies x = y] \]

*Proof.* We can create a set \( S \) of \( n \) elements by taking the alphabetical sentence of a language with \( n \) symbols, then posing all permutation of the alphabetical sentence to be equal to \( S \). Taking as an example the case where \( n = 3 \):

\[
S := 012 \\
:= 021 \\
:= 102 \\
:= 120 \\
:= 201 \\
:= 210
\]

We can iterate over the different formulations of the same set switching to \( S \) and back, such as

\[
012 = S = 102
\]

\[\square\]

**Lemma 1.76. Subset**

*Proof.* Subsets are simply the possible joins and disjoins of \( S \), where we replace the "." with \( \cup \).

\[
S := 01.2 = 01 \cup 2 = 0.12 = 0 \cup 12 \\
:= 02.1 = 02 \cup 1 = 0.21 = 0 \cup 21 \\
:= 10.2 = 10 \cup 2 = 1.02 = 1 \cup 02 \\
:= 12.0 = 12 \cup 0 = 1.20 = 1 \cup 20 \\
:= 20.1 = 20 \cup 1 = 2.01 = 2 \cup 01 \\
:= 21.0 = 21 \cup 0 = 2.10 = 2 \cup 10
\]

\[\square\]

A question we might have at this point is, can sentence theory "talk" about any theory? We will see that the answer is yes. Have we got ourselves something too general to be useful? In the next two sections, we will see that this is not the case at all.
1.12 Representation of any theory of knowledge

We define two languages $Q$ and $A$. Without loss of generality, suppose that the languages are binary.

\[
\begin{array}{ccc}
Q & A \\
q_1 := e_2 & a_1 := e_2 \\
q_2 := 0_2 & a_2 := 0_2 \\
q_3 := 1_2 & a_3 := 1_2 \\
q_4 := 00_2 & a_3 := 00_2 \\
& \\
\vdots & \
\end{array}
\]

Although these languages look identical, no equality for languages has been defined in sentence theory. If a definition of equality for languages is desired, then it must be derived from the law of unrestricted definitions. From $Q$, $A$ and this law, we can define theories of knowledge.

**Definition 1.79 (Theory of knowledge).** We create a theory of knowledge by associating to every $q_i$ either an $a_j$, or we mark it as undefined $\notin$.

**Remark 1.80.** Although for languages with the same number of symbols we can pose the trivial case $\forall n(q_n := a_n)$, it gets more interesting for other cases.

**Primitive Theorem 1.81.** Each $q_i := a_j$ definition that we pose is a claim that there exists a theory of knowledge $k$ composed of a group of axioms such that for all defined $q_i$ and its associated answer $a_j$, this holds

\[
\exists k[\forall q_i (k \vdash (q_i = a_j))]
\]

**Proof.** It is insightful when we can compress $k$ in such a way that very few axioms prove all associations. For the proof of this theorem however, it suffices to say that each $q_i$ to $a_j$ association could itself be an axiom of $k$. \qed

**Definition 1.82 (Q2A-sentence).** A Q2A-sentence is a translation sentence in binary that maps each $q_i$ to an $a_j$. Every binary string can be interpreted as a theory of knowledge.

A Q2A-sentence contains all the knowledge otherwise obtainable from the axiomatic representation of the theory. Taking the analogy where $Q$ represents the question and $A$ is the answer, the Q2A-sentence encodes the answer to every question of the language.
Remark 1.83. Each Q2A-sentence is a binary sentence. Since we identify sentences by assigning them a natural number (alphabetical position), we can only identify countably infinitely many sentences. However, there are uncountably infinitely many theories of knowledge. Neglecting to include these theories will however not be a problem. Indeed, being able to express countably infinitely many theories of knowledge is enough to recover all of those which can be expressed with recursively enumerable axioms or better. The ones which are left out are "random junk".

Remark 1.84. If a specific $q_i$ would be mapped to an infinitely long $a_j$, we would claim that $q_i$ is undefined rather than map to it. This ensures that each unitary encoded segment of the Q2A-sentence is finite and well behaved.

Remark 1.85. We reserve the first unitary segment $0_2$ to correspond to the undefined state such that any $q_i$ mapped by it is undefined.

As an example, consider the following binary sentence

$$11010010001110110 \ldots$$

(1.86)

Its unitary encoded segments are

$$110, 10, 0, 10, 0, 0, 1110, 110, \ldots$$

(1.87)

From this sentence, $Q$ is mapped to $A$ as

$$q_1 := a_2 \quad (110)$$

(1.88)

$$q_2 := a_1 \quad (10)$$

$$q_3 := \not{a} \quad (0)$$

$$q_4 := a_1 \quad (10)$$

$$q_5 := \not{a} \quad (0)$$

$$q_6 := a_3 \quad (1110)$$

$$q_7 := a_2 \quad (110)$$

$$\vdots$$

Definition 1.89 (UTK). A universal theory of knowledge (UTK) is a theory of knowledge $k$ which embeds every recursively enumerable theory of knowledge. By embedding, we mean that the following first order sentence holds,

$$\forall k \forall t [(k \vdash t) \rightarrow U \vdash (k \vdash t)]$$

In plain English, it means that for all set of axioms $k$, if $k$ proves a theorem $t$, then a universal theory of knowledge is able to prove that $k$ proves $t$. 
Primitive Theorem 1.90. Sentence theory is a UTK.

Proof. This is a trivial consequence of the existence of arbitrary Q2A-sentences. ☐

Theorem 1.91. First order arithmetic is a UTK.

Proof. Consistent with the Gödel numbering method and for each well formed formula of arithmetic, we associate to it a natural number. Then, for each recursively enumerable theory of knowledge \( k \) we pose a series of equations to be used on the natural numbers. The equations are posed such that they transform the natural number to another natural number in a manner consistent with the rules of inference of \( k \).

This can be repeated for all possible recursively enumerable axiomatic theories of knowledge. As a result, arithmetic is a UTK. ☐

2 The Theory of Everything

Primitive theorems are provable only by direct construction of the object the existence of which we want to prove or by invoking other primitive theorems. As a result and in a certain sense, the pen and the paper we use to write becomes a laboratory where that which we can construct can be used as evidence of its existence. Of course writing about pink elephants does not prove their existence. But writing very carefully about language using language does. Indeed and as shown, in a properly formulated pure theory of language (e.g. the laws of sentence theory) this “exception” is sufficient to prove all of its axiom-like laws as primitive theorems.

As a result a theory which only concerns itself with language (and no other abstract entities) is a special case of logic for which an axiomless derivation is possible.

2.1 Elimination of physically impossible theories

In this work, our goal is to find the theory of everything in physics. So far, we have obtained a way to list all possible theories of knowledge that can be formulated in a finite language. Since they are all in the list, then necessarily, one of them is the theory of everything describing our universe. What we need now is an argument to eliminate all non-ToEs from consideration.

Since this work is done within the universe we are trying to explain, it must be the case that the ToE explaining the universe cannot contradict the existence of that which we primitively prove in it. Stated differently, any primitive theorem of sentence theory must be theorems of the theory of everything.
For example, if we prove the existence of a symbol by writing 1, then it cannot be the case that \( \neg \exists x (\text{isSymbol}(x) \land (x = 1)) \) is a theorem of the theory of everything. As a result we can eliminate all of the theories which would produce such a theorem from being ToE candidates on the grounds that they contradict the primitive knowledge herein derived. The more ToE-candidate we filter out using primitive knowledge, the closer we get to the real ToE.

What if we use absolutely every available primitive theorems, of which there are infinitely many, to filter out as many ToE-candidates as possible? We will see that we exactly recover a condition which is both necessary and sufficient for a ToE. It is in fact enough by itself to recover the physics, all of the physics and nothing but the physics.

**Primitive Theorem 2.1.** The Theory of Everything must be a universal theory of knowledge.

**Proof.** If the ToE is not a UTK, then a mathematician cannot do his job. Indeed, using a pen and a paper a mathematician is free to pose any axiomatic theory \( k \) and take it to its logical conclusions. Therefore, all theorems of the form

\[
k \vdash t
\]

must be theorems of the theory of everything, or the ToE would contradict that which can provably be done with a pen and paper. Since mathematicians presumably reside in the universe governed by this ToE, and mathematicians can in principle recover any theorem provable within a UTK, then the ToE must be a UTK. And this sentence must hold.

\[
\forall k \forall t [(k \vdash t) \rightarrow \text{ToE} \vdash (k \vdash t)]
\]  

\[\square\]

All we have to do now is show that the laws of physics are a necessary consequence of any UTK. Then we have an axiomless derivation of the ToE.

### 2.2 The universal theory of knowledge

To make the laws of physics come out of the UTK, we will reformulate the UTK within the framework of a universal Turing machine (UTM). Doing this will unlock the formalism of algorithmic information theory and other UTM-related theorems to help us out.

To achieve this, recall definition 1.82 on the representation of a theory of knowledge from a \( Q2A \)-sentence of the binary language.
In this scenario, the Q2A-sentence is interpreted as a translation sentence from language Q to language A such that each translated sentence poses a single of definition from Q to A, or leaves it undefined.

The Q2A-sentence representing an axiomatic theory does in fact contain all the information of the theory. If by analogy we suppose that Q stands for the question and A stands for the answer, we can find the answer to question n by looking up the n\textsuperscript{th} translation segment. Knowledge of the specific axiomatic theory k is not required as a compatible formulation of it is necessarily recoverable.

If, however, we do have knowledge of k, we can simplify the Q2A-sentence greatly to a new sentence. This new sentence will be called the Ω-sentence of the theory. This sentence will be how we will recover a link to the UTM formulation.

**Definition 2.4 (Ω-sentence).** An Ω-sentence is a binary sentence where b(i), the i\textsuperscript{th} bit of the sentence, corresponds to q\textsubscript{i} such that:

\[
b(i) = \begin{cases} 
0 & q_i \text{ is undefined} \\
1 & \text{otherwise}
\end{cases}
\]

**Theorem 2.5.** The Ω-sentence together with the axiomatic formulation of the theory k are sufficient to correctly recover the Q2A-sentence in finite time.

**Proof.** To see how, first consider that we obtain the definition of a UTM by performing the following replacements;

- k is the program.
- q\textsubscript{i} is the input.
- a\textsubscript{j} is the output.

Then for all identifiable k, there exists an encoding such that a UTM can read k and q\textsubscript{i} as the input and produce a\textsubscript{j} as the output. Furthermore, knowing the Ω-sentence, the UTM can skip over non-halting problems. This allows it to recover any number (\(< \infty\)) of leading bits of the Q2A-sentence in finite time. As a result, the Q2A-sentence is computable from a UTM having knowledge of the Ω-sentence.

**Remark 2.6.** It is much faster to find answers with the Q2A-sentence than it is from the Ω-sentence. Indeed,

- Finding the answer to q\textsubscript{i} using the Q2A-sentence is a simple lookup of the i\textsuperscript{th} unitary encoded segment of the sentence. This segment points to the answer a\textsubscript{j}. This is a relatively fast search algorithm.
• Using the $\Omega$-sentence involves considerably more effort. First we find the bit corresponding to the question we are interested in. If it is 0, we know there is no answer and we can stop immediately. If instead it is 1, we begin the search. To find the answer, it is necessary to first try to find a proof, in dovetail, that the answer is $a_1$, then $a_2$, then $a_3$ and so on until we find it. We know that eventually we will find an answer because its $\Omega$-sentence bit is 1 hence the answer exists.

Here we assume that speed is not relevant. If we are interested only in logical completeness (and we are) then both sentences offer that. We will use the $\Omega$-sentence formulation as it is more convenient for the theorems we will later prove.

2.3 The halting probability

The formalism presented so far was useful to prove, axiomlessly, that the ToE is a UTK. As all UTK will yield the same ToE, we are free to choose any UTK to work with. Let us select the very uncontroversial first-order arithmetic as our choice. The implication will be that the laws of physics are a necessary consequence of first-order arithmetic.

We will now introduce a formalism, within first-order arithmetic, to facilitate the mathematical treatment.

**Definition 2.7.** We define $S(i)$ as the $i^{th}$ sentence of the alphabetical enumeration of language $L$.

We know that some sentences are provable, but we do not know which ones. To represent this formally, we will define a two-state identification function $T(i)$ as

**Definition 2.8.** For each sentence $S(i)$

$$T(i) = \begin{cases} 0 & S(i) \text{ is a theorem} \\ \infty & S(i) \text{ is not a theorem} \end{cases}$$

The reason for the choice of $\infty$ or 0 as the states, as opposed to say 0 and 1, will become clear when we start to write equations in the form of a sum.

**Theorem 2.9.** If we take the number $i$ corresponding to each sentence and weigh it according to a power probability distribution in base 2, we can write

$$1 = \sum_{i=1}^{\infty} 2^{-i} \quad (2.10)$$

**Proof.** Expanding the sum in binary, we get
\[
\sum_{i=1}^{\infty} 2^{-i} = 2^{-1} + 2^{-2} + 2^{-3} + \ldots \\
= 0.1_b + 0.01_b + 0.001_b + 0.0001_b + \ldots \\
= 0.1_b \\
= 1_b
\]  
\(2.11\)

\(2.12\)

\(2.13\)

\(2.14\)

**Theorem 2.15.** Further refining this equation we add the term \(T(i)\) to capture only the theorems while excluding all other sentences from the sum,

\[\Omega = \sum_{i=1}^{\infty} 2^{-T(i)}2^{-i}\]

**Proof.** Again expanding the sum into binary, we get

\[\Omega = \sum_{i=1}^{\infty} 2^{-T(i)}2^{-i}\]

\(2.16\)

\(2.17\)

\(2.18\)

\(2.19\)

We obtain a number \(\Omega\) where its bits are in a one-to-one correspondence with the sentences given by \(S(i)\). If \(S(i)\) is a theorem of \(L\), the \(i^{th}\) bit of \(\Omega\) is 1, otherwise it is 0. The reason why \(T(i)\) uses 0 and \(\infty\) as its two states is now clear. It is to remove the terms that are not theorems from the sum by making them vanish to 0.

This definition of \(\Omega\) is equivalent to the Gregory Chaitin’s \(\Omega\) number construction of algorithmic information theory associated with the halting probability of a prefix-free universal Turing machine.

**Remark 2.20** (The Universe). The number \(\Omega\) represents all the knowledge obtainable for a language \(L\) according to some rules of inference. If we define the universe by its knowledge content, there is a \(\Omega_U\) corresponding to our universe. This definition will be made explicit in the next sections as we consider that it is more convenient to represent the universe by its knowledge content, rather than it is by its rules of inference.

### 2.4 The universal Turing machine

Switching to the universal Turing machine at this stage gives us the benefit of unlocking the formalism of algorithmic information theory.
Since we are now working with Turing machines it is more appropriate to use the word ‘program’ instead of ‘sentence’, and to use the word ‘halts’ instead of ‘is a theorem’. For example, instead of saying ‘this sentence is a theorem’, we will say ‘this program halts’. It is equivalent but more appropriate when referring to Turing machines.

**Theorem 2.21.** \( \Omega \), if calculated from a UTM, is the halting probability of a prefix-free UTM.

**Proof.** The proof will be divided into 2 lemmas.

**Lemma 2.22.** \( 0 \leq \Omega \leq 1 \).

**Proof.** If all programs halt, \( \Omega = 0.\overline{1} = 1 \). And if no program halts, \( \Omega = 0 \). In the general case, some programs will halt and some will not. Hence, \( \Omega \) is inclusively between 0 and 1. Therefore it meets the definition of a probability.

**Lemma 2.23.** The program encoding is prefix free

**Proof.** In the case of unitary encoding, it is easy to see that no program is the prefix of another. A UTM reading unitary encoding will end at the first 0 it encounters and will assume this is the full program. Since the encoding enforces a single 0 bit per program at the very end, this guarantees that the UTM cannot mistake one program for another.

**Definition 2.24.** \( \Omega \) is a Chaitin omega number \(^4\) (or halting probability). Each halting probability is a normal and transcendental real number that is not computable, which means that there is no algorithm to compute its digits. Indeed, each halting probability is Martin-Löf random, meaning there is not even any algorithm which can reliably guess its digits.

Why is \( \Omega \) not computable? Because of the halting problem. Many sentences exist that do not contradict the rules of inference, but cannot be proven by them. When attempting to prove these sentences, the UTM will run forever without halting.

### 2.5 The halting partition

I have chosen to sum the sentences as an exponential distribution of base 2 in equation 2.10 because it is the most conceptually simple way to do it. It is not however the only way as other choices which preserve the information are available.

In this final subsection of this section, we generalize the halting probability to the halting partition. We note that key properties of the halting probability are preserved under the following generalization.
Theorem 2.25. A multiplication factor $\beta \in [0, \infty]$ can be added to $T(i)$ without changing the result of the sum

$$Z = \sum_{i=1}^{\infty} 2^{-\beta T(i)} 2^{-i}$$

Proof. expanding the sum into binary, we get

$$Z = \sum_{i=1}^{\infty} 2^{-\beta T(i)} 2^{-i}$$

$$= 2^{-\beta \infty} 2^{-1} + 2^{-\beta 0} 2^{-2} + 2^{-\beta 0} 2^{-3} + 2^{-\beta 0} 2^{-4} + 2^{-\beta \infty} 2^{-5} + ...$$

$$= 0 + 0.01 + 0.001 + 0.0001 + 0 + ...$$

$$= 0.01110...$$

The constant $Z$ does not change because $\beta$ it is always multiplied by $\infty$ or 0 in the sum, which erases its effects.

Theorem 2.30. A multiplication factor, $F \geq 1$ can be added to $i$ while preserving the non-computable nature of the sum, as well as the halting information content of each sentence

$$Z = \sum_{i=1}^{\infty} 2^{-F_i}$$

Proof. Tadaki has shown that Gregory Chaitin’s constant can be extended to include a compression term $F$ on $2^{-i} \implies 2^{-Fi}$ such that the Takadi constant $\Omega^F$ remains non-computable $F$-random. Furthermore, he goes to show that $\Omega^F$’s first $n$ bits contain $|n - F|$ halting bit. For example, take the case where $F = 2$, then expanding the sum into binary, we get

$$Z = \sum_{i=1}^{\infty} 2^{-Fi}$$

$$= 2^{-2 \times 1} + 2^{-2 \times 2} + 2^{-2 \times 3} + 2^{-2 \times 4} + 2^{-2 \times 5} + ...$$

$$= 2^{-2} + 2^{-4} + 2^{-6} + 2^{-8} + 2^{-10}...$$

$$= 0.01 + 0.0001 + 0.000001 + 0.000000001 + 0.000000000001 + ...$$

$$= 0.0101010101...$$

The compression factor $F$ "decompresses" the information by inserting some 0 in between the bits. It does not erase data. For the full proof, refer to Takadi’s paper.  

Part II

The Theory of Everything

We complete the halting partition by giving it an interpretation as the Theory of Everything in physics from the perspective of algorithmic information theory. The exponential terms are given the roles of macroscopic observables and are program output, program length, program runtime, program memory usage, etc. We complete the halting partition with other possible observables taking the form of a sum of conjugate variable to macroscopic observable pair as \( \sum_k \sigma_k C_k(i) \).

**Theorem 2.36.** This equation must hold for any possible ToE-candidate;

\[
Z = \sum_{i=1}^{\infty} e^{-(\ln 2)\beta[T(i) + F(i) + \sum_k \sigma_k C_k(i)]}
\]

**Remark 2.37.** In the special case where \( \beta = 1, F = 1 \) and \( -\sum_k \sigma_k C_k(i) = 0 \), we recover the halting probability of a prefix-free UTM and \( Z = \Omega \).

**Proof.** This equation is applicable to all UTK. Indeed, since the ToE must be a UTK (theorem 2.1), it follows that the equation must hold for any ToE.

We note the following risk; perhaps the equation is too general and as a result is unable to derive any physical laws.

To show that physical laws can indeed be derived from this equation, in part III we will derive from this equation the major equations of physics including but not limited to statistical physics, general relativity, quantum mechanics, the holographic principle, the speed of light as a maximum speed, and more. As these derivations will be done from this single equation and with no appeal to physical observation, they are derived from pure reason. This would imply that all possible UTK-universe share the same physic.

**Remark 2.38.** We do not claim to derive all of physics in this paper. For example, a derivation of the standard model of particles is not yet provided. We hope to show sufficient promising evidence to encourage members of the scientific community to further contribute to the theory herein proposed.
Part III

Derivation of physical laws

3 Thermodynamics

It has been said that thermodynamics is the most general of all the disciplines of physics. Hence it is expected to be the first derived from a ToE.

A theory is the more impressive the greater the simplicity of its premises, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression that classical thermodynamics made upon me. It is the only physical theory of universal content which I am convinced will never be overthrown, within the framework of applicability of its basic concepts.

–Albert Einstein

Introduction. In statistical physics, we are interested in the distribution that maximizes entropy

$$S = - \sum_{x \in X} p(x) \ln p(x) \quad (3.1)$$

subject to the fixed macroscopic observables. The solution is the Gibbs ensemble. As an example we take Table 4 as the observables.

<table>
<thead>
<tr>
<th>Observable</th>
<th>Conjugate variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy $E$</td>
<td>Temperature $\beta = 1/(k_b T)$</td>
</tr>
<tr>
<td>Volume $V$</td>
<td>Pressure $\gamma = p / (k_b T)$</td>
</tr>
<tr>
<td>Number of particles $N$</td>
<td>Chemical potential $\delta = -\mu / (k_b T)$</td>
</tr>
</tbody>
</table>

Table 4: Typical observables of statistical mechanics.

then the partition function becomes

$$Z = \sum_{x \in X} e^{-\beta E(x) - \gamma V(x) - \delta N(x)} \quad (3.2)$$

The probability of occupation of a micro-state is

$$p(x) = \frac{1}{Z} e^{-\beta E(x) - \gamma V(x) - \delta N(x)} \quad (3.3)$$

the average values and their variance for the observables are
The laws of thermodynamics can be recovered from the partition function by taking the derivatives

\[
\begin{align*}
\left. \frac{\partial S}{\partial E} \right|_{V,N} &= \frac{1}{T} & \left. \frac{\partial S}{\partial V} \right|_{E,N} &= p & \left. \frac{\partial S}{\partial N} \right|_{E,V} &= -\frac{\mu}{T}
\end{align*}
\] (3.7)

We summarize these equations to

\[
dE = TdS - pdV + \mu dN
\] (3.8)

**Related work on algorithmic thermodynamics.** In their paper \(^6\), John C. Baez and Mike Stay suggest an interpretation of algorithmic information theory based on thermodynamics, where the characteristics of programs are considered to be observables. Starting from Gregory Chaitin’s \(\Omega\) number, the halting probability

\[
\Omega = \sum_{p \text{ halts}} 2^{-|p|} \tag{3.9}
\]

is extended with algorithmic observables to obtain

\[
\Omega' = \sum_{x \in \mathcal{X}} e^{-\beta E(x) - \gamma V(x) - \delta N(x)} \tag{3.10}
\]

Noting the similarity between equation 3.2 and 3.10, they suggest an interpretation where \(E\) is the expected value of the logarithm of the program’s runtime, \(V\) is the expected value of the length of the program and \(N\) is the expected value of the program’s output. Furthermore, they interpret the conjugate variables as (quoted verbatim from their paper);

1. \(T = 1/\beta\) is the *algorithmic temperature* (analogous to temperature). Roughly speaking, this counts how many times you must double the runtime in order to double the number of programs in the ensemble while holding their mean length and output fixed.

2. \(p = \gamma/\beta\) is the *algorithmic pressure* (analogous to pressure). This measures the tradeoff between runtime and length. Roughly speaking, it counts how much you need to decrease the mean
length to increase the mean log runtime by a specified amount, while holding the number of programs in the ensemble and their mean output fixed.

3. \( \mu = -\delta / \beta \) is the \textit{algorithmic potential} (analogous to chemical potential). Roughly speaking, this counts how much the mean log runtime increases when you increase the mean output while holding the number of programs in the ensemble and their mean length fixed.

From equation 3.10, they derive analogues of Maxwell’s relations and they consider thermodynamic cycles such as the Carnot cycle or Stoddard cycle. For this they introduce the concepts of \textit{algorithmic heat} and \textit{algorithmic work}.

The authors then claim that the choice of correspondence between thermodynamic observables and algorithmic observables is somewhat arbitrary and reference other authors\(^7\) who have used completely different correspondences.

In this work, we study \textit{algorithmic thermodynamics} for the purpose of explaining \textit{physical thermodynamics}. As a result, our preferred choice of correspondence will be one that recovers the conventional language of physics. We will map the observables as follows.

- The \textbf{program-runtime} is the number of \textit{Iterations} a UTM needs to perform until a program halts. It is therefore natural to associate it with the physical \textit{Time} in \textit{seconds}. Indeed, a program requiring more iterations to halt will also require more time to terminate. If a system performs iterations at a faster or slower rate, the conjugate variable to time, the \textit{Power} in \textit{Watts}, can be adjusted to account for this variation.

- Its inverse, the \textbf{algorithmic-frequency}, is associated with the reverse of the second, \( s^{-1} \), and its conjugate variable is the \textit{Action} in \textit{Joules-seconds}.

- The \textbf{program-size} is expressed in number of \textit{bits}. Writing the bits one after the other on any medium (paper, disk drive, etc.) will require a certain physical size for each bit. As the line is the lowest dimensional geometry to spread bits, the program-size is naturally associated with the physical \textit{length} as its simplest case. Furthermore, if an encoding medium would allow greater or lesser “packing-tightness” of the bits, it can be modelled with its conjugate variable the \textit{Force} in \textit{Newtons} pushing the bits together or pulling them apart. If one wishes instead to investigate geometries of higher dimensions, one can use different units. For the 3D case, the program-size can be mapped to a \textit{Volume} in \( m^3 \) and its conjugate variable will be the \textit{Pressure} in \( N/m^2 \). For the 2D case, it can

be mapped to an Area in $m^2$ and its conjugate variable will be the Surface tension in $N/m$.

- Only the halting event remains. As it is the only quantity with no units, it is natural to map it to the Energy in Joules. Indeed, in the Gibb’s ensemble, the energy is the only observable not multiplied by a conjugate variable. Adding extra units to the halting event only to have them cancelled out by a conjugate variable would be futile.

Summarizing the points above, we obtain table 5 as our mapping of choice between algorithmic thermodynamics and physical thermodynamics.

<table>
<thead>
<tr>
<th>Observable</th>
<th>Variable</th>
<th>Units</th>
<th>Conjugate</th>
<th>Variable</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halting event</td>
<td>$E$</td>
<td>$J$</td>
<td>Temperature</td>
<td>$T$</td>
<td>$K$</td>
</tr>
<tr>
<td>Program-size (length)</td>
<td>$x$</td>
<td>$m$</td>
<td>Force</td>
<td>$F$</td>
<td>$N$</td>
</tr>
<tr>
<td>Program-size (area)</td>
<td>$A$</td>
<td>$m^2$</td>
<td>Stiffness</td>
<td>$\gamma$</td>
<td>$N/m$</td>
</tr>
<tr>
<td>Program-size (volume)</td>
<td>$V$</td>
<td>$m^3$</td>
<td>Pressure</td>
<td>$p$</td>
<td>$N/m^2$</td>
</tr>
<tr>
<td>Program-frequency</td>
<td>$\tau$</td>
<td>$1/s$</td>
<td>Action</td>
<td>$S$</td>
<td>$J \times s$</td>
</tr>
<tr>
<td>Program-runtime</td>
<td>$t$</td>
<td>$s$</td>
<td>Power</td>
<td>$P$</td>
<td>$W$</td>
</tr>
</tbody>
</table>

Table 5: The preferred correspondence between algorithmic thermodynamics and statistical physics.

**Theorem 3.11.** The general halting partition is the Boltzmann distribution used in statistical physics.

**Proof.** We try to explain thermodynamics and physics starting from algorithmic thermodynamics. First, we recall theorem 2.36. We recover thermodynamics by introducing an energy associated with each halting event, denoted by $E$. Adding the rest of the conjugate-observable pairs, we obtain

$$Z = \sum_x e^{-(\ln 2)\beta(E + Fx + Pt + \gamma A + pV...)}$$  \hspace{1cm} (3.12)

where the triple dots represent other possible observables. We interpret the program $x$ as a micro-state of the set of all prefix-free programs that are run on the UTM. It is easy to see that the function for $Z$ is the partition function of the Gibbs ensemble of thermodynamics.

Both the running frequency and the runtime are associated with time and are the converse of each other. As a result we only need to select one of them as our conjugate-observable pair whenever we want to take into account the effects of time. In this work, we will select $S\tau$ over $Pt$ (and vice-versa) whenever it leads to conceptually simpler results.
3.1 Energy

**Theorem 3.13.** The law of conservation of energy.

*Proof.* Since we have recovered a thermodynamic partition function, we can define a conserved energy quantity. This is a direct consequence of taking the thermodynamic state equation of the partition function.

\[
dE = TdS - Pdt - Fdx - \gamma dA - pdV - \ldots
\]  

(3.14)

The equation for the proposed ToE was obtained with no appeal to physical observations. As a result, it follows that any theorem derived from it necessarily is an experimental prediction. Therefore, deriving the law of conservation of energy is a prediction on the universe - it must have a law of conservation of energy or it cannot be a UTK!

**Remark 3.15.** Note that the law of conservation of energy (despite its name) does not imply that the total energy must remain the same. For example, the total energy could increase (or decrease) with time. But if it does it will vary according to a sum over the individual contributions of each changing thermodynamics observables. In other words, the energy does not change for no reason.

**Theorem 3.16.** When the values of the halting bits of \( Z \) are not known to an observer, each bit of information has the following energy

\[
E = k_B T \ln 2
\]

*Proof.* Consider an observer not aware of the bit values of \( Z \). To the observer, \( Z \) looks like

\[
Z_N = 0.\omega_1\omega_2\omega_3\ldots\omega_N
\]  

(3.17)

There are \( W = 2^N \) different possibilities, or micro-states. Since each bit has two possible values, the entropy of the system is \( S = k_B \ln 2^N \). Adding or removing a bit changes the entropy and the energy by

\[
\Delta S = S_{N+1} - S_N
\]  

(3.18)

\[
= k_B \ln 2^{N+1} - k_B \ln 2^N
\]  

(3.19)

\[
= k_B \ln 2
\]  

(3.20)

\[
\Delta E = T \Delta S
\]  

(3.21)

\[
= T (k_B \ln 2)
\]  

(3.22)
This result agrees with the well known Landauer limit\(^8\).

### 3.2 Speed of light

We now investigate the halting partition with the power-time observable \(P \times t\).

\[ Z = \sum e^{-\frac{(\ln 2)^2}{\beta}(E-Pt-Fx)} \]  

(3.23)

The units of the term \(P \times t\) are \((J/s) \times s = J\) and the units of \(F \times x\) are \((J/m) \times m = J\). Here \(J\) is joules, \(P\) is a power in Watts, \(F\) is a force in newtons, \(t\) is a time in seconds and \(x\) is a distance in meters.

The fundamental thermodynamics state equation for this partition functions becomes,

\[ dE = TdS - Pdt - Fdx \]  

(3.24)

**Theorem 3.25.** The maximum speed of any object is a constant \(c\), the speed of light.

**Proof.** Starting from equation 3.24 and posing \(dE = 0\),

\[ 0 = TdS - Pdt - Fdx \]  

(3.26)

\[ Fdx = TdS - Pdt \]  

(3.27)

\[ \frac{dx}{dt} = \frac{T}{F} dS - \frac{P}{F} \]  

(3.28)

\[ v = \frac{T}{F} dS - \frac{P}{F} \]  

(3.29)

Note that the units for each term of equation 3.29 are meters per seconds. The equation therefore describes a speed.

Let us look at three cases:

1. If \(|v| > |P/F|\), then \(dS/dt < 0\) and the entropy decreases with time. This violates the second law of thermodynamics.

2. If \(|v| < |P/F|\), then \(dS/dt > 0\) and the entropy increases with time. This is fine.

3. If \(|v| = |P/F|\), then \(dS/dt = 0\) and the entropy remains constant. This is also fine.

Since, according to the second law of thermodynamics, the average entropy cannot decrease with time, it follows that \(P/F\) is the fastest speed possible for a given system. Hence,

\[ \frac{P}{F} = c \]  

(3.30)
where $c$ is a characteristic speed of the universe dependant on $P$, the characteristic pressure and $F$ the characteristic force.

Taking $P$ to be the characteristic Planck pressure, and $F$ to be the characteristic Planck force of the universe, we do in fact recover the speed of light.

$$P \left( \frac{1}{F} \right) = \frac{c^3}{G} \left( \frac{G}{c^4} \right) = c$$ (3.31)

This is another prediction. The proposed ToE predicts that, to be a UTK, a universe must have a maximal speed. In fact, the axioms of the universe cannot override this fact without destroying its UTK features. This is an improvement over the formulation of special relativity as we have now proven one of its axioms from first principles. Indeed, in special relativity, a maximal speed has to be posited and it is justified by experimental observations.

3.3 Light-cone

We look at the thermodynamic cycle of the system transiting through time and space starting at $P_x$ to $P_0$ to $P_t$ to $P_x$ as illustrated on Figure 1. During the transitions and to keep the energy constant, trade-offs must be made between time, distance and entropy. This cycle is reminiscent of other thermodynamic cycles such as those involving pressure and volume, etc. The cycle presented here is reminiscent of relativistic light cones.

We work in the quasi static approximation

$$\Delta E = T \Delta S - P \Delta t - F \Delta x$$ (3.32)

and we pose that $\Delta E = 0$ throughout the cycle

$$T \Delta S = P \Delta t + F \Delta x$$ (3.33)

$P_x$ to $P_0$: As we translate $P_x$ closer in space to $P_0$ while keeping the time fixed, the entropy must decrease to compensate. This situation occurs when $\Delta x < 0$ and when $\Delta t = 0$.

$$(T \Delta S = P \Delta t + F \Delta x)|_{\Delta t = 0}$$ (3.34)

$$\implies \Delta S = \frac{F}{T} \Delta x$$ (3.35)

From the equation above, we note that $\Delta S$ is negative when $\Delta x < 0$. Since entropy tends to increase, we conclude that objects have
a tendency to resist being returned to the origin and are instead encouraged to expand away from each other.

\(P_0 \to P_t:\) As we translate \(P_0\) backward in time to \(P_t\) while keeping the distance fixed, the entropy must decrease to compensate. This situation occurs when \(\Delta t < 0\) and when \(\Delta x = 0\).

\[
(T \Delta S = P \Delta t + F \Delta x)_{\Delta x = 0} \\
\Rightarrow \Delta S = \frac{P}{T} \Delta t
\]  

From the equation above, we note that \(\Delta S\) is negative when \(\Delta t < 0\). We conclude that an object is encouraged by entropic considerations to evolve forward in time and is discouraged from evolving backward in time.

\(P_x \to P_t:\) As we translate \(P_x\) forward in time and backward in space to \(P_0\) keeping the entropic constant \((\Delta S = 0)\), we have movement at the speed \(c\).

\[
(T \Delta S = P \Delta t + F \Delta x)_{\Delta S = 0} \\
\Rightarrow \frac{\Delta x}{\Delta t} = \frac{P}{F} = c
\]  

From the equation above, an object travelling at speed \(c\) is neither encouraged nor discouraged by entropic considerations.

From the results of this section, we conclude that objects traveling forward in time, and objects travelling further apart are two sides of the same entropic coin. The natural direction for the evolution of time and space are justified by the same entropic argument.

Hence, this derivation predicts an expanding universe where time moves in the forward direction. Reversing this natural direction would require reversing the second law of thermodynamics in both cases.

**Remark 3.40.** Already we can see a connection between the rate of change of entropy and the passage of time. It is known from special relativity that a photon experiences no passage of time. Indeed, here we see that the maximum speed \(c\) occurs when \(\Delta S/\Delta t = 0\). Furthermore, \(c\) is enforced by \(\Delta S/\Delta t\) becoming negative if the speed exceeds it, which would violate the second law of thermodynamics. As a result, if the change of entropy is connected to the passage of time, its direction should reverse above \(c\). A converse conclusion can be obtained in opposite case. Indeed an object moving slower than \(c\) will see a positive \(\Delta S/\Delta t\). This does indicate that it moves forward in time. The slower its speed, the higher the rate of increase in entropy.
is, hence the faster in time it moves. This behaviour is reminiscent of special relativity. More on that in section 3.5.

3.4 Time

In this section, we will analyze the proposed ToE from the perspective of algorithmic information theory. We will further provide a definition of time and of observers and will prove that no physical observers can compute its future. From this, we will demonstrate the existence of a hard arrow of time.

We consider a simplified subset of the halting partition involving only the frequency and the length observable. Note the use of the conjugate-observable $S\tau$ (as opposed to $Pt$) as it is conceptually simpler here. Also note that $\tau = 1/t$.

$$Z = \sum_x e^{-(\ln 2)\beta(F_x + S\tau)} \quad (3.41)$$

The first step is to show that $Z$ is related to $\Omega$ which is a provably non-computable real number.

**Theorem 3.42.** At the limit of $t \to \infty$, we recover $\Omega$

**Proof.** A program $p$ can have any value of $S$ within $[0, \infty]$. If the program halts immediately, $S = 0$. If it never halts, $S = \infty$. If it halts after a certain time, $S \in \mathbb{N}$. A program that never halts will not contribute to the halting partition. This will be the case if $S = \infty$. As a result we obtain,

$$\lim_{\tau \to 0^+} \tau S_x = \lim_{t \to \infty} \frac{S_x}{t} = \begin{cases} 0 & \text{x halts} \\ \infty & \text{otherwise} \end{cases} \quad (3.43)$$

As this is the definition of $T(i)$ (see 2.8), we obtain

$$\lim_{t \to \infty} \frac{S_x}{t} = T(x) \quad (3.44)$$

Therefore,

$$\lim_{t \to \infty} Z = \lim_{t \to \infty} \left( \sum_x e^{-(\ln 2)\beta(F_x + S\tau)} \right) \quad (3.45)$$

$$= \sum_x e^{-(\ln 2)\beta(F_x + T(x))} \quad (3.46)$$

$$= \Omega \quad (3.47)$$

At $t \to \infty$ the halting programs, i.e. $S \in [0, \infty[$, have all halted, whereas the programs where $S = \infty$ have not. This is the definition of $\Omega$. □
Theorem 3.48. At the limit of $t \to 0^+$, we obtain initial conditions.

Proof. We study the limit of $t \to 0^+$. We obtain

$$
\lim_{t \to 0^+} \frac{S_x}{t} = \begin{cases} 
0 & S_x = 0 \\
\infty & \text{otherwise}
\end{cases}
$$  \hspace{1cm} (3.49)

At that limit, the only programs that contribute to $Z$ are those that halt immediately. These are the initial conditions. \qedhere

Theorem 3.50. For $0 < t < \infty$, the partition function $Z$ is

$$
Z(t) = \Omega - 2^{-k(t)}
$$

where $2^{-k(t)}$ is an error rate that is monotonically decreasing to 0 as $t \to \infty$.

Proof. 9

Definition 3.51. For any $k \geq 0$ and time $t \geq 0$, let $k(t)$ be the location of the first zero bit after position $k$ in the estimation of $\Omega$.

Then because $-\frac{S_x}{t}$ is a monotonically decreasing function of the running frequency and decreases faster than $k(t)$, there will be a time step where the total contribution of all the programs that have not halted yet is less than $2^{-k(t)}$.

\qedhere

For example, say

$$
\Omega = 0.0111100 \ldots
$$  \hspace{1cm} (3.52)

To keep it simple we consider, in isolation, a single program and assume that all other programs have long halted (at $t \to 0^+$). Let us take the values $x = 5$ and $S_x = 50$ for this program. We obtain,

$$
Z_x(t) = 2^{-x} 2^{-\frac{S_x}{t}}
$$  \hspace{1cm} (3.53)

$$
Z_5(t) = 2^{-52} 2^{-\frac{50}{t}}
$$  \hspace{1cm} (3.54)

$$
= 0.00001 \times 2^{-\frac{50}{t}}
$$  \hspace{1cm} (3.55)

The halting probability $\Omega$ is,

$$
\Omega = 0.0111000 \ldots + Z_5(t)
$$  \hspace{1cm} (3.56)

Let us look at what happens as we vary $t$.

1. If $t \to 0^+$, then $Z_5(0^+) = 0$. $Z$ differs from $\Omega$ by the maximum uncertainty of $2^{-5}$. Therefore $\Omega - Z_5(0^+)$ is accurate only in its first 5 bits.
2. As \( t \to \infty \), then \( Z_5(\infty) = 0.00001 \).

3. Between 0 and \( \infty \), \( Z_5(t) \) varies from \( 2^{-5} \) at \( t = 0 \) to 0 at \( t \to \infty \).
   Since \(- (S_5/t)\) is monotonically decreasing, the uncertainty \( 2^{-k(t)} \) must decrease monotonically to 0 as \( t \) increases.

4. At distances further than \( 2^{-k(t)} \), the partition function contains bits of programs that have yet to halt. So, in a sort, a reversal of time occurs where halting information is available before the time \( t \) is long enough for the program to have halted. \(^{10}\)

We are now ready to define directional time.

**Definition 3.57** (Directional Time). As \( t \) varies, more program halts. Hence the value of \( Z \) changes as bits flip from 0 to 1. Each value of \( Z \) corresponds to a time slice which is distinct from the others. The halting information of \( Z \) is valid up to the error rate of \( 2^{-k(t)} \) which monotonically decreases with time. Future time slices contain the halting information of all past slices, but the reverse is not true.

For example, consider the following time slices each corresponding to a different value of \( Z \). As we move down along the different \( Z_i \), more and more bits are flipped from 0 to 1. Each \( Z_i \) corresponds to the state of the universe at a different time of its history. \(^{11}\)

\[
Z_0 = 00000000000000000000000000000000 \ldots \\
Z_1 = 00000000000000110000000000000000 \ldots \\
Z_2 = 00000010000000100000000000000000 \ldots \\
Z_3 = 00000110000000100000000000000000 \ldots \\
Z_4 = 00010110000010000000000000000000 \ldots \\
Z_5 = 00010110010010000000000000000000 \ldots \\
Z_6 = 01110110010010001000110011000110 \ldots \\
Z_7 = 11110110010010001000110011001100 \ldots \\
\vdots
\]

We will now define rigorously what we mean by an experiment and by an observer.

**Definition 3.66** (Experiment). An experiment can be any program of \( Z \) that halts. As a result, an experiment is bounded both in memory and in time. It can be executed on a finite state machine provided it has sufficient memory.

**Definition 3.67** (Abstract observer). An abstract observer is defined as a set of experiments. The set comprises all experiments that it performs over its lifetime.

\(^{10}\) Here again we recover a behaviour reminiscent of special relativity.

\(^{11}\) Note that this example contains a slight simplification. In the calculation of \( Z \), halting bits are shifted leftwards along \( Z \) as time increases hence there is a possibility of shifting bits left to right. However, this detail can be ignored for this example.
Definition 3.68 (Physical observer). A physical observer is defined in a similar manner as an abstract observer but with an additional restriction. None of its experiments have a Kolmogorov complexity higher than the Kolmogorov complexity of the set of axioms $k$ of the universe.

Remark 3.69. The distinction between abstract and physical observer is by nomenclature only and is not necessarily enforced. Hypothetically, if a physical observer does get its hands on a finite state machine of greater complexity than $k$, the observer would seamlessly be upgraded to an abstract observer.

Remark 3.70. That being said, we will see that the Kolmogorov complexity of $k$ is so high that computing power in an amount at least equal to that of the whole universe would be required of an observer to be considered an abstract one. Hence, all observers (except perhaps the universe itself) are physical observers.

Theorem 3.71. For all physical observers, the future is non-computable.

Proof. The partition function dovetails programs. As a result, it is possible to imagine a simple algorithm—the dovetailing algorithm—that, knowing the set of axioms $k$ of the universe, is able to calculate future time slices before they occur.

To calculate future time slices as an experiment, the observer must encode $k$ in a finite state machine and run the program until it calculates a time slice newer than his. But in the case of a physical observer, this cannot be done. Indeed, all of its experiments have less Kolmogorov complexity than what is required to encode $k$.

Remark 3.72. If a physical observer does acquire possession of a finite state machine of sufficient complexity to encode $k$, it would be able to pre-calculate its future. Philosophical questions of fatalism\textsuperscript{12} would then enter the picture which are otherwise not applicable to a physical observer.

Theorem 3.73. For all physical observers, directional time is asymmetric.

Proof. From the example above, we note that $Z_7$ contains all the halting information in $Z_6$. The reverse however is not true. Indeed, looking at $Z_6$, a physical observer could not compute the next bits that will flip before they flip. Hence time is asymmetric.

Since a physical observer cannot compute its future before it occurs, but does retain past knowledge in the present, it is experiencing a hard arrow of time.

Remark 3.74. The physical observer can compute approximations of his future. For example, it can determine that the probability of raining tomorrow is 10%. This is not computing the future, it is computing a probability.

\textsuperscript{12} Can someone who know his future change it?
3.5 Representation of the Lorentz group

Armed with this definition of time, we now return to special relativity. As we have proven the existence of a maximum speed in the universe and further presented a thermodynamic cycle as a light-cone, it follows that the Lorentz group is the natural representation to continue our investigation with. Let us investigate this connection into more detail.

The Lorentz group is represented by the Lie Group $O(1,3)$. The Lorentz group embeds multiple other representations as subgroups. Some of them are:

1. The subgroup of transformations that preserves the direction of time is called **orthochronous** and is represented by $O^+(1,3)$

2. The subgroup of transformations that preserves the orientation, having a determinant of $+1$, is called **proper** and is represented by $SO(1,3)$.

3. The subgroup which includes transformations that are both proper and orthochronous is denoted by $SO^+(1,3)$.

4. The set of all rotations also forms a subset and is denoted by $SO(3)$.

We will now investigate how the halting partition behaves under transformation of these subgroups. Let us inject the simplest of the subgroups listed above, the rotations $SO(3)$, and investigate the results. We leave the more complicated subgroups as an exercise to the reader.

First, we recall the existence of an exponential map between a Lie Group and its corresponding Lie Algebra such that,

\[
SL(2, \mathbb{C}) = \exp(sl(2, \mathbb{C})) \tag{3.75}
\]

\[
SO(3) = \exp(so(3)) \tag{3.76}
\]

\[
SU(2) = \exp(su(2)) \tag{3.77}
\]

etc.

The existence of such a correspondence allows us to inject the Lie algebra $so(3)$ in a exponential function, such as the Gibb’s ensemble, and still recover a workable Lie Group after the exponentiation is executed.

Instead of using the $SO(3)$ directly, it will be more convenient to use the $SU(2)$ and $su(2)$ which are a double cover of it. Since the Lie algebra represents the rotations, it can be represented as a unit
quaternion using the notation $r = x_i + y_j + z_k$, and where $|r| = 1$. Let us inject the Lie algebra into the ToE.

$$Z = \sum_x e^{-\frac{(ln2)\beta \left[ (a_x x + a_y y + a_z z) + E(x) \right]}{|r|}}$$  \hspace{1cm} (3.78)

where $1/|r|$ is a normalization constant with the proper unit conversions. Injecting $su(2)$ into the partition function produces a thermodynamic observable of the rotation of the system comprising the halting bit. In other words, injecting $su(2)$ is simply a claim that a rotation of the system can be observed as a macroscopic thermodynamic variable.

To see what this rotation is internally, we now unpack the sum and execute the exponential. We obtain a Lie group expression where the halting bit of each program is prefixed with an $SU(2)$ matrix.

$$Z = SU(2)_1 \omega_1 + SU(2)_2 \omega_2 + SU(2)_3 \omega_3 + \ldots$$  \hspace{1cm} (3.79)

The $SU(2)$ matrices are defined as

$$SU(2) = \left\{ \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1 \right\}$$  \hspace{1cm} (3.80)

Finally, replacing the $SU(2)$ matrices in $Z$ by their definitions yields

$$Z = \begin{pmatrix} a_1 & -\beta_1 \\ \beta_1 & \alpha_1 \end{pmatrix} \omega_1 + \begin{pmatrix} a_2 & -\beta_2 \\ \beta_2 & \alpha_2 \end{pmatrix} \omega_2 + \ldots$$  \hspace{1cm} (3.81)

The only term left to investigate are the $\omega_i$. Here, each $\omega_i$ represents a bit of the halting probability with possible values 0 and 1. Since each of these bit is preceded by a $2 \times 2$ multiplication matrix, then it is natural to represent each $\omega_i$ as a two state system, namely as a $2 \times 1$ matrix to facility the matrix multiplication. We obtain,

$$Z = \begin{pmatrix} a_1 & -\beta_1 \\ \beta_1 & \alpha_1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} a_2 & -\beta_2 \\ \beta_2 & \alpha_2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \ldots$$  \hspace{1cm} (3.82)

We conclude that each halting bit becomes a spin when it is imbued with a 3D-rotation observable. The transformation of the $\omega_i$ from a bit to a $2 \times 1$ matrix is simply a claim that we do not know the value of the bit. Hence it describes the spin prior to the "measurement" of the bit.

How controversial is this result? The existence of 3D-rotations are not very controversial. They are present everywhere in physics, from Galileo mechanics to string theory. Here we of course take the $SO(3)$ as a subgroup of the Lorentz transformation as we have recovered
a maximum speed earlier. Since a maximum speed implies special relativity which itself implies the Lorentz group, the existence of \( SO(3) \) is implied.

Nevertheless, injecting its Lie algebra into the ToE does allow us to recover the spin. Let us study it in more detail in the following section.

### 3.6 Spin

Equation 3.82 is enough to recover the usual spin, but with significant improvements. Let us compare the state pre-measurement to the state post-measurement to notice the differences on \( Z \).

Pre-measurement, the value of \( \omega \) is unknown. Hence it takes the form of a \( 2 \times 1 \) matrix. It is therefore possible to define a matrix of the \( SU(2) \) group and to multiply it by \( \omega \). Macroscopically the \( SU(2) \) matrix is perceived as a thermodynamic observable related to the rotation of the system. This description corresponds to the usual spin.

Post-measurement, \( \omega \) ceases to be a matrix and is instead fixed to a specific bit. The \( SU(2) \) matrix can no longer be multiplied with \( \omega \), now a simple scalar, and the macroscopic observable of rotation must be eliminated from the description along with the \( SU(2) \) matrix. This system is no longer a spin, but a classical bit.

This is where we notice the improvement over the usual description of the spin. Indeed, the conventional theory of spins is unable to predict the value of the quantum measurement and only state that it is random as per experimental evidence. Here, however the value of a spin measurement is well defined and even deterministic but nonetheless provably non-computable. It is the value of the corresponding bit of the halting probability. Although its value is deterministic and reproducible, it can be shown to be related to the non-halting problem in such a way as to be non-computable and algorithmically random to any physical observer.

This is a proof from first principle that the quantum measurement is algorithmically random. To the authors knowledge, this is the only such proof in the literature.

**Remark 3.83.** According to these results, as the halting probability is deterministic, if the universe were to be run again from the same set of axioms, the quantum measurement outcomes will repeat themselves exactly. We stress however that the outcomes cannot be pre-calculated by a physical observer as they appear algorithmically random to it.

**Remark 3.84.** Suppose that the spin corresponding to a specific program is implemented by an electron. After measurement, the spin system disappears from \( Z \) and \( \omega \) becomes a single bit. What happens to the electron’s
spin - does this not imply that it too would disappear? No if the electron is perturbed by the measurement. The spin orientation of the electron will be changed by the measurement. In future sections, we will see that space coordinates (which are modified under measurement) encodes different programs (Theorem 4.1 for rotations and Theorem 4.67 for linear position). Measuring an electron’s spin perturbs it, hence it encodes a different program.

In the previous section, we have seen that the halting information is connected to time. Indeed, distinguishable slices of time are produced when programs halt and some of the classical bits of $Z$ flip. A universe with 3D-rotations is able to define a spin and this definition also includes its full behaviour under quantum measurement. This measurement induced bit flipping contributes to the physical observer’s perception of irreversible time.

What happens if we try to define an observable for 4D-rotations or higher? These rotations do not produce $SU(2)$ matrices in the partition function, hence their multiplication with the $2 \times 1$ matrix of $\omega$ is undefined. Therefore, they are unable to define a spin nor a mechanism for the quantum measurement. A physical observer does see the produce of time slices from them. If these higher dimensional rotations exists, they cannot be granted the status of a macroscopic thermodynamic observable and they cannot irreversibly impact nature for a physical observer. This status is reserved for rotations representable with $SU(2)$.

**Remark 3.85.** We do not necessarily excludes higher dimensions defined microscopically or as a mathematical aid. But for rotations definable as a macroscopic thermodynamic observable with time-irreversibility, these results would seem to indicate that 3D-rotations is all there is to it. Here we take the term ”macroscopic thermodynamic observable with time-irreversibility” to be essentially a synonym of “normal day-to-day life”.

**Remark 3.86.** We have not had to construct the spin by hand like it is sometimes done in elementary quantum mechanics. The spin came out naturally from the equation. Hence we would argue that it has been derived from first principles.

We can now define a quantum measurement.

**Definition 3.87** (Quantum measurement of qubits). For a physical observer, a quantum measurement occurs when the observer acquires enough information to compute the halting bit of the qubit being measured. Halting bits that are non-computable to the observer are seen as qubits.

For the spin to exist, it must be the case that the value of the corresponding halting bit $\omega_i$ of $Z$ is non-computable by the physical observer. In contrast, an abstract observer can compute the value of
Z for each time slice of the universe future and past, hence it cannot be the case that an $\omega_j$ is non-computable for it. The case where $\omega_j$ is non-computable can only occur from the perspective of a physical observer lacking the ability to independently compute $Z$. Hence, only physical observers can observe a spin.

**Remark 3.88.** As the halting partition was obtained with no appeal to physical observations, the existence of the spin is a prediction of the theory.
4 Space

We have derived the spin as a result of injecting 3D-rotations into the ToE. In this chapter, we will extend this result to define a linear position away from an origin. This result will be strong enough to derive black holes, the force of gravity, Schrödinger’s equation and general relativity.

4.1 Maximum entropy

Theorem 4.1. The maximal entropy of a section of a halting probability represented in space by spins is

\[ S = k_B \frac{c^3 A}{4Gh} \]

where \( A \) is the area of the sphere enclosing the volume used for the measurements.

Proof. In the previous section, we have seen that the spin of a halting bit leads to the statistical mechanic observable of a rotation in 3D. It is therefore natural to relate the spin to the sphere. We start with empty space, define an origin and add spins one by one, until we reach a maximum. We ask; how many can we fit and what shape does the system take?\(^{13}\)

Lemma 4.2. The spin requires four degrees of freedom to be fully represented.

Proof. We use the Bloch sphere representation of the qubit. Two degrees are taken up by the yaw and pitch angle of the wave-function pointing to a point on the top-half surface of the Bloch sphere. These are respectively angle \( \phi \) and \( \theta \) shown on Figure 2. The Bloch sphere represents a pure quantum 2 state system with wave-function

\[ |\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \text{ where} \]
\[ \alpha = \cos \frac{\theta}{2} \]
\[ \beta = e^{i\phi} \sin \frac{\theta}{2} \]

The third degree is associated with the result for the measurement value of the spin along the \( z \) axis. Finally, the fourth degree is associated to the measurement value along the \( xy \) plane. \(^{14}\) This completes the full description of a spin, including all observables and measurement outcomes.

\(^{13}\) We take all spins to be centred at the origin. Introducing a position away from the origin will require bits to describe it. As we introduce a proper definition of position in the next subsection, we will see that this balances the entropy to eliminate any gains.

\(^{14}\) The third and the fourth degrees are not available from the wave-function but are instead read by a quantum measurement.
Remark 4.6. Why must the measurement outcome of orthogonal quantum states both be described pre-measurement, if only one of them can be measured? After measurement and wave-function collapse, the system has one degree of freedom. The pure-state is destroyed and is replaced with a classical bit. However, the system does not know which axis will be measured first. Therefore they must both be described.

The two degrees of freedom of the yaw and pitch angle of the wave-function are real numbers and require up to infinitely many classical bits to be expressed to the desired precision. It would be a bad idea to encode the orientation of spins using real numbers because there is no limit to their precision. A single spin could consume arbitrarily many bits to describe just one of its real numbered degrees of freedom.

However, we consider the Pauli exclusion principle, and note that a non-integer spin cannot occupy the same quantum state. We also note that the precision required to define a spin is wasted beyond the precision of the instrument measuring it. This instrument cannot be precisely rotated with an angle smaller than the Planck angle. As a result, the spins need not be expressed more accurately than the Planck angle.

Lemma 4.7. Since the distinguishable rotation angles are finite, there exists an efficient way to encode spins in a sphere, where one classical bit maps to one degree of freedom.

Proof. We associate classical bits to a position on the surface of the enclosing sphere. It is natural to pose that the number of bits that can fit on its surface is given by

$$N = \frac{A}{L_p^2} = \frac{c^3 A}{G\hbar}$$

where $L_p^2$ is the Planck length. We can maximize the number of spins in a volume, and therefore maximize the entropy, if we minimize the number of classical bits required to express each spin. Figure 3 shows a spherical geodesic divided in equal areas, each occupied by a classical bit. Each bit on the surface encodes the presence or absence of a pure state wave-function pointing to the bit and centered at the origin.

From the Bloch representation, we easily see that only half of the bits on the surface are actually needed to express all permissible spin orientations in the volume. For each second bit, a one-valued bit...
on the surface of the sphere corresponds to the presence of a wave-
function $|\psi\rangle$ pointing to it, while a zero-valued bit indicates a vacant
spot. Figure 4 illustrates the encoding of the orientation in 2D.

The remaining half of bits are used for the third and fourth de-
grees of freedom. This encoding is maximally efficient and each
classical bit encodes a single degree of freedom.

We now want to determine the maximum entropy of the spin
system describable by those bits. Since the general halting partition
has a factor $\ln 2$ on the exponential term, the derivative $dS/dE$ gives
us a pseudo-entropy related to the real entropy by a factor $1/\ln 2$.

\[
\frac{dS'}{dE} = \frac{\ln 2}{k_B T} \tag{4.9}
\]
\[
dE = k_B T \frac{dS'}{\ln 2} \tag{4.10}
\]
\[
dE = k_B T dS \quad \text{(where } S \ln 2 = S')
\]

The factor $\ln 2$ is obtained because the statistical mechanics parti-
tion is derived from the base-2 sum of the halting probability. This
implies that,

\[
S = \frac{1}{\ln 2} S' \tag{4.11}
\]
\[
= \frac{1}{\ln 2} k_B \ln W \tag{4.12}
\]

where $W$ is the number of micro-states for the system. Since each
micro-state (e.i. spin) requires four classical bits to be described, a
factor of $\frac{1}{4}$ is added, such that

\[
W = 2^{\frac{1}{4} N} \tag{4.13}
\]
\[
= 2^{\frac{3A}{4 \ln 2}} \tag{4.14}
\]
\[
\ln W = \frac{c^3 A}{4G} \ln 2 \tag{4.15}
\]
\[
S = \frac{1}{\ln 2} k_B \frac{c^3 A}{4G} \ln 2 \tag{4.16}
\]

Simplifying the constants, we get

\[
S = k_B \frac{c^3 A}{4G} \tag{4.17}
\]

which is the Bekenstein-Hawking entropy.

\[
\]
Theorem 4.18. At most, half of the permissible orientations are occupied by spins.

Proof. First, note that 4 classical bits are required to describe a spin. Second, note that 2 of these classical bits are used to mark occupied orientations. This means that half of the orientations must be vacant.

\[
\text{The halting probability of a UTM is a normal real number that is algorithmically random. This implies that, statistically, it contains as many zeros as ones in its binary digit representation. Conveniently, the maximal entropy of the sphere occurs when the bits on its surface are composed of equally many zeros and ones.}
\]

4.2 Position

We recall the general halting partition,

\[
Z = \sum_x e^{-(\ln 2)\beta(E + Fx + S\tau + \ldots)}
\]

(4.19)

and we specifically focus our attention on the second term in the sum of the exponential $Fx$, where $F$ is a force with units Joules per meter and $x$ is the length of the program with units meter. This term has the same form and units as an entropic force of the form $Fdx = TdS$. For example, entropic forces also describe polymer tension, osmotic force, etc.

Since $x$ represents the program’s length in meters, this result strongly suggests that it is encoding spatial lengths. To encode length using bits we will consider an efficient algorithm with the following characteristics:

1. It must have the observer as its origin.
2. It must be able to express positions arbitrarily far away from the origin.
3. It must be able to express positions with arbitrary precision.

Naively, we might be tempted to use a sequence of bits to represent the quantity of discreet steps required to reach a point in space from the present location. For example, suppose the Planck’s length is the smallest unit of space. Then an object of 1 m away from the origin would have its position encoded with the number $6.25 \times 10^{34}$ in binary. A rather large bit requirement for something that is very simple to describe in layman’s terms.

We can make significant better use of our bits by specifying a scale before listing the bits. For example, let’s say we only care about
specifying the position within a centimetre. We could initially specify a scale, then list the number of repetition of the scale unit to reach the location of interest.

The price to pay for such algorithmic compression is that the position of an object is definable up to a certain precision, and it must have a characteristic scale. I will conjecture that this is the most efficient position encoding algorithm that meets the requirements above.

**Definition 4.20** (Positioning via program length, or PVPL for short). Suppose a Cartesian coordinate system with orthogonal axis x y z and an origin at (0, 0, 0). The position of a point will be encoded via four digits. One scaling constant \( \lambda \), and 3 scale-repetition values \( l_x, l_y, l_z \). It’s position is noted as \( p = (\lambda, l_x, l_y, l_z) \). The point can be identified within an error margin of \( \pm (1/2)\lambda \). The scale \( \lambda \) can be made as small as we want it by increasing the bit count of \( l_x, l_y \) and \( l_z \).

**Theorem 4.21** (Conversion formulas). PVPL encoding can be converted back to cartesian coordinates by multiplying the scaling factor \( \lambda \) with the repetition value for each coordinate and keeping the uncertainties.

\[
\begin{align*}
x &= l_x \lambda \pm \Delta \lambda \\
y &= l_y \lambda \pm \Delta \lambda \\
z &= l_z \lambda \pm \Delta \lambda
\end{align*}
\]

where \( \Delta \lambda = (1/2)\lambda \).

For example, a point at cartesian coordinate (0, 0, 1) with an error margin \( \pm 0.1 \) is encodable in PVPL with the following 4-digits: \( p_1 = (0.1, 0, 0, 1010_b) \). To double the precision without changing the position, we must add an extra bit to the repetition value in \( z \) and halve the scaling constant such that \( p_2 = (0.05, 0, 0, 10100_b) \). Using PVPL to describe near objects of similar scales has the advantage of requiring a very low bit count.

**Theorem 4.25** (Positional entropy). Using the positional encoding of definition 4.20, we can pose a relation between \( \Delta S \) and \( \Delta x \) of

\[
\Delta S = 2\pi k_B \frac{1}{\lambda} \Delta x
\]

**Proof.** Naively, we might be tempted to allow the PVPL values \( l_x, l_y \) and \( l_z \) to take any bit sequence. If we do so, doubling \( \Delta x \) from 10 meters to 20 meters (or in binary; 1010\(_b\) meters to 10100\(_b\) meters) increases the bit count by 1. Since this does not double the entropy, it contradicts theorem 4.25.
However, it would be a mistake. PVPL values are not the bit sequence themselves, but the length of the program. What is the entropy of a program of length $L$? A physical observer will measure the entropy of the program to be equal to its length. A physical observer does not and cannot know the precise prefix-free encoding because he sees a degenerate general halting partition. Hence, he must accept that all possible bit sequences could represent a program of length $L$. Since there are $2^L$ such sequences, $S = k_B L = k_B \Delta x$.

Why multiply $\Delta x$ with $1/\lambda$? For example, suppose we have $\Delta x = 5$ meters. If we encode the position using the meter scale, we will require a program of length 5. However, if we use the centimetre scale, the required program length will now be 500. This increases the entropy by a factor of 100. The smaller the scale the higher the entropy.

Why multiply $\Delta x$ with a factor of $2\pi$? Our previous result from theorem 4.1 shows that the number of bits depends on the area of the sphere. This results suggests that the bits rest of the surface of the sphere defining a volume in space.

Suppose a circle of radius $r$ with $N$ points uniformly distributed on its perimeter. The points on the perimeter maps to a segment on a line of length $L$. Keeping the scale intact, the length of the segment will be $L = 2\pi r$ and each point will be separated by a distance $d = 2\pi r / N$. The factor $2\pi$ multiplying $\Delta x$ is a consequence of the scale preserving mapping of linear distances encoded by equidistant bits on a circle.

Why go to such lengths to avoid rescaling the line coordinates and to keep the factor $2\pi$? We are not allowed to rescale the axis as a rescale would increase the precision of all positions and therefore consume more bits.

$\lambda$ is one of the most fundamental quantity in the universe. Any physical object having a position in space needs to have a specific value of $\lambda$ defined for it. For pedagogical reason, we suggest that any physics involving position should be conceptually understood in terms of $\lambda$.

However, to recover physics equations in a form that we are familiar with, it will help to rewrite $\lambda$ using other variables. To do it, we simply need to relate the wavelength of a physical object to another convenient quantity. For example, we might want to express equation 4.25 in terms of mass instead of wavelength. Let us do that know.

**Theorem 4.26.** The algorithmic scaling factor $\lambda$ for massive particles is the
reduced Compton wavelength

\[ \lambda = \frac{\hbar}{mc} \]

**Proof.** To express equation 4.25 in terms of a mass, instead of a wavelength, we simply need to find an equation relating the wavelength of a physical object to its mass. To obtain it, we start from the characteristic wave equation \( E = \hbar \omega \), and we get;

\[
\begin{align*}
E &= \hbar \omega \\
    &= \hbar 2\pi f
\end{align*}
\]

\( mc^2 = \hbar 2\pi \lambda / \lambda_c \) (4.29)

\[
\lambda_c = \frac{\hbar}{mc}
\]

where \( \lambda_c \) is the Compton wavelength. Since we have already reduced the algorithmic scaling factor by \( 2\pi \) in equation 4.25, the Compton wavelength is related to \( \lambda \) by a \( 2\pi \) factor. Hence,

\[
2\pi \lambda = \lambda_c
\]

\[
\lambda = \frac{\hbar}{mc}
\]

We can now express equation 4.25 in terms of mass instead of wavelength. We obtain

\[
\Delta S = 2\pi k_B \frac{mc}{\hbar} \Delta x
\]

4.3 **Schwarzschild radius**

**Theorem 4.34.** Using all the classical bits on the surface of a sphere encoding a halting probability, we encode, using entropic positioning, the longest distance that can be expressed for the black hole. We find that this distance is equal to the Schwarzschild radius,

\[
r_s = \frac{2GM}{c^2}
\]

**Proof.** According to theorem 4.1, the maximum entropy in a volume of space is

\[
S = k_B \frac{Ac^3}{4Gh} = k_B \frac{\pi r^2 c^3}{\hbar G}
\]
We will convert all of these bits to entropic positioning so that we can express the point furthest away from the center of the sphere. We recall equation 4.54.

\[ \Delta S = 2\pi k_B \frac{1}{\lambda} \Delta x \]  

(4.37)

Equating the two entropies (\(\Delta S = S\)), we get

\[ k_B \frac{\pi r^2 c^3}{hG} = 2\pi k_B \frac{1}{\lambda} \Delta x \]  

(4.38)

Since we are dealing with a black hole, it should not be able to express programs of length longer than its horizon, or information might leak out. Therefore we pose \(\Delta x = r\), and we obtain,

\[ k_B \frac{\pi r^2 c^3}{hG} = 2\pi k_B \frac{1}{\lambda} (r) \]  

(4.39)

Solving for \(r\), we obtain

\[ r = \frac{1}{k_B} \frac{hG}{\pi c^3} \frac{2\pi k_B}{1} \frac{1}{\lambda} \]  

(4.40)

Then, we take the algorithmic scaling factor \(\lambda\) to be, in fact, the reduced Compton wavelength. We obtain

\[ r = \frac{1}{k_B} \frac{hG}{\pi c^3} 2\pi k_B \left( \frac{Mc}{\hbar} \right) \]  

(4.41)

Reducing the constants, we obtain

\[ r = \frac{2GM}{c^2} \]  

(4.42)

Which is the Schwarzschild radius.

\[ \square \]

4.4 Hawking radiation

**Theorem 4.43.** A sphere encoding bits of a halting probability will radiate at a temperature inversely proportional to its mass. The temperature will be consistent with Hawking radiation;

\[ T = \frac{hc^3}{8\pi GMk_B} \]

**Proof.** According to theorem 4.1, the maximum entropy in a volume of space is

\[ S = k_B \frac{Ac^3}{4Gh} = k_B \frac{\pi r^2 c^3}{hG} \]  

(4.44)
Using the thermodynamic relation $dE = TdS$, we want to obtain the energy, then the mass via $E = mc^2$.

$$dS = 2k_B \frac{\pi r c^3}{\hbar G} dr$$  \hfill (4.45)

multiplying by $T$, we obtain the derivative of energy

$$dE = TdS = 2k_B T \frac{\pi r c^3}{\hbar G} dr$$  \hfill (4.46)

dividing by $c^2$, we obtain the mass

$$dM = \frac{1}{c^2} dE = 2k_B \pi T \frac{c}{\hbar G} r dr$$  \hfill (4.47)

posing $r$ to be the Schwarzschild radius calculated in theorem 4.34 and $dr$ to be its derivative with respect to the mass, we obtain

$$dM = 2k_B \pi T \left(\frac{2GM}{c^2} \right) \frac{c}{\hbar G} \frac{2G}{c^2} dM$$  \hfill (4.48)

Solving for $T$ and reducing the constants, we obtain

$$T = \frac{\hbar c^3}{8\pi GMk_B}$$  \hfill (4.49)

, the temperature of Hawking radiation.

\[\square\]

4.5 Newton’s law of inertia

**Theorem 4.50.** The positional entropy leads to Newton’s law of inertia and to the proof of the existence of inertia.

$$F = ma$$

**Proof.** The laws of statistical physics state that we can take the derivative of the entropy with respect to $x$ and obtain $F/T$. For an entropic force, the derivative is

$$\frac{dS}{dx} = \frac{F}{T}$$  \hfill (4.51)

In the quasi-static approximation, the derivative becomes

$$F \Delta x = T \Delta S$$  \hfill (4.52)

Let us replace $\Delta S$ with theorem 4.25. We obtain
\[ F \Delta x = T \left( 2\pi k_B \frac{1}{\lambda} \Delta x \right) \]  
(4.53)

\[ F = 2\pi k_B T \frac{1}{\lambda} \]  
(4.54)

Inspired by Erik Verlinde’s paper on entropic gravity \(^{15}\), we take \( T \) to be Unruh’s temperature \(^{16}\) and we obtain

\[ F = 2\pi k_B \left( \frac{1}{2\pi k_B} \frac{\hbar a}{c} \right) \frac{1}{\lambda} \]  
(4.55)

Then, we take the algorithmic scaling factor \( \lambda \) to be, in fact, the reduced Compton wavelength \((\lambda = \frac{\hbar}{m c})\). We obtain

\[ F = 2\pi k_B \frac{1}{2\pi k_B} \frac{\hbar a}{c} \left( \frac{mc}{\hbar} \right) \]  
(4.56)

Finally, it reduces to

\[ F = ma \]  
(4.57)

\[ 4.6 \quad \text{Newton’s law of gravity} \]

**Theorem 4.58.** Bits spread across the surface of a sphere will produce an entropic force governed by Newton’s law of gravity. Namely, a force with the following formula.

\[ F = \frac{G m M}{r^2} \]

**Proof.** Here we abandon the volume entropy and consider that gravity is an entropic force created from the very bits that encode the universe at its most fundamental level. The number of classical bits on the surface is

\[ N = \frac{c^3 4\pi r^2}{G\hbar} \]  
(4.59)

From this equation, we can obtain the energy via the equipartition theorem which maps to the energy of a system to its number of degrees of freedom.\(^ {17}\)

\[ E = \frac{1}{2} k_B TN \quad \text{(equipartition theorem)} \]

\[ = \frac{1}{2} k_B T \left( \frac{c^3 4\pi r^2}{G\hbar} \right) \]  
(4.60)


\(^{16}\) Unruh’s temperature: \( k_B T = \frac{\hbar}{2 m c} \).

The use of Unruh’s temperature is justified by the same reasons given by Erik Verlinde. Unruh has shown that an observer in an accelerated frame of reference will measure a vacuum temperature of the given form.

\(^{17}\) We will reproduce the proof by Erik Verlinde
We can obtain the mass by dividing by $c^2$

$$M = \frac{1}{c^2} E = k_B T \frac{2\pi^2 c}{\hbar G}$$  \hspace{1cm} (4.61)

solving for $T$, we obtain

$$T = \frac{\hbar GM}{2\pi k_B cr^2}$$  \hspace{1cm} (4.62)

We recall equation 4.54 and insert the $T$ we just found in it and we obtain

$$F = 2\pi k_B T \frac{1}{\lambda}$$  \hspace{1cm} (4.63)

$$F = 2\pi k_B \left( \frac{\hbar GM}{2\pi k_B cr^2} \right) \frac{1}{\lambda}$$  \hspace{1cm} (4.64)

Again, posing that the algorithmic scale factor $\lambda$ is in fact the reduced Compton wavelength, we obtain

$$F = 2\pi k_B \frac{\hbar GM}{2\pi k_B cr^2} \left( \frac{mc}{\hbar} \right)$$  \hspace{1cm} (4.65)

Note here that we introduce $m$ as the mass of the reduced Compton wavelength as opposed to $M$ because we want to quantify the force felt between two objects of different masses. Finally, it reduces to

$$F = \frac{GmM}{r^2}$$  \hspace{1cm} (4.66)

Erik Verlinde further generalizes entropic gravity to account for arbitrary matter distributions.

### 4.7 Schrödinger’s equation

**Theorem 4.67.** A position described by entropic positioning (theorem 4.25) will evolve in time according to Schrödinger’s equation.

$$i\hbar \frac{\partial}{\partial t} \psi(x, t) = \left[ \frac{-\hbar^2}{2m} \nabla^2 + V(x, t) \right] \psi(x, t)$$

The proof is slightly more involved than the preceding theorems. First, here is a sketch

1. We will show that entropic position encoding using the bits produced by the general halting partition leaves holes in space where a position cannot be expressed.
2. We will show that these holes are causing a Brownian motion of the encoded position.

3. We will derive its diffusion coefficient to be $\frac{\hbar}{2m}$.

4. We will consider that the presence of any external field is experienced as acceleration via $F = ma$.

5. Using the well known Brownian motion equations of Langevin, we show that the above reproduces Schrödinger’s equation exactly.

Lemma 4.68. Holes.

Proof. We recall the general halting partition

$$Z = \sum_x e^{-(\ln 2)\beta(E + Fx)} \quad (4.69)$$

We have previously seen how positions can be encoded with PVPL. We have also seen that the observable $x$ denotes program lengths. However, not all programs halt hence some lengths are missing from the sum. These missing programs are holes in space the position of which cannot be expressed by the general halting partition. Since $\Omega$ is a normal number, we can predict certain randomness related properties of these holes.

Lemma 4.70. A particle in space will experience Brownian motion due to the holes.

Proof. We will calculate the average displacement $\Delta x$ of a particle subjected to entropic positioning and space holes. Since $Z$ is a normal number, we conclude that half of the program’s lengths are available to describe position and half are not. Therefore, to describe a particle at position $x$, there is a 50% chance there is a halting program available to express it. And in the case where there is no program at exactly $x$, then there is a 50% chance that there will be one at position $x + 1$, and so on. In other words, a particle at $x$ has 50% chance of being at $x$, 25% chance of being at $x + 1$, 12.5% chance of being at $x + 2$, etc. Expressed as a sum, we obtain

$$\Delta x = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots \quad (4.71)$$

$$= \sum_{i=0}^{\infty} \frac{i}{2^{i+1}} \quad (4.72)$$

$$= 1 \quad (4.73)$$

On average, as it moves through space, a position will shift by $\Delta x = 1$ at each iteration of the Brownian motion. But will a stationary point also experience Brownian motion? The answer is yes.
A stationary point will experience Brownian motion because of the degeneracy of the general halting partition. The general halting partition is degenerate to an observer because he does not know the specific $\Omega$-value used by the universe. Therefore he must assume that all $\Omega$ are possible and that each forms a degenerate state. As the micro-states switch around the possible degeneracies, the holes are moved around. The probabilities are the same as the sum above and $\Delta x = 1$ for a stationary point.

**Lemma 4.74.** The diffusion coefficient of the described Brownian motion is

$$D = \frac{\hbar}{2m}$$

*Proof.* It is well known that in general the diffusion coefficient of Brownian motion is given by

$$D = \frac{l^2}{2\tau} \quad (4.75)$$

where $l$ is the length of the random step and $\tau$ is the frequency of the occurrence of the steps. Entropic position uses the scale factor $\lambda$ for each unit of length. When $\lambda$ is the reduced Compton wavelength, we get a scaling factor of

$$\lambda = \frac{\hbar}{mc} \quad (4.76)$$

Since entropic positioning can only express position as multiples of $\lambda$, we take it as the Brownian step of length $l$. The diffusion coefficient becomes

$$D = \left(\frac{\hbar}{mc}\right)^2 \frac{1}{2\tau} \quad (4.77)$$

This leaves of us with the need to define $\tau$. For $\tau$, we take the characteristic frequency of the wave $E = \hbar\omega$. Solving for $\tau = 1/\omega$, we obtain

$$\omega = \frac{E}{\hbar} \quad (4.78)$$

$$\tau = \omega^{-1} = \frac{\hbar}{E} \quad (4.79)$$

Replacing $\tau$ in the equation for $D$, we obtain

$$D = \frac{\hbar^2}{m^2c^2} \left(\frac{E}{2\hbar}\right) \quad (4.80)$$

Using $E = mc^2$, and reducing the constants, we obtain our final expression of $D$. 

\[ D = \frac{\hbar^2}{m^2c^2} \left( \frac{mc^2}{2\hbar} \right) \]  
(4.81)

\[ = \frac{\hbar}{2m} \]  
(4.82)

**Lemma 4.83.** The Langevin equations for Brownian motion with a diffusion coefficient of \( \hbar/(2m) \) and an external field \( F = ma \) reproduces Schrödinger's equation.

**Proof.** We recall the well known Langevin equation,

\[ d [x(t)] = v(t)dt \]  
(4.84)

\[ d [v(t)] = -\frac{\gamma}{m} v(t)dt + \frac{1}{m} W(t)dt \]  
(4.85)

where \( W(t) \) is a random force and a stochastic variable giving the effect of a background noise to the motion of the particle.

From \( F = ma \) and replacing the acceleration \( d[v(t)]/dt \) with \( F/m \), Edward Nelson \(^{18}\) is able to show that the Langevin equation becomes,

\[ \frac{1}{2} \nabla u^2 + D \nabla^2 u = \frac{1}{m} \nabla V \]  
(4.86)

where \( D \) is the diffusion coefficient of \( \hbar/(2m) \) obtained in lemma 4.74, where \( F = -\nabla V \), where \( u = v \nabla \ln \rho \) and \( \rho \) is the probability density of \( x(t) \). For brevity, the proof of 4.86 is omitted here but can be reviewed in his paper. Eliminating the gradients on each side and simplifying the constants, we obtain

\[ \frac{m}{2} u^2 + \frac{\hbar}{2} \nabla u = V - E \]  
(4.87)

where \( E \) is the arbitrary integration constant. This equation in non-linear because of the term \( u^2 \) but it can be made linear by a change of dependant variable. To make it linear, let us pose

\[ u = \frac{\hbar}{m} \frac{1}{\psi} \nabla \psi \]  
(4.88)

and replace it into equation 4.87, we obtain

\[ \frac{m}{2} \left( \frac{\hbar}{m} \frac{1}{\psi} \nabla \psi \right)^2 + \frac{\hbar}{2} \nabla \left( \frac{\hbar}{m} \frac{1}{\psi} \nabla \psi \right) = V - E \]  
(4.89)

taking the gradients and the exponents, we obtain

\[ \frac{\hbar^2}{2m} \frac{1}{\psi^2} \nabla^2 \psi + \frac{\hbar^2}{2m} \left[ -\frac{1}{\psi^2} \nabla^2 \psi + \frac{1}{\psi} \nabla^2 \psi \right] = V - E \]  
(4.90)

The first two terms cancel each other.

$$\frac{\hbar^2}{2m} \frac{1}{\psi} \nabla^2 \psi = V - E$$ \hfill (4.91)

Finally, it simplifies to

$$\left[ -\frac{\hbar^2}{2m} \nabla^2 + V - E \right] \psi = 0$$ \hfill (4.92)

which is the time independent Schrödinger’s equation.

We are now ready to derive the time dependent Schrödinger’s equation and prove theorem 4.67.

**Proof.** We use the same proof used by Edward Nelson in the same paper. Starting from the time dependent Schrödinger’s equation and show that a replacement of $\psi = e^{R+iS}$ leads to the Langevin equation of Brownian motion. We write the time dependent Schrödinger’s equation, perform the replacement and obtain the Langevin equations of Brownian motion.

$$\frac{\partial \psi}{\partial t} = i \frac{\hbar}{2m} \nabla^2 \psi - i \frac{1}{\hbar} V \psi$$ \hfill (4.93)

Replacing $\psi$ with $e^{R+iS}$, we obtain

$$\frac{\partial}{\partial t} (e^{R+iS}) = i \frac{\hbar}{2m} \nabla^2 (e^{R+iS}) - i \frac{1}{\hbar} V (e^{R+iS})$$ \hfill (4.94)

Taking the derivatives and the gradients, we obtain

$$\left[ \frac{\partial R}{\partial t} + i \frac{\partial S}{\partial t} \right] (e^{R+iS}) = i \frac{\hbar}{2m} \left[ \nabla^2 R + i \nabla^2 S + (\nabla (R + iS))^2 \right] (e^{R+iS}) - i \frac{1}{\hbar} V (e^{R+iS})$$ \hfill (4.95)

Eliminating $e^{R+iS}$ from each side and simplifying, we obtain

$$\begin{align*}
\frac{\partial R}{\partial t} + i \frac{\partial S}{\partial t} &= i \frac{\hbar}{2m} \left[ \nabla^2 R + i(\nabla (R + iS))^2 \right] - i \frac{1}{\hbar} V \quad \text{(eliminating } e^{R+iS}) \\
\frac{\partial R}{\partial t} + i \frac{\partial S}{\partial t} &= i \frac{\hbar}{2m} \left[ \nabla^2 R + i \nabla^2 S + (\nabla R)^2 + 2i \nabla R \nabla S - (\nabla S)^2 \right] - i \frac{1}{\hbar} V \quad \text{(taking the product)} \\
\frac{\partial R}{\partial t} + i \frac{\partial S}{\partial t} &= \frac{\hbar}{2m} \left[ i \nabla^2 R - \nabla^2 S + i(\nabla R)^2 - 2\nabla R \nabla S - i(\nabla S)^2 \right] - i \frac{1}{\hbar} V \quad \text{(distributing the } i) \\
\end{align*}$$

We obtain two equations by separating the real and the imaginary parts

$$\begin{align*}
\frac{\partial R}{\partial t} &= \frac{\hbar}{2m} \left[ -\nabla^2 S - 2\nabla R \nabla S \right] \quad \text{(4.96)} \\
\frac{\partial S}{\partial t} &= \frac{\hbar}{2m} \left[ \nabla^2 R + (\nabla R)^2 - (\nabla S)^2 \right] - \frac{1}{\hbar} V \quad \text{(4.97)}
\end{align*}$$
This is equivalent to the Langevin equations with some replacements:

\[
\begin{align*}
\frac{\partial u}{\partial t} &= -\frac{\hbar}{2m} \nabla^2 v - \nabla (v \cdot u) \quad (4.98) \\
\frac{\partial v}{\partial t} &= \frac{\hbar}{2m} \nabla^2 u + \frac{1}{2} \nabla (u^2) - \frac{1}{2} \nabla (v^2) - \frac{1}{m} \nabla V \quad (4.99)
\end{align*}
\]

**Lemma 4.100.** Equation 4.96 with the replacements \( \nabla R = (m/\hbar)u \) and \( \nabla S = (m/\hbar)v \) produces 4.98

**Proof.**

\[
\begin{align*}
\frac{\partial R}{\partial t} &= \frac{\hbar}{2m} \left[ -\nabla^2 S - 2\nabla R \nabla S \right] \quad \text{(equation 4.96)} \\
\nabla \frac{\partial R}{\partial t} &= \nabla \frac{\hbar}{2m} \left[ -\nabla^2 S - 2\nabla R \nabla S \right] \quad \text{(multiplying by \( \nabla \))}
\end{align*}
\]

\[
\begin{align*}
\frac{\partial \nabla R}{\partial t} &= \nabla \frac{\hbar}{2m} \left[ -\nabla^2 S - 2\nabla R \nabla S \right] \quad (4.101) \\
\frac{m}{\hbar} \frac{\partial u}{\partial t} &= \nabla \frac{\hbar}{2m} \left[ -\nabla \left( \frac{m}{\hbar} v \right) - 2 \left( \frac{m}{\hbar} u \right) \left( \frac{m}{\hbar} v \right) \right] \quad \text{(replacing \( \nabla R \) and \( \nabla S \))}
\end{align*}
\]

\[
\begin{align*}
\frac{\partial u}{\partial t} &= \nabla \frac{\hbar}{2m} \left[ -\nabla^2 S - \nabla (u \cdot v) \right] \quad \text{(eliminating \( m/\hbar \))}
\end{align*}
\]

\[
\frac{\partial v}{\partial t} = \nabla \frac{\hbar}{2m} \left[ -\nabla (u^2) - \frac{1}{2} \nabla (v^2) - \frac{1}{2} \nabla (v^2) - \frac{1}{m} \nabla V \right] \quad \text{(equation 4.99)}
\]

This completes the proof of theorem 4.67.
4.8 General Relativity

Analogous to the Newtonian case, Erik Verlinde \(^1\) shows that, assuming special relativity and the holographic principle, an entropic argument can be used to recover general relativity.

In the present paper, we have demonstrated special relativity by proving that the speed of light is a maximum, and we have obtained the holographic principle in our derivation of the Bekenstein-Hawking entropy. This augments Erik Verlinde’s derivation to a theorem of the proposed ToE.

Here we present a sketch of his derivation. He starts with a gradient of bits over the surface of a sphere

\[
dN = \frac{c^3 dA}{G\hbar} \tag{4.103}
\]

This result, proven in theorem 4.1, was assumed in Erik Verlinde’s paper. The mass of these bits over a surface is obtainable via integration

\[
M = \frac{1}{2} \int_S TdN \tag{4.104}
\]

\[
= \frac{1}{4\pi G} \int_S e^\phi \nabla \phi dA \tag{4.105}
\]

where \(e^\phi \Delta \phi\) is the red shift that he derives earlier in his paper. This equation is known to be the natural generalization of Gauss’s law to general relativity. He then argues that this is enough to recover the full description of general relativity. For brevity the proof will not be included here. Instead we will refer to Erik Verlinde’s paper for the rest.

The theory herein presented cast some light on Erik Verlinde’s theory and patches a few holes.

In his paper it is not explained why equation 4.103 is 4 times higher than the Bekenstein-Hawking entropy, the alleged highest possible entropy in the universe. Indeed, if the number of bits had been posed to be equal to the Bekenstein-Hawking entropy, the gravity equations would have been 4 times less than their actual values

\[
F = (0.25)GmM/r^2 \quad \text{and} \quad M = (0.25)1/(4\pi G) \int e^\phi \nabla \phi dA.
\]

This discrepancy is not explained in his paper.

Here we have a clear explanation of the 4 to 1 ratio. The Bekenstein-Hawking entropy links entropy to geometry. It is derived from packing as much spins as possible (which are 3 dimensional) in space. The 4 to 1 factor is obtain as we consider that 4 classical bits per spin are required to fully describe them, provided that such bits are on the surface of a sphere.

General relativity is not derived from the Bekenstein-Hawking entropy as this is just one of many possible encoding for the classical bits. General relativity is instead derived from the classical bits on the surface which encode the quantum information. As a result it is a purely classical theory (e.g. not quantum). Since those bits also encode position, it implies that, unlike the other forces, general relativity must necessarily apply to all systems and all particles.
5 Discussion

So what does this say about the universe?

5.1 The ToE is a Universal Theory of Knowledge (UTK)

From pure reason, we can pose any axioms $k$ and take them to their logical conclusions. This implies that the following first order sentence holds in the universe.

$$\forall k \forall t [(k \vdash t) \rightarrow \text{ToE} \vdash (k \vdash t)]$$

As shown in theorem 2.1, this sentence necessitate that the ToE is a UTK. To illustrate why this is important consider the reverse case. If the ToE was not a UTK, a mathematician could, using only a pen and paper, prove theorems that should not be provable in the universe - a contradiction. In a non-UTK universe, proving the wrong theorem on paper could literally end up destroying the universe.

5.2 Physics of UTK universes

A universe explained by a UTK has two logical parts; 1) an $\Omega$-sentence related to the provability of its theorems and 2) a set of axioms required to derive the answer to any of its sentences. Using these two we can skip over non-halting problems. Hence the answer to arbitrarily many questions can be derived in finite time (theorem 2.5).

In section 2.4, we reformulated the UTK as a universal Turing machine (UTM) where the $\Omega$-sentence becomes the halting probability of the UTM - its $\Omega$ number. From this reformulation, we noticed that we obtained a Gibb’s ensemble equation. Completing the ensemble with additional program-observables, we obtained an equation for the proposed ToE (part II) which is understood from an algorithmic information perspective.

In section 3, we have reinterpreted this equation within the language of thermodynamics. Indeed, each program-observable can be matched to a corresponding thermodynamic observable via its units. Physical laws obtained from this equation are applicable to every UTK-universe. Specifically, we have obtained;

1. Statistical physics (Section 3)

2. A law of conservation of energy (Theorem 3.13)

3. A maximum speed (Theorem 3.25)

4. Quantum mechanical spins (Equation 3.82)
5. A mechanism for the quantum measurement (Section 3.6)

6. The Schrödinger equation (Section 4.7)

7. General Relativity (Section 4.8)

8. A maximal entropy given by the Bekenstein-Hawking entropy and suggesting an holographic principle (Theorem 4.1)

9. A few more minor theorems spread around the text.

10. The list is obviously not complete and the author hopes that contributions from others will had to it.

From these results, it is reasonable to conjecture that the laws of physics are the same for all UTK-universes.

So why would a UTK-universe need axioms at all - they all have the same laws anyways? The axioms of a UTK-universe encode the outcome of quantum measurements. If a physical observer measures a certain spin, will it be up or will it be down? This information is encodable as a single bit along the halting probability $\Omega$ of the UTM-formulation of the UTK. $\Omega$ is maximally compressed and algorithmically random and for a physical observer, the bits cannot be derived from more fundamental principles. Hence, the result of each quantum measurement is necessarily an axiom of the universe.

If the axioms of a UTK-universe would encode anything else such that the laws of physics would change, even if only slightly, the resulting universe would cease to be a UTK and mathematicians would become weapons of mass destruction. So in summary, the laws of physics are common between all UTK and they differ only in terms of their quantum measurement outcomes. These outcomes are its axioms.

This is why formulating the ToE as a small set of axioms has been elusive... A proper formulation of it requires arbitrarily many axioms - one axiom to fix each quantum measurement outcome! The physical laws are in fact derivable from this formulation and from considering what is common between all such axiomatic constructions.

As a result, the axiomless derivation herein presented is the only possible formulation for the ToE. The axiomless derivation allows us to recover what is common between all UTK each having arbitrarily-many quantum measurement outcome as axioms. It is doable within unrestricted language, a UTK, without introducing any unprovable statements and without any appeal to experimental evidence.
5.3 What is the Quantum measurement?

The theory obtained herein provides an explanation of the quantum measurement and subsequent collapse.

Quantum mechanics only applies to physical observers. Such an observer will perceive a quantum mechanical system for all bits of $\Omega$ that are non-computable to him. Once an halting bit becomes computable to him, the bit can no longer be represented as a two-state system. This forces the elimination of the quantum system from the partition function. Hence, it seizes to be a quantum system. This is interpreted as the quantum mechanical collapse.

For an abstract observer, the situation is different. As the Kolmogorov complexity of $\Omega$ is infinite, it cannot be compressed. However, any finite subset of the bits of $\Omega$ can potentially be compressed via a smaller algorithm. An abstract observer able to run this algorithm as an experiment will have access to the bits of $\Omega$ before the universe produces them. The abstract observer sees a deterministic universe and has a fatalistic existence. He does not see a quantum mechanical universe as all bits of $\Omega$ are computable by him. The computational requirement to be an abstract observer are equal to the Kolmogorov complexity of the axioms of the universe. Hence only the universe itself can act as an abstract observer.

The physical observer sees a local theory of quantum mechanics that respects the Bell inequality. In contrast, an abstract observer sees a non-local universe with no quantum mechanics.

5.4 What is time?

The partition function contains the action-frequency observable. This observable pushes programs with very long run time deep into $\Omega$ such that the first digits of $\Omega$ are accurate within a small error rate (see theorem 3.50). As time passes, the error rate is reduced and more correct bits of $\Omega$ are known. Each additional bit gained provides us a way to define a distinct and uniquely identifiable time slice in an irreversible manner. Each time slice contains all past information and no future information.

5.5 Quantum mechanics plus general relativity

We have derived the Schrödinger equation, the spin, special and general relativity from the proposed ToE. This is one of the strongest evidence we present that the halting partition obtained is indeed the ToE. Deriving both QM and GR from it is an indication that they are unified within it.
5.6 What does it tell us about black holes?

We have considered the limiting case where all halting bits are non-computable to a physical observer. Hence each are described as a spin. We then asked how many such spins can we fit in space such that each is distinguishable (occupies a different quantum state). We found that we can fit as many as required to obtain the Bekenstein-Hawking entropy associated with black holes (theorem 4.1) but no more. Furthermore, we have shown that such a structure will have the Schwarzschild radius (theorem 4.34) and will emit Hawking radiation (theorem 4.43). It is a black hole in the usual sense.

On the premise that deriving both QM and GR is suggestive that the halting partition unifies them, we analyze the purely quantum mechanical black hole obtained and made exclusively of spins. To see how quantum effects changes the classical conception of the black hole, let us consider the case of a physical observer outside the black hole, then the case where he is inside.

To an outside physical observer, no information can exit the black hole. As a result a physical observer does not know any halting bit associated with the black hole. Hence, all halting bits in the halting partition associated with said black hole is represented as a two state spin system. For him, the black hole is a agglomeration of unmeasurable spins. The agglomeration has the Bekenstein-Hawking entropy and it emits Hawking radiation. Its radius, equal to the Schwarzschild radius is solely determined by the Pauli exclusion principle pushing the spins apart.

To an inside observer, things are more interesting. The results derived herein would seem to suggest the following;

- The spins in the sphere are encoded as classical bits on the surface of the sphere in a 4 to 1 ratio. Those bits encodes the spins which encode the halting knowledge of the observer’s universe.

- The mass of the black hole comes from the bits on its surface (equation 4.61). By analogy to a hollow sphere of mass $M$, we should expect the force of gravity inside the sphere to cancel out as it pulls equally from all directions on the objects inside of it.

- The physical observer inside the black hole can define a relationship between the entropy on the surface of the sphere and the position of an object away from this surface. This relationship recovers numerous laws of physics including, Newton’s law of inertia, Newton’s law of gravitation, Schrödinger equation and General relativity. All of these laws are derived inside the black hole.
General relativity is derived entirely from the classical bits encoding the content of the black hole. The bits are distributed on the surface of the sphere enclosing it. In this scenario, gravity is an entropic force that rearranges those bits so as to increase entropy over time. Hence gravity applies to all objects of the universe.

Since these results are derived as a consequence of the presence of a sphere of maximal entropy surrounding the observable universe (a.k.a. a black hole), the outside observer will likely experience a fraction of these laws unless he is also inside an even larger black hole.
6 Conclusion

So why does the axiomomless derivation work at all?

Sufficient reason, by virtue of which we consider that we can find no true or existent fact, no true assertion, without there being a sufficient reason why it is thus and not otherwise, although most of the time these reasons cannot be known to us.

–Gottfried Wilhelm Leibniz

The principle of sufficient reason of Leibniz suggests that for any fact to be true, there must be a sufficient explanation for it. If the principle holds, then it follows that positing an axiom as a requirement to derive the ToE violates the principle of sufficient reason. Therefore, if the universe does exist, then there must be an axiomless derivation of the laws of physics.

The axiomless derivation in this paper is a constructive proof of Leibniz’s principle of sufficient reason.

Remark 6.1. What about the axioms of quantum measurement (QM) outcomes, are there not arbitrarily many of them (end of section 5.2)? The axioms encoding QM outcomes are special cases. Each one can be true or false without contradicting the logical consistency of the universe. QM outcomes are the only physical facts whose true or false value are not determined by the ToE equation. Hence they must be taken as algorithmically random axioms. As a result the principle of sufficient reason is powerless to make a ruling as to their truth values. However, the existence of such axioms is provable by sufficient reason (the ToE equation requires these axioms) but the specific outcome is not enforced by any possible algorithm (also required by it). Hence the principle of sufficient reason of Leibniz holds for QM outcome axioms.

A summary of the argument presented goes as follows:

1. The axiomless theory (def 1.25) obtained from the primitive existence of language allows us to produce a universal theory of knowledge (def 1.89).

2. Using an antiquated term, a UTK essentially contains the “world of reason” envisioned by Plato where one is free to pose any assumption and correctly follow these assumptions to their logical conclusions. This “world” is valid for any and all assumptions posed. It comprises all the theorems that require at least one axiom to prove.

3. This “world of reason” taken as a whole provides the necessary condition to recover the ToE. Indeed since the “world of reason” is provable primitively, it follows that any ToE must be a UTK.


21 An appeal to falsifiability is considered to be an insufficient reason.
4. Finally, part III of this work is the derivation of the laws of physics from a UTK, as a convenience taken to be first-order arithmetic. The proof holds for any theory sufficiently descriptive to be a UTK.

5. From the restrictions imposed on a ToE by being a UTK, we have recovered the halting partition; an expression of the ToE in its most general form, and we have shown that it is sufficient by itself and with no appeal to physical observations, to derive quantum mechanic, statistical physics, general relativity, a proof of a maximal speed, etc.

This is presented as evidence of the existence of an axiomless derivation of physics producing the ToE.

We iterate the equation corresponding to the proposed ToE

\[
Z = \sum_{i=1}^{\infty} e^{-(\ln 2)\beta[T(i)+F_i+\sum_k c_k C_k(i)]}
\]

where each \(i\) is a program running on a UTM with the corresponding program-observables.

From arguments on the axiomless nature of the derivation herein presented and backed by the derivation of the laws of physics, it follows that it must be the ToE.

References


