The Experiments, as the Irreducible Basis of All Science, and the Observer, as the Probability Space of All Experiments, Are Found Sufficient to Entail All Physics

Alexandre Harvey-Tremblay[1]

[1]Independent scientist, aht@protonmail.ch

February 19, 2022

Abstract

While there exist in the wild a process to derive the laws of physics —namely, the practice of science— such has not been formalized in mathematics as of yet. Here, we understand this lack as an unrealized opportunity to investigate the relationship between experiments, science and physics entirely formally, and to report key findings. The first step in the program will be to eliminate all ambiguities from our language by expressing experiments using Turing complete languages and halting programs. A listing of such experiments via a machine or algorithm is recursively enumerable and, if understood as an *incremental contribution* to knowledge, then serves as a formulation of mathematics that models the practice of science purely formally. In turn this formulation leads to a definition of the observer as the probability space of all experiments, and the laws of physics are found by solving an optimization problem on said probability space. The final product is a comprehensive theory of physics entirely sourced from the practice of science, formulated from the perspective of the observer, exactly derived from having maximized what represents the limits of "*experimenter freedom*" in nature ("freedom" in the Bell Inequality sense), and free of all informal physical and metaphysical language.

Contents

1 Sketch

Before we go into details, let us first sketch the main ideas of the thesis, naturally without making any claim to rigour.

This work is primarily motivated by our identification of a missed opportunity to formulate the laws of physics without ambiguities, without informal physical or metaphysical language (sometimes called physical baggage — Max Tegmark[1]), and to derive them as the result of a theorem of some formal prescription, rather that having them be merely posited. The opportunity is realized by formalizing the practice of science itself, and then capitalizing on increased rigour and clarity. Ultimately, the program demystifies the origins of the laws of physics.

To support our main result, four systems will need to be introduced. The last, being the formal derivation of the laws of physics, constitute the main result. The systems are:

1. A Formal System of Knowledge

   In this section, we select the appropriate mathematical tool to model knowledge itself, and we justify the choice. Specifically, a unit of knowledge will be modelled as a halting program. Then, each discovery of a halting program will be taken as a contribution to the lexicon of knowledge. The lexicon, as it includes all halting programs, is Turing complete, and consequently will not be decidable, but rather recursively enumerable. Furthermore, as it is Turing-complete, the lexicon is maximally expressive, and on par with any other Turing complete language. Finally, we show that the lexicon meets, exactly, the epistemological definition of knowledge.

2. A Formal System of Science (used to "map" knowledge)

   Here, we introduce the notion of an experimental protocol, comprised of an initial preparation along with a series of steps to be performed on the preparation such that it terminates with a result. We then further introduce the notion of a universal experimenter; a machine that can carry out any experimental protocol in nature, and we find that it admits an equivalent mathematical definition to that of a halting program or Turing machine. Finally, we argue that any physical system can be described equivalently, and without ambiguities, by a corresponding collection of terminated experimental protocols (in lieu of, say, dynamical equations).

   In this system, a scientific method is defined as a function that recursively enumerates experimental protocols. A specific listing of halting programs or terminating experimental protocols produces an incremental contribution of mathematical or experimental knowledge, respectively, and such can be used to validate or invalidate predictive models of knowledge — yielding an epistemologically-complete formalization of the practice of science that is entirely free of informal physical or metaphysical language.

3. A Formal Theory of the Observer (used to practice science)

   Here, we produce who or what practices science, and we answer how it does so. We attack the problem from this angle: If the observer *deterministically* produces a recursive enumeration of knowledge, then it is merely a machine and the physics it is subjected to is super-deterministic; If however, the observer *probabilistically* produces a recursive enumeration of knowledge, then it is a probility space of experiments, and the corresponding physics models 'experimenter freedom' (in the sense of the Bell inequality). The term freedom is to be interpreted similarly to the strict sense given by Conway and Kochen[2] and the Bell inequality as freedom from being determined by past history, thus allowing a probabilistic enumeration in the present. It is not meant to be interpreted in the metaphysical sense of free will used in philosophy. In the probabilistic enumeration case, the observer would represent an execution of scientific participation in nature under the assumption of experimenter freedom, and it is in this context that the laws of physics are derived.

4. A Formal Model of Physics (used to model what the observer can or cannot do as it practices science).

   Here, we find that physics is intimately connected to, and ultimately derived from, the definition of the observer as a probability space of experiments. Indeed, to obtain the laws of physics, it now only suffices to maximize the quantity of information of the measure (i.e. find a solution to an optimization problem), and this is sufficient to uniquely entail, as we found: quantum field theory, general relativity, and up to and including quantum gravity, and to offer resolution to enduring problems such as the origin of the Born rule and the mechanism of the wave-function collapse.

   The result is a theory of physics best interpreted as an information-theoretical claim regarding what the observer can or cannot do from its participation in nature. In fact, the model of physics, as it is obtained by maximizing the entropy/information of the probability space (which represents experimenter freedom), thus describe the maximally permissive model of such participation. An inviolability of said laws is therefore realized with respect to the observer's own actions, as the laws are recovered at the limit what the observer can or cannot do. This narrows the scope of the laws of physics to its most fundamental expression, whilst nonetheless remaining sufficiently comprehensive to account for not only the laws of physics and their origins, but also for their uniqueness, universality and inviolability.

We pass now to the detailed and rigorous execution of the argument sketched above.

2 The Formal System of Knowledge

Mathematics tends to be formulated on the backbone of *theories of truth*, for instance propositional logic or first order logic, and their aim are to correctly *propagate* truth from statement to statement; whereas in the sciences, we tend to find *theories of knowledge* whose aim are to produce *incremental contributions* to validate (or invalidate) an ever more complete model thereof.

Knowledge is similar to truth in many ways. For instance, both quantitatively relate to a binary state: knowledge is either known (1) or unknown (0), and truth is either true (1) or false (0). But, theories of truth tend to view incompleteness as a weakness since it signals obscurities, whereas those of knowledge seek it as it signals an opportunity for progress. Furthermore, axiomatic theories formulated in terms of truth tend to clash with one another (incompatible premises entail contradictions), whereas those formulated based on knowledge contribute to one another (knowledge is closed under union).

We have a plurality of theories of truth in mathematics, but so far we have not captured these differences and intuitions into a formal system of knowledge.

Let us begin by stating that attempts to find a complete logical basis for truth have been made ad nauseam in the past but they failed for primarily two reasons. First, they were attempted before Gödel-type theorems were known and appreciated, and attempts were directed at constructing *decidable* logical bases for truth. Secondly, instead of directing efforts to recursively enumerable bases following the discovery of said incompleteness theorems, efforts simply felt out of favour as it was understood that any sufficiently expressive system of truth would contain obscurities, and this made them philosophically unattractive. It is however possible to construct recursively enumerable bases (provided they are not decidable), and further the limitations of recursive enumeration ought to instead be seen as an opportunity; in this case, to create a formal system to map out knowledge, such that it may serve as the foundation to a formalization of science. In this case, the theory challenges us to discover new knowledge, rather than to merely fix truth definitionally only to bail out at the first obscurity, and in this context we call it a theory of knowledge to distinguish it from a theory of truth. Theories of knowledge, as recursively enumerable systems, are a more general concept than theories of truth which are subsets thereof. Indeed, for all statements that are either true (1) or false (0), it is the case that we can know (1) its truth value; but if a statement is such that it is undecidable, a binary state of knowledge still applies to it, in this case its truth value is unknown (0).

To help appreciate the utility and to fix the intuition, consider the following "amusing" construction which we will call rotting arithmetic. In logic we are allowed to inject any sentence as a new axiom, and to investigate its consequences. Rotting arithmetic will be defined as the union of the axioms of Peano's arithmetic and of the axiom of rot, which we define as follows:

$$\text{Axiom of rot} := \left(2^{2^{82,589,933}} - 1\right) \text{ is a prime} \tag{1}$$

$$\text{Rotting Arithmetic} := \{\text{Peano's Arithmetic}\} \cup \{\text{Axiom of rot}\} \tag{2}$$

The axiom of rot claims that a very large is number is a prime. If it's true, then it has no effects on the system, but if it's false, the system is inconsistent. Comparatively, the largest known prime (at the time of this writing) is $2^{82,589,933} - 1$ which is orders of magnitude smaller than the number referenced in the axiom of rot. Since we have used randomness to generate the axiom of rot, odds are minuscules that it is a prime... or perhaps we did hit the jackpot and it is a prime. A theory of knowledge can assign the state unknown (0) to the axiom of rot until such a time as we find out if the proposed number is or isn't a prime; whereas a theory of truth expects true or false right now, as it's truth-value is fixed in principle.

It may be that it takes us a century (or maybe less, maybe more, maybe never) until we find out if the axiom of rot is or isn't true, as our computing capacities may need to improve before we can know. As time goes by the "freshness" of the theory slowly diminishes, until such a time as it is revealed to be rotten at which time it is discarded (or it keeps perpetually fresh if we did hit the jackpot and the number is a prime).

The example of rotten arithmetic may appear convoluted or unnatural — after-all why would we take the chance with an axiom of rot, when we can easily do arithmetic without it —, but now consider what often happens in science. For nearly a century before Einstein produced the theory of special relativity (Einstein, 20th century), the union of both classical mechanics (Newton, 17th century) and electromagnetism (19th century) was "fresh":

$$\begin{aligned} \text{Law of Inertia} :=& \quad F = ma && (3) \\ \text{Maxwells' equation} :=& \quad \nabla \cdot \mathbf{E} = \rho/\epsilon_0, \nabla \cdot \mathbf{B} = 0, ... && (4) \\ \text{Union} :=& \quad \{F = ma\} \cup \{\text{Maxwells' equation}\} && (5) \end{aligned}$$

The discovery of "rot" in their union (Maxwell's equations reports a constant speed of light independently of the observer's velocity, whereas velocities in F=ma are additive) had to wait for nearly a century to be noticed and corrected. In the mean time, most were happy to use both theories, and the problem remained unnoticed. Similarly to the case of rotten arithmetic, the state of knowledge of "rot" in the union had to go from unknown (0) to known (1), before a new model was to be produced.

Falsification in general can be manipulated in a similar fashion. But instead of having two axiomatic theories, we have an empirical statement along with an axiomatic theory:

$$\begin{aligned} \text{Observation} :=& \quad \text{Precession of Mercury's orbit} && (6) \\ \text{Law of Gravitation} :=& \quad F = GmM/r^2 && (7) \\ \text{Falsification?} :=& \quad \{\text{P[...] of Mercury's orbit}\} \cup \{F = GmM/r^2\} && (8) \end{aligned}$$

The statement "Precession of Mercury's orbit" would plausibly indicate a sequence of measurements, observations or experiments, and falsification occurs if they are not solutions to the law of gravitation.

So far, we have discussed the intuition only informally, and so the next step is to ask what mathematical tools are the best to describe knowledge formally? To find out we must be a bit more technical. Let us look at the philosophical discipline that study knowledge: epistemology — What does it tell us about knowledge, that we can use?

Epistemology, at least historically and dating all the way back to Plato, has considered knowledge to be that which is a justified true belief. For instance "I know Bob is from Arkansas (as a justified true belief), because his driver's license is from Arkansas (justification), and he is from Arkansas (true)". However, the Gettier problem[3] is a well known objection to this definition. Essentially, if the justification is not loophole free, there exists a case where one is right by pure luck, even if the claim were true and believed to be justified. For instance, if one glances at a field and sees a shape in the form of a dog, one might think he or she is justified in the belief that there is a dog in the field. Now suppose there is a dog elsewhere in the field, but hidden from view. The belief "there is a dog in the field" is justified and true, but it is a hard sale to call it knowledge because it is only true by pure luck.

Richard Kirkham[4] proposed to add the criteria of infallibility to the justification. Knowledge, previously *justified true belief*, would now be *infallible true belief*. Merely seeing the shadow of a dog in a field would not be enough to qualify as infallible true belief, as all claims will have to be exactly proportional to the evidence. This is generally understood to eliminate the loophole, but it is an unpopular solution because adding it is assumed to reduce knowledge to radical skepticism in which almost nothing is knowledge, thus rendering knowledge non-comprehensive.

Here, we will adopt the insight of Kirkham regarding the requirement of infallibility whilst resolving the non-comprehensiveness objection, and also retaining the intuitive characteristics of knowledge as we have described them in this introduction. To do so, we will structure our statements such that they are individually infallible, yet as a group form a Turing complete language.

Our tool of choice for this will be halting programs, and they will act as the building blocks of knowledge in our system. Here, we understand halting programs as a descriptive language, similar in expressive power to any other Turing complete language, such as say english. But unlike english, using halting programs makes the description of each unit of knowledge completely free of ambiguities. And ambiguities are of course antithetical to knowledge. General translations between all Turing languages exists, and so we do not lose any expressive power by using them, over any other choice of language. For instance, any mathematical problem can be reformulated as a statement regarding the halting status of a program via the Curry–Howard correspondence.

The primary advantage in using a listing of halting programs to represent our state of knowledge is that it will allow us to union all new discoveries of knowledge with older ones, without any risk of the new ones invalidating the previous ones, thus making knowledge closed under union. Indeed, if a program is known to halt, then no other halting programs discovered afterwards can contradict that. Rather, it will be *explanatory models of knowledge* that would or could be invalidated (falsified) by new knowledge. Contributions of new knowledge to a Turing-complete lexicon will thus be incremental by guarantee.

Halting programs are of course subject to the halting problem and this will entail our system to be a trial and error system. Consequently, acquiring new knowledge will be difficult, even arbitrarily difficulty, and may even contain dead-ends (non-halting programs). This trial and error effect will in turn become the basis of a formal model of science that is entirely formalized, yet comprehensive.

Let us inform the reader that information regarding the connection between mathematics, science and programs, is available in the seminal works of Gregory Chaitin[5, 6, 7]. A familiarity with his work is assumed.

2.1 Halting Programs as Knowledge

How do we construct an infallible statement, so that it qualifies as an epistemic statement in the sense of Kirkham?

Let us take the example of a statement that may appear as an obvious true statement such as "$1+1=2$", but is in fact not infallible. Here, we will provide the definition of an infallible statement, but equally important, such that the set of all such statements is Turing complete, thus forming a language of maximum expressive power.

Specifically, the sentence "$1+1=2$" halts on some Turing machine, but not on others and thus is not infallible. Instead consider the sentence $\text{PA} \vdash [s(0)+s(0)=s(s(0))]$ to be read as "Peano's axioms prove that $1+1=2$". Such a statement embeds as a prefix the set of axioms in which it is provable. One can deny that $1+1=2$ (for example, an adversary could claim binary numbers, in which case $1+1=10$), but if one specifies the exact axiomatic basis in which the claim is provable, said adversary would find it harder to find a loophole to fail the claim. Nonetheless, even with this improvement, an adversary can fail the claim by providing a Turing machine for which $\text{PA} \vdash [s(0)+s(0)=s(s(0))]$ does not halt.

The key is to structure the statement so that all context required to prove the statement is provided along with the statement itself; then it is the claim that the context entails the statement that is infallible. If we use the tools of theoretical computer science we can produce statements free of all loopholes, thus ensuring they are infallible. Those statements, which are mathematical theorems, are also —via Curry–Howard correspondence— the halting programs. The value in the knowledge acquired by knowing that a specific programs halts is associated to the "difficulty" of running the program until termination. Let us now introduce a few definitions.

Let Σ be a set of symbols; called an alphabet. A word is a sequence of symbols from Σ. The empty word is represented as $\varnothing$. The set of all finite words is given as:

$$\mathbb{W} := \bigcup_{i=0}^{\infty} \Sigma^i \tag{9}$$

Finally a language $\mathbb{L}$ is as a subset of $\mathbb{W}$.

As an example, the sentences of the binary alphabet $\Sigma = \{0, 1\}$ are the binary words $\{\varnothing, 0, 1, 00, 01, 10, 11, 000, \ldots\}$.

There exists multiple models of computation, such a Turing machines, μ-recursive functions, Lambda calculus, etc. Here, to retain generality we will use computable functions without requiring a specific model.

Instead of a Turing machine, we will consider a Turing-computable function and its definition is as follows:

A Turing machine Φ computes a partial function $\text{TM}\colon \mathbb{W} \to \mathbb{W}$ iff:

1. For each $d \in \text{Dom}(\text{TM})$, $\Phi(d)$ halts and equals $\text{TM}(d)$.

2. For each $d \notin \mathrm{Dom}(\mathrm{TM})$, $\Phi(d)$ never halts.

Then, TM is a Turing-computable function (or simply, a computable function). We denote $\mathbb{TM}$ as the set of all computable partial functions from $\mathbb{W} \to \mathbb{W}$.

Likewise, and instead of a universal Turing machine as a specific implementation, we will prefer to use a universal Turing-computable partial function of two inputs $\mathrm{DM}\colon \mathbb{W} \times \mathbb{W} \to \mathbb{W}$. To use the elements of $\mathbb{TM}$ in this function, we must introduce a bijective function, which we call an interpreter, as: $\langle\cdot\rangle\colon \mathbb{TM} \leftrightarrow \mathbb{W}$ specific to the DM. Then, if forall $\mathrm{TM} \in \mathbb{TM}$ and forall $d \in \mathrm{Dom}(\mathrm{TM})$, it is the case that if $\mathrm{DM}(\langle \mathrm{TM} \rangle, d) \simeq \mathrm{TM}(d)$, then DM is a universal function, and we denote it as UTM.

Definition 1 (Halting Program)**.** *A halting program p is a pair $\mathbb{TM} \times \mathbb{W}$:*

$$p := (\mathrm{TM}, d) \tag{10}$$

such that $\mathrm{TM}(d) = r$.

With this definition, d can be considered as the statement, TM is its context, and if $\mathrm{TM}(d)$ halts, then both are paired as a context-free halting claim:

$$p = (\mathrm{TM}, d); \mathrm{UTM}(\langle \mathrm{TM} \rangle, d) \text{ halts} \tag{11}$$

Since a translation exists between universal Turing machine, a claim that d halts on TM, if known, entails "p halts" is verifiable on all universal Turing machines, and requires no specific context for this to be verified.

For instance, the following (trivial) program halts:

```
fn one_plus_one_equals_two(){
 if 1+1==2{
  return;
 }
 loop{};
}
```

The claim "$p = (\text{cargo run}, \text{one\_plus\_one\_equals\_two}); \mathrm{UTM}(p)$ halts" is a unit of knowledge, and I can contribute it to the lexicon.

A less trivial example is shown in Annex C which presents a formal proof of the commutativity of addition for natural numbers written in COQ[8]. Thus, the claim "$p = (\text{COQ}, \text{plus\_comm})$; $\mathrm{UTM}(p)$ halts" would be another unit of knowledge.

The second objection is that the infallibility requirement is too demanding, preventing knowledge from being comprehensive by making it able at most to only tackle a handful of statements. However, the set of all halting programs constitutes the entire domain of the universal Turing machine, and thus the expressive power of halting programs must be on par with any Turing complete language. Since there exists no greater expressive power for a formal language than that of Turing completeness, then no reduction takes place. The resulting construction is both element-wise infallible, and comprehensive as a set:

Definition 2 (Lexicon (of Knowledge))**.** *The set of all programs* $\mathbb{TM} \times \mathbb{W}$ *that halts constitutes the lexicon of knowledge* $\mathbb{K}$*.*

- $\mathbb{K}$ *constitute the set of all knowledge-bearing statements.*
- $\mathbb{K}$ *is non-computable, but is recursively enumerable.*
- $\mathbb{K}$ *contains countably infinitely many elements.*
- *We can definite* $\mathbb{K}$*, we can also contribute to it, but we cannot complete it.*
- *Unlike the hyperwebster[9] which includes all possible words from* Σ *regardless of halting status and thus is without knowledge, here each entry is a halting program and thus bears knowledge from its context.*

Definition 3 (Translation (of $\mathbb{K}$))**.** *A translation* $\mathbb{T}$ *of* $\mathbb{K}$ *is a map from* $\mathbb{TM} \times \mathbb{W}$ *to* $\mathbb{W} \times \mathbb{W}$ *such that the interpreter function* $\langle \cdot \rangle$ *is applied to each element of* $\mathbb{TM}$*. Each translation of* $\mathbb{K}$ *corresponds to the domain of a universal Turing machine* UTM*.*

$$\mathbb{T}_{\mathrm{UTM}} := \mathrm{Dom}(\mathrm{UTM}) \tag{12}$$

And contains all pairs $\mathbb{W} \times \mathbb{W}$ *that halt on* UTM*.*

Theorem 1 (Incompleteness Theorem)**.** *Since a translation of* $\mathbb{K}$ *is the domain of a* UTM*, then it is undecidable. The proof follows from the domain of a universal Turing machine being undecidable. Finally, since* $\langle \cdot \rangle$ *is bijective, it follows that* $\mathbb{K}$ *is also undecidable.*

The theorem implies that we will never run out of new knowledge to discover, and can thus perpetually contribute to the lexicon.

Theorem 2 ($\mathbb{K}$ is recursively enumerable)**.** *We will list* $\mathbb{K}$ *by dovetailing.*

Proof. First, let us recursively enumerate the translation $\mathbb{T}$ of $\mathbb{K}$. Consider a dovetail program scheduler which works as follows.

1. Sort the columns of $\mathbb{W} \times \mathbb{W}$ in shortlex:

$$\begin{array}{lllll}
 & d_1 & d_2 & d_3 & \dots \\
\langle \mathrm{TM}_1 \rangle & (\langle \mathrm{TM}_1 \rangle, d_1) & (\langle \mathrm{TM}_1 \rangle, d_2) & (\langle \mathrm{TM}_1 \rangle, d_3) & \dots & (13) \\
\langle \mathrm{TM}_2 \rangle & (\langle \mathrm{TM}_2 \rangle, d_1) & (\langle \mathrm{TM}_2 \rangle, d_2) & (\langle \mathrm{TM}_2 \rangle, d_3) & \dots & (14) \\
\langle \mathrm{TM}_3 \rangle & (\langle \mathrm{TM}_3 \rangle, d_1) & (\langle \mathrm{TM}_3 \rangle, d_2) & (\langle \mathrm{TM}_3 \rangle, d_3) & \dots & (15) \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

then trace a line across the pairs starting at $(\langle \mathrm{TM}_1\rangle, d_1)$ then $(\langle \mathrm{TM}_2\rangle, \mathrm{d}_1)$, $(\langle \mathrm{TM}_1\rangle, \mathrm{d}_2)$, $(\langle \mathrm{TM}_2\rangle, \mathrm{d}_2)$, $(\langle \mathrm{TM}_3\rangle, \mathrm{d}_1)$ and so on. This produces an order which grabs all pairs.

2. Take the first element of the sort, $\mathrm{DM}(\langle \mathrm{TM}_1\rangle, \mathrm{d}_1)$, then run it for one iteration.

3. Take the second element of the sort, $\mathrm{DM}(\langle \mathrm{TM}_2\rangle, \mathrm{d}_1)$, then run it for one iteration.

4. Go back to the first element, then run it for one more iteration.

5. Take the third element of the sort, $\mathrm{DM}(\langle \mathrm{TM}_1\rangle, \mathrm{d}_2)$, then run it for one iteration.

6. Continue with the pattern, performing iterations one by one, with each cycle adding a new element of the sort.

7. Make note of any pair $(\langle \mathrm{TM}_i\rangle, \mathrm{d}_j)$ which halts.

Finally, use the interpreter function to convert $\mathbb{W} \times \mathbb{W}$ to $\mathbb{TM} \times \mathbb{L}$, yielding the lexicon. □

This scheduling strategy is called dovetailing and allows one to enumerate the domain of a universal Turing machine recursively, without getting stuck by any singular program that may not halt. Progress will eventually be made on all programs... thus producing a recursive enumeration.

Definitionally, the domain of a recursively enumerable function is a set; however in practice and implemented as an algorithm, a dovetailer and other implementations of recursive enumerations produces a sequence of *incremental contributions* to knowledge, as each new element that halts gets added to a list; the order of which depends on the implementation.

2.2 Incremental Contributions

We will now use the lexicon of knowledge and halting programs to redefine the foundations of mathematics in terms of *incremental contributions* to knowledge, replacing *formal axiomatic systems*.

In principle, one can use any Turing complete language to re-express mathematics. The task is not particularly difficult. One generally has to build a translator between the two formulation, whose existence is interpreted as a proof of equivalence. For instance, one can write all of mathematics using the english language, or using set theory with arbitrary equipments, or using a computer language such as c++, or using arithmetic with multiplication, etc. If the language is Turing complete, then it is as expressive as any other Turing complete language, and a translator is guaranteed to exist. So why pick a particular system over another? This is often due to conveniences and constraints other than pure expressive power. For instance, sets allow us to intuitively express a very large class of mathematical problems quite conveniently. Typical selection criterions are; can we express the problem at hand clearly?, elegantly?, are the solutions also clear and easier to formulate, than in the alternative system?

Here we will use and introduce the incremental contribution formulation of mathematics, and, as we will see, its advantages are stunning. An incremental contribution comprises a group of programs known to halt, and this group of programs defines a specific instance of accumulated mathematical knowledge.

Definition 4 (Incremental Contribution (to Knowledge))**.** *Let $\mathbb{K}$ be the lexicon of knowledge. An incremental contribution $\mathbf{m}$ of n halting program is an element of the n-fold Cartesian product of $\mathbb{K}$:*

$$\mathbf{m} \in \mathbb{K}^n \tag{16}$$

The tuple, in principle, can be empty $\mathbf{m} := ()$, finite $n \in \mathbb{N}$ or countably infinite $n = \infty$.

- *Note on the notation: we will designate $p_i = (\mathrm{TM}_i, d_i)$ as an halting program element of $\mathbf{m}$, and $\mathrm{proj}_1(p_i)$ and $\mathrm{proj}_2(p_i)$ designate the first and second projection of the pair p_i, respectively. Thus $\mathrm{proj}_1(p_i)$ is the TM_i associated with p_i, and $\mathrm{proj}_2(p_i)$ is the input d_i associated with p_i. If applied to a tuple or set of pairs, then $\mathrm{proj}_1(\mathbf{m})$ returns the set of all TM in $\mathbf{m}$ and $\mathrm{proj}_2(\mathbf{m})$ returns the set of all inputs d in $\mathbf{m}$.*

The programs comprising the incremental contribution adopt the normal role of both axioms and theorems and form a single verifiable atomic concept constituting a unit of mathematical knowledge. Let us explicitly point out the difference between the literature definition of a formal system and ours: for the former, its theorems are a subset of the sentences of $\mathbb{L}$ provable from the axioms — whereas for a sequence of incremental contributions, its elements are pairs of $\mathbb{TM} \times \mathbb{W}$ which halts on a UTM.

Let us now explore some of the advantages of using incremental contributions versus formal axiomatic systems. Sequences of incremental contributions are more conductive to a description of the scientific process, including the accumulation of experimental knowledge, than formal axiomatic systems are. Let us take an example. Suppose we wish to represent in real-time, and with live updates, the set of all knowledge produced by a group of, say 50,000 mathematicians working in a decentralized manner (perhaps from their offices) over the course of at least many decades, and perhaps even for an indefinite amount of time into the future. Some of the work they produced may build on each others', but it will also be the case that part of their work is incompatible. For instance, some might find contradictions in their assumptions and abandon large segments of their work. As one learns primarily from his or her errors, we may wish to catalogue these contradictions for posterity. Let us first try with formal axiomatic systems. Finding the 'correct' and singular formal axiomatic system to describe the totality of what they have discovered, including abandoned work and contradictions, will be quite a challenge. One challenge occurs whenever a new contradiction is found, as one would need to further isolate it within a wrapper of para-consistent logic, before inclusion within a formal axiomatic system. Another challenge occurs when mathematicians invent new, possibly more elegant, axiomatic basis outright. One would constantly need to adjust his or her proposed formal axiomatic system to account for new discoveries as they are made. Such an axiomatic basis would eventually grow to an unmaintainable level, not unlike the spaghetti codes of the early days of software engineering. And we have not even mentioned the problems spawned by general incompleteness theorems such as those of Gödel and Gregory Chaitin, and the negative resolution to Hilbert's second problem. What if someone proves a statement (using a new axiomatic basis) that is not provable from the "master" axiomatic basis; in this case re-adjustments are perpetually necessary. As mathematicians are a creative bunch, one would never be able to settle on a final axiomatic system as they could always decide to explore a sector of mathematical space not covered by the current system. Comparatively, using an incremental contribution, the task is much easier: One simply need to push each new discovery at the end of the sequence; no adjustment is ever required after insertion, we never run out of space, and halting programs do not undermine each other even if they internally represent a contradiction. An incremental contribution is the equivalent of an empirical notebook of raw mathematical knowledge.

Formal axiomatic systems do not excel at pure description because they are more akin to an *interpretation* of mathematical knowledge based on a preference of some patterns or tools (we like sets, thus ZFC!, or we prefer categories, thus category theory!). New knowledge and new problems will eventually force one to challenge this preference. Not so with incremental contributions! Incremental contributions are a true and final representation of pure unadulterated mathematical knowledge.

2.3 Connection to Formal Axiomatic Systems

We can, of course, connect our incremental contributions formulation of mathematics to the standard formal axiomatic system (FAS) formulation:

Definition 5 (Enumerator (of a FAS)). *Let* FAS *be a formal axiomatic system and let s be a valid sentence of* FAS*. A function* $\text{enumerator}_{\text{FAS}}$ *is an enumerator for* FAS *if it recursively enumerates the theorems of* FAS*:*

$$\text{enumerator}_{\text{FAS}}(s) = \begin{cases} 1 & \text{FAS} \vdash s \\ \nexists/\ \text{does-not-halt} & \text{otherwise} \end{cases} \tag{17}$$

Definition 6 (Domain (of a FAS)). *Let* FAS *be a formal axiomatic system and let* $\text{enumerator}_{\text{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then the domain of* FAS*, denoted as* Dom(FAS)*, is the set of all sentences $s \in \mathbb{L}$ which halts for* $\text{enumerator}_{\text{FAS}}$*.*

Definition 7 (Formal Axiomatic Representation (of a sequence of incremental contributions)). *Let* FAS *be a formal axiomatic system, let* $\mathbf{m}$ *be a sequence of* *incremental contributions and let* $\text{enumerator}_{\text{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then* FAS *is a formal axiomatic representation of* $\mathbf{m}$ *iff:*

$$\text{Dom}(\text{FAS}) = \text{proj}_2(\mathbf{m}) \tag{18}$$

Definition 8 (Episto-morphism)**.** *Let* $\mathbb{FAS}$ *be the set of all formal axiomatic systems; an episto-morphism is a map* $M\colon \mathbb{FAS} \to \mathbb{FAS}$ *such that* $\forall\, \text{FAS} \in \mathbb{FAS}\colon \text{Dom}(\text{FAS}) = \text{Dom}(M(\text{FAS}))$*. Two formal axiomatic systems* FAS_1 *and* FAS_2 *are said to be episto-morphic if and only if* $\text{Dom}(\text{FAS}_1) = \text{Dom}(\text{FAS}_2)$*.*

2.4 Discussion — The Mathematics of Knowledge

Each element of an incremental contribution is a program-input pair representing an algorithm which is known to halt. Let us see a few examples.

How does one know how to tie one's shoes? One knows the algorithm required to produce a knot in the laces of the shoe. How does one train for a new job? One learns the internal procedures of the shop, which are known to produce the result expected by management. How does one impress management? One learns additional skills outside of work and applies them at work to produce results that exceed the expectation of management. How does one create a state in which there is milk in the fridge? One ties his shoes, walks to the store, pays for milk using the bonus from his or her job, then brings the milk back home and finally places it in the fridge. How does a baby learn about object permanence? One plays peak-a-boo repeatedly with a baby, until it ceases to amuse the baby — at which point the algorithm which hides the parent, then shows him or her again, is learned as knowledge. How does one untie his shoes? One simply pulls on the tip of the laces. How does one untie his shoes if, after partial pulling, the knot accidentally tangles itself preventing further pulling? One uses his fingers or nails to untangle the knot, and then tries pulling again.

Knowledge can also be in more abstract form — for instance in the form of a definition that holds for a special case. How does one know that a specific item fits a given definition of a chair? One iterates through all properties referenced by the definition of the chair, each step confirming the item has the given property — then if it does for all properties, it is known to be a chair according to the given definition.

In all cases, knowledge is an algorithm along with an input, such that the algorithm halts for it, lest it is not knowledge. The set of all known pairs form an incremental contribution to knowledge.

Let us consider a few peculiar cases. What if a sequence contains both "A" and "not A" as theorems? For instance, consider:

$$\mathbf{m} := \big((\text{TM}_1, A), (\text{TM}_1, \neg A)\big) \tag{19}$$

Does such a contradiction create a problem? Should we add a few restrictions to avoid this unfortunate scenario? Let us try an experiment to see what happens — specifically, let me try to introduce $A \wedge \neg A$ into my personal knowledge, and then we will evaluate the damage I have been subjected to by this insertion. Consider the following implementation of TM_1:

```
fn main(input: String){
 if p=="A" {
  return;
 }
 if p=="not A"{
  return;
 }
 loop();
}
```

It thus appears that I can have knowledge that the above program halts for both "A" and "not A" and still survive to tell the tale. A-priori, the sentences "A" and "not A" are just symbols. Our reflex to attribute the law of excluded middle to these sentences requires the adoption of a deductive system. This occurs one step further at the selection of a specific formal axiomatic representation of the sequence of incremental contributions, and not at the level of the sequence itself.

The only inconsistency that would create problems for this framework would be a proof that a given halting program both [HALTS] and [NOT HALTS] on a UTM. By definition of a UTM, this cannot happen lest the machine was not a UTM to begin with. Thus, we are expected to be safe from such contradictions.

Now, suppose one has a sizeable sequence of incremental contributions which may contain a plurality of pairs:

$$\mathbf{m} := \big((\mathrm{TM}_1, d_1), (\mathrm{TM}_2, \neg d_1), (\mathrm{TM}_1, d_2), (\mathrm{TM}_2, d_1), (\mathrm{TM}_2, \neg d_3)\big) \tag{20}$$

Here, the negation of some, but not all, is also present across the pairs: in this instance, the theorems d_1 and d_3 are negated but for different premises. What interpretation can we give to such elements of a sequence? For our example, let us call the sentences d_1, d_2, d_3 the various flavours of ice cream. It could be that the Italians define ice cream in a certain way, and the British define it in a slightly different way. Recall that halting programs are pairs which contain a computable function and a premise. The computable function contains the 'definition' under which the flavour qualifies as real ice cream. A flavour with a large spread is considered real ice cream by most definitions (i.e. vanilla or chocolate ice cream), and one with a tiny spread would be considered real ice cream by only very few definitions (i.e. tofu-based ice cream). Then, within this example, the presence of p_1 and its negation associated with another definition, simply means that tofu-based ice cream is ice cream according to one definition, but not according to another.

Reality is of a complexity such that a one-size-fits-all definition does not work for all concepts, and further competing definitions might exist: a chair may be a chair according to a certain definition, but not according to another. The existence of many definitions for one concept is a part of reality, and a mathematical framework which correctly describes it ought to be sufficiently flexible to handle this, without itself exploding into a contradiction.

Even in the case where both A and its negation $\neg A$ were to be theorems of $\mathbf{m}$ while also having the same premise, is still knowledge. It means one has verified that said premise is inconsistent. One has to prove to oneself that a given definition is inconsistent by trying it out against multiple instances of a concept, and those 'trials' are each incremental contributions.

2.5 Axiomatic Information

Let us introduce axiomatic information. If any account for the elements of any particular incremental contribution is relegated to having been 'randomly picked', according to a probability measure ρ, from the set of all possible halting programs, then we can quantify the information of the pick using the entropy.

Definition 9 (Axiomatic Information)**.** *Let $\mathbb{Q}$ be a set of halting programs. Then, let $\rho : \mathbb{Q} \to [0,1]$ be a probability measure that assigns a real in $[0,1]$ to each program in $\mathbb{Q}$. The axiomatic information of a single element of $\mathbb{Q}$ is quantified as the entropy of ρ:*

$$S = -\sum_{p\in\mathbb{Q}} \rho(p) \ln \rho(p) \tag{21}$$

For instance, a well-known (non-computable) probability measure regarding a sum of prefix-free programs is the Halting probability[10] of computer science:

$$\Omega = \sum_{p\in\mathrm{Dom(UTM)}} 2^{-|p|} \implies \rho(p) = 2^{-|p|} \tag{22}$$

The quantity of axiomatic information (and especially its maximization), rather than any particular set of axioms, will be the primary quantity of interest for the production of a maximally informative theory in this framework. A strategy to gather mathematical knowledge which picks halting programs according to the probability measure which maximizes the entropy will be a maximally informative strategy.

3 The Formal System of Science

We now assign to our re-formulation of mathematics in terms of incremental contributions, the interpretation of a purely mathematical system of science. As hinted previously, the primary motivation for constructing a system of science follows from the set of knowledge being recursively enumerable (as opposed to decidable) making its enumeration subject to the non-halting problem. Notably, in the general case, halting programs can only be identified by trial and error and this makes the approach irreducibly experimental.

At this point in the paper, I must now warn the reader that, based on my previous experience, almost any of the definitions I choose to present next will likely either quickly induce at least a feeling of uneasiness, or may even trigger an aversion in some readers. First and foremost, let me state that the definitions are, we believe, mathematically correct, scientifically insightful and productive, and thus we elected to fight against this aversion, rather than to abandon the project. This uneasiness would present itself to a similar intensity regardless of which definition I now choose to present first, and so we might as well pick the simplest one. For instance, let us take the relatively simple definition of the scientific method, which will be:

Definition 10 (Scientific method)**.** *An algorithm which recursively enumerates knowledge, or a subset thereof, is called a scientific method.*

Mathematically speaking, this is a very simple definition. We have previously defined knowledge as halting programs (this made it comprehensible) and it's domain as that of a universal Turing-function (this made it comprehensive). Now we simply give a name, the scientific method, to any algorithm which recursively enumerates its domain, or part thereof. The notion of the scientific method, a previously informal construction, is now imported into pure mathematics and as such we presume to have produced a net gain for science, compared to not having it.

The features of the scientific method are found implicitly in the definition. Indeed, implicit in said definition lies a requirement for the algorithm to verify the input to be knowledge by running its corresponding program to completion, and reporting success once proven to halt. That it may or may not halt is the hypothesis, and the execution of the function is the 'experiment' which verifies the hypothesis. If an input runs for an abnormally long time, one may try a different hypothesis hoping to reach the conclusion differently. Since knowledge is element-wise infallible, each terminating experiments are formally reproducible as many times as one needs to, to be satisfied of its validity. All of the tenets of the scientific method are implicit in the definition, and its domain is that of knowledge itself, just as we would expect from the scientific method. Finally, the domain of knowledge is arbitrarily complex and countably infinite, therefore we never run out of new knowledge allowing for a perpetual and never ending application of the scientific method. Mathematically, it is a remarkably simple definition for such an otherwise rich concept.

But outside of mathematical land, the tone gets a bit more grim. Some readers may need a few more definitions before they start feeling the full weight induced by a *total commitment to formalization* on their worldview, but for many this definition will mark that point. Let us give a few comments to illustrate the type and intensity of the aversions that can plausibly be experienced:

1. Those who previously believed, or even nurtured the hope that, reality admitted elements of knowledge that are outside the scientific method must now find a flaw, or correct their worldview. As scientific as most people claim to be, a surprisingly large group seem to have an aversion to this. The unbiased response is, rather, to appreciate that what they thought was knowledge was in fact fallible (and thus simply a guess), whereas the scientific method does not output guesses, it outputs knowledge (which is infallible).

2. Those who nurture a worldview which is not "reducible" to our definition of knowledge in terms of halting program, must now argue that our definition contains gaps of knowledge, lest they have to correct their worldview. But our definition is simply the unique logical construction of knowledge with is both comprehensible and comprehensive. Thus, as comprehensiveness implies no gaps, their worldview is revealed to necessarily contain at least some elements that are incurably ambiguous, or *it would be* reducible to our definition.

3. The elimination of all naive concepts or notions (no more "magic" or "handwaving") is now required. If one has a worldview that relies upon a plurality of non-formalizable ambiguities, then one's worldview will not survive this formalization. For many, this is interpreted as killing the "fun" or the "imagination" from reality. It is unlikely that anyone's pre-existing worldview survives without some changes to accommodate total formalization.

Does one even stand a chance at maintaining an informal worldview, when facing such definitions? Many of our base definitions were carefully chosen to merely *match* and *rebrand* pre-existing and well respected mathematical definitions, and this was a strategic choice to make it incredibly unlikely to find fatal flaws. In our experience the battery of aversion we typically receive boils down to an equivalent formulation of "I can't find a specific error, but it must be wrong because [my worldview] requires [certain informal physical or metaphysical language], and here there is no support for that". The other possibility, however, is that one could be simply wrong in assuming that the world needs such informals to be defined. Furthermore, a fatal flaw has so far not been identified otherwise we would either correct the source of the error if possible, or immediately abandon the project altogether depending on the nature of the error presented, and would clearly state so to avoid wasting anyone's time.

Consider the alternative for a moment and let us try to be a crowd pleaser. How could we leave room for ambiguities so that people to not feel constrained by formalism, while remaining mathematically precise? Should we define the scientific method as a function that recursively enumerates 95% of knowledge, leaving a sympathetic 5% out for love, beauty and poetry? How would we possibly justify this mathematically. Functions which recursively enumerate one hundred percent of the domain do exists; should we just lie to ourselves and pretend they don't? Of course, we cannot. Whether a painting is or isn't beautiful, if not the result of an instantiation of infallible knowledge, is merely a guess. The scientific method does not output guesses, it outputs knowledge.

Now, there is a way to discuss, for instance, beauty scientifically: if one actually works out a precise definition of beauty, such as:

```
fn is_beautiful(painting: Object) -> bool{
   if (painting.colors.count()>=3){
    return true;
   }
   return false;
}
```

Then congratulations, one now has a definition of beauty that is actually comprehensible for the scientific method! The function returns true if the painting has 3 or more colours, otherwise it returns false. The scientific method can now use this definition to output all objects which are "beautiful" according to *said definition.*

Good luck getting everyone to agree to accept *this* definition as the be-all-end-all of beauty. However, all hope is not lost: the set of all halting programs includes the totality of all possible comprehensible definitions of beauty and therefore if a 'good-one' does exists then by necessity of having them all it must be in there, otherwise it simply means the concept is fundamentally non-comprehensible. Picking the 'good-one' from the set of all comprehensible definitions of beauty could merely be a social convention based on what everyone concept of beauty coalesces to. Even under this more challenging description, which references a social convention, comprehensible definitions are still found in the purview of the scientific method, as one can use a function such as this:

```
fn is_beautiful(painting: Object, people: Vec<Person>) -> bool{
  for person in people{
   if person.is_beautiful(painting)==true{
    return true;
   }
  }
  return false;
}
```

This function returns true if at least one person thinks it's beautiful. In this case, the scientific method 'polls' every 'person' in 'people' and asks if the painting is beautiful, and as soon as one says yes, then it returns true, otherwise it returns false at the end of the loop. In this case the definition of beauty is comprehensible provided that each 'person' in 'people' also produced a comprehensible implementation of the function **is_beautiful**. The scientific method a-priori has no preference for which definition we end up agreeing (or disagreeing) upon, it simply verifies that which can be verified comprehensibly.

The scientific method's sole purpose is to convert comprehensible questions or definitions into knowledge.

Let us return to our discussion on aversion. At the other end of the aversion spectrum, we find some readers (it would be overly optimistic to expect it from all readers, but hopefully some) that accept and understand that the proposed system induces what amounts to a checkmate position for informal (naive) worldview. Of those readers, most will then condition themselves to accept a re-adjustment of their worldview such that it becomes conductive to complete formalization. For these readers, their desire for formalization is greater than their attachment to an informal worldview, and they are willing to make the necessary sacrifices to work completely formally.

Let us now reprise our more neutral tonality to introduce and complete the formal system of science. Although the "magic" is now gone, we hope that the reader can find the will to smile again by immersing himself or herself in the cheerful world of formal terminating protocols, in lieu of said "magic".

3.1 Terminating Protocols as Knowledge about Nature

Both *Oxford Languages* and the *Collins dictionary* defines a protocol as

> [Protocol]: A procedure for carrying out a scientific experiment

Comparatively, Wikipedia, interestingly more insightful in this case, describes it as follows:

> [Protocol]: In natural and social science research, a protocol is most commonly a predefined procedural method in the design and implementation of an experiment. Protocols are written whenever it is desirable to standardize a laboratory method to ensure successful replication of results by others in the same laboratory or by other laboratories. Additionally, and by extension, protocols have the advantage of facilitating the assessment of experimental results through peer review.

The above description precisely hits all the right cords, making it especially delightful as an introduction of the concept. We will now make the case for a new description of nature, or natural processes, which is conductive to complete formalization. Of course, as we did for knowledge, we will require this description of nature to also be comprehensible and comprehensive in the same mathematical sense.

The proposed description will essentially require that one describes nature via the set of all protocols known to have terminated thus far. This type of description has a similar connotation to our previous formulation of mathematics in terms of halting programs. In fact, the tools introduced for the former will also be usable for the later. The proposed description is further similar to a requirement well-known to peer-review, and should be already familiar to most readers. In the peer-reviewed literature, the typical requirement regarding the reproducibility of a protocol is that an expert of the field be able to reproduce the experiment, and this is of course a much lower standard than formal reproducibility which is a mathematically precise definition, but nonetheless serves as a good entry-level example.

Hinkelmann, Klaus and Kempthorne, Oscar in 'Design and Analysis of Experiments, Introduction to Experimental Design'[11] note the following:

> If two observers appear to be following the same protocol of measurement and they get different results, then we conclude that the specification of the protocol of measurement is incomplete and is susceptible to different implementation by different observers. [...] If a protocol of measurement cannot be specified so that two trained observers cannot obtain essentially the same observation by following the written protocol of measurement, then the measurement process is not well-defined.

In practice it is tolerated to reference undefined, even informal, physical language, as long as 'experts in the field' understand each other. For instance, one can say "take a photon-beam emitter" or one can reference an "electric wire", etc, without having to provide a formal definition of either of these concepts. Those definitions of physical objects ultimately tie to a specific product ID, as made by a specific manufacturer, and said ID is often required to be mentioned in the research report explicitly. For the electric wire, a commonly used product, it is perhaps sufficient that the local hardware store sells them, and for more complex products, such as a specific laser or protein solutions, an exact ID from the manufacturer will likely be required for the paper to pass peer-review. If we attempted to explain to, say, an alien from another universe what an electric wire is, we would struggle unless our neighbourhood chain of hardware stores also as a local office in its universe for it to buy the same type of wire. In computer language terms, we would say we pass the concept of the electric wire to another expert by reference.

Appeal to the concept of 'expert' is a way for us to introduce and to tolerate informality into a protocol without loosing face; as that which is understood by 'experts' does not need to be specified. In a formal system of science we will require a much higher standard of protocol repeatability than merely being communicable to a fellow expert. We aim for mathematically precise definitions. For a protocol to be completely well-defined, the protocol must specify all steps of the experiment including the complete inner workings of any instrumentation used for the experiment. The protocol must be described as an effective method equivalent to an abstract computer program.

Let us now produce a thought experiment to help us understand how this will be done.

3.2 The Universal Experimenter

Suppose that an industrialist, perhaps unsatisfied with the abysmal record of irreproducible publications in the experimental sciences (i.e. replication crisis), or for other motivations, were to construct what we would call a *universal experimenter*; that is, a machine able to execute in nature the steps specified by any experimental protocol.

A universal experimenter shares features with the universal constructor of Von-Neumann, as well as some hint of constructor theory concepts, but will be utilized from a different stand-point, making it particularly helpful as a tool to formalize the practice of science and to investigate its scope and limitations self-reflectively. Von-Neumann was particularly interested in the self-replicating features of such a construction, but self-replication will here not be our primary focus of interest. Rather, the knowledge producible by such a machine will be our focus.

The Universal Constructor of Von-Neumann is a machine that is able to construct any physical item that can be constructed, including copies of itself. Whereas, a Universal Experimenter is a machine that can execute any scientific protocol, and thus perform any scientific experiment. Of course, both machines are subject to the halting problem, and thus a non-terminating protocols (or an attempt to construct the non-constructible in the case of the Universal Constructor) will cause the machine to run forever.

Both the machine and the constructor can be seen as the equivalent of each other. Indeed, it is the case that a Universal Constructor is also a Universal Experimenter (as said constructor can build a laboratory in which an arbitrary protocol is executed), and a Universal Experimenter is also a Universal Constructor (as a protocol could call for the construction of a Universal Constructor, or even for a copy of itself, to experiment on).

Specifically, a Universal Experimenter produces a result if the protocol it is instructed to follows terminates. A realization of such a machine would comprise possibly wheels or legs for movement, robotic arms and fingers for object manipulation, a vision system and other robotic appendages suitable for both microscopic and macroscopic manipulation. It must have memory in sufficient quantity to hold a copy of the protocol and a computing unit able to work out the steps and direct the appendages so that the protocol is realized in nature. It must be able to construct a computer, or more abstractly a Turing machine, and run computer simulation or other numerical calculation as may be specified by the protocol. The machine can thus conduct computer simulations as well as physical experiments. Finally, the machine must have the means to print out, or otherwise communicate electronically, the result (if any) of the experiment. Such result may be in the form of a numerical output, a series of measurements or even binary data representing pictures where appropriate.

Toy models are easily able to implement an universal experimenter; for instance Von Neumann, to define an implementation universal constructor, created a 2-D grid 'universe', allocated a state to each element of the grid, then defined various simple rules of state-transformations, and showed that said rule applied on said grid allowed for various initial grid setups in which a constructor creates copies of itself. Popular games, such as Conway's Game of Life are able to support self-replication and even the implementation of a universal Turing machine, and thus would admit specific implementations of a universal experimenter. In real life, the human body (along with its brain) is the closest machine I can think of that could act as a general experimenter.

How would a theoretical physicist work with such a machine?

To put the machine to good use, a theoretical physicist must first write a protocol as a series of steps the machine can understand. For instance, the machine can include **move** instructions, using it to move its appendages in certain ways as well as a **capture** instruction to take snapshots of its environment, etc. In any case, the physicist will produce a sequence of instructions for the machine to execute. The physicist would also specify an initial setup, known as the *preparation*, such that the protocol is applied to a well-defined initial condition. The initial condition is specified in the list of instructions, as such it is created by the machine making the full experiment completely reproducible. Finally, the physicist would then upload the protocol to the machine, and wait for the output to be produced.

The mathematical definition of the protocol is as follows:

Definition 11 (Protocol)**.** *A protocol is defined as a partial computable function:*

$$\begin{array}{rcrcl} \text{prot} & : & \mathbb{W} & \longrightarrow & \mathbb{W} \\ & & \text{prep} & \longmapsto & r \end{array} \tag{23}$$

- *The domain of the protocol* Dom(prot) *includes the set of all preparations which terminates for it.*

Let us now define the universal experimenter. A universal experimenter is able to construct any preparation and execute any protocol on it. If a protocol does not terminate, then the universal experimenter will run forever, hence it is subject to the non-halting problem.

Definition 12 (Universal Experimenter)**.** *Let* $\langle\text{prot}\rangle$ *be the description of a protocol* prot *interpreted into the language of a universal experimenter* UE*, and* prep*, the preparation, both be sentences of a* $\mathbb{W}$*, called the instructions. Then a universal experimenter is defined as:*

$$\text{UE}(\langle\text{prot}\rangle, \text{prep}) \simeq \text{prot}(\text{prep}) \tag{24}$$

for all protocols and all preparations.

Definition 13 (Experiment)**.** *Let* $\mathbb{PROT}$ *be the set of all protocols, and let* $\mathbb{W}$ *be the set all preparations. An experiment* p *is a pair* $\mathbb{PROT} \times \mathbb{W}$*:*

$$p := (\text{prot}, \text{prep}) \tag{25}$$

that terminates, such that $\text{prot}(\text{prep}) = r$*.*

Definition 14 (Domain of Science). *We note $\mathbb{D}$ as the domain of science. The domain of science is the set of all experiments.*

Definition 15 (Experimental Contribution (to Knowledge)). *An experimental contribution to knowledge is a tuple of n elements of $\mathbb{D}$:*

$$\mathbf{m} := \mathbb{D}^n \tag{26}$$

- *An experimental contribution to knowledge only contains protocol-preparation pairs that have terminated.*
- *An experimental contribution to knowledge corresponds, intuitively, to a sequence of related or unrelated experiments, that have been verified by a universal experimenter.*
- *An experimental contribution to knowledge corresponds to an instance of natural knowledge (knowledge about nature).*
- *Finally, as the set of knowledge is comprehensive, then all systems which admits knowledge, physical or otherwise, can be represented in the form of a specific experimental contribution associated to a specific experimenter, and said contribution constitutes a complete representation of the knowledge the machine has produced thus far for its operator.*

For a universal experimenter to execute a protocol, both the protocol and its preparation must be described without ambiguity. Physical language such as a camera cannot be referenced informally in the specifications of the protocol, otherwise the universal experimenter cannot construct it. If the protocol calls for the usage of a camera, then the behaviour of the camera must also be specified without ambiguity in formal terms within the instructions. Consequently, all rules and/or physical laws which are required to be known, including any initial conditions, must be precisely provided in the description, so that the universal experimenter can construct the experiment. For some highly convoluted experiments, such as : "is this a good recipe for apple pie?"... the aphorism from Carl Sagan "If you wish to make an apple pie from scratch, you must first invent the universe" is adopted quite literally by the universal experimenter. The universal experimenter must create (or at least simulate) the universe, let interstellar matter accretes into stars, let biological evolution run its course, then finally conduct the experiment once the required actors are in play by feeding them apple pie. For a universal experimenter, certain protocols, due to their requirement for arbitrary complex contexts or general protocol complexity, cannot be created more efficiently than from literal scratch and by going through the full sequence of events until the end of the experiment.

3.3 Classification of Scientific Theories

Definition 16 (Scientific Theory). *Let* $\mathbf{m}$ *be an experimental contribution by* UE*, and let* ST *be a formal axiomatic system. If*

$$\mathrm{proj}_2(\mathbf{m}) \cap \mathrm{Dom}(\mathrm{ST}) \neq \varnothing \tag{27}$$

then ST *is a scientific theory of* $\mathbf{m}$*.*

Definition 17 (Empirical Theory)**.** *Let* $\mathbf{m}$ *be an experimental contribution by* UE *and let* ST *be a scientific theory. If*

$$\mathrm{proj}_2(\mathbf{m}) = \mathrm{Dom}(\mathrm{ST}) \tag{28}$$

then ST *is an empirical theory of* $\mathbf{m}$*.*

Definition 18 (Scientific Field)**.** *Let* $\mathbf{m}$ *be an experimental contribution by* UE *and let* ST *be a scientific theory. If*

$$\mathrm{Dom}(\mathrm{ST}) \subset \mathrm{proj}_2(\mathbf{m}) \tag{29}$$

then ST *is a scientific field of* $\mathbf{m}$*.*

Definition 19 (Predictive Theory)**.** *Let* $\mathbf{m}$ *be an experimental contribution by* UE *and let* ST *be a scientific theory. If*

$$\mathrm{proj}_2(\mathbf{m}) \subset \mathrm{Dom}(\mathrm{ST}) \tag{30}$$

then ST *is a predictive theory of* $\mathbf{m}$*.*
Specifically, the predictions of ST *are given as follows:*

$$\mathbb{S} := \mathrm{Dom}(\mathrm{ST}) \setminus \mathrm{proj}_2(\mathbf{m}) \tag{31}$$

Scientific theories that are predictive theories are supported by experiments, but may diverge outside of this support.

3.4 The Fundamental Theorem of Science

With these definitions we can prove, from first principle, that the possibility of falsification is a necessary consequence of the scientific method.

Theorem 3 (The Fundamental Theorem of Science)**.** *Let* $\mathbf{m}_1$ *and* $\mathbf{m}_2$ *be two experimental contributions to knowledge, such that the premises of the former are a subset of the later:* $\mathrm{proj}_2(\mathbf{m}_1) \subset \mathrm{proj}_2(\mathbf{m}_2)$*. If* ET_2 *is an empirical theory of* $\mathbf{m}_2$*, then it follows that* ET_2 *is a predictive theory of* $\mathbf{m}_1$*. Finally, up to episto-morphism,* $\mathrm{Dom}(\mathrm{ET}_2)$ *has measure* 0 *over the set of all distinct domains spawned by the predictive theories of* $\mathbf{m}_2$*.*

Proof. $\mathrm{Dom}(\mathrm{ET}_2)$ is unique. Yet, the number of distinct domains spawned by the set of all possible predictive theories of $\mathbf{m}_1$ is infinite. Finally, the measure of one element of an infinite set is 0. □

Consequently, the fundamental theorem of science leads to the concept of falsification, as commonly understood in the philosophy of science and as given in the sense of Popper. It is almost certain (measure of 1) that a predictive scientific theory will eventually be falsified.

4 A Formal Theory of the Observer

Biology has the organism, microbiology the cell and chemistry the molecule, but what about physics, what is its fundamental object of study? Is it the planets ($\sim$16th century), is it mechanics (17th century), is it thermodynamics (18th century), is it electromagnetism (19th century), is it quantum mechanics and special relativity (early 20th century) or is it general relativity, quantum field theory, the standard model and cosmology (20th century). Is it broadly what we haven't figured out about nature yet? Or is it permissively anything physicists do?

In our model of physics, it will be the observer, and specifically its participation in nature, that will be the fundamental object of study.

Let us first attempt to fix the intuition by taking the example of a generic theory of the electron. To understand the electron, one must experiment on the electron. For instance, in a lab, one could power electricity into a wire, undertake spin measurements, perform double-slits experiments or magnetism experiments, etc. All of these experiments build up the knowledge of the electron's behaviour and properties. Eventually with enough accumulated knowledge, one can formulate a theory of the electron, which describes its behaviour and properties. The theory of the electron is considered a physical theory by association, because it applies to the electron, which by definition is a physical particle.

We now invite the reader to think of our theory of the observer along the same lines, except we replace the word 'electron' with the word 'observer'. Instead of experimenting on the electron, we experiment on the observer. Instead of a few targeted experiment in the lab, we target all possible experiments this observer could do in nature. Instead of recovering, say, the Schrödinger equation which governs the behaviour of the electron, we get a comprehensive theory of fundamental physics which governs the comprehensive participation of the observer in nature.

But where the electron only knows a few tricks, the scope of possible observer participation is a coalescence of three mathematically related but philosophically distinct concepts: the universal Turing machine, the universal constructor and the universal experimenter, and thus is able to account for all construction and verification rules whether physical, simulated or mathematical and over all possible knowledge-bearing states of any possible systems.

4.1 A Theory of Reality

We have so far reformulated mathematics in terms of an incremental contribution to knowledge, and this automatically produced an epistemological model of the practice of science. But this is not yet the end of the story as the reformulation is able to take it one step further.

For instance, let us consider a universal experimenter and recall that it is able to produce any valid incremental contribution from the domain of science. To allow investigation of whole of domain, the model of investigation of said experimenter must of course be Turing-complete, otherwise there will be gaps in knowledge that cannot be investigated by it. Now, let us consider the collection of all possible experimental contributions. Let us further consider that this collection forms a "space" which we call the "investigable space". It follows that, whatever the experimenter does or doesn't, it will be confined to remain in this space because producing experimental contributions is all the experimenter can do. This "confinement" bounds and delimits said investigative model; and whatever rules may be required to enforce or support the confinement would then automatically be perceived as inviolable laws by the experimenter.

For these inviolable laws to be the familiar laws of physics, we found that the model of investigation must support a *probability measure* over the possible incremental contributions, rather than merely be a deterministic implementation of a particular scientific method algorithm. In this case, we may qualify said ability as the experimenter having the "freedom" to investigate the space. This property is reminiscent of (and ultimately connects to) the Bell inequality experimenter freedom in familiar quantum mechanics; and consequently, a probability space of experiments along with a message of incremental contributions produced under the assumption of experimenter freedom will be used to define the observer, and as we will see this corresponds to what we understand an observer to be in physics.

4.2 Nature

The "investigable space" of experimental contributions the observer is confined to operate in will be called nature, and it will be defined as follows:

Definition 20 (Nature)**.** *Let* $\mathbf{m}$ *be an experimental contribution comprised of* n *terminating protocols, and let* $\mathbb{M}_{\mathbf{m}} = \bigcup_{i=1}^{n} \mathrm{proj}_i(\mathbf{m})$ *be the set comprised of the elements of* $\mathbf{m}$*. Then, Nature (*$\mathcal{N}_{\mathbf{m}}$*) is the "powertuple" of* $\mathbf{m}$*:*

$$\mathcal{N}_{\mathbf{m}} := \bigcup_{i=0}^{n} (\mathbb{M}_{\mathbf{m}})^i \tag{32}$$

- *Conceptually, a powertuple is similar to a powerset where the notion of the set is replaced by that of the tuple.*
- *Put simply, nature (*$\mathcal{N}_{\mathbf{m}}$*) is the set of all possible experimental contributions (including the empty experimental contribution) that can be built from* $\mathbf{m}$*.*
- *All elements of nature are experimental contributions, and all "sub-tuples" of an experimental contributions are elements of nature.*

4.3 Definition of the Observer

The departure here from typical practice and intuition is exceptional; let us note that the observer in modern theoretical physics is considered by many to be the last element of quantum physics that is not yet mathematically integrated into the formalism. Whereas here, is it the *only* (physical) axiom that we define, and is sufficient by itself to entail fundamental physics.

Axiom 1 (Observer)**.** *An observer, denoted as $\mathcal{O}$, is the probability space over all experiments in nature:*

$$\mathcal{O} := (\mathbf{m}, \mathcal{N}_{\mathbf{m}}, \rho : \mathcal{N}_{\mathbf{m}} \to [0,1]) \tag{33}$$

where ρ is a probability measure, $\mathbf{m}$ is an experimental contribution, and $\mathcal{N}_{\mathbf{m}}$, nature, is the space of all experimental contributions over $\mathbf{m}$, and where the measure sums to one.

We note that, unlike traditional measure theory in mathematics, here our definition of the measure is over tuples rather than sets. A prescription to tackle such a measure will be given in the main result section.

Just like we did earlier with a minimalistic definition of the scientific method as a recursive enumeration of the domain of science, and then showed that the richness of the concept was implicit in the relatively simple definition, here a similar richness will be recovered for fundamental physics as a consequence of this definition.

To obtain the laws of physics, in an exact formulation, we have found the second and final step to be to maximize the information associated with the result of the measure referenced in axiom 1, and this will be achieved by maximizing the entropy using the familiar tools of statistical mechanics. Fundamental physics will consequently be a specialization of the definition of the observer, in the sense that an observer is a probability space, and the laws of physics will be its entropy-maximized version. Maximizing the entropy of a measure over a power-tuple rather than a power-set requires a technical prescription which is given in the main result section. As for the context, we will not think of the entropy in terms of the typical notion common in introductory physics as a 'measure of disorder', rather we will think of it as a quantification of information in the sense of Claude Shannon. In this context, the information acquired by the observer following a measure adopts the role of a message that fixes the newly acquired knowledge into a new state describing the knowledge of the observer.

Furthermore, as the probability space represents the mathematical definition of experimental freedom in nature, then maximizing its entropy is further interpreted as minimizing the constraints on experimenter freedom.

Thesis 1 (Fundamental Physics)**.** *Physics is the product of an optimization problem on experimenter freedom. Specifically, the probability measure that maximizes experimenter freedom in nature constitutes the fundamental physics, or simply 'the laws of physics'.*

We note that our definition of fundamental physics will entail a probability measure. It is not given as a pre-formulated law such as $F = ma$. That is not to say that laws do not come into play; but when they do they are derived from this measure, and not brutely postulated. Taking an example of statistical mechanics, the ideal gas law $PV = nRT$ can be derived from the Gibbs measure as an equation of state under the appropriate energy and volume constraint on entropy. In the present case, the derived laws become the logical equivalent of a statement on what the observer can or cannot do while remaining consistent with its own definition and state.

Let us also clarify that axiomatic information does not represent knowledge itself, rather it encodes the *state* of knowledge of the observer. We distinguish knowledge (which is infallible once known) from information (which encodes the result of the measure).

5 Intermission

5.1 Science

To introduce falsification within a formal system of science, the notion of knowledge being infallible is critical. It is the reason why we can be certain that acquiring new knowledge does in fact necessarily falsify any conflicting models. If our knowledge was uncertain, we would simply be perpetually juggling the probabilistic weights of various hypotheses and models, and no model could ever be falsified. With this in mind, let us correct a terminology error made by Karl Popper. A core tenet of Karl Popper's philosophy is that scientific *knowledge* is always transitory, and so a scientific theory would be subject to falsification. The correction is minor, but nonetheless leads to substantial clarifications. The correction is on the usage of the term knowledge; knowledge is not transitory rather it is the models that are. Models are entailed by knowledge, as such they do not entail it in return. In fact, when acquiring new knowledge, if the model conflicts with it, then the model always loses the tug of war because the former is infallible while the later isn't. The correct terminology is that scientific *models* (not knowledge) are transitory because knowledge (which isn't transitory) takes precedence over the conflicting model.

Karl Popper's extended philosophy is correct in regards to scientific theories (e.g. biology, economic science, psychology, etc.), but physics as it would is a different beast altogether. The difference between the two stands out when we investigate their relationship to our newly formalized observer. For instance, if an observer "violates" a scientific theory, then said theory is simply falsified. This happens every once in a while, and other than perhaps a bruised ego, not much harm is done. Whereas, if an observer were to violate the laws of physics, presumably all hell would break loose. Why the difference? Of course, without a formal system of science, we have historically constructed our laws of physics the same way as any other scientific theory assuming they are of the same category, and thus the difference was unnoticed, but with a formal system of science we can pinpoint the difference. A scientific theory involves a choice of formal axiomatic representation of an experimental contribution, and it is this choice that is falsified when facing conflicting knowledge. Whereas, the observer cannot violate the fundamental physics without ceasing to be the probability space over all experiments in nature, and thus violating its own definition. There is a "self-referential component" from the observer's participation onto physics, which is absent from mere scientific theories, that makes the fundamental physics inviolable to the observer, whereas the scientific theory is only falsifiable to the observer hence not inviolable.

In mathematics we typically welcome newfound clear-cut delimitations between previously overlapping concepts. For instance, chemistry overlaps with physics significantly, and so does biology via bio-physics. What is the exact split, if any, or is everything ultimately physics? With our system, we now know the difference; scientific theories are entailed by knowledge, whereas fundamental physics is entailed by the definition of the observer as the maximally permissive model of its participation. A word of caution however; in practice one could always demand that we subject Axiom 1, and its predictions, to the falsification process, and thus physics, via the (technical) definition of the observer remains falsifiable and predictive despite being of a different class. Thus and although they may appear as the same due to their many similarities, physics is the unique member of a special class of falsifiable theories.

This difference carries over with respect to the techniques used to falsify physics. Physics, although falsifiable as we just said, is not subject to the fundamental theorem of science which applies only to formal axiomatic representations of experimental contributions and is responsible for a common scientific theory being falsifiable. For physics, a special falsification theorem must be created, and such must start with the definition of the observer rather than with the elements of the experimental contribution. The resulting falsification theorem will be more challenging than the first, simply because the observer is a probability space and this is a more challenging mathematical object to work with than a mere enumeration. To falsify a common empirical theory via the fundamental theorem of science, it suffices to identify a halting program within the experimental contribution to knowledge that is not entailed by it. For instance, J.B.S. Haldane one of the founders of evolutionary biology reportedly stated that finding the fossils of a rabbit in the Precambrian would falsify the theory of evolution. This is a binary yes/no type of falsification. Whereas, since physics is entailed by a probability space, falsifying physics will involve the use of probabilities. Specifically, we will find that repeat experiments over multiple copies of identical preparations, such that a probability distribution can be extracted from a plurality of similar measurements, will be required to test or falsify physics by comparing it to the predicted expectation values.

5.2 The Observer

The reader will notice that Axiom 1 does not reference a plurality of observers, rather it postulates what amounts to a *singleton observer*. The system is intended to be formulated from the perspective of the observer. This should be less surprising than it tends to be as it avoids a battery of observer-related paradoxes, and captures the philosophically safest possible foundation, but ouch if the intent is misunderstood. Let us explain the term, and then we will discuss its motivation and attempt to address the concerns. The term singleton is imported from software engineering, where the singleton pattern refers to a design pattern of object-oriented programming in which a class can only be instantiated once. Singleton does not mean that the program itself can only be ran once, it only means that each running copy has only one instantiation of its singleton variables within its memory. Our system supports the idea of "running" multiple times in parallel, thus admitting multiple observers —or more precisely formulated; it allows other observers to claim singleton status from their perspective—, and the singleton observer axiom is not designed to prevent that; it simply means that for each execution, the theory is formulated from the perspective of its singleton observer. The singleton observer is "I" from my perspective, and "you" from your perspective, and "him" from his perspective, etc. To be explicit, it is not a universal observer neither it is God — just you, him, her or I. The singleton observer is a mathematical description of who "I" am that also conveniently formalizes the set of tools "I" have access to in order to understand or participate in nature.

First, let us explain exactly how the theory is intended to support other observers from the perspective of the singleton observer. Their existence will be evidence-derived rather than postulated. Other observers, if they exist, can and will be derived by the singleton observer the same way any other facts are investigated, by merely inspecting the experimental contributions and weighting the evidence for them, and thus do not need to be postulated. Do we also need to postulate rocks, trees, or bees — or can we accept that their existence will be derived conditional upon the scientific evidence, and if so why not demand the same in regards to evidence for other observers? Indeed, psychologically and developmental-wise, this is what happens naturally as an infant matures and over time develop a theory of the mind to assess the motivations and the decision-making strategies of others — i.e. Infant solipsism (Piaget). Evidence for other observers is identified by inspection of all available evidence and builds over time and is the subject of the scientific method and to falsification. To include other observers via postulation would be to erase said developmental steps from the domain of the scientific method, or at least eliminate the necessity of a laborious but insightful derivation thereof by virtue of having reduced them to mere postulation, and would result in a representation of reality missing those parts.

Let us further point out that if we find it reasonable to expect that physics ought to work for any number of observers between 1 and infinity; then physics must also work with just one — and just like Peano's axiom posits only the first natural number and the others are derived by the successor function, here the singleton observer is the base axiom and others are to be entailed by the framework.

Secondly, one must remember the role of axioms. Axioms are the *logical minimal* required to derive a theory and are intended to be free of any redundancy. They are not a collection of desiratas, nor are they designed to make the world a better place than it is. Not only do we not need a plurality thereof to complete the theory, if we made the world conditional upon multiple axiomatic observers, the thesis becomes nonsensical:

- We would be claiming that at least two observers are needed to entail the laws of physics... can an observer, when working alone, violate the laws of physics, but can't if working as a team... ?

Just like quantum theory should work for one or any number of particles, the laws of physics should logically be definable against only one observer if need be, or any other number, because they limit what each observer can individually do or cannot do. Team work, although perhaps socially beneficial, does not in this case prevail against the laws of physics.

Let us now discuss the singleton observer within the context of a relevant physics experiment. The Wigner's friend experiment puts forward a paradox in which two observers appears to witness the collapse of a wave-function at different times. The Wigner's friend experiment supposes that an observer F measures a wave-function $|\psi\rangle = \alpha\,|\phi_1\rangle + \beta\,|\phi_2\rangle$ to be in state $|\phi_1\rangle$ or $|\phi_2\rangle$, with probability $|\alpha|^2$ and $|\beta|^2$ respectively, that F notes the result somewhere in his laboratory, but refrains from advising another observer W of the result. This other observer then understands the wave-function of the laboratory in which F performed his measurement to retain the superposition. Whether the system is or isn't in a superposition appears to be resolved at different times for each observer; F sees the collapse at the instant of measure, but W sees it only after F choose to share his notes with him. This difference in collapse timing is the paradox. A commonly proposed resolution is that superposition does not occur in macroscopic objects, and the reproduction of this experiment in a microscopic system would appear less paradoxical. In actuality however, as soon as observer F notes then hides the result, F begins to act as a glorified hidden variable theory with respect to W and this is ruled out by Bell's inequality; thus F cannot cause $|\psi\rangle$ to collapse at any time other than simultaneously for all observers. In his original paper Wigner focused on another possible resolution: "All this is quite satisfactory: the theory of measurement, direct or indirect, is logically consistent so long as I maintain my privileged position as ultimate observer". Historically, this has not been the preferred interpretation because of the obvious resistance to the connotation associated with a privileged ultimate observer. In our system, the observer is not assumed to be "privileged ultimate" but merely to be formulated as a singleton, and other observers can proclaim the same. Specifically, the wave-function will be formulated from the perspective of a singleton observer; and since other observers are derived by inspection of the evidence "gathered" by said wave-function, they will obviously be unable to support a different wave-function behaviour than that reported by the singleton. Furthermore, as we said, other observers can proclaim the same, and thus, symmetry oblige and as we found guaranteed by the laws of physics, all observers, singleton and evidence-derived, will observe the same wave-function behaviour as each other.

Finally, why did we present the singleton observer as an axiom, and not say, as a definition? An axiom implies one could claim it to be false, and technically speaking this is indeed possible. For instance, one simply has to state they do not believe they exist as an observer, and as we would only have their proclamation of such to go by since the singleton observer is postulated, the scientific method would be powerless to prove the claim wrong. The question "what if axiom 1 is false" is answered amicably with "then you are not an observer", and we move on. The other, slightly more challenging, reformulation of the question is "what if axiom 1 is *incorrect*" (in the sense that our probability space definition of the singleton observer is the wrong one to use, but the concepts we put forward might still otherwise be the correct overall approach). In this case we would simply get, proportionally to a how wrong our definition is, the wrong laws of physics, which is why we claimed earlier that physics is also subject to falsification. We do not exclude that, in principle, future experiments may confirm, or force us to adjust, the mathematical form of axiom 1 along with its accompanying entropy-maximization prescription. Consequently, our physical theory of the observer is in principle falsifiable.

5.3 Ontology of Quantum Physics

In the (Discussion — Science) section, we have stated that a scientific model is entailed by knowledge, but that it does not itself entail knowledge. Phrasing it like this however makes it sounds like models are vacuous... This is not the case as they do provide value, but we do not call this value knowledge, rather we call it <u>insight</u>. For instance, it is the case that natural selection is an insightful model of bio-diversity, but does it give us knowledge of bio-diversity? — or, it is knowledge of bio-diversity that entails natural selection? Consider these statements:

1. (The model of) natural selection entails knowledge of bio-diversity.
2. Knowledge of bio-diversity entails (the model of) natural selection.

In our system, the entailment is always as follows:

$$\text{Knowledge} \implies \text{Model} \tag{34}$$

Consequently, it is only the second statement that is correct.

Our framework demands that one takes this perspective for all scientific models. Let us take a more counter-intuitive example:

1. (The model of) gravity entails knowledge of objects falling.
2. Knowledge of objects falling entails (the model of) gravity.

Here we are dealing with that is traditionally considered a law of physics rather than a purely scientific theory, and thus some more easily confuse the model with reality. It is very common to encounter the reflex to say that it is the first that is true, and the second isn't because gravity obviously causes objects to fall. However, here as well, only the second one can be sustained. It is knowledge of objects falling that caused Newton to produce the model of gravity as a scientific theory. Gravity may appear as logically equivalent to the sum-total of all falling objects; but nature is not gravity, rather nature is the sum-total of all falling objects.

It is in regards to the interpretation of quantum physics that understanding and accepting the correct entailment pays the most unrealized gains. Now, consider the following statements; which one is true?

1. Measuring a wave-function $|\psi\rangle = \alpha |\phi_1\rangle + \beta |\phi_1\rangle$ caused it to collapse to $|\phi_1\rangle$ or to $|\phi_2\rangle$.
2. Registering 'clicks' such as $|\phi_1\rangle$ or $|\phi_2\rangle$ on an incidence counter causes us to derive $|\psi\rangle = \alpha |\phi_1\rangle + \beta |\psi_1\rangle$ as a statistical model of the clicks.

As before, our system demands that the second be the correct entailment. This will in fact constitute the formulation of quantum mechanics within our system: 'clicks' are registered *then* the wave-function is derived. Let us now investigate in the following section the consequences of this formulation.

5.4 Formulation of Quantum Physics

In our main result we will see that maximizing the entropy of our definition of the observer produces (a generalization of) the wave-function along with the Born rule as the measure, and further automatically yields the definition of a quantum Turing machine as the model of computation. We elected to discuss the interpretation before the main result because our feedback was that it is too abstract without conceptual decorations. It is probably beneficial for the reader to read this section twice; once right now, and another time after reading the main result.

First let us review statistical mechanics which also maximizes its entropy to obtain its measure. In statistical mechanics, constraints on the entropy are associated to instruments acting on the system. For instance, an energy constraint on the entropy:

$$\overline{E} = \sum_{q \in \mathbb{Q}} \rho(q) E(q) \tag{35}$$

is interpreted physically as an energy-meter measuring the system and producing a series of energy measurement $E_1, E_2, \ldots$ converging to an average value $\overline{E}$.

Another common constraint is that of the volume:

$$\overline{V} = \sum_{q \in \mathbb{Q}} \rho(q) V(q) \tag{36}$$

associated to a volume-meter acting on the system and converging towards an average volume value $\overline{V}$, also by producing a sequence of measurements of the volume $V_1, V_2, \ldots$.

With these two constraints, the typical system of statistical mechanics is obtained by maximizing the entropy using its corresponding Lagrange equation, and the method of the Lagrange multipliers:

$$\mathcal{L} = -k_B \sum_{q\in\mathbb{Q}} \rho(q)\ln\rho(q) + \lambda\left(1 - \sum_{q\in\mathbb{Q}}\rho(q)\right) + \beta\left(\overline{E} - \sum_{q\in\mathbb{Q}}\rho(q)E(q)\right) + \gamma\left(\overline{V} - \sum_{q\in\mathbb{Q}}\rho(q)V(q)\right) \tag{37}$$

and then solving $\frac{\partial\mathcal{L}}{\partial\rho} = 0$ for ρ, we get the Gibbs measure:

$$\rho(q,\beta,p) = \frac{1}{Z}\exp\bigl(-\beta E(q) + \gamma V(q)\bigr) \tag{38}$$

$$= \frac{1}{Z}\exp\bigl(-\beta(E(q) + pV(q))\bigr) \tag{39}$$

We will now introduce a formulation of quantum mechanics produced by extending statistical mechanics with a larger class of instruments. The entropy will now be maximized under the constraint of *measurement-events* (or 'clicks') collected by *phase-invariant instruments* (or 'click' recorders), and this will yield the wave-function along with the Born rule automatically as the statistical model of the 'clicks'.

Instead of an energy-meter or a volume-meter, consider a phase-invariant instrument, such that the constraint it induces on the entropy is given as follows:

$$\operatorname{tr}\begin{pmatrix}\overline{a} & -\overline{b}\\ \overline{b} & \overline{a}\end{pmatrix} = \sum_{q\in\mathbb{Q}}\rho(q)\operatorname{tr}\begin{pmatrix}a(q) & -b(q)\\ b(q) & (a)\end{pmatrix} \tag{40}$$

where $\begin{pmatrix}a(q) & -b(q)\\ b(q) & a(q)\end{pmatrix} \cong a(q) + ib(q)$ is the matrix representation of the complex numbers. Using the trace follows the prescription of the main result in order to define a measure over a space of tuples instead of sets.

Here, the purpose of the trace is to enforce the phase-invariance of the instrument. The corresponding Lagrangian equation that maximizes the entropy will in this case be:

$$\mathcal{L} = -\sum_{q\in\mathbb{Q}}\rho(q)\ln(q) + \alpha\left(1 - \sum_{q\in\mathbb{Q}}\rho(q)\right) + \tau\left(\operatorname{tr}\begin{pmatrix}\overline{a} & -\overline{b}\\ \overline{b} & \overline{a}\end{pmatrix} - \sum_{q\in\mathbb{Q}}\rho(q)\operatorname{tr}\begin{pmatrix}a(q) & -b(q)\\ b(q) & a(q)\end{pmatrix}\right) \tag{41}$$

Maximizing the entropy under such constraints does produce the probability measure of the wave-function along with the Born rule. But here it is derived using the same technique as we would for any other system of statistical mechanics, and thus we inherit its tools, interpretation and will be able to account for the origin of the measure. Indeed, solving ρ for $\frac{\partial \mathcal{L}}{\partial \rho(q)} = 0$ gives:

$$\rho(q) = \frac{1}{Z} \exp \operatorname{tr} \tau \begin{pmatrix} a(q) & -b(q) \\ b(q) & a(q) \end{pmatrix} \tag{42}$$

$$= \frac{1}{Z} \det \exp \tau \begin{pmatrix} a(q) & -b(q) \\ b(q) & a(q) \end{pmatrix} \tag{43}$$

$$\cong \exp 2\tau a(q) |\exp i\tau b(q)|^2 \qquad \text{Born rule} \tag{44}$$

In this formulation, the interpretation of quantum mechanics will simply become that of an instrument producing a sequence of measurements on a system such that an average value is obtained, but instead of the simpler scalar instruments typically used in statistical mechanics, here we have a phase-invariant instrument; and maximizing its entropy yields the wave-function along with the Born rule. What is an example of such a detector; quite simply an incidence-counter or a single-photon detector would be one. Such an instrument produces a sequence of incidences ('clicks') as photons are detected and "advanced features" such as an interference pattern is a consequence of this phase-invariance.

Let us state that elements of this interpretation connect to the ensemble interpretation of quantum mechanics, and others appear very similar to what John A. Wheeler had in mind when he wrote "Information, Physics, Quantum; The Search for Links.". For instance, consider the following statement by him:

> It from bit symbolizes the idea that every item of the physical world has at bottom — at a very deep bottom, in most instances — an immaterial source and explanation; that what we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and this is a participatory universe.

and also;

> Three examples may illustrate the theme of it from bit. First, the photon. With polarizer over the distant source and analyzer of polarization over the photodetector under watch, we ask the yes or no question, "Did the counter register a click during the specified second?" If yes, we often say, "A photon did it." We know perfectly well that the photon existed neither before the emission nor after the detection. However, we also have to recognize that any talk of the photon "existing" during the intermediate period is only a blown-up version of the raw fact, a count

Raw fact; a count — in our scheme this is encoded in the form of axiomatic information and relate to the phase-invariant 'clicks' of instrumentation. The other part, also identified by Wheeler and necessary to complete the description, is equipment-evoke response — in our scheme rather than equipment we use the ambiguity-free notion of a terminating protocol along with its preparation but its role is similar. The 'knowledge' corresponds to the steps required to construct an experiment in which photons are sent according to a repeatable and well-defined preparation. When a 'click' is registered, it yields more than just a bit; it is also associated to a unit of 'knowledge' given in the form of a protocol-preparation pair, and associated to the unitary transformations comprising the protocol applied to a given preparation.

5.5 The Measurement Problem

In our system, we have crossed through the measurement problem but from the other side. Indeed, we started with a *singular* incremental contribution describing the participation of the observer in nature and then we have *derived* the wave-function (which supports a plurality, or superposition, of measurement results) by having used the Lagrange-multiplier equation to maximize the entropy of a measure constrained by 'click' recorders. This is the opposite direction of the standard formulation which applies a measurement to the wave-function to get the 'clicks':

$$\text{Standard:} \quad \text{Measurement}(\underbrace{\text{wave-function}}_{\text{Axiomatic}}) \implies \underbrace{\text{'click'}}_{\text{collapse problem?}} \tag{45}$$

$$\text{Ours:} \quad \text{Max-Entropy}(\underbrace{\text{'clicks'}}_{\text{Axiomatic}}) \implies \underbrace{\text{wave-function}}_{\text{derived}} \tag{46}$$

Let us investigate the distinction in more details.

All statistical systems, classical or quantum, admit two formulations; a pre-facto and a post-facto construction, or sometimes referred to as, a frequentist or a bayesian formulation. Our formulation of quantum mechanics from statistical mechanics is *entirely frequentist*, whereas the standard formulation mix and match elements of both the bayesian and the frequentist schools and specifically, the wave-function collapse in the standard formulation is treated using the bayesian school. Specifically, the standard formulation assumes a prior (the wave-function) and updates it with the result of each measurement as it occurs, but the bayesian update (the collapse) is not understood and constitute what is known as the wave-function collapse problem.

To better understand how the formulations differ and to illustrate the nature of problem, let us now consider the archetypical example of a fair coin toss with 50%/50% probability of landing on either head or tail, as the distinction is clearer in this simple model. We can produce each of the two formulations by picking different points in the history of the system as the origin; either before or after the coin lands:

1. If formulated before the coin lands:

   If we are to formulate the statistical model before the coin lands, we would first posit a probability measure 50% head /50% tail (bayesian prior), then we would find that sampling the system (throwing the coin) yields either head or tail whenever it lands (this is the bayesian update). In this case, the statistical model, if it were proclaimed to be the complete physical description of the system, would be insufficient to support, as a mere probability measure, a physical mechanism for the bayesian update. This is similar to how quantum mechanics is typically formulated.

2. If formulated after the coin lands:

   Now, consider the discovery of a coin on the floor with head facing up. To work with probabilities, we must first assume an appropriate history for this coin making the present result due to chance; and specifically the choice will be that the coin landed as a result of fair coin toss. Unlike the previous case, we did not witness this history. We merely assumed it following the discovery of the coin on the floor, and theoretically speaking, we could be wrong about our assumption. This is the post-facto formulation. Then, under the fair coin toss history assumption, we can construct a probability measure to account for the state the coin was discovered in. The probability measure assigns a 50/50 probability to each outcome: head or tail. We end up with the same statistical model as the previous case, but no sampling of the probability measure has occurred and the probability interpretation is readily understood to merely be a model of the history of the system. We no longer have to look for a physical mechanism to account for the bayesian update, as there are no such updates.

Likewise, in our system, the wave-function along with the Born rule are no longer fundamental, but derived like any other measures of statistical mechanics, and following the registration of 'clicks'. It is a post-facto formulation. The wave-function in our system is the optimized model used to understand reality; it is not a physical object that is acted on by an observer.

Another angle that may help understanding the value of the post-facto formulation is to say that it catches interpretational errors at compile time, rather than at runtime. For instance, it tells us explicitly and in no uncertain terms that the assumption of a random history is in fact an assumption. Consequently, it has to be posited (Axiom 1) before progress is allowed. Whereas in the pre-facto formulation, this hint is not available at compile time. In the pre-facto formulation, the wave-function is first postulated, then it is only when we try to apply the model to nature that we encounter a "runtime error"; and to redress its applicability to reality, such is corrected by postulating the Born rule and eventually the measurement collapse.

In our interpretation, we maximize the entropy of the clicks to get the wave-function, whereas in the standard interpretations we measure the wave-function to get the clicks. However, the measurement operation is problematic, and introduces the problem of the wave-function collapse; whereas the maximization of entropy is a non-problematic operation. The ontology of our interpretation matches that which actually occurs in nature: we first register the 'clicks', then we derive the measure to be a wave-function. Likewise, if we register no clicks, we assume no wave-function.

Some may have the impression that this may be too simple to be the solution, but we assure you it is not the case. This interpretation was not possible before our main result relating statistical mechanics (and phase-invariant instruments) to the wave-function as its origin was otherwise too obscure.

Finally, since we cannot have knowledge of the existence of a wave-function (the prior) before a click is registered (the update), then the update always predates the prior, and consequently the bayesian formulation cannot be appropriate. So, if clicks are not caused by measuring a wave-function, then what causes them, and why do they occur? Let us now investigate the model of reality further.

5.6 The Model of Reality

Our system admits the following correspondence with statistical mechanics:

	Statistical Mechanics	*Statistical System of 'clicks'*	
Entropy	Boltzmann	Shannon	(47)
Measure	Gibbs	Born Rule	(48)
Constraint	Energy meter	Phase-invariant instrument	(49)
Micro-state	Energy values	Sub-Contributions	(50)
Macro-state	Equation of state	Evolution of the wave-function	(51)
Experience	Ergodic	Message	(52)

Let us recall that $\mathcal{N}$ (nature) is the power tuple of $\mathbf{m}$ (the incremental contribution), and therefore comprises multiple elements. Let us now call the elements of $\mathcal{N}$ the possible sub-contributions. These elements will be the micro-states of our system. As for the macroscopic state, it will acquire the form of the laws of physics (via the entropy maximization prescription of the main result), and they will describe the evolution of the possible micro-states (as the evolution of the wave-function). Finally, the message connects to the macro-state via a specific sequence of registered 'clicks', and this is quantified by entropy.

In statistical mechanics it is often assumed that the system, say the molecules of air in a box, permute over the possible micro-states of the system (i.e. ergodic hypothesis); whereas in our system as we use the Shannon entropy, there is no such permutation — how the observer experiences the system is in the form of a message fixing the realized incremental contribution from the possible sub-contributions available in nature. And the information gained by the observer from this message is equal and opposite to that of the entropy of the system. Thus, our formulation contains no unmeasurable redundancies. There is no many-worlds; only a unique message of incremental contributions that the observer experiences as reality.

Here, it is the experience that is considered *real and foundational*, and it serves as the justification for the rest of the framework. In our system, the experience is defined under the assumption of an observer with experimenter freedom (axiom 1; the probability space over all experiments in nature) receiving or producing a message as an incremental contribution: the observer *could* have produced a different incremental contribution but it so happen to have produced *this* particular one, and the difference between what is and what could have been constitutes entropy. As for the rest of the table, it simply follows from (maximizing the quantity of information/entropy of) the experience.

Finally, new contributions by the observer resulting from its continued participation in nature, will automatically trigger new 'clicks' to be registered in the system so as to preserve the general form of the experience as a message and to do so consistently over the addition of those new contributions. The clicks are therefore directly produced by observer participation in nature and serves as evidence of experimenter freedom, and this is more fundamental than the wave-function which is derived afterwards as a statistical model entailed by the registered clicks.

The interpretation connects with elements of Conway and Kochen[2], where they describe "new bits of information coming into existence in the universe" as a result of experimenter freedom. They say:

> ...there will be a time t_0 after x, y, z are chosen with the property that for each time $t < t_0$ no such bit is available, but for every $t > t_0$ some such bit is available.
>
> But in this case the universe has taken a free decision at time t_0, because the information about it after t_0 is, by definition, not a function of the information available before t_0!

In our work, we have taken it a step further and revealed that the unit of information of experimenter freedom is not the bit but actually the 'click', and that unlike the bit, the 'click' contains enough complexity and flexibility (and as we found; even geometry) to encode the experience of the observer such that its assumed freedom of action is precisely delimited by the laws of physics.

Finally, as a disclaimer we state that the term freedom is to be interpreted similarly to the strict sense given by Conway and Kochen[2] as freedom from being determined by past history, thus allowing a probabilistic enumeration in the present. It is not meant to be interpreted in the metaphysical sense of free will in philosophy.

5.7 Generalization

In the main result we will actually obtain a generalization of quantum mechanics to general linear transformations. In this case the interpretation of quantum mechanics takes its simplest and most visualizable interpretation. The phase-invariant instruments are upgraded from a complex phase to a general linear phase. The probability will now be associated with a sequence of 'clicks' recorded in space-time as events. Thus, the framework describes reality as a sequence of space-time 'clicks' (or events) which, under entropy maximization, are associated to a general linear wave-function in lieu of the Gibbs ensemble. As we note, general relativity is primarily a theory of events in space-time, and the extension to quantum theory assigns a probability and an entropy to said events, such that the measure over said events is a wave-function able to support the transformations required by general relativity while preserving the invariance of the probability measure. This generalization yields a quantum theory of gravity whose equations of motions are exactly the Einstein field equations. Standard quantum field theory will also be shown to be a special case of the presented quantum theory.

5.8 Context

Finally, let us address a common objection:

> Surely there is more to reality than simply 'clicks' — what about objects such as chairs or kitchen tables, what about colours or literature?

Here we give a "first-approximation answer". It is not designed to be a complete answer, but merely to guide the intuition.

This is where context sneaks in. Each halting program contains the context to the claim it makes. An observer with a brain will be influenced by biological evolution and other environmental factors, and will pick a preferred context to clarify the set of halting program it knows (this could be colours instead of wavelength, or set theory instead of quantum Turing machine programs). The mind understands reality as a collection of scientific fields, which are choices of formal axiomatic bases, formulated in a matter it finds efficient or convenient; each representing a falsifiable choice of context. Why we think of a particular cloud of 'clicks' as a chair, or another cloud as a wood club which is a potential weapon, is simply the result of natural selection influencing these contextual choices and committing them to the biological wiring of the brain.

The laws of physics, however, are without context and this is required for them to be universal. The lack of context to the laws of physics relegates them to a "fraction" of all reality (the dynamics of 'clicks'), and the "rest" of reality is relegated to scientific theories which are contextual.

6 Main Result

Let us now use the definition of the observer (Axiom 1) to derive the fundamental physics.

Our starting point will be the definition of the observer. We will then maximize the entropy of ρ using the method of the Lagrange multipliers. We recall that our definition of the observer is:

$$\mathcal{O} := (\mathbf{m}, \mathcal{N}, \rho : \mathcal{N} \to [0,1]) \tag{53}$$

where $\mathbf{m}$ is a n-tuple, $\mathcal{N}$ is a "powertuple" and ρ is a (probability) measure over $\mathcal{N}$.

Note the similarity between our definition of the observer to that of a measure space in mathematics. Comparatively, the definition of a measure space is:

$$M := (X, \Sigma, \mu(X)) \tag{54}$$

where X is a set, Σ is (often) taken to be the powerset of X, and μ is a measure over Σ. The difference with our measure is simply that sets have been replaced by tuples. Consequently, we must adapt the standard definition of a measure space from set to tuples. To do so, we will use the following prescription:

1. We assign a non-negative number to each element of $\mathcal{N}$.
2. We equip said numbers with the addition operation, converting the construction to a vector space.
3. We maximize the entropy of a single element under the effect of constraints, by using the method of the Lagrange multipliers.
4. We prescribe that any and all constraints on said entropy must remain invariant with respect to a change of basis of said vector space.
5. We use the tensor product n-times over said vector space to construct a probability measure of n-tuples of halting programs.
6. We use the direct sum to complete the measure over the whole of tuple-space by combining the measures of different sizes as a single measure.

Explicitly, we maximize the entropy:

$$S = -\sum_{p \in \mathbf{m}} \rho(p) \ln \rho(p) \tag{55}$$

subject to these constraints:

$$\sum_{p \in \mathbf{m}} \rho(p) = 1 \tag{56}$$

$$\sum_{p \in \mathbf{m}} \rho(p) \operatorname{tr} \mathbf{M}(p) = \operatorname{tr} \overline{\mathbf{M}} \tag{57}$$

where the notation $\sum_{p\in\mathbf{m}}$ designates a sum over the elements of the experimental contribution $\mathbf{m}$, where $\mathbf{M}(p)$ are a matrix-valued maps from the elements of $\mathbf{m}$ to $\mathbb{C}^{n\times n}$ representing the linear transformations of the vector space and where $\overline{\mathbf{M}}$ is a element-by-element average matrix.

Usage of the trace of a matrix as a constraint imposes an invariance with respect to a similarity transformation, accounting for all possible linear reordering of the elements of the tuples of the sum, thus allowing the creation of a measure of a tuple or group of tuples from within a space of tuples, invariantly with respect to the order of the elements of the tuples.

Similarity transformation invariance on the trace is the result of this identity:

$$\operatorname{tr}\mathbf{M} = \operatorname{tr}\mathbf{B}\mathbf{M}\mathbf{B}^{-1} \tag{58}$$

We now use the Lagrange multiplier method to derive the expression for ρ that maximizes the entropy, subject to the above mentioned constraints. Maximizing the following equation with respect to ρ yields the answer:

$$\mathcal{L} = -k_B\sum_{p\in\mathbf{m}}\rho(p)\ln(p) + \alpha\left(1-\sum_{p\in\mathbf{m}}\rho(p)\right) + \tau\left(\operatorname{tr}\overline{\mathbf{M}} - \sum_{p\in\mathbf{m}}\rho(p)\operatorname{tr}\mathbf{M}(p)\right) \tag{59}$$

where α and τ are the Lagrange multipliers. The explicit derivation is made available in Annex B. Except for the presence of the trace and matrices, using the Lagrangian multiplier method on the entropy is standard and shown in most introductory textbooks of statistical mechanics to derive the Gibbs measure, where the quantities are simple scalars. With the trace and matrices, the result of the maximization process is:

$$\rho(p,\tau) = \frac{1}{Z(\tau)}\det\exp\bigl(-\tau\mathbf{M}(p)\bigr) \tag{60}$$

where

$$Z(\tau) = \sum_{p\in\mathbf{m}}\det\exp\bigl(-\tau\mathbf{M}(p)\bigr) \tag{61}$$

Prior: A probability measure requires a prior. The prior, which accounts for an arbitrary preparation of the ensemble, ought to be —for purposes of preserving the scope of the theory— of the same kind as the elements of the probability measure. Let us thus introduce the prior as the map $\mathbf{P}$ from the elements of $\mathbf{m}$ to $\mathbb{C}^{n\times n}$ and inject it into the probability measure as well as into the partition function:

$$\rho(p) = \frac{1}{Z} \det \exp\big(\mathbf{P}(p)\big) \det \exp\big(-\tau \mathbf{M}(p)\big) \tag{62}$$

where

$$Z = \sum_{p \in \mathbf{m}} \det \exp\big(\mathbf{P}(p)\big) \det \exp\big(-\tau \mathbf{M}(p)\big) \tag{63}$$

6.1 Completing the Measure

We have produced a measure over a sum of single experiments. Whereas the measure we are after is a sum over the whole space of experiments spawned by an experimental contribution, which contains all sub-tuples of the experimental contribution. Completing the measure over said space will require us to sum over differently-sized tuples. To do so, first, we will use the tensor product to produce measures summing over multiple elements, and second, we will use the direct sum to combine the differently-sized measures into a single final measure.

6.1.1 Split to Amplitude / Probability Rule

Before we are able to proceed with both the tensor product and the direct sum, it helps with familiarity to split the measure into two operations.

We begin by splitting the probability measure into a first step, which is linear with respect to a 'probability amplitude', and a second which connects the amplitude to the probability. We thus write the probability measure as:

$$\rho(p, \tau) = \frac{1}{Z} \det \psi(p, \tau) \tag{64}$$

where

$$\psi(p, \tau) = \exp\big(\mathbf{P}(p)\big) \exp\big(-\tau \mathbf{M}(p)\big) \tag{65}$$

Here, the determinant is interpreted as a generalization of the Born rule and reduces to exactly it when $\mathbf{M}$ is the matrix representation of the complex numbers. In the general case where $\mathbf{M}$ are arbitrary $n \times n$ matrices, $\psi(p, \tau)$ will be called the *general linear probability amplitude*.

We can write $\psi(p, \tau)$ as a column vector:

$$|\psi\rangle := \begin{pmatrix} \psi(p_1, \tau) \\ \psi(p_2, \tau) \\ \vdots \\ \psi(p_n, \tau) \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \tag{66}$$

6.1.2 Tensor Product

How do we extend the measure to experimental contributions containing multiple experiments? We have to use a Cartesian product on the sets of experimental images and a tensor product on the probability amplitudes. For instance, let us consider the following sets of experiments:

$$\mathbf{M}_1 = \{p_{1a}, p_{1b}\} \tag{67}$$
$$\mathbf{M}_2 = \{p_{2a}, p_{2b}\} \tag{68}$$

The Cartesian product produces experimental images comprised of two elements:

$$\mathbf{m} \in \mathbb{M}_1 \times \mathbb{M}_2 = \{(p_{1a}, p_{2a}), (p_{1a}, p_{2b}), (p_{1b}, p_{2a}), (p_{1b}, p_{2b})\} \tag{69}$$

At the level of the probability amplitude, the Cartesian product of sets translates to the tensor product. For instance, we start with a column vector where each entry is one experiment;

$$|\psi_1\rangle = \begin{pmatrix} \exp \mathbf{P}(p_{1a}) \\ \exp \mathbf{P}(p_{1b}) \end{pmatrix} \tag{70}$$

Adding a program-step via a linear transformation produces:

$$\mathbf{T}\,|\psi_1\rangle = \begin{pmatrix} T_{00} \exp \mathbf{P}(p_{1a}) + T_{01} \exp \mathbf{P}(p_{1b}) \\ T_{10} \exp \mathbf{P}(p_{1a}) + T_{11} \exp \mathbf{P}(p_{1b}) \end{pmatrix} \tag{71}$$

We then introduce another column vector:

$$|\psi_2\rangle = \begin{pmatrix} \exp \mathbf{P}(p_{2a}) \\ \exp \mathbf{P}(p_{2b}) \end{pmatrix} \tag{72}$$

along with a program-step:

$$\mathbf{T}'\,|\psi_2\rangle = \begin{pmatrix} T'_{00} \exp \mathbf{P}(p_{2a}) + T'_{01} \exp \mathbf{P}(p_{2b}) \\ T'_{10} \exp \mathbf{P}(p_{2a}) + T'_{11} \exp \mathbf{P}(p_{2b}) \end{pmatrix} \tag{73}$$

Then the tensor product of these states produces the probability measure of an experimental contribution as follows:

$$\mathbf{T}\,|\psi_1\rangle \otimes \mathbf{T}'\,|\psi_2\rangle = \begin{pmatrix} (T_{00} \exp \mathbf{P}(p_{1a}) + T_{01} \exp \mathbf{P}(p_{1b}))(T'_{00} \exp \mathbf{P}(p_{2a}) + T'_{01} \exp \mathbf{P}(p_{2b})) \\ (T_{00} \exp \mathbf{P}(p_{1a}) + T_{01} \exp \mathbf{P}(p_{1b}))(T'_{10} \exp \mathbf{P}(p_{2a}) + T'_{11} \exp \mathbf{P}(p_{2b})) \\ (T_{10} \exp \mathbf{P}(p_{1a}) + T_{11} \exp \mathbf{P}(p_{1b}))(T'_{00} \exp \mathbf{P}(p_{2a}) + T'_{01} \exp \mathbf{P}(p_{2b})) \\ (T_{10} \exp \mathbf{P}(p_{1a}) + T_{11} \exp \mathbf{P}(p_{1b}))(T'_{10} \exp \mathbf{P}(p_{2a}) + T'_{11} \exp \mathbf{P}(p_{2b})) \end{pmatrix} \tag{74}$$

Now, each element of the resulting vector is an experimental contribution of two programs, but its probability is a sum over a path. One can repeat the process n times.

6.1.3 Direct Sum

In the previous section, we have introduced a way to produce measures of fixed sizes n by using the tensor product. Here, we wish to produce a measure with elements of different sizes. Taking the direct sum of the measures of different sizes (where each individual size is produced from the tensor product), accomplishes the goal and yields an amplitude given has follows:

$$|\psi\rangle = |\psi_1\rangle \oplus (|\psi'_1\rangle \otimes |\psi'_2\rangle) \otimes (|\psi''_1\rangle \otimes |\psi''_2\rangle \otimes |\psi''_3\rangle) \oplus ... \tag{75}$$

In quantum field theory, in the limiting case $n \to \infty$ and when $\mathbf{M}(p)$ is reduced to the complex field, these are the states of a Fock Space, which we have obtained here simply my maximizing the entropy of the measure associated with our simple definition of the observer (Axiom 1). In the case for the measure space to be on all possible experiments, it requires $n \to \infty$.

6.2 Discussion — Fock Spaces as Measures over Tuples

Some may consider our result from the angle of measure theory in the sense that an entropy-maximized measure over the tuples of a tuple-space (as an extension to typical measure theory defined for the subsets of a set) induces a Fock Space, along with the appropriate probability rule (Born rule) for use in quantum mechanics. The measures used in quantum mechanics would thus result quite intuitively from this simple extension of measure theory, previously defined for sets, to tuples, and then simply maximizing the entropy.

We should mention that, although tuples can represent anything, in our system Axiom 1 requires the tuples to represent experimental contributions (or halted programs). But this is a minimal constraint, enforcing, while introducing no other constraints, that all experimental preparations or protocols must be describable using computable steps, and thus be comprehensible to the scientific method.

6.3 Connection to Computation

In the section on the formal system of knowledge, we have stated that our definitions will reference the concept of the computable function, and that such will hold irrespectively of the underlaying model of computation. Can we now say more? Here, we will see that having maximized the entropy of the measure space over tuples has automatically generated the underlaying model of computation.

Let us begin by reviewing the basics of quantum computation. One starts with a state vector:

$$|\psi_a\rangle = \begin{pmatrix} 0 \\ \vdots \\ n \end{pmatrix} \tag{76}$$

Which evolves unitarily to a final state:

$$|\psi_b\rangle = U_0 U_1 \dots U_m |\psi_a\rangle \tag{77}$$

Clever use of the unitary transformations, often arranged as simple 'gates', allows one to execute a program, but technically speaking any arrangements of unitary transformations qualify abstractly as a program (without or without gates). The input to the program is the state $|\psi_a\rangle$ and the output is the state $|\psi_b\rangle$. One would note that, so defined and if the sequence of unitary transformation is finite, such a program must always halt, and thus its complexity must be bounded. One can however get out of this predicament by taking the final state $|\psi_b\rangle$ to instead be an intermediary state, and then to add more gates in order continue with a computation:

$$\begin{aligned}
&\text{step 1} \qquad && |\psi_b\rangle = U_0 U_1 \dots U_p |\psi_a\rangle && (78)\\
&\text{step 2} \qquad && |\psi_c\rangle = U'_0 U'_1 \dots U'_q |\psi_b\rangle && (79)\\
&\vdots && && \\
&\text{step k} \qquad && |\psi_{k'}\rangle = U'_0 U'_1 \dots U'_v |\psi_k\rangle && (80)\\
&\vdots && &&
\end{aligned}$$

For a quantum computation to simulate a universal Turing machine it must be able to add more steps until a halting state is reached (or continue to add steps indefinitely if the program never halts). But note, that each step represents a valid quantum mechanical state of nature and is itself a completed program.

Comparatively, the linear transformations $\mathbf{T}_1, \mathbf{T}_2, \dots$ of our main result are here interpreted in the same manner as those used in quantum computations, but extended to the general linear group. Protocols are executed by chaining transformations on a preparation:

$$\underbrace{|\psi_b\rangle}_{\text{final state}} = \underbrace{\mathbf{T}_1 \mathbf{T}_2 \dots \mathbf{T}_n}_{\text{protocol}} \underbrace{|\psi_a\rangle}_{\text{preparation}} \tag{81}$$

And a quantum computation involves a sequence of unitary transformations on the unit vectors of a complex Hilbert space:

$$\underbrace{|\psi_b\rangle}_{\text{final state}} = \underbrace{\mathbf{U}_1\mathbf{U}_2\ldots\mathbf{U}_n}_{\text{computing steps}} \underbrace{|\psi_a\rangle}_{\text{initial state}} \tag{82}$$

We are now ready to begin investigating the main result as a general linear quantum theory.

7 Foundation of Physics

Based on our main result, we will now introduce an *algebra of natural states* and we will use it to classify the linear transformations on said amplitude. We will start with the 2D case, then the 4D case. In all cases, the probability amplitude transforms linearly with respect to general linear transformations and the probability measure, obtained from the determinant, is positive-definite. In the application section, we will then see that the 4D case automatically reduces to the Dirac theory of relativistic quantum mechanics when the general linear group is reduced to the spinor group. Finally, we will show that the general linear wave-function entails the Einstein field equations as its evolution equation in the general case.

7.1 Matrix-Valued Vector and Transformations

To work with the general linear probability amplitude, we will use vectors whose elements are matrices. An example of such a vector is:

$$|\psi\rangle = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_m \end{pmatrix} \tag{83}$$

Likewise a linear transformation of this space will expressed as a matrix of matrices:

$$\mathbf{T} = \begin{pmatrix} \mathbf{M}_{00} & \ldots & \mathbf{M}_{0m} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{m0} & \ldots & \mathbf{M}_{mm} \end{pmatrix} \tag{84}$$

Note: The scalar element of the vector space are given as:

$$a\,|\psi\rangle = \begin{pmatrix} a\mathbf{M}_1 \\ \vdots \\ a\mathbf{M}_m \end{pmatrix} \tag{85}$$

7.2 Algebra of Natural States, in 2D

The notation of our upcoming definitions will be significantly improved if we use a geometric representation for matrices. Let us therefore introduce a geometric representation of 2×2 matrices.

7.2.1 Geometric Representation of 2×2 matrices

Let $\mathbb{G}(2,\mathbb{R})$ be the two-dimensional geometric algebra over the reals. We can write a general multi-vector of $\mathbb{G}(2,\mathbb{R})$ as follows:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{B} \tag{86}$$

where A is a scalar, $\mathbf{X}$ is a vector and $\mathbf{B}$ is a pseudo-scalar. Each multi-vector has a structure-preserving (addition/multiplication) matrix representation. Explicitly, the multi-vectors of $\mathbb{G}(2,\mathbb{R})$ are represented as follows:

Definition 21 (Geometric representation of a matrix (2×2))**.**

$$A + X\hat{\mathbf{x}} + Y\hat{\mathbf{y}} + B\hat{\mathbf{x}} \wedge \hat{\mathbf{y}} \cong \begin{pmatrix} A+X & -B+Y \\ B+Y & A-X \end{pmatrix} \tag{87}$$

And the converse is also true, each 2×2 real matrix is represented as a multi-vector of $\mathbb{G}(2,\mathbb{R})$.

We can define the determinant solely using constructs of geometric algebra[12].

Definition 22 (Clifford conjugate (of a $\mathbb{G}(2,\mathbb{R})$ multi-vector))**.**

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 \tag{88}$$

Then the determinant of $\mathbf{u}$ is:

Definition 23 (Geometric representation of the determinant (of a 2×2 matrix))**.**

$$\begin{array}{rcrcl} \det & : & \mathbb{G}(2,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{u} \end{array} \tag{89}$$

For example:

$$\det \mathbf{u} = (A - \mathbf{X} - \mathbf{B})(A + \mathbf{X} + \mathbf{B}) \tag{90}$$

$$= A^2 - X^2 - Y^2 + B^2 \tag{91}$$

$$= \det \begin{pmatrix} A+X & -B+Y \\ B+Y & A-X \end{pmatrix} \tag{92}$$

Finally, we define the Clifford transpose:

Definition 24 (Clifford transpose (of a matrix of 2×2 matrix elements))**.** *The Clifford transpose is the geometric analogue to the conjugate transpose. Like the conjugate transpose can be interpreted as a transpose followed by an element-by-element application of the complex conjugate, here the Clifford transpose is a transpose, followed by an element-by-element application of the Clifford conjugate:*

$$\begin{pmatrix} \mathbf{u}_{00} & \dots & \mathbf{u}_{0n} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}_{mn} \end{pmatrix}^{\ddagger} = \begin{pmatrix} \mathbf{u}_{00}^{\ddagger} & \dots & \mathbf{u}_{m0}^{\ddagger} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}_{nm}^{\ddagger} \end{pmatrix} \tag{93}$$

If applied to a vector, then:

$$\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix}^{\ddagger} = \begin{pmatrix} \mathbf{v}_1^{\ddagger} & \dots \mathbf{v}_m^{\ddagger} \end{pmatrix} \tag{94}$$

7.2.2 Axiomatic Definition of the Algebra, in 2D

Let $\mathbb{V}$ be an m-dimensional vector space over $\mathbb{G}(2,\mathbb{R})$. A subset of vectors in $\mathbb{V}$ forms an algebra of natural states $\mathcal{A}(\mathbb{V})$ iff the following holds:

1. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, the bilinear map:

$$\begin{array}{rcrcl} \langle \cdot, \cdot \rangle & : & \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(2,\mathbb{R}) \\ & & \langle \mathbf{u}, \mathbf{v} \rangle & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{v} \end{array} \tag{95}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{96}$$

2. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, then for each element $\psi(q) \in \boldsymbol{\psi}$, the function:

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \psi(q)^{\ddagger} \psi(q) \tag{97}$$

   is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{98}$$

We note the following comments and definitions:

- From (1) and (2) it follows that $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, the probabilities sum to unity:

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{99}$$

- $\boldsymbol{\psi}$ is called a *natural* (or physical) state.
- $\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle$ is called the *partition function* of $\boldsymbol{\psi}$.
- $\rho(q, \boldsymbol{\psi})$ is called the *probability measure* (or generalized Born rule) of $\psi(q)$.
- The set of all matrices $\mathbf{T}$ acting on $\boldsymbol{\psi}$, as $\mathbf{T}\boldsymbol{\psi} \to \boldsymbol{\psi}'$, which leaves the sum of probabilities normalized (invariant):

$$\sum_{\psi(q)\in\boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q)\in\boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{100}$$

  are the *natural* transformations of $\boldsymbol{\psi}$.

- A matrix $\mathbf{O}$ such that $\forall \mathbf{u} \forall \mathbf{v} \in \mathcal{A}(\mathbb{V})$:

$$\langle \mathbf{O}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v} \rangle \tag{101}$$

  is called an observable.

- The expectation value of an observable $\mathbf{O}$ is:

$$\langle \mathbf{O} \rangle = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi} \rangle \tag{102}$$

7.2.3 Observable, in 2D — Self-Adjoint Operator

Let us now investigate the general case of an observable is 2D. A matrix $\mathbf{O}$ is an observable iff it is a self-adjoint operator; defined as:

$$\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle = \langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle \tag{103}$$

$\forall \mathbf{u} \forall \mathbf{v} \in \mathbb{V}$.

Setup: Let $\mathbf{O} = \begin{pmatrix} O_{00} & O_{01} \\ O_{10} & O_{11} \end{pmatrix}$ be an observable. Let $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ be 2 two-state vectors $\boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ and $\boldsymbol{\psi} = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$. Here, the components ϕ_1, ϕ_2, ψ_1, ψ_2, O_{00}, O_{01}, O_{10}, O_{11} are multi-vectors of $\mathbb{G}(2, \mathbb{R})$.

Derivation: 1. Let us now calculate $\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle$:

$$\begin{aligned} 2\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle &= (O_{00}\phi_1 + O_{01}\phi_2)^\ddagger \psi_1 + \psi_1^\ddagger (O_{00}\phi_1 + O_{01}\phi_2) \\ &\quad + (O_{10}\phi_1 + O_{11}\phi_2)^\ddagger \psi_2 + \psi_2^\ddagger (O_{10}\phi_1 + O_{11}\phi_2) \qquad (104) \\ &= \phi_1^\ddagger O_{00}^\ddagger \psi_1 + \phi_2^\ddagger O_{01}^\ddagger \psi_1 + \psi_1^\ddagger O_{00}\phi_1 + \psi_1^\ddagger O_{01}\phi_2 \\ &\quad + \phi_1^\ddagger O_{10}^\ddagger \psi_2 + \phi_2^\ddagger O_{11}^\ddagger \psi_2 + \psi_2^\ddagger O_{10}\phi_1 + \psi_2^\ddagger O_{11}\phi_2 \qquad (105) \end{aligned}$$

2. Now, $\langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle$:

$$
\begin{aligned}
2\langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi}\rangle &= \phi_1^\ddagger(O_{00}\psi_1 + O_{01}v_2) + (O_{00}\psi_1 + O_{01}\psi_2)^\ddagger \phi_1 \\
&\quad + \phi_2^\ddagger(O_{10}\psi_1 + O_{11}\psi_2) + (O_{10}\psi_1 + O_{11}\psi_2)^\ddagger \phi_1 && (106) \\
&= \phi_1^\ddagger O_{00}\psi_1 + \phi_1^\ddagger O_{01}\psi_2 + \psi_1^\ddagger O_{00}^\ddagger \phi_1 + \psi_2^\ddagger O_{01}^\ddagger \phi_1 \\
&\quad + \phi_2^\ddagger O_{10}\psi_1 + \phi_2^\ddagger O_{11}\psi_2 + \psi_1^\ddagger O_{10}^\ddagger \phi_1 + \psi_2^\ddagger O_{11}^\ddagger \phi_1 && (107)
\end{aligned}
$$

For $\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi}\rangle = \langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi}\rangle$ to be realized, it follows that these relations must hold:

$$O_{00}^\ddagger = O_{00} \tag{108}$$

$$O_{01}^\ddagger = O_{10} \tag{109}$$

$$O_{10}^\ddagger = O_{01} \tag{110}$$

$$O_{11}^\ddagger = O_{11} \tag{111}$$

Therefore, it follows that it must be the case that $\mathbf{O}$ must be equal to its own Clifford transpose. Thus, $\mathbf{O}$ is an observable iff:

$$\mathbf{O}^\ddagger = \mathbf{O} \tag{112}$$

which is the equivalent of the self-adjoint operator $\mathbf{O}^\dagger = \mathbf{O}$ of complex Hilbert spaces.

7.2.4 Observable, in 2D — Eigenvalues / Spectral Theorem

Let us show how the spectral theorem applies to $\mathbf{O}^\ddagger = \mathbf{O}$, such that its eigenvalues are real. Consider:

$$\mathbf{O} = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{113}$$

In this case, it follows that $\mathbf{O}^\ddagger = \mathbf{O}$:

$$\mathbf{O}^\ddagger = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{114}$$

This example is the most general 2×2 matrix $\mathbf{O}$ such that $\mathbf{O}^\ddagger = \mathbf{O}$. The eigenvalues are obtained as follows:

$$0 = \det(\mathbf{O} - \lambda I) = \det \begin{pmatrix} a_{00} - \lambda & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} - \lambda \end{pmatrix} \tag{115}$$

implies:

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a - xe_1 - ye_2 - be_{12})(a + xe_1 + ye_2 + be_{12} + a_{11}) \tag{116}$$

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a^2 - x^2 - y^2 + b^2) \tag{117}$$

finally:

$$\lambda = \{\frac{1}{2}\left(a_{00} + a_{11} - \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right), \tag{118}$$

$$\frac{1}{2}\left(a_{00} + a_{11} + \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right)\} \tag{119}$$

We note that in the case where $a_{00} - a_{11} = 0$, the roots would be complex iff $a^2 - x^2 - y^2 + b^2 < 0$, but we already stated that the determinant of real matrices must be greater than zero because the exponential maps to the orientation-preserving general linear group— therefore it is the case that $a^2 - x^2 - y^2 + b^2 \geq 0$, as this expression is the determinant of the multi-vector. Consequently, $\mathbf{O}^\ddagger = \mathbf{O}$ — implies, for orientation-preserving[1] transformations, that its roots are real-valued, and thus constitute a 'geometric' observable in the traditional sense of an observable whose eigenvalues are real-valued.

7.3 Algebra of Natural States, in 4D

We will now consider the general case for a vector space over 4×4 matrices.

7.3.1 Geometric Representation (in 4D)

The notation will be significantly improved if we use a geometric representation of matrices. Let $\mathbb{G}(4, \mathbb{R})$ be the two-dimensional geometric algebra over the reals. We can write a general multi-vector of $\mathbb{G}(4, \mathbb{R})$ as follows:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{120}$$

where A is a scalar, $\mathbf{X}$ is a vector, $\mathbf{F}$ is a bivector, $\mathbf{V}$ is a pseudo-vector, and $\mathbf{B}$ is a pseudo-scalar. Each multi-vector has a structure-preserving (addition/multiplication) matrix representation. Explicitly, the multi-vectors of $\mathbb{G}(4, \mathbb{R})$ are represented as follows:

[1] We note the exception that a geometric observable may have real eigenvalues even in the case of a transformation that reverses the orientation if the elements $a_{00} - a_{11}$ are not zero and up to a certain magnitude, whereas transformations in the natural orientation are not bounded by a magnitude — thus creating an orientation-based asymmetry.

Definition 25 (Geometric representation of a matrix (4×4)).

$$
\begin{aligned}
&A + T\gamma_0 + X\gamma_1 + Y\gamma_2 + Z\gamma_3 \\
&\quad + F_{01}\gamma_0 \wedge \gamma_1 + F_{02}\gamma_0 \wedge \gamma_2 + F_{03}\gamma_0 \wedge \gamma_3 + F_{23}\gamma_2 \wedge \gamma_3 + F_{13}\gamma_1 \wedge \gamma_3 + F_{12}\gamma_1 \wedge \gamma_2 \\
&\quad + V_t\gamma_1 \wedge \gamma_2 \wedge \gamma_3 + V_x\gamma_0 \wedge \gamma_2 \wedge \gamma_3 + V_y\gamma_0 \wedge \gamma_1 \wedge \gamma_3 + V_z\gamma_0 \wedge \gamma_1 \wedge \gamma_2 \\
&\quad + B\gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3 \\
&\qquad \cong \begin{pmatrix}
A + X_0 - iF_{12} - iV_3 & F_{13} - iF_{23} + V_2 - iV_1 & -iB + X_3 + F_{03} - iV_0 & X_1 - iX_2 + F_{01} - iF_{02} \\
-F_{13} - iF_{23} - V_2 - iV_1 & A + X_0 + iF_{12} + iV_3 & X_1 + iX_2 + F_{01} + iF_{02} & -iB - X_3 - F_{03} - iV_0 \\
-iB - X_3 + F_{03} + iV_0 & -X_1 + iX_2 + F_{01} - iF_{02} & A - X_0 - iF_{12} + iV_3 & F_{13} - iF_{23} - V_2 + iV_1 \\
-X_1 - iX_2 + F_{01} + iF_{02} & -iB + X_3 - F_{03} + iV_0 & -F_{13} - iF_{23} + V_2 + iV_1 & A - X_0 + iF_{12} - iV_3
\end{pmatrix}
\end{aligned} \tag{121}
$$

And the converse is also true, each 4×4 real matrix is represented as a multi-vector of $\mathbb{G}(4, \mathbb{R})$.

We can define the determinant solely using constructs of geometric algebra[12].

Definition 26 (Clifford conjugate (of a $\mathbb{G}(4, \mathbb{R})$ multi-vector)).

$$\mathbf{u}^\ddagger := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 + \langle \mathbf{u} \rangle_3 + \langle \mathbf{u} \rangle_4 \tag{122}$$

and $\lfloor \mathbf{m} \rfloor_{\{3,4\}}$ *as the blade-conjugate of degree 3 and 4 (flipping the plus sign to a minus sign for blade 3 and blade 4):*

$$\lfloor \mathbf{u} \rfloor_{\{3,4\}} := \langle \mathbf{u} \rangle_0 + \langle \mathbf{u} \rangle_1 + \langle \mathbf{u} \rangle_2 - \langle \mathbf{u} \rangle_3 - \langle \mathbf{u} \rangle_4 \tag{123}$$

Then, the determinant of $\mathbf{u}$ is:

Definition 27 (Geometric representation of the determinant (of a 4×4 matrix)).

$$
\begin{array}{rcrcl}
\det & : & \mathbb{G}(4, \mathbb{R}) & \longrightarrow & \mathbb{R} \\
 & & \mathbf{u} & \longmapsto & \lfloor \mathbf{u}^\ddagger \mathbf{u} \rfloor_{3,4} \mathbf{u}^\ddagger \mathbf{u}
\end{array} \tag{124}
$$

7.3.2 Axiomatic Definition of the Algebra, in 4D

Let $\mathbb{V}$ be a m-dimensional vector space over the 4×4 real matrices. A subset of vectors in $\mathbb{V}$ forms an algebra of natural states $\mathcal{A}(\mathbb{V})$ iff the following holds:

1. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, the quadri-linear form:

$$
\begin{array}{rcrcl}
\langle \cdot, \cdot, \cdot, \cdot \rangle & : & \mathbb{V} \times \mathbb{V} \times \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(4, \mathbb{R}) \\
 & & \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle & \longmapsto & \lfloor \mathbf{u}^\ddagger \mathbf{v} \rfloor_{3,4} \mathbf{w}^\ddagger \mathbf{x}
\end{array} \tag{125}
$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{126}$$

2. $\forall \boldsymbol{\psi} \in \mathcal{A}(\mathbb{V})$, then for each element $\psi(q) \in \boldsymbol{\psi}$, the function:

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \lfloor \psi(q)^{\ddagger} \psi(q) \rfloor_{3,4} \psi(q)^{\ddagger} \psi(q) \tag{127}$$

is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{128}$$

We note the following properties, features and comments:

- $\boldsymbol{\psi}$ is called a *natural* (or physical) state.
- $\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle$ is called the *partition function* of $\boldsymbol{\psi}$.
- $\rho(\psi(q), \boldsymbol{\psi})$ is called the *probability measure* (or generalized Born rule) of $\psi(q)$.
- The set of all matrices $\mathbf{T}$ acting on $\boldsymbol{\psi}$ such as $\mathbf{T}\boldsymbol{\psi} \to \boldsymbol{\psi}'$ which leaves the sum of probabilities normalized (invariant):

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{129}$$

  are the *natural* transformations of $\boldsymbol{\psi}$.
- A matrix $\mathbf{O}$ such that $\forall \mathbf{u} \forall \mathbf{v} \forall \mathbf{w} \forall \mathbf{x} \in \mathbb{V}$:

$$\langle \mathbf{Ou}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{Ov}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{Ow}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{Ox} \rangle \tag{130}$$

  is called an observable.
- The expectation value of an observable $\mathbf{O}$ is:

$$\langle \mathbf{O} \rangle = \frac{\langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \tag{131}$$

7.4 Probability-Preserving Transformation

7.4.1 Left Action in 2D

A left action on a wave-function : $\mathbf{T} |\psi\rangle$, connects to the bilinear form as follows: $\langle\psi| \mathbf{T}^{\ddagger}\mathbf{T} |\psi\rangle$. The invariance requirement on $\mathbf{T}$ is as follows:

$$\langle\psi| \mathbf{T}^{\ddagger}\mathbf{T} |\psi\rangle = \langle\psi|\psi\rangle \tag{132}$$

We are thus interested in the group of matrices such that:

$$\mathbf{T}^{\ddagger}\mathbf{T} = I \tag{133}$$

Let us consider a two-state system. A general transformation is:

$$\mathbf{T} = \begin{pmatrix} u & v \\ w & x \end{pmatrix} \tag{134}$$

where u, v, w, x are multi-vectors of 2 dimensions. The expression $\mathbf{G}^\ddagger\mathbf{G}$ is:

$$\mathbf{T}^\ddagger\mathbf{T} = \begin{pmatrix} v^\ddagger & u^\ddagger \\ w^\ddagger & x^\ddagger \end{pmatrix} \begin{pmatrix} v & w \\ u & x \end{pmatrix} = \begin{pmatrix} v^\ddagger v + u^\ddagger u & v^\ddagger w + u^\ddagger x \\ w^\ddagger v + x^\ddagger u & w^\ddagger w + x^\ddagger x \end{pmatrix} \tag{135}$$

For the results to be the identity, it must be the case that:

$$v^\ddagger v + u^\ddagger u = 1 \tag{136}$$
$$v^\ddagger w + u^\ddagger x = 0 \tag{137}$$
$$w^\ddagger v + x^\ddagger u = 0 \tag{138}$$
$$w^\ddagger w + x^\ddagger x = 1 \tag{139}$$

This is the case if

$$\mathbf{T} = \frac{1}{\sqrt{v^\ddagger v + u^\ddagger u}} \begin{pmatrix} v & u \\ -e^{\varphi} u^\ddagger & e^{\varphi} v^\ddagger \end{pmatrix} \tag{140}$$

where u, v are multi-vectors of 2 dimensions, and where e^{φ} is a unit multi-vector. Comparatively, the unitary case is obtained with $\mathbf{X} \to 0$, and is:

$$\mathbf{U} = \frac{1}{\sqrt{|a|^2 + |b|^2}} \begin{pmatrix} a & b \\ -e^{i\theta} b^\dagger & e^{i\theta} a^\dagger \end{pmatrix} \tag{141}$$

We can show that $\mathbf{G}^\ddagger\mathbf{G} = I$ as follows:

$$\implies \mathbf{T}^\ddagger\mathbf{T} = \frac{1}{v^\ddagger v + u^\ddagger u} \begin{pmatrix} v^\ddagger & -e^{-\varphi} u \\ u^\ddagger & e^{-\varphi} v \end{pmatrix} \begin{pmatrix} v & u \\ -e^{\varphi} u^\ddagger & e^{\varphi} v^\ddagger \end{pmatrix} \tag{142}$$

$$= \frac{1}{v^\ddagger v + u^\ddagger u} \begin{pmatrix} v^\ddagger v + u^\ddagger u & v^\ddagger u - v^\ddagger u \\ u^\ddagger v - u^\ddagger v & u^\ddagger u + v^\ddagger v \end{pmatrix} \tag{143}$$

$$= I \tag{144}$$

In the case where $\mathbf{T}$ and $|\psi\rangle$ are n-dimensional, we can find an expression for it starting from a diagonal matrix:

$$\mathbf{D} = \begin{pmatrix} e^{x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1} & 0 \\ 0 & e^{x_2\hat{\mathbf{x}}+y_2\hat{\mathbf{y}}+ib_2} \end{pmatrix} \tag{145}$$

where $\mathbf{T} = P\mathbf{D}P^{-1}$. It follows quite easily that $D^{\ddagger}D = I$, because each diagonal entry produces unity: $e^{-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1}e^{x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1} = 1$.

7.4.2 Adjoint Action in 2D

The left action case can recover at most the special linear group. For the general linear group itself, we require the adjoint action. Since the elements of $|\psi\rangle$ are matrices, in the general case, the transformation is given by adjoint action:

$$\mathbf{T}\,|\psi\rangle\,\mathbf{T}^{-1} \tag{146}$$

The bilinear form is:

$$(\mathbf{T}\,|\psi\rangle\,\mathbf{T}^{-1})^{\ddagger}(\mathbf{T}\,|\psi\rangle\,\mathbf{T}^{-1}) = (\mathbf{T}^{-1})^{\ddagger}\,\langle\psi|\,\mathbf{T}^{\ddagger}\mathbf{T}\,|\psi\rangle\,\mathbf{T}^{-1} \tag{147}$$

and the invariance requirement on $\mathbf{T}$ is as follows:

$$(\mathbf{T}^{-1})^{\ddagger}\,\langle\psi|\,\mathbf{T}^{\ddagger}\mathbf{T}\,|\psi\rangle\,\mathbf{T}^{-1} = \langle\psi|\psi\rangle \tag{148}$$

With a diagonal matrix, this occurs for general linear transformations:

$$\mathbf{D} = \begin{pmatrix} e^{a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1} & 0 & 0 \\ 0 & e^{a_2+x_2\hat{\mathbf{x}}+y_2\hat{\mathbf{y}}+ib_2} & 0 \\ 0 & 0 & \ddots \end{pmatrix} \tag{149}$$

where $\mathbf{T} = P\mathbf{D}P^{-1}$.

Taking a single diagonal entry as an example, the reduction is:

$$e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1^{\ddagger}e^{a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1}e^{a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1 e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{150}$$

$$= e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1^{\ddagger}e^{2a_1}\psi_1 e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{151}$$

We note that $\psi^{\ddagger}\psi$ is a scalar, therefore

$$= \psi_1^{\ddagger}\psi_1 e^{2a_1}e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{152}$$

$$= \psi_1^{\ddagger}\psi_1 e^{2a_1}e^{-a_1}e^{-a_1} = \psi_1^{\ddagger}\psi_1 \tag{153}$$

8 Applications

8.1 Dirac Current and the Bilinear Covariants

The general linear wave-function is:

$$\psi = \exp(A + \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B}) \tag{154}$$

In this application, let us take a group reduction from the general linear group to the spinor group. As such we pose $\mathbf{X} \to 0$ and $\mathbf{V} \to 0$. The wave-function becomes:

$$\psi = \exp(A + \mathbf{F} + \mathbf{B}) \tag{155}$$

We recall that in 4D, the probability is given as follows:

$$\det \psi = \lfloor \psi^{\ddagger}\psi \rfloor_{3,4} \psi^{\ddagger}\psi = \exp 4A = \rho \tag{156}$$

but, since we eliminated $\mathbf{X} \to 0$ and $\mathbf{V} \to 0$, we can drop the blade inversion of degree 3, and the rule reduces to:

$$\det \psi = (\psi^{\ddagger})^* \psi^* \psi^{\ddagger}\psi = \exp 4A = \rho \tag{157}$$

Let us now recover the familiar Dirac theory.

First, we will expand the probability rule, while injecting γ_0 and γ_μ as follows:

$$(\psi^{\ddagger})^* \gamma_0 \psi^* \psi^{\ddagger} \gamma_\mu \psi = (e^A e^{-\mathbf{B}} e^{-\mathbf{F}})\gamma_0 (e^A e^{-\mathbf{B}} e^{\mathbf{F}})(e^A e^{\mathbf{B}} e^{-\mathbf{F}})\gamma_\mu (e^A e^{\mathbf{B}} e^{\mathbf{F}}) \tag{158}$$

But before we continue, let us introduce the notation of David Hestenes. We write $e^{\mathbf{F}} = R$, a rotor, and $e^{-\mathbf{F}} = \widetilde{R}$, its reverse. The pseudo-scalar term will also be written as $e^{\mathbf{B}} = e^{ib}$. Finally, we write $e^{4A} = \rho$. Consequently, we obtain:

$$= \rho^{\frac{1}{4}} e^{-ib} \widetilde{R} \gamma_0 \rho^{\frac{1}{4}} e^{-ib} R \rho^{\frac{1}{4}} e^{ib} \widetilde{R} \gamma_\mu \rho^{\frac{1}{4}} e^{ib} R \tag{159}$$

$$= \rho e^{-ib} \widetilde{R} \gamma_0 \gamma_\mu e^{ib} R \tag{160}$$

$$= \rho \widetilde{R} \gamma_0 \gamma_\mu R \tag{161}$$

$$= (\rho, \vec{J}) \tag{162}$$

This is simply the Dirac current expressed with Tetrads. The Dirac equation describes the dynamics which preserve this current. The base wave-function in canonical form is:

$$\psi = \rho^{\frac{1}{4}} e^{ib} R \tag{163}$$

Comparatively, David Hestenes' wave-function is $\psi = \rho^{\frac{1}{2}} e^{ib} R$ which is very similar. To make the full Dirac theory standout, we can introduce an intermediary form of the wave-function, as follows:

$$(\psi^\ddagger)^* \gamma_0 \psi^* \psi^\ddagger \gamma_\mu \psi = \rho^{\frac{1}{4}} e^{-ib} \widetilde{R} \gamma_0 \rho^{\frac{1}{4}} e^{-ib} R \rho^{\frac{1}{4}} e^{ib} \widetilde{R} \gamma_\mu \rho^{\frac{1}{4}} e^{ib} R \tag{164}$$

$$= \underbrace{\rho^{\frac{1}{2}} e^{-ib} \widetilde{R} \gamma_0}_{\bar{\phi}} \gamma_\mu \underbrace{\rho^{\frac{1}{2}} e^{ib} R}_{\phi} \tag{165}$$

Specifically,

$$\bar{\phi} := \rho^{\frac{1}{2}} e^{-ib} \widetilde{R} \gamma_0 \tag{166}$$

$$\phi := \rho^{\frac{1}{2}} e^{ib} R \tag{167}$$

and thus

$$\det \psi = (\psi^\ddagger)^* \psi^* \psi^\ddagger \psi = \bar{\phi} \gamma_0 \phi = \rho \tag{168}$$

The full list of bilinear covariants are:

	Determinant	ϕ-notation	Standard Form	Result
scalar	$(\psi^\ddagger)^* \gamma_0 \psi^* \psi^\ddagger \psi$	$\bar{\phi}\phi$	$\langle \bar{\psi} \vert \psi \rangle$	$e_0 \rho \cos b$
vector	$(\psi^\ddagger)^* \gamma_0 \psi^* \psi^\ddagger \gamma_\mu \psi$	$\bar{\phi}\gamma_\mu\phi$	$\langle \bar{\psi} \vert \gamma_\mu \vert \psi \rangle$	J_μ
bivector	$(\psi^\ddagger)^* \gamma_0 \psi^* \psi^\ddagger I \gamma_\mu \gamma_\nu \psi$	$\bar{\phi} I \gamma_\mu \gamma_\nu \phi$	$\langle \bar{\psi} \vert i\gamma_\mu \gamma_\nu \vert \psi \rangle$	S
pseudo-vector	$(\psi^\ddagger)^* \gamma_0 \psi^* \psi^\ddagger \gamma_\mu I \psi$	$\bar{\phi}\gamma_\mu I \phi$	$\langle \bar{\psi} \vert \gamma_\mu \gamma_5 \vert \psi \rangle$	s_μ
pseudo-scalar	$(\psi^\ddagger)^* \gamma_0 \psi^* \psi^\ddagger I \psi$	$\bar{\phi} I \phi$	$\langle \bar{\psi} \vert i\gamma_5 \vert \psi \rangle$	$-e_0 \rho \sin b$

(169)

One might have been sceptical that our extension from the Born rule on complex-valued wave-functions to the determinant of matrices could yield any relevance for physics, but this result along with the dynamical result for $U(1)$ and the physical interpretation in the next sections, show that it is equivalent to the full Dirac theory, in 4D. We would argue, however, that our approach conceptually much simpler... as it essentially only involves applying the determinant to a sum of matrices and noting the emergence of the geometric elements manifest in the geometric algebra representation of matrices. No import of physical language is required to get the Dirac theory, as this result follows directly from our prescription.

8.2 Yang-Mills Theories — Unitary Gauge/Recap

The typical gauge theory in quantum electrodynamics is obtained by the production of a gauge covariant derivative over a $U(1)$ invariance associated with the use of the complex norm in any probability measure of quantum mechanics. A parametrization of ψ over a differentiable manifold is required to support the derivation. Localizing the invariance group $\theta \to \theta(x)$, over said parametrization, yields the corresponding covariant derivative:

$$D_\mu = \partial_\mu + iqA_\mu(x) \tag{170}$$

Where $A_\mu(x)$ is the gauge field. The $U(1)$ invariance results from the usage of the complex norm to construct a probability measure in a quantum theory, and the presence of the derivative is the result of constructing said probability measure as the Lagrangian of a Dirac field. If one then applies a gauge transformation to ψ and A_μ:

$$\psi \to e^{-iq\theta(x)}\psi \quad \text{and} \quad A_\mu \to A_\mu + \partial_\mu\theta(x) \tag{171}$$

Then, applies the covariant derivation, one gets:

$$\begin{aligned} D_\mu\psi &= \partial_\mu\psi + iqA_\mu\psi && (172)\\ &\to \partial_\mu(e^{-iq\theta(x)}\psi) + iq(A_\mu + \partial_\mu\theta(x))(e^{-iq\theta(x)}\psi) && (173)\\ &= e^{-iq\theta(x)}D_\mu\psi && (174) \end{aligned}$$

Finally, the field is given as follows:

$$F_{\mu\nu} = [\mathcal{D}_\mu, \mathcal{D}_\nu] \tag{175}$$

where $\mathcal{D}_\mu$ is the covariant derivative with respect to the potential one-form $A_\mu = A_\mu^{\alpha}T_\alpha$, and where T_α are the generators of the lie algebra of $U(1)$.

8.3 Quantum Gravity — General Linear Gauge

When our wave-function is extended to the general linear group (and the Born rule to the determinant), then its fundamental invariance group is the orientation-preserving general linear group $\mathrm{GL}^+(4,\mathbb{R})$, rather than $U(1)$.

Gauging the $\mathrm{GL}^+(4,\mathbb{R})$ group is known to automatically entail the Einstein field equations, as the resulting $\mathrm{GL}^+(4,\mathbb{R})$-valued field can be viewed as the Christoffel symbols Γ^μ, and the commutator of the covariant derivatives as the Riemann tensor. Expressing gravity via the general linear gauge is not a new result: This is a result dating back from the 1956 with Utiyama[13], in 1961 with Kibble[14], as well as the more recent work of David Hestenes[15] specifically with geometric algebras.

A general linear transformation of ψ:

$$\psi'(x) \to g\psi(x)g^{-1} \tag{176}$$

leaves the probability measure invariant.
The gauge-covariant derivative is:

$$D_\mu \psi = \partial_\mu \psi - [iqA_\mu, \psi] \tag{177}$$

Finally, the field is given as follows:

$$R_{\mu\nu} = [D_\mu, D_\nu] \tag{178}$$

where $R_{\mu\nu}$ is the Riemann tensor.
The Lagrangian is of course the Einstein-Hilbert action which, up to numerical constant, is:

$$S = \int \epsilon_{abcd} R^{ab} \wedge e^c \wedge e^d = \int \mathrm{d}^4 x \sqrt{-g} R \tag{179}$$

Here we have an extra backbone to this gauge formulation of general relativity; specifically a quantum theory able to accommodate and normalize the general linear group as a probability measure. Let us now discuss the physical interpretation of quantum theory.

8.4 Physical Interpretation

Typically to insert gravity into a quantum field theory, one would take the Einstein Field equation, then would linearize the metric: $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$. After expansion in powers of h, the Einstein-Hilbert action becomes:

$$\left(S = \frac{1}{16\pi G} \int \mathrm{d}^4 x \sqrt{-g} R \bigg|_{g_{\mu\nu} \to \eta_{\mu\nu} + h_{\mu\nu}} \to \frac{1}{16\pi G} \int \mathrm{d}^4 x (\partial h \partial h + h \partial h \partial h + h^2 \partial h \partial h + \dots) \right. \tag{180}$$

The indices on h has been dropped for brevity. The terms $\partial h \partial h$ are similar in role to scalar field theory $\partial \phi \partial \phi$, and the higher order terms are interaction terms. Since the higher order terms cannot be normalized, and one typically drops the higher order terms from the theory yielding at best an approximation of the full theory of quantum gravity to the first order approximation.

Comparatively in our theory, the Einstein field equations are automatically entailed by the theory as the equation of motion of the wave-function. They do not have to be imported from general relativity, nor to they have to be made linear to fit in.

Before we tackle the general linear case, let us first review the interpretation of David Hestenes for the spinor case and in flat space-time. For this interpretation, we pose $\mathbf{X} \to 0, \mathbf{V} \to 0$ and use the setting of flat spacetime. The wave-function would be parametrized as follows:

$$\psi(x_0,x_1,x_2,x_3)=\rho(x_0,x_1,x_2,x_3)^{1/4}\exp\bigl(\mathbf{F}(x_0,x_1,x_2,x_3)+\mathbf{B}(x_0,x_1,x_2,x_3)\bigr) \tag{181}$$

In this case the wave-function assigns a rotor $R=\exp(\mathbf{F})$, a probability density ρ and a phase $\exp(\mathbf{B})$ to each point of $\mathbb{R}^4$. R as a rotor accounts for a Lorentz rotation at each point: $e_\mu = R\gamma_u\widetilde{R}$. Thus, it assigns at each points on the manifold an instruction to rotate in addition to giving it a statistical weight and a phase. As argued by David Hestenes is his seminal paper[16], this description is informationally complete, and equivalent, to other interpretations of quantum physics; and is framed in terms of a relativistic kinematic theory of fermions, such as that of the electron.

How does the general linear case compare? In this case, our wave-function assigns an element of the $\mathrm{GL}^+(4,\mathbb{R})$ group to each point on the manifold. The wave-function, using the multi-vector notation, in this case would be parametrized as follows:

$$\begin{aligned}\psi(x_0,x_1,x_2,x_3) &= \\ &\rho(x_0,x_1,x_2,x_3)^{1/4}\exp\bigl(\mathbf{X}(x_0,x_1,x_2,x_3)+\mathbf{F}(x_0,x_1,x_2,x_3)+\mathbf{V}(x_0,x_1,x_2,x_3)+\mathbf{B}(x_0,x_1,x_2,x_3)\bigr)\end{aligned} \tag{182}$$

where $(x_0,x_1,x_2,x_3)\in\mathcal{M}_4$.

In this case the wave-function dynamics are given by the complete Einstein field equations, and the wave-function assigns an instruction to transform the frame bundle at each point via a general linear transformation, and gives it a statistical weight (instead of merely an instruction to rotate the comoving frame of reference as in the case of electron kinematics above). Finally, the general linear wave-function can be extended to the Fock space via simple tensoring and direct sums, and thus the interpretation is able to account for multiple general linear frame bundles transformations at each point on $\mathcal{M}_4$, as well as superpositions thereof — while still easily connecting to a normalized probability measure via the application of the determinant as a generalization of the Born rule. The quantum theory works with general relativity as-is, without the need to linearize it or modify it in any way, or even importing it from elsewhere (as it is entailed from the theory directly), and gives a real valued-probability in 4D for all superpositions or combinations of gravitational fields. In our framework not only is gravity and quantum physics compatible; they are the same.

Testable predictions regarding general linear interference patterns are proposed in the Annex.

8.5 The General Linear Probability Amplitude in Other Dimensions

Let us investigate the application of this system to dimensions other than 4. In this section, we simply list various observations and do not reach any conclusions.

8.5.1 Zero-dimension case

0. In 0D, the "geometric algebra" is: $\psi = \exp A$, where A is a scalar. In our system, this is equivalent to classical probabilities.

Obviously, there is no geometry in a 0D system.

8.5.2 Odd-dimension cases

1. In 1D, the geometric algebra is: $\psi = \exp A + \mathbf{B}$, where A is a scalar and $\mathbf{B} = e_0 B = \mathbf{I}B \cong iB$ is a pseudo-scalar. In 1D, the multi-vector is a 1×1 matrix. The probability measure is given as the determinant $\det\big(a + ib\big) = a + ib$. This is not a real number, so naturally we flag the 1D case.

3. In 3D, the geometric algebra is $\psi = \exp A + \mathbf{X} + \mathbf{V} + \mathbf{B}$, where A is a scalar, $\mathbf{X}$ is a vector, $\mathbf{V}$ is a pseudo-vector and $\mathbf{B}$ is a pseudo-scalar. Here the multi-vector is a representation of the complex 2×2 matrices. Taking the determinant produces a complex value and not a real, so naturally we flag 3D for the same reason we flagged 1D.

2n+1 In (2n+1)D, the same here happens as the 1D and 3D cases. We flag all odd dimensions for the same reason: the determinant produces a complex value instead of a real.

For all odd-dimension cases, the probability in our system maps to a complex value instead of a real. We are not necessarily claiming that these are not relevant for physics, but if they are then one needs an explanation to account for complex probabilities (two probably measures required to describe the whole space... ?)

8.5.3 Even-dimension cases

2. In 2D, the geometric algebra is $\psi = \exp A + \mathbf{X} + \mathbf{B}$, where A is a scalar, $\mathbf{X}$ is a vector and $\mathbf{B}$ is a pseudo-scalar. The probability normalizes to a real value $\det \psi = \psi^\ddagger \psi$. As we did in the 4D case, let us now use spinors ($\mathbf{X} \to 0$). We get:

$$\psi^\ddagger \hat{\mathbf{x}} \psi = e^A e^{-\mathbf{B}} \hat{\mathbf{x}} e^A e^{\mathbf{B}} = e^{2A} e^{\mathbf{B}} \hat{\mathbf{x}} e^{-\mathbf{B}} \tag{183}$$

It supports a particle kinematics as we get one rotor: $e_\mu = e^{\mathbf{B}} \hat{\mathbf{x}}_\mu e^{-\mathbf{B}}$. The 2D theory also supports a gravity theory as it invariably admits the general linear group. In 2D, since the determinant is a polynomial of degree 2 of ψ then the QFT and the quantum gravity are the same as they are entailed from the same probability rule. This equivalence between QFT and quantum gravity is a feature unique to 2D.

4. In 4D, we have recovered in the previous section the kinematic relativistic theory of the electron that we are familiar with (or at least an interpretation –David Hestenes' interpretation– equivalent to it). But unlike the 2D case, the QFT is a sub-construction of the quantum gravity theory. Specifically, the familiar QFT comes out with these replacements:

$$\overline{\phi} = \rho^{\frac{1}{2}} e^{-ib} \widetilde{R} \gamma_0 \tag{184}$$

$$\phi = \rho^{\frac{1}{2}} e^{ib} R \tag{185}$$

Field theories over $\overline{\phi}$ and ϕ do not capture the full invariance group of the determinant. Trying to make quantum gravity fit on a $\overline{\phi}$ and ϕ frame, that is to say make quantum gravity a QFT, in 4D is bound to fail since the $\overline{\phi}$ and ϕ frame is a subconstruction of quantum gravity. Quantum gravity requires the 4 degree polynomial $\lfloor \psi^{\ddagger} \psi \rfloor_{3,4} \psi^{\ddagger} \psi$ to be defined in 4D.

6. In 6D, the wave-function is:

$$\psi = \exp A + \mathbf{X} + \mathbf{F} + \mathbf{T} + \mathbf{Q} + \mathbf{V} + \mathbf{B} \tag{186}$$

where A is a scalar, $\mathbf{X}$ is a vector, $\mathbf{F}$ is a bivector, $\mathbf{T}$ is a trivector, $\mathbf{Q}$ is a quadrivector, $\mathbf{V}$ is a pseudo-vector and finally $\mathbf{B}$ is a pseudo-scalar. Taking the even-sub-algebra, the spinor is:

$$\psi = \exp A + \mathbf{F} + \mathbf{Q} + \mathbf{B} \tag{187}$$

The wave-function will produce a probability measure (after simplifications), and the transformation of the comoving frame would be as follows:

$$e^{6A} e^{-\mathbf{F}} e^{-\mathbf{Q}} e^{-\mathbf{B}} \mathbf{x}_\mu e^{\mathbf{F}} e^{\mathbf{Q}} e^{\mathbf{B}} \tag{188}$$

Here we have more than just the rotors $e^{\mathbf{F}}$ as we also have the "spin-rotors" $e^{\mathbf{Q}}$. A typical QFT is extended to more dimensions than 4 by adding the rotations instructions over the extra dimensions. However, in 6D the pure geometric interpretation in terms of rotations is not complete, and the expected QFT construction requires the extra spin-rotation terms $e^{\mathbf{Q}}$.

2n In even dimensions ($2n > 6$) the same happens as the 6D case, but with even more extra terms for the spinors.

The constraints of our probability measure are such that the rotors required to produce a pure kinematical and geometric interpretation of the wave-function only show up by themselves in 2D and 4D, and this seems to suggest that a geometric interpretation purely in terms of a wave-function that assigns an instruction to rotate at each point in space-time, is only appropriate for these dimensions. In 6D or above, the formulation still admits structures that feel similar enough to a QFT; but such formulation contains additional terms that are not purely Lorentz rotation, thus making any geometric interpretation above 4D more challenging.

Even in 4D the geometric interpretation still contains the small increased challenge of a spin around the pseudo-scalar I with parameter B, whereas in the 2D version it is purely a rotor. Only the 2D version is therefore purely geometric and without spin.

9 Discussion

In 1543, Copernicus published *De revolutionibus orbium coelestium* in which he proposed the heliocentric model as a replacement for Ptolemy's geocentric model. The geocentric model required a plurality of "circles within circles" to be accurate, and was significantly more complicated than the heliocentric model. Copernicus' publication was mostly ignored by the Church until 1616 when it was added to the "Index"; a list of banned publications. Yet, interestingly and despite the ban, the Church did not object to the use of the theory by others provided it was presented as merely a mathematical device to simplify the calculations of planetary motions.

Here, we have presented our framework in four steps; each one leading into the next:

1. A Formal System of Knowledge
2. A Formal System of Science (used to map out knowledge)
3. A Formal Theory of the Observer (used to practice science)
4. A Formal Model of Physics (used to model what the observer can or cannot do as it practices science).

This breakdown is our attempt to present our theory as a typical physical theory of some physical object; in this case the observer. This breakdown is the result of trial and errors, asking for feedback and doing revisions to the presentation of the theory; gauging what people are receptive to and what they are not, to produce this explanation that we hope to be palatable to the majority.

However, although palatable, this breakdown does not provide the real reason this framework has been created, and neither does it address the main novelty. The peculiarity of our system is to be constructed as an *introspective* model of the observer. Although we presented our system abstractly as a model of the observer, it is better qualified as a model of "I" or "you" due to said introspective formulation. And, in this domain, having retained full mathematical rigour, comprehensiveness and relevance for physics, is quite notable. In fact, axiom 1 is as close as we could get to a mathematical definition of the observer as "I [do], therefore I am", while still retaining a minimal degree of generality and precision sufficient to entail fundamental physics. As we will now argue, it is in this context that a mathematical device that cures reality of its interpretative difficulties and paradoxes, is obtained.

9.1 Modelling The Self-Reflective Observer

Quite notably, the observer that we wish to model is also the one that creates the model.

This self-reflectivity effect induces significant constraints on the type of models that are satisfactory for a complete description of the observer. This effect is in fact the reason why it is insufficient to merely pick an axiomatic basis to model ourselves. Indeed, as we have the freedom to investigate any section of mathematical space, we can step outside any axiomatic basis by merely enumerating a section not covered by it. To successfully model ourselves, we must construct a model that we cannot step out of.

For instance, assume that the eventual and ultimate theory of everything, once and if discovered, will be formulated as a formal axiomatic system with either finitely many axioms, or perhaps via one or more axiom schema. Let us call this formal axiomatic system the ToE. This ToE would have a domain, Dom(ToE), that comprises the set of all sentences provable from it. Now, unless the ToE is able to prove every sentences (e.g. explosive inconsistency), then there will exist sentences that it cannot prove. Furthermore, as I presumably am a product of this ToE (i.e. I am an element of nature), then it follows that I should <u>not</u> be able to falsify it via my actions. However, I can! In fact, I can prove some (and perhaps all) of those "missing" sentences by picking an appropriate axiomatic basis (changing the context), or by constructing its Gödel sentence, G(ToE), and proving it within the extension ToE $\wedge$ Con(ToE) $\vdash$ G(ToE). Then, as per the fundamental theorem of science, as I have proved a sentence that cannot be proven from ToE, while demanding that I am a product of this ToE, I have hence falsified the ToE as the complete model of reality (i.e. I would be outside of it). Since I can apply this argument to any formal axiomatic systems, then it follows that any ToE, if constructed as a formal axiomatic system, is falsifiable by my actions. Consequently, no formal axiomatic system can completely model my actions, and to do so we must look at something else.

Does that mean we have to abandon mathematics? No! Consider instead that I am using the incremental contribution of mathematics we have introduced. In such a formulation, anything I think or do simply becomes an element of an incremental contribution, and is thus absorbed into the framework automatically. The sentence ToE $\wedge$ Con(ToE) $\vdash$ G(ToE) simply become a word of the lexicon, and the framework has resisted falsification.

> "Behind it all is surely an idea so simple, [...], that when we grasp it [...] we will all say to each other, how could it have been otherwise?"
>
> Wheeler, J. A., 1986, p. 304

That our system absorbs any possible action is the primary reason why it sustains an equivalence with what we understand as reality, and why in turn the laws of physics are self-reflectively entailed as the rules of what "I" or "you" can or cannot do from these actions.

Based on our results, we suggest that the primary aim of physics is not to explain, say, how the electron or the planets behave, as these are incidental to its main objective — that is, physics is not a theory extrinsic to the observer. Rather, physics is the model that describes my *experience* in nature as an observer. This experience admits a formal definition in terms of the production or receipt of a message of incremental contributions taken from a probability space over all experiments in nature, and we understand it as a model that maximizes experimenter freedom. Then, as the laws of physics are derived as the result of maximizing experimenter freedom in nature, I thus perceive the laws of physics as inviolable because they limit my freedom of action in nature.

Specifically, the fundamental expression of physics is given in three steps:

1. (Reformulation) Mathematics, re-formulated as an incremental contribution of knowledge, is automatically an epistemological model of the practice of science. It contains no informal physical or metaphysical language and no axioms (yet).

2. (Axiom) A recursive enumeration of knowledge is a message (i.e. Claude Shannon's theory of information) that describe the reality I see as a result of my observation or participation in it. Said message defines my experience of reality.

3. (Optimization) Maximizing the entropy of this message entails the fundamental physics as the maximally permissive model of said "experience". The measure and information which fixes the pick of the message from nature acquires automatically the probabilistic structure of the laws of physics (i.e. a wave-function and a Born rule) and the geometry is entailed simply as a consequence of the scope of this new larger Born rule.

Specifically (1) is an "introspective" and participatory formulation of mathematics (i.e. it centres around the production of an incremental contribution), (2) is the assumption of experimenter freedom (as the model of said participation) and (3) is the maximally permissive version of said participatory model. The observer is thus an introspectively formulated model of experimenter freedom in nature, and the laws of physics constitute its maximally permissive delimitation.

10 Conclusion

We believe the proposed definition of the observer as a probability space over all experiments in nature to be a more powerful formulation of fundamental physics as it is automatically entailed from a mere reformulation of mathematics sufficient to make it coherent and complete with respect to the experience of the observer — in other words; it is a self-evident, self-sufficient, and "philosophically safest" construction that automatically organizes itself into a model of fundamental physics. Physics was thus found as the product of an optimization problem on experimenter freedom.

References

[1] Max Tegmark. *Our mathematical universe: My quest for the ultimate nature of reality.* Vintage, 2014.

[2] John Conway and Simon Kochen. The free will theorem. *Foundations of Physics*, 36(10):1441–1473, 2006.

[3] Edmund L Gettier. Is justified true belief knowledge? *analysis*, 23(6):121–123, 1963.

[4] Richard L Kirkham. Does the gettier problem rest on a mistake? *Mind*, 93(372):501–513, 1984.

[5] Gregory J Chaitin. Meta math! the quest for omega. *arXiv preprint math/0404335*, 2004.

[6] G. J. Chaitin. Foundations of Mathematics. *ArXiv Mathematics e-prints*, February 2002.

[7] Gregory Chaitin. Mathematics, complexity and philosophy. `https://www.academia.edu/31320410/Mathematics_Complexity_and_Philosophy_full_bilingual_text_`, 2010.

[8] Wikipedia contributors. Curry–howard correspondence — Wikipedia, the free encyclopedia, 2021. [Online; accessed 6-December-2021].

[9] Ian Stewart et al. *From here to infinity.* Oxford Paperbacks, 1996.

[10] Gregory J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, July 1975.

[11] Klaus Hinkelmann and Oscar Kempthorne. *Design and analysis of experiments, volume 1: Introduction to experimental design*, volume 592. John Wiley & Sons, 2007.

[12] Douglas Lundholm and Lars Svensson. Clifford algebra, geometric algebra, and applications. *arXiv preprint arXiv:0907.5356*, 2009.

[13] Ryoyu Utiyama. Invariant theoretical interpretation of interaction. *Physical Review*, 101(5):1597, 1956.

[14] Tom WB Kibble. Lorentz invariance and the gravitational field. *Journal of mathematical physics*, 2(2):212–221, 1961.

[15] David Hestenes. Spacetime calculus for gravitation theory, 1996.

[16] David Hestenes. Spacetime physics with geometric algebra. *American Journal of Physics*, 71(7):691–714, 2003.

[17] Bohdan I Lev. Wave function as geometric entity. *Journal of Modern Physics*, 3:709–713, 2012.

A Notation

Sets, unless a prior convention assigns it another symbol, will be written using the blackboard bold typography (ex: $\mathbb{L}, \mathbb{W}, \mathbb{Q}$, etc.). Matrices will be in bold upper case (ex: $\mathbf{A}, \mathbf{B}$), whereas tuples, vectors and multi-vectors will be in bold lower case (ex: $\mathbf{u}, \mathbf{v}, \mathbf{g}$) and most other constructions (ex.: scalars, functions) will have plain typography (ex. a, A). The identity matrix is I, the unit pseudo-scalar (of geometric algebra) is $\mathbf{I}$ and the imaginary number is i. The Dirac gamma matrices are $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ and the Pauli matrices are $\sigma_x, \sigma_y, \sigma_z$. The basis elements of an arbitrary curvilinear geometric basis will be denoted $\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ (such that $\mathbf{e}_\nu \cdot \mathbf{e}_\mu = g_{\mu\nu}$) and if they are orthonormal as $\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_n$ (such that $\hat{\mathbf{x}}_\mu \cdot \hat{\mathbf{x}}_\nu = \eta_{\mu\nu}$). The asterisk z^* denotes the complex conjugate of z, and the dagger $\mathbf{A}^\dagger$ denotes the conjugate transpose of $\mathbf{A}$. A geometric algebra of m dimensions over a field $\mathbb{F}$ is noted as $\mathbb{G}(m, \mathbb{F})$. The grades of a multi-vector will be denoted as $\langle \mathbf{v} \rangle_k$. Specifically, $\langle \mathbf{v} \rangle_0$ is a scalar, $\langle \mathbf{v} \rangle_1$ is a vector, $\langle \mathbf{v} \rangle_2$ is a bi-vector, $\langle \mathbf{v} \rangle_{n-1}$ is a pseudo-vector and $\langle \mathbf{v} \rangle_n$ is a pseudo-scalar. Furthermore, a scalar and a vector $\langle \mathbf{v} \rangle_0 + \langle \mathbf{v} \rangle_1$ is a para-vector, and a combination of even grades $(\langle \mathbf{v} \rangle_0 + \langle \mathbf{v} \rangle_2 + \langle \mathbf{v} \rangle_4 + \ldots)$ or odd grades $(\langle \mathbf{v} \rangle_1 + \langle \mathbf{v} \rangle_3 + \ldots)$ are even-multi-vectors or odd-multi-vectors, respectively. The commutator is defined as $[\mathbf{A}, \mathbf{B}] := \mathbf{AB} - \mathbf{BA}$ and the anti-commutator as $\{\mathbf{A}, \mathbf{B}\} := \mathbf{AB} + \mathbf{BA}$. We use the symbol $\cong$ to relate two sets that are related by a group isomorphism. We use the symbol $\simeq$ to relate two expressions that are equal if defined, or both undefined otherwise. We denote the Hadamard product, or element-wise multiplication, of two matrices using $\odot$, and is written for instance as $\mathbf{M} \odot \mathbf{P}$, and for a multivector as $\mathbf{u} \odot \mathbf{v}$; for instance: $(a_0 + x_0\hat{\mathbf{x}} + y_0\hat{\mathbf{y}} + b_0\hat{\mathbf{x}} \wedge \hat{\mathbf{y}}) \odot (a_1 + x_1\hat{\mathbf{x}} + y_1\hat{\mathbf{y}} + b_0 1\hat{\mathbf{x}} \wedge \hat{\mathbf{y}})$ would equal $a_0 a_1 + x_0 x_1 \hat{\mathbf{x}} + y_0 y_1 \hat{\mathbf{y}} + b_0 b_1 \hat{\mathbf{x}} \wedge \hat{\mathbf{y}}$.

B Lagrange equation

The Lagrangian equation to maximize is:

$$\mathcal{L}(\rho, \alpha, \tau) = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q)\right) + \tau \operatorname{tr}\left(\overline{\mathbf{M}} - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}(q)\right) \tag{189}$$

where α and τ are the Lagrange multipliers. We note the usage of the trace operator for the geometric constraint such that a scalar-valued equation is maximized. Maximizing this equation for ρ by posing $\frac{\partial \mathcal{L}}{\partial \rho(p)} = 0$, where $p \in \mathbb{Q}$, we obtain:

$$\frac{\partial \mathcal{L}}{\partial \rho(p)} = -k_B \ln \rho(p) - k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p) \tag{190}$$

$$0 = k_B \ln \rho(p) + k_B + \alpha + \tau \operatorname{tr} \mathbf{M}(p) \tag{191}$$

$$\implies \ln \rho(p) = \frac{1}{k_B} \left(-k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p)\right) \tag{192}$$

$$\implies \rho(p) = \exp\left(\frac{-k_B - \alpha}{k_B}\right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(p)\right) \tag{193}$$

$$= \frac{1}{Z} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(p)\right) \tag{194}$$

where Z is obtained as follows:

$$1 = \sum_{q \in \mathbb{Q}} \exp\left(\frac{-k_B - \alpha}{k_B}\right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q)\right) \tag{195}$$

$$\implies \left(\exp\left(\frac{-k_B - \alpha}{k_B}\right)\right)^{-1} = \sum_{q \in \mathbb{Q}} \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q)\right) \tag{196}$$

$$Z := \sum_{q \in \mathbb{Q}} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(q)\right) \tag{197}$$

We note that the Trace in the exponential drops down to a determinant, via the relation $\det \exp A \equiv \exp \operatorname{tr} A$.

B.1 Multiple constraints

Consider a set of constraints:

$$\overline{\mathbf{M}}_1 = \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_1(q) \tag{198}$$

$$\vdots$$

$$\overline{\mathbf{M}}_n = \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_n(q) \tag{199}$$

Then the Lagrange equation becomes:

$$\mathcal{L} = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q)\right) + \tau_1 \operatorname{tr} \left(\overline{\mathbf{M}}_1 - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_1(q)\right) + \dots$$
$$+ \tau_n \operatorname{tr} \left(\overline{\mathbf{M}}_n - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_n(q)\right) \tag{200}$$

and the measure references all n constraints:

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{\tau_1}{k_B} \mathbf{M}_1(q) - \dots - \frac{\tau_n}{k_B} \mathbf{M}_n(q)\right) \tag{201}$$

B.2 Multiple constraints - General Case

In the general case of a multi-constraint system, each entry of the matrix corresponds to a constraint:

$$\overline{M}_{00} \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = \sum_{q \in \mathbb{Q}} \rho(q) M_{00}(q) \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \tag{202}$$

$$\vdots$$

$$\overline{M}_{01} \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \sum_{q \in \mathbb{Q}} \rho(q) M_{01}(q) \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \tag{203}$$

$$\vdots$$

$$\overline{M}_{nn} \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = \sum_{q \in \mathbb{Q}} \rho(q) M_{nn}(q) \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} \tag{204}$$

For a $n \times n$ matrix, there are n^2 constraints.
The probability measure which maximizes the entropy is as follows:

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{1}{k_B} \boldsymbol{\tau} \odot \mathbf{M}(q)\right) \tag{205}$$

where $\boldsymbol{\tau}$ is a matrix of Lagrange multipliers, and $\odot$, the element-wise multiplication, assigns the corresponding Lagrange multiplier to each constraint.

C Formal Proof (Example)

The following program[8] is a formal proof of the commutativity of addition for natural numbers written in COQ:

```
plus_comm =
fun n m : nat =>
nat_ind (fun n0 : nat => n0 + m = m + n0)
  (plus_n_O m)
  (fun (y : nat) (H : y + m = m + y) =>
   eq_ind (S (m + y))
     (fun n0 : nat => S (y + m) = n0)
     (f_equal S H)
     (m + S y)
     (plus_n_Sm m y)) n
     : forall n m : nat, n + m = m + n
```

D A Step Towards Testable Predictions (Spacetime interference)

Certain transformations of the wave-function in quantum gravity, under the general linear group or some of its subgroups, would produce richer interference patterns than what is possible merely with complex interference in standard QFT. This offer a difference in predictions between ordinary QFT and our system, that can be used to test our system. The possibility of interference patterns resulting from geometric algebra representation of the wave-function has been proposed before; specifically, I note the work of B. I. Lev.[17] which suggests (theoretically) the possibility of an interference pattern associated with the David Hestenes form of the relativistic wave-function and for the subset of rotors.

Here we derive a number of these possible interference patterns.

In the case of the general linear group, the interference pattern is much more complicated than the simple cosine of the standard Born rule, but that is to be expected as it comprises the full general linear group and not just the unitary group. It accounts for the group of all geometric transformations which preserves the probability distribution ρ for a two-state general linear system.

General linear interference can be understood as a generalization of complex interference, which is recovered under a "shallow" phase rotation in 4D and under just a plain normal phase rotation in 2D. Furthermore, when all elements of the odd-sub-algebra are eliminated (by posing $\mathbf{X} \to 0$, $\mathbf{V} \to 0$), then the wave-function reduces to the geometric algebra form of the relativistic wave-function identified by David Hestenes, in terms of a spinor field.

Such reductions entails a series of interference patterns of decreasing complexity, and as such they provide a method to experimentally identify which group of geometric transformations physical reality allows in the most general case of quantum gravity, using interference experiments as the identification tool. Identification of the full general linear interference pattern (with all the elements $A, \mathbf{X}, \mathbf{F}, \mathbf{V}, \mathbf{B}$) in a lab experiment would suggest a general linear gauge, whereas identification of a reduced interference pattern (produced by $A, \mathbf{F}, \mathbf{B}$)

and subsequently showing a failure to observe the full general linear interference ($\mathbf{X} \to 0, \mathbf{V} \to 0$) would suggest the Lorentz gauge instead of full quantum gravity.

Let us start by introducing a notation for a dot product, then we will list the various possible interference patterns.

D.1 Geometric Algebra Dot Product

Let us introduce a notation. We will define a bilinear form using the dot product notation, as follows:

$$\begin{array}{rrcl} \cdot \quad : & \mathbb{G}(2n,\mathbb{R}) \times \mathbb{G}(2n,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & \mathbf{u} \cdot \mathbf{v} & \longmapsto & \frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) \end{array} \tag{206}$$

For example,

$$\mathbf{u} = A_1 + X_1 e_1 + Y_1 e_2 + B_1 e_{12} \tag{207}$$
$$\mathbf{v} = A_2 + X_2 e_1 + Y_2 e_2 + B_2 e_{12} \tag{208}$$
$$\implies \mathbf{u} \cdot \mathbf{v} = A_1 A_2 + B_1 B_2 - X_1 X_2 - Y_1 Y_2 \tag{209}$$

Iff $\det \mathbf{u} > 0$ and $\det \mathbf{v} > 0$ then $\mathbf{u} \cdot \mathbf{v}$ is always positive, and therefore qualifies as a positive inner product (over the positive determinant group), but no greater than either $\det \mathbf{u}$ or $\det \mathbf{v}$, whichever is larger. This definition of the dot product extends to multi-vectors of 4 dimensions.

2D: In 2D the dot product is equivalent to this form:

$$\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) = \frac{1}{2}\left((\mathbf{u}+\mathbf{v})^\ddagger(\mathbf{u}+\mathbf{v}) - \mathbf{u}^\ddagger\mathbf{u} - \mathbf{v}^\ddagger\mathbf{v}\right) \tag{210}$$
$$= \mathbf{u}^\ddagger\mathbf{u} + \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} + \mathbf{v}^\ddagger\mathbf{v} - \mathbf{u}^\ddagger\mathbf{u} - \mathbf{v}^\ddagger\mathbf{v} \tag{211}$$
$$= \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} \tag{212}$$

4D: In 4D it is substantially more verbose:

$$
\begin{aligned}
&\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det\mathbf{u} - \det\mathbf{v}) && (213)\\
&= \frac{1}{2}\left(\lfloor(\mathbf{u}+\mathbf{v})^{\ddagger}(\mathbf{u}+\mathbf{v})\rfloor_{3,4}(\mathbf{u}+\mathbf{v})^{\ddagger}(\mathbf{u}+\mathbf{v}) - \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} - \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\right) && (214)\\
&= \frac{1}{2}\left(\lfloor\mathbf{u}^{\ddagger}\mathbf{u}+\mathbf{u}^{\ddagger}\mathbf{v}+\mathbf{v}^{\ddagger}\mathbf{u}+\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}(\mathbf{u}^{\ddagger}\mathbf{u}+\mathbf{u}^{\ddagger}\mathbf{v}+\mathbf{v}^{\ddagger}\mathbf{u}+\mathbf{v}^{\ddagger}\mathbf{v}) - \dots\right) && (215)\\
&= \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\\
&\quad + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\\
&\quad + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\\
&\quad + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} - \dots && (216)\\
&= \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\\
&\quad + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\\
&\quad + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\\
&\quad + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} && (217)
\end{aligned}
$$

D.2 Geometric Interference (General Form)

A multi-vector can be written as $\mathbf{u} = a + \mathbf{s}$, where a is a scalar and $\mathbf{s}$ is the multi-vectorial part. In general, the exponential $\exp \mathbf{u}$ equals $\exp a \exp \mathbf{s}$ because a commutes with $\mathbf{s}$.

One can thus write a general two-state system as follows:

$$\psi = \psi_1 + \psi_2 = e^{A_1} e^{\mathbf{S}_1} + e^{A_2} e^{\mathbf{S}_2} \tag{218}$$

(219)

The general interference pattern will be of the following form:

$$
\begin{aligned}
\det \psi_1 + \psi_2 &= \det \psi_1 + \det \psi_2 + \psi_1 \cdot \psi_2 && (220)\\
&= e^{nA_1} + e^{nA_2} + \psi_1 \cdot \psi_2 && (221)
\end{aligned}
$$

where $\det \psi_1 + \det \psi_2$ is a sum of probabilities and where $\psi_1 \cdot \psi_2$ is the interference pattern, and where n is the number of dimensions of the geometric algebra.

D.3 Complex Interference (Recall)

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} \tag{222}$$

The interference pattern familiar to quantum mechanics is the result of the complex norm:

$$
\begin{aligned}
\psi^\dagger\psi &= \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 && (223)\\
&= e^{A_1}e^{-\mathbf{B}_1}e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2}e^{A_2}e^{\mathbf{B}_2} + e^{A_1}e^{-\mathbf{B}_1}e^{A_2}e^{\mathbf{B}_2} + e^{A_2}e^{-\mathbf{B}_2}e^{A_1}e^{\mathbf{B}_1} && (224)\\
&= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}(e^{-\mathbf{B}_1+\mathbf{B}_2} + e^{-(-\mathbf{B}_1+\mathbf{B}_2)}) && (225)\\
&= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2}\cos(B_1 - B_2)}_{\text{interference}} && (226)
\end{aligned}
$$

D.4 Geometric Interference in 2D

Consider a two-state wave-function:

$$
\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{X}_1+\mathbf{B}_1} + e^{A_2}e^{\mathbf{X}_2+\mathbf{B}_2} \tag{227}
$$

To lighten the notation we will write it as follows:

$$
\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{S}_1} + e^{A_2}e^{\mathbf{S}_2} \tag{228}
$$

where

$$
\mathbf{S} = \mathbf{X} + \mathbf{B} \tag{229}
$$

The interference pattern for a full general linear transformation on a two-state wave-function in 2D is:

$$
\begin{aligned}
\psi^\dagger\psi &= \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 && (230)\\
&= e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_1}e^{\mathbf{S}_1} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_1}e^{\mathbf{S}_1} && (231)\\
&= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}((e^{\mathbf{S}_1})^\ddagger e^{\mathbf{S}_2} + (e^{\mathbf{S}_2})^\ddagger e^{\mathbf{S}_1}) && (232)\\
&= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{e^{A_1+A_2}(e^{-\mathbf{X}_1-\mathbf{B}_1}e^{\mathbf{X}_2+\mathbf{B}_2} + e^{-\mathbf{X}_2-\mathbf{B}_2}e^{\mathbf{X}_1+\mathbf{B}_1})}_{\text{interference}} && (233)
\end{aligned}
$$

D.5 Geometric Interference in 4D

Consider a two-state wave-function:

$$
\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{X}_1+\mathbf{F}_1+\mathbf{V}_1+\mathbf{B}_1} + e^{A_2}e^{\mathbf{X}_2+\mathbf{F}_2+\mathbf{V}_2+\mathbf{B}_2} \tag{234}
$$

To lighten the notation we will write it as follows:

$$
\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{S}_1} + e^{A_2}e^{\mathbf{S}_2} \tag{235}
$$

where

$$\mathbf{S} = \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{236}$$

The geometric interference patterns for a full general linear transformation in 4D is given by the product:

$$\begin{aligned}
\lfloor \psi^{\ddagger}\psi \rfloor_{3,4}\psi^{\ddagger}\psi &= \lfloor \psi_1^{\ddagger}\psi_1 \rfloor_{3,4}\psi_1^{\ddagger}\psi_1 + \lfloor \psi_2^{\ddagger}\psi_2 \rfloor_{3,4}\psi_2^{\ddagger}\psi_2 + \psi_1 \cdot \psi_2 && (237)\\
&= e^{4A_1} + e^{4A_2} + \left(e^{A_1}e^{\mathbf{S}_1}\right) \cdot \left(e^{A_2}e^{\mathbf{S}_2}\right) && (238)
\end{aligned}$$

In many cases of interest, the pattern simplifies. Let us see some of these cases now.

D.6 Geometric Interference in 4D (Shallow Phase Rotation)

If we consider a sub-algebra in 4D comprised of even-multi-vector products $\psi^{\ddagger}\psi$, then a two-state system is given as:

$$\psi = \psi_1 + \psi_2 \tag{239}$$

where

$$\begin{aligned}
\psi_1 &= (e^{A_1}e^{\mathbf{F}_1}e^{\mathbf{B}_1})^{\ddagger}(e^{A_1}e^{\mathbf{F}_1}e^{\mathbf{B}_1}) = e^{2A_1}e^{2\mathbf{B}_1} && (240)\\
\psi_2 &= (e^{A_2}e^{\mathbf{F}_2}e^{\mathbf{B}_2})^{\ddagger}(e^{A_2}e^{\mathbf{F}_2}e^{\mathbf{B}_2}) = e^{2A_2}e^{2\mathbf{B}_2} && (241)
\end{aligned}$$

Thus

$$\psi = e^{2A_1}e^{2\mathbf{B}_1} + e^{2A_2}e^{2\mathbf{B}_2} \tag{242}$$

The quadri-linear map becomes a bilinear map:

$$\begin{aligned}
\psi^{\dagger}\psi &= (e^{2A_1}e^{-2\mathbf{B}_1} + e^{2A_2}e^{-2\mathbf{B}_2})(e^{2A_1}e^{2\mathbf{B}_1} + e^{2A_2}e^{2\mathbf{B}_2}) && (243)\\
&= e^{2A_1}e^{-2\mathbf{B}_1}e^{2A_1}e^{2\mathbf{B}_1} + e^{2A_1}e^{-2\mathbf{B}_1}e^{2A_2}e^{2\mathbf{B}_2} + e^{2A_2}e^{-2\mathbf{B}_2}e^{2A_1}e^{2\mathbf{B}_1} + e^{2A_2}e^{-2\mathbf{B}_2}e^{2A_2}e^{2\mathbf{B}_2} && (244)\\
&= \underbrace{e^{4A_1} + e^{4A_2}}_{\text{sum}} + \underbrace{2e^{2A_1+2A_2}\cos(2B_1 - 2B_2)}_{\text{complex interference}} && (245)
\end{aligned}$$

D.7 Geometric Interference in 4D (Deep Phase Rotation)

A phase rotation on the base algebra (rather than the sub-algebra) produces a difference interference pattern. Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} \tag{246}$$

The sub-product part is:

$$\begin{aligned}
\psi^\ddagger \psi &= (e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2})(e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2}) && (247)\\
&= e^{A_1} e^{\mathbf{B}_1} e^{A_1} e^{\mathbf{B}_1} + e^{A_1} e^{\mathbf{B}_1} e^{A_2} e^{\mathbf{B}_2} + e^{A_2} e^{\mathbf{B}_2} e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} e^{A_2} e^{\mathbf{B}_2} && (248)\\
&= e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_2} e^{2\mathbf{B}_2} + 2e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} && (249)
\end{aligned}$$

The final product is:

$$\begin{aligned}
\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi &= (e^{2A_1} e^{-2\mathbf{B}_1} + e^{2A_2} e^{-2\mathbf{B}_2} + 2e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2})\\
&\quad \times (e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_2} e^{2\mathbf{B}_2} + 2e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2}) && (250)\\
&= e^{2A_1} e^{-2\mathbf{B}_1} e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_1} e^{-2\mathbf{B}_1} e^{2A_2} e^{2\mathbf{B}_2} + e^{2A_1} e^{-2\mathbf{B}_1} 2e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2}\\
&\quad + e^{2A_2} e^{-2\mathbf{B}_2} e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_2} e^{-2\mathbf{B}_2} e^{2A_2} e^{2\mathbf{B}_2} + e^{2A_2} e^{-2\mathbf{B}_2} 2e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2}\\
&\quad + 2e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} e^{2A_1} e^{2\mathbf{B}_1}\\
&\qquad + 2e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} e^{2A_2} e^{2\mathbf{B}_2}\\
&\qquad + 2e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} 2e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} && (251)\\
&= e^{4A_1} + e^{4A_2} + 2e^{2A_1+2A_2} \cos(2B_1 - 2B_2)\\
&\quad + e^{2A_1} e^{-2\mathbf{B}_1} 2e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2}\\
&\quad + e^{2A_2} e^{-2\mathbf{B}_2} 2e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2}\\
&\quad + 2e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} e^{2A_1} e^{2\mathbf{B}_1}\\
&\quad + 2e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} e^{2A_2} e^{2\mathbf{B}_2}\\
&\quad + 4e^{2A_1+2A_2} && (252)\\
&= \underbrace{e^{4A_1} + e^{4A_2}}_{\text{sum}} + \underbrace{2e^{2A_1+2A_2} \cos(2B_1 - 2B_2)}_{\text{complex interference}}\\
&\quad \underbrace{+2e^{A_1+A_2}(e^{2A_1} + e^{2A_2}) \cos(B_1 - B_2) + 4e^{2A_1+2A_2}}_{\text{deep phase interference}} && (253)
\end{aligned}$$

D.8 Geometric Interference in 4D (Deep Spinor Rotation)

Consider a two-state wave-function (we note that $[\mathbf{F}, \mathbf{B}] = 0$):

$$\psi = \psi_1 + \psi_2 = e^{A_1} e^{\mathbf{F}_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{F}_2} e^{\mathbf{B}_2} \tag{254}$$

The geometric interference pattern for a full general linear transformation in 4D is given by the product:

$$\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi \tag{255}$$

Let us start with the sub-product:

$$\begin{aligned}
\psi^\ddagger \psi &= (e^{A_1} e^{-\mathbf{F}_1} e^{\mathbf{B}_1} + e^{A_2} e^{-\mathbf{F}_2} e^{\mathbf{B}_2})(e^{A_1} e^{\mathbf{F}_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{F}_2} e^{\mathbf{B}_2}) && (256) \\
&= e^{A_1} e^{-\mathbf{F}_1} e^{\mathbf{B}_1} e^{A_1} e^{\mathbf{F}_1} e^{\mathbf{B}_1} + e^{A_1} e^{-\mathbf{F}_1} e^{\mathbf{B}_1} e^{A_2} e^{\mathbf{F}_2} e^{\mathbf{B}_2} \\
&\quad + e^{A_2} e^{-\mathbf{F}_2} e^{\mathbf{B}_2} e^{A_1} e^{\mathbf{F}_1} e^{\mathbf{B}_1} + e^{A_2} e^{-\mathbf{F}_2} e^{\mathbf{B}_2} e^{A_2} e^{\mathbf{F}_2} e^{\mathbf{B}_2} && (257) \\
&= e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_2} e^{2\mathbf{B}_2} + e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} (e^{-\mathbf{F}_1} e^{\mathbf{F}_2} + e^{-\mathbf{F}_2} e^{\mathbf{F}_1}) && (258) \\
&= e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_2} e^{2\mathbf{B}_2} + e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) && (259)
\end{aligned}$$

where $R = e^{\mathbf{F}}$, and where $\tilde{R} = e^{-\mathbf{F}}$.
The full product is:

$$\begin{aligned}
\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi &= \Big(e^{2A_1} e^{-2\mathbf{B}_1} + e^{2A_2} e^{-2\mathbf{B}_2} + e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)\Big) \\
&\quad \times \Big(e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_2} e^{2\mathbf{B}_2} + e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)\Big) \\
& && (260) \\
&= e^{2A_1} e^{-2\mathbf{B}_1} e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_1} e^{-2\mathbf{B}_1} e^{2A_2} e^{2\mathbf{B}_2} + e^{2A_1} e^{-2\mathbf{B}_1} e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \\
&\quad + e^{2A_2} e^{-2\mathbf{B}_2} e^{2A_1} e^{2\mathbf{B}_1} + e^{2A_2} e^{-2\mathbf{B}_2} e^{2A_2} e^{2\mathbf{B}_2} + e^{2A_2} e^{-2\mathbf{B}_2} e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \\
&\quad + e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{2A_1} e^{2\mathbf{B}_1} \\
&\qquad + e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{2A_2} e^{2\mathbf{B}_2} \\
&\qquad + e^{A_1+A_2} e^{-\mathbf{B}_1-\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{A_1+A_2} e^{\mathbf{B}_1+\mathbf{B}_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \\
& && (261) \\
&= e^{4A_1} + e^{4A_2} + 2e^{2A_1+2A_2} \cos(2B_1 - 2B_2) && (262) \\
&\quad + e^{A_1+A_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)(&& (263) \\
&\qquad e^{2A_1} (e^{-\mathbf{B}_1+\mathbf{B}_2} + e^{\mathbf{B}_1-\mathbf{B}_2}) && (264) \\
&\qquad + e^{2A_2} (e^{\mathbf{B}_1-\mathbf{B}_2} + e^{-\mathbf{B}_1+\mathbf{B}_2})) && (265) \\
&\quad + e^{2A_1+2A_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)^2 && (266) \\
&= \underbrace{e^{4A_1} + e^{4A_2}}_{\text{sum}} + \underbrace{2e^{2A_1+2A_2} \cos(2B_1 - 2B_2)}_{\text{complex interference}} \\
&\quad + \underbrace{2e^{A_1+A_2} (e^{2A_1} + e^{2A_2})(\tilde{R}_1 R_2 + \tilde{R}_2 R_1)(\cos(B_1 - B_2)) + e^{2A_1+2A_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)^2}_{\text{deep spinor interference}} \\
& && (267)
\end{aligned}$$