A Mathematical Proof of Physics, Obtained by Formalizing the Scientific Method

Alexandre Harvey-Tremblay[1]

[1]Independent scientist, aht@protonmail.ch

June 18, 2021

Abstract

It is generally expected that the laws of nature are obtained as the end-product of the scientific process. In this paper, consistently with said expectation, I produce a model of science using mathematics, then I use it to derive the laws of nature by applying the (formalized) scientific method to the model. Modern notions relating to mathematical undecidability are utilized to create a 'trial and error' foundation to the discovery of new mathematical truths, such that one is required to run programs to completion —essentially to perform 'mathematical experiments'— to acquire knowledge about mathematics. The 'laws of nature' are then derived as the probability measure that maximizes the quantity of information produced by the scientific method as it traces a path in the space of all possible (mathematical) experiments. In this model, said laws have the same probabilistic structure and domain (the set of all experiments) as the modern laws of physics, and quite remarkably, the two are identical. Since the definitions start at the level of science and experiments, are purely mathematical, yet are nonetheless able to derive the laws of nature and do so from first principles, we argue that the present derivation of said laws, as it is ultimately the product of the (formalized) scientific method, is a plausible, conceptually minimal and purely mathematical foundation to the laws of physics. We end with applications of the model to fundamental open problems of physics and produce testable predictions.

Contents

1 Introduction **3**
- 1.1 Universal Facts . . . 4
- 1.2 The Mathematics of Knowledge . . . 5
  - 1.2.1 Connection to Formal Axiomatic Systems . . . 6
  - 1.2.2 Axiomatic Information . . . 7
- 1.3 Discussion — The Mathematics of Knowledge . . . 7
  - 1.3.1 Special Cases, Inconsistencies, etc. . . . 8

2 Formal Science **9**
2.1 Axiomatic Foundation of Science . . . 9
2.1.1 The Fundamental Theorem of Science . . . 12
2.2 Axiomatic Foundation of Reality . . . 12
2.2.1 The Fundamental Assumption of Science . . . 14
2.3 Axiomatic Foundation of Nature . . . 14
2.3.1 Laws of Nature . . . 15

3 Main Result **16**
3.1 Overview . . . 16
3.1.1 Halting Probability Ω . . . 16
3.1.2 Quantum Computing . . . 18
3.1.3 Program-Steps and Programs are indistinguishable . . . 18
3.1.4 Manifest-to-Manifest Computing... ? . . . 19
3.2 Derivation . . . 19
3.2.1 General Linear Ensemble . . . 19
3.2.2 Prior . . . 21

4 Foundation **21**
4.1 Overview . . . 21
4.1.1 Matrix-Valued Vector and Transformations . . . 21
4.1.2 Linear Transformations as Computations . . . 22
4.1.3 General Linear Group . . . 23
4.2 Algebra of Natural States, in 2D . . . 23
4.2.1 Geometric Representation of 2×2 matrices . . . 23
4.2.2 Axiomatic Definition of the Algebra, in 2D . . . 25
4.2.3 Reduction to Complex Hilbert Spaces . . . 26
4.2.4 Observable, in 2D — Self-Adjoint Operator . . . 26
4.2.5 Observable, in 2D — Eigenvalues / Spectral Theorem . . . 27
4.3 Algebra of Natural States, in 3D (brief) . . . 28
4.3.1 Axiomatic Definition of the Algebra, in 3D . . . 29
4.3.2 Reduction to Complex Hilbert Spaces . . . 30
4.4 Algebra of Natural States, in 4D . . . 30
4.4.1 Geometric Representation (in 4D) . . . 30
4.4.2 Axiomatic Definition of the Algebra, in 4D . . . 31
4.4.3 Reduction to Complex Hilbert Space . . . 32

5 Applications (Physics) **33**
5.1 Paths in $\mathbb{E}$ are Path Integrals . . . 33
5.2 Gauge Completeness . . . 35
5.2.1 General Linear Invariance . . . 35
5.2.2 Lagrangian . . . 36
5.2.3 Example Gauge . . . 37

6 Testable Prediction **38**
6.1 Geometric Interference 39
6.1.1 Geometric Algebra Dot Product 39
6.1.2 Geometric Interference (General Form) 40
6.1.3 Complex Interference (Recall) 40
6.1.4 Geometric Interference in 2D 41
6.1.5 Geometric Interference in 4D 41
6.1.6 Geometric Interference in 4D (Shallow Phase Rotation) . 42
6.1.7 Geometric Interference in 4D (Deep Phase Rotation) . . . 42
6.1.8 Geometric Interference in 4D (Deep Spinor Rotation) . . 43
6.1.9 Geometric Interference Experiment (Sketch) 45

7 Discussion (Science + Physics) **46**
7.1 Observer Bias — An Origin Story 46
7.2 Interpretation of Quantum Mechanics 48
7.2.1 Interpretation of the Collapse 49

A Notation **54**

B Lagrange equation **54**
B.1 Multiple constraints 55
B.2 Multiple constraints - General Case 56

1 Introduction

In classical philosophy an axiom is a statement which is self-evidently true such that it is accepted without controversy or question. But this definition has been retired in modern usage. Any so-called "self-evident" axiom can also be posited to be false and either choice of its truth-value yields a different model; the archetypal example being the parallel line postulate of Euclid, allowing for hyperbolic/spherical geometry when it is false. Consequently, in modern logic an axiom is simply a starting point for a premise, and in mathematics an axiom is a sentence of a language that is held to be true by definition.

A long standing goal of philosophy has been to find necessarily true principles that could be used as the basis of knowledge. For instance, the universal doubt method of Descartes had such a goal in mind. The 'justified true belief' theory of epistemology is another attempt with a similar goal. But, so far, all such attempts have flaws and loop-holes; the elimination of which is assumed, at best, to reduce the theory to a handful of statements, rendering it undesirable as a foundation to all knowledge.

In epistemology, the Gettier problem[1] is a well known objection to the belief that knowledge is that which is both true and justified, relating to a family of counter-examples. All such counter-examples rely on the same loop-hole: if the justification is not 'air-tight' then 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 not knowledge because it is only true by pure luck.

Richard Kirkham[2] proposed to add the criteria of infallibility to the justification. This eliminates the loop-hole, but it is an unpopular solution because adding it is assumed to reduce epistemology to radical skepticism in which almost nothing is knowledge.

Since the primary purpose of a scientific process is to gather knowledge (about the world), then any serious attempt at the formalization of such will require a theory of knowledge that is also equally rigorous. Here, we will propose the concept of the universal fact as a new candidate to serve as the foundation to knowledge. As we will see in a moment, and due to their construction, universal facts are sufficiently strong to be infallible, yet have sufficient expressive power to form a Turing complete theory thus they resolve the Gettier problem without reducing epistemology. Universal facts will be the primary subject matter of our mathematical model of science and they are revealed and verified by the (formalized) scientific method.

1.1 Universal Facts

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, I will provide what I believe to be the correct 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 (*universal* in the computer-theoretical sense). I will use the term *universal fact* to refer to the concept.

Specifically, the sentence "$1 + 1 = 2$" halts on some Turing machine, but not on others and thus is not a universal fact. 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, a trickster could claim binary numbers, in which case $1+1 = 10$), but if one specifies the exact axiomatic basis in which the claim is provable, a trickster would find it harder to find a loop-hole to fail the claim. Nonetheless, even with this improvement, a trickster can fail the claim by providing a Turing machine for which $\text{PA} \vdash [s(0) + s(0) = s(s(0))]$ does not halt.

If we use the tools of theoretical computer science and observe an equivalence with programs, we can produce statements free of all loop-holes, thus ensuring they are infallible:

Definition 1 (Universal Fact)**.** *Let $\mathbb{L}$ be the set of all sentences with alphabet Σ. A universal fact is a pair* (TM, p) *of sentences from* $\mathbb{L} \times \mathbb{L}$ *such that a universal Turing machine* UTM *halts for it:*

$$\textit{iff } \mathrm{UTM}(\mathrm{TM}, p) \textit{ halts, then } (\mathrm{TM}, p) \textit{ is a universal fact} \tag{1}$$

A universal Turing machine UTM which takes a Turing machine TM and a sentence p as inputs, will halt if and only if p halts on TM. Thus a claim that p halts on TM, if true, is a universal fact because it is verifiable on all universal Turing machines.

The second objection is that infallible justified true beliefs collapse epistemology to radical skepticism, where at best only a handful of statements constitute knowledge. However, the set of all universal facts constitute the entire domain of the universal Turing machine, and thus the expressive power of universal facts 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.

1.2 The Mathematics of Knowledge

We can use universal facts to re-define the foundations of mathematics to be *knowledge-based*. When it comes to formulating a model of science whose goal is to acquire more knowledge, one can intuit why that would be a desirable reconstruction. A knowledge-based foundation further works well with theories having only finitely-many theorems, whereas working with such theories using the typical tools of mathematics is mostly ineffective, because all such theories are decidable and thus completely solvable in principle. Furthermore, even tools such as complexity theory require the size of the input to be n, allowing for arbitrarily large sizes of input to produce an effective classification system. Instead of defining a mathematical theory as a finite deductive system of axioms, which typically entails infinitely many theorems, let us define it as a finite (or in some cases even infinite) set of universal facts.

Definition 2 (Knowledge base)**.** *A knowledge base* $\mathbb{KB}$ *is defined as a set of universal facts:*

$$\mathbb{KB} := \{(\mathrm{TM}_1, p_1), \ldots, (\mathrm{TM}_n, p_n)\} \tag{2}$$

The set, in principle, can be empty ($\mathbb{KB} := \{\}$)*, finite* ($n \in \mathbb{N}$) *or countably infinite* ($n = \infty$)*, but, as we will see, finite non-empty sets will be more interesting for us.*

For a knowledge base, universal facts replace the normal role of both axioms and theorems and instead form a single verifiable atomic concept constituting a unit of mathematical knowledge. Let me explicitly point out the difference between the literature definition of a formal theory and ours: for the former, its theorems are a subset of the sentences of $\mathbb{L}$ — whereas for a knowledge base, its elements are pairs of $\mathbb{L} \times \mathbb{L}$ which halts on a UTM.

Note on the upcoming notation: we will designate k as elements of $\mathbb{KB}$, and $\pi_1(k)$ and $\pi_2(k)$ designate the first and second projection of the pair k, respectively. Thus $\pi_1(k)$ is the TM associated with k, and $\pi_2(k)$ is the input p associated with k. If applied to a set of pairs, then $\pi_1(\mathbb{KB})$ return the set of all p in $\mathbb{KB}$ and $\pi_2(\mathbb{KB})$ returns the set of all TM in $\mathbb{KB}$.

Theorem 1 (Incompleteness Theorem). *Let $\mathbb{KB}$ be a knowledge base. If $\mathbb{KB} = \mathrm{Dom(UTM)}$, then $\mathbb{KB}$ is recursively enumerable (and non-decidable). The proof follows from the domain of a universal Turing machine being non-decidable.*

Definition 3 (Theorems). *The theorems of a knowledge base $\mathbb{KB}$ are defined as the set of all p in $\mathbb{M}$:*

$$\mathbb{T} := \pi_2(\mathbb{KB}) \tag{3}$$

Definition 4 (Atomic Enumerator). *The atomic enumerators of $\mathbb{KB}$ are defined as the set of all* TM *in $\mathbb{KB}$:*

$$\mathbb{A} := \pi_1(\mathbb{KB}) \tag{4}$$

Definition 5 (Spread (of a theorem)). *The set of all atomic enumerators in $\mathbb{KB}$ in which a theorem is repeated is called the spread of the theorem. For instance if $\mathbb{KB} = \{(\mathrm{TM}_1, p_1), (\mathrm{TM}_2, p_1)\}$, then the spread of p_1 is $\{\mathrm{TM}_1, \mathrm{TM}_2\}$.*

Definition 6 (Scope (of an enumerator)). *The set of all theorems in $\mathbb{KB}$ in which an enumerator is repeated is called the scope of the enumerator. For instance if $\mathbb{KB} = \{(\mathrm{TM}_1, p_1), (\mathrm{TM}_1, p_2)\}$, then the scope of TM_1 is $\{p_1, p_2\}$.*

1.2.1 Connection to Formal Axiomatic Systems

We can, of course, connect our construction to a formal axiomatic system:

Definition 7 (Formal Axiomatic Representation). *Let* FAS *be a formal axiomatic system, let $\mathbb{KB}$ be a knowledge base and let* $\mathrm{enumerator}_{\mathrm{FAS}}$ *be a function which recursively enumerates the theorems of* FAS. *Then* FAS *is a formal axiomatic representation of $\mathbb{KB}$ iff:*

$$\forall (s_1, s_2) \in \mathbb{L} \times \mathbb{L} \quad [\mathrm{enumerator}_{\mathrm{FAS}}(s_1, s_2) \textit{ halts iff } (s_1, s_2) \in \mathbb{KB}] \tag{5}$$

Definition 8 (Domain (of FAS)). *Let* FAS *be a formal axiomatic system, let $\mathbb{KB}$ be a knowledge base and let* $\mathrm{enumerator}_{\mathrm{FAS}}$ *be a function which recursively enumerates the theorems of* FAS. *Then the domain of* FAS, *denoted as* Dom(FAS), *is the set of all pairs $(s_1, s_2) \in \mathbb{L} \times \mathbb{L}$ which halts for* $\mathrm{enumerator}_{\mathrm{FAS}}$.

Definition 9 (Factual Isomorphism). *Two formal axiomatic systems* FAS_1 *and* FAS_2 *are factually-isomorphic if and only if* $\mathrm{Dom}(\mathrm{FAS}_1) = \mathrm{Dom}(\mathrm{FAS}_2)$.

Theorem 2 (Principle of Computational Equivalence[3])**.** *Let* $\mathbb{KB}$ *be a knowledge base. If* $\mathbb{KB} = \mathrm{Dom}(\mathrm{UTM})$ *then all Turing complete formal axiomatic systems are factually-isomorphic representations of* $\mathbb{KB}$*. Furthermore, their enumerator function is a universal Turing machine, under an appropriate re-encoding of inputs. The proof follows because* $\mathrm{Dom}(\mathrm{UTM})$ *includes all universal facts.*

1.2.2 Axiomatic Information

Although we *can* connect the formulation of a knowledge base to a formal axiomatic representation, we will find that it is more advantageous for the purposes of constructing a model of science to study a knowledge base using the formalism of universal facts we have introduced. We can understand the elements of any particular knowledge base as having been 'picked', in some sense, from the set of all possible universal facts. If the pick is random and described as a probability measure ρ, we can quantify the information of the pick using the entropy.

Definition 10 (Axiomatic Information)**.** *Let* $\mathbb{Q}$ *be the domain of a universal Turing machine* $\mathbb{Q} = \mathrm{Dom}(\mathrm{UTM})$ *(full theory) or of a subset thereof* $\mathbb{Q} \subset \mathrm{Dom}(\mathrm{UTM})$ *(toy theory). Then, let* $\rho : \mathbb{Q} \to [0, 1]$ *be a probability measure over* $\mathbb{Q}$*. Finally, let* $\mathbb{KB}$ *be a knowledge base subset of* $\mathbb{Q}$*. The axiomatic information of a single element of* $\mathbb{KB}$ *is quantified as the entropy of* ρ*:*

$$S = -\sum_{q\in\mathbb{Q}} \rho(q) \ln \rho(q) \tag{6}$$

For instance, a well-known (non-computable) probability measure regarding a sum of prefix-free programs is the Halting probability[4] of computer science:

$$\Omega = \sum_{p\in\mathrm{Dom}(\mathrm{UTM})} 2^{-|p|} \implies \rho(p) = 2^{-|p|} \tag{7}$$

The quantity of axiomatic information of a given knowledge base (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 universal facts according to the probability measure which maximizes the entropy is a maximally informative strategy.

1.3 Discussion — The Mathematics of Knowledge

Each element of a knowledge base is a program-input pair representing an algorithm which is known to produce a specific result. 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 impresses 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 unties his shoes? One simply pulls on the tip of the laces. How does one unties his shoes if, after partial pulling, the knot accidentally mangles itself preventing further pulling? One uses his fingers or nails to un-mangle 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 knows 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 forms a knowledge base.

1.3.1 Special Cases, Inconsistencies, etc.

What if a knowledge base contains both "A" and "not A" as theorems? For instance, consider:

$$\mathbb{KB} := \{(\mathrm{TM}_1, A), (\mathrm{TM}_1, \neg A)\} \tag{8}$$

Does allowing contradictions at the level of the theorems of $\mathbb{KB}$ 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 $A \land \neg A$ into my personal knowledge base, and then we will evaluate the damage I have been subjected to by this insertion. Consider the following program TM_1:

1. If (p="A" or p="not A") then
2. return 1;
3. else (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" just symbols. Our reflex to attribute a special exclusionary requirement to these symbols requires the adoption of a deductive system. This occurs one step further at the selection of a specific formal axiomatic representation of the knowledge base, and not at the level of the knowledge base itself.

The only inconsistency that would create problems for this framework would be a proof that a given universal fact both [HALT] and [NOT HALT] on a UTM. By definition of a UTM, this cannot happen lest the machine was not a UTM to begin with. Thus, we should be safe from these contradictions.

Now, suppose one has a sizeable knowledge base which may contain a plurality of pairs:

$$\mathbb{KB} := \{(\mathrm{TM}_1, p_1), (\mathrm{TM}_2, \neg p_1), (\mathrm{TM}_1, p_2), (\mathrm{TM}_2, p_1), (\mathrm{TM}_2, \neg p_3)\} \quad (9)$$

Here, the negation of some, but not all, are also present across the pairs: in this instance, the theorems p_1 and p_3 are negated but for different atomic enumerators. What interpretation can we give to such elements of a knowledge base? For our example, let us call the sentences p_1, p_2, p_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 universal facts are pairs which contain an atomic enumerator and a theorem. The atomic enumerator is 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 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 the mathematical framework which correctly describes its knowledge base 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 $\mathbb{KB}$ while also having the same atomic enumerator, is still knowledge. It means one has verified that said atomic enumerator 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 all part of the knowledge base.

2 Formal Science

2.1 Axiomatic Foundation of Science

The fundamental object of study of science is not the electron, the quark or even super-strings, but the reproducible experiment. An experiment represents an 'atom' of verifiable knowledge.

Definition 11 (Experiment)**.** *Let* (TM, p) *be a pair comprising of two sentences of a language* $\mathbb{L}$*. The first sentence,* TM*, is called the protocol. The second sentence, p, is called the hypothesis. Let* UTM *be a universal Turing machine. If* $\mathrm{UTM}(\mathrm{TM}, p)$ *halts then the pair* (TM, p) *is said to be an experiment. In this case, we say that the protocol verifies the hypothesis. If* $\mathrm{UTM}(\mathrm{TM}, p)$ *does not halt, we say that the pair fails to verify the hypothesis.*

$$\mathrm{UTM}(\mathrm{TM}, p) \begin{cases} \mathrm{HALT} & \implies \textit{the experiment verifies } p \\ \neg\,\mathrm{HALT} & \implies \textit{the pair fails verification} \end{cases} \tag{10}$$

Of course, in the general case, as per the halting problem there exists no decidable function which can determine which pair is an experiment and which pair fails verification. In the general case, one must try them out to see which one halts — this is why they are called experiments.

An experiment, so defined, is formally reproducible. I can transmit, via fax or other telecommunication medium, the pair (TM, p) to another experimentalist, and I would know with absolute certainty that he or she has everything required to reproduce the experiment to perfection.

Theorem 3 (Formal Reproducibility)**.** *Experiments are formally reproducible.*

Proof. Let UTM and UTM' each be a universal Turing machine. For each pair $\mathrm{UTM}(\mathrm{TM}, p)$ which halts on UTM, there exists a computable function, called an encoding function, which maps said pairs as $encode(\mathrm{TM}, p) \to (\mathrm{TM}', p')$ such that (TM', p') halts for UTM'. The existence of such function is guaranteed by (and equivalent to) the statement that any UTM can simulate any other. □

In the peer-reviewed literature, the typical requirement regarding the reproducibility of an experiment 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. Here, for the protocol TM to be a Turing machine, 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. Should the protocol fail to verify the hypothesis, the entire experiment (that is the group comprising the hypothesis, the protocol and its complete description of all instrumentation) is rejected. For these reasons and due to the generality of the definition, I conjecture that the above definition is the only (sensible) definition of the experiment that is formally reproducible (as opposed to say "sufficiently reproducible for practical purposes").

Definition 12 (Empirical Evidence)**.** *The set of all pairs whose protocol* TM *verifies p is defined as the empirical evidence* Ev *of p:*

$$\mathrm{Ev}(p) := \mathrm{Dom}(\mathrm{UTM}, p) \tag{11}$$

where

$$\mathrm{Dom}(\mathrm{UTM}, p) := \{(\mathrm{TM}, p) : \text{ where } \mathrm{UTM}(\mathrm{TM}, p) \text{ halts } \} \tag{12}$$

If we are referencing the empirical evidence of a set of hypothesis $\mathbb{H}$*, then we note it as* $\mathrm{Ev}(\mathbb{H})$.

Definition 13 (Scientific method)**.** *An algorithm which recursively enumerates the empirical evidence, or parts thereof, of an hypothesis or a set thereof, is called a scientific method.*

Empirical evidence is thus produced by the application of the scientific method to an hypothesis.

Theorem 4 (Scientific method (Existence of))**.** *Existence of the scientific method.*

Proof. Consider a dovetail program scheduler which works as follows.

1. Sort all pairs of sentences of $\mathbb{L} \times \mathbb{L}$ in shortlex. Let the ordered pairs (TM_1, p_1), (TM_2, p_1), (TM_1, p_2), (TM_2, p_2), (TM_3, p_1), $\ldots$ be the elements of the sort.

2. Take the first element of the sort, $\mathrm{UTM}(\mathrm{TM}_1, p_1)$, then run it for one iteration.

3. Take the second element of the sort, $\mathrm{UTM}(\mathrm{TM}_2, p_1)$, 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, $\mathrm{UTM}(\mathrm{TM}_1, p_2)$, 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 (TM_i, p_j) which halts.

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. □

Dovetailing is of course a simple/non-creative approach to the scientific method. The point here was only to show existence of such an algorithm, not to find the optimal one. This will be done in the upcoming section regarding nature.

Definition 14 (Validated theory)**.** *Let* VT *be a formal axiomatic representation, and let* $\mathbb{H}$ *be a set of hypothesis. If* $\mathrm{Ev}(\mathbb{H}) = \mathrm{Dom}(\mathrm{VT})$*, then* VT *is a validated theory of* $\mathbb{H}$.

2.1.1 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.

Definition 15 (Scientific theory)**.** *Let* $\mathbb{D}$*, called the 'collected scientific data' or just 'the data', be a subset of the empirical evidence of a set of hypothesis* $\mathbb{H}$*:*

$$\mathbb{D} \subset \mathrm{Ev}(\mathbb{H}) \tag{13}$$

A finitely axiomatized representation is called a scientific theory ST *of* $\mathbb{D}$ *iff* $\mathbb{D} \subset \mathrm{Dom}(\mathrm{ST})$*. The set* $\mathbb{P}$*, called the predictions of* ST*, is defined as:*

$$\mathbb{P} := \mathrm{Dom}(\mathrm{ST}) \setminus \mathbb{D} \tag{14}$$

Definition 16 (Predictive theory)**.** *Let* ST *be a scientific theory of* $\mathbb{D}$*. If the predictions of* ST *are not empty, then* ST *is a predictive theory.*

Definition 17 (Empirical theory)**.** *Let* ST *be a scientific theory of* $\mathbb{D}$*. If the predictions of* ST *are empty, then* ST *is a empirical theory.*

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

Theorem 5 (The Fundamental Theorem of Science)**.** *If the empirical evidence of* $\mathbb{H}$ *is recursively enumerable, but not decidable, then the empirical evidence of* $\mathbb{H}$ *has measure* 0 *over the set of all possible predictive scientific theories of* $\mathbb{H}$*.*

Proof. The empirical evidence of $\mathbb{H}$ is unique, yet —excluding atypical cases $\mathrm{Ev}(\mathbb{H})$ where it is decidable— there exists countably infinitely many predictive theories of $\mathbb{H}$, for any set of data $\mathbb{D}$. Finally, the measure of one element of a countably 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 that a non-decidable predictive scientific theory will eventually be falsified.

2.2 Axiomatic Foundation of Reality

Definition 18 (Domain of science)**.** *We note* $\mathbb{S}$ *as the domain* (Dom) *of science. We can define* $\mathbb{S}$ *in reference to a universal Turing machine* UTM *as follows:*

$$\mathbb{S} := \mathrm{Dom}(\mathrm{UTM}) \tag{15}$$

Thus, for all pairs of sentences (TM, p)*, if* $\mathrm{UTM}(\mathrm{TM}, p)$ *halts, then* $(\mathrm{TM}, p) \in \mathbb{S}$*. It follows that all experiments are elements of the domain of science.*

Definition 19 (Experimental Manifest)**.** *An experimental manifest is a 'constructed' knowledge base. Specifically, an experimental manifest* $\mathbf{m}$*, or just a 'manifest', is a n-tuple constructed as an element of the cartesian product applied to the domain of science.*

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

A manifest is therefore a tuple of experiments:

$$\mathbf{m} := ((\mathrm{TM}_1, p_1), \ldots, (\mathrm{TM}_n, p_n)) \tag{17}$$

Note how similar this definition is to that of the knowledge base (repeated here):

$$\mathbb{KB} := \{(\mathrm{TM}_1, p_1), \ldots, (\mathrm{TM}_n, p_n)\} \tag{18}$$

We note that since a manifest may contain repetitions (experiments are formally reproducible) we have not defined $\mathbf{m}$ *as a set, but instead as a n-tuple to retain said repetitions. Quite remarkably, this tuple vs set (experimental manifest vs knowledge base) definition is the primary difference between formal theories in math versus those in science — we will investigate the consequences of this difference in great detail in the main results section.*

For a given manifest, the possibility exists that some hypotheses or, likewise, some protocols be repeated in the other tuples of the manifest. For instance it could be the case that within a manifest $p_2 = p_5$, or that $\mathrm{TM}_1 = \mathrm{TM}_6 = \mathrm{TM}_{21}$, etc. The set of all hypotheses for a given protocol is called the *scope* (of the protocol), and the set of all protocols for a given hypothesis is called the *spread* (of the hypothesis).

One use for repetitions, for instance, is as a quality check on the UTM. Indeed, if one repeats many different experiments, and one finds they are indeed correctly reproduced, then one has a high degree of confidence in the reliability in one's machine. Comparatively a knowledge base includes no repetitions (it is a set), the existence of a reliable UTM is merely assumed and in fact conditional to the framework, and as such cannot be inferred or proven. To prove with absolute certainty that one's machine is a (perfect) universal Turing machine using a manifest, one has to repeat every pair infinitely many times. Consequently, in any practical case one only infers, to a finite degree of certainty, that one has access to such a machine. Using a tuple of experiments, as opposed to a set, grants us the ability to discover the limits to computation of one's machine (in nature), should such limits exist.

Definition 20 (Experimental Space)**.** *Experimental space* $\mathbb{E}$ *is the "powertuple" of the domain of science:*

$$\mathbb{E} := \bigcup_{i=0}^{\infty} \mathbb{S}^i \tag{19}$$

All elements of experimental space are manifests, and all manifests are elements of experimental space.

Definition 21 (Toy Model)**.** *A subset of experimental space is called a 'toy model'. Some toy models, but not all, are be decidable.*

2.2.1 The Fundamental Assumption of Science

Definition 22 (State of Affairs)**.** *The axiomatic information associated to a manifest (in addition to the manifest itself) constitute a state of affairs.*

1. *The state of affairs (in philosophy) refers to the state the world must be in for a proposition about it to be true.*
2. *Information (in information theory, and say in base 2) relates to the number of yes/no questions one must ask to identify an element randomly selected from a set.*
3. *A yes/no statement referring to the inclusion of an experiment to be part of a manifest, connects axiomatic information to the state of affairs.*
4. *By maximizing the entropy, redundancies in the yes/no questions are eliminated.*

Definition 23 (The Fundamental Assumption of Science)**.** *There exists a manifest, called the reference manifest, which corresponds to the knowledge base of the state of affairs of reality.*

1. *The fundamental assumption of science essentially implies that there are no gaps of knowledge in the state of affairs — science can capture everything the state of affairs has to offer in terms of knowledge.*
2. *If the assumption of science would be false, it would mean that there are elements of the knowledge base of the state of affairs that are outside the domain of science... (hyper-computation by an oracle guiding our life?)*

2.3 Axiomatic Foundation of Nature

Recall that earlier we used a dovetailing algorithm in Theorem 4 as an implementation of the scientific method, and we claimed that although it was a possible strategy, it was not necessarily the optimal one. So what then is the optimal implementation of the scientific method applicable to a tuple of experiments? Well, I suppose it depends on what we mean by optimal. One might be tempted to search along the lines of an efficient algorithm, perhaps the most elegant one, or the one that uses the least amount of memory, etc., but thinking in those terms would be a trap — we must think a bit more abstractly than postulating or arguing for a specific implementation. Potentially, every manifest could in principle have its own best strategy. It is therefore more strategic overall to identify a condition applicable to all cases, which produces the implementation as a maximization problem.

The best strategy will be to *maximize the axiomatic information* gained from the scientific method, as an experimental manifest is produced; and this means in the technical sense to maximize the entropy of a probability measure on the paths in experimental space:

Definition 24 (The Fundamental Assumption of Nature)**.** *Let ρ be a probability measure on the points in experimental space (full model) or a subset thereof (toy model). An observer, denoted as $\mathcal{O}$, is a point randomly selected from experimental space. Thus, with a probability measure that spawns the set $\mathbb{E}$, every point in experimental space qualifies as an observer. Specifically, an observer is:*

$$\mathcal{O}_i := (\mathbf{m}_i, \rho : \mathbb{E} \to [0,1]) \tag{20}$$

The definition of the observer is a specialization of the definition of the manifest in the sense that a manifest is a point in experimental space, and the observer is a *randomly selected point* in experimental space (and thus the notion of information is associated to it). Note that typically in physics, the observer (which is not mathematically integrated into the formalism... leading to a family of open problems regarding the 'observer effect') is associated to a random selection of an element from a set of possible observations. This 'effect' will eventually be revealed to be a consequence of the present definition. Here, the observer 'has information' of a randomly selected point in experimental states and consequently has the opportunity to implement an information producing version of the scientific method so as to realize its optimal implementation.

2.3.1 Laws of Nature

Definition 25 (Laws of Nature)**.** *The laws of nature are the probability measure that maximizes the entropy of a path in $\mathbb{E}$ between observers.*

The laws of nature are thus defined formally as the *information-theoretical maximum* of the scientific method, for an observer. The best strategy to maximize ones axiomatic knowledge of reality produces the laws of nature as the solution.

The axioms required to formally define the practice of science may be minimal, but they nonetheless require a minimal structure for the practice to be possible, and the laws of nature simply emerge as the rules which preserves this structure over the transformation of the path of an observer. Each stop along the path of the observer must be a knowledge base about reality. The appellation 'laws of nature' in this context refers to the theory of all possible transformations which preserve the structure necessary to accommodate the scientific method — essentially, the laws of nature are the laws that preserve nature.

We note a distinction between a scientific theory obtained by inspection of the manifest followed by subsequent inductive generalizations and the laws of nature, as defined above. Scientific theories, as they represent a complete formal axiomatic representation of the knowledge base, can 'explain' up to the totality of the manifest. However, the price to pay for this completeness is that the theory is subject to future falsification as per the fundamental theorem of science, as early as the very next transformation of the manifest, if one is unlucky.

The laws of nature, in contrast, hold for all possible stops along experimental space. To gain this resilience, the laws of nature must 'distance' themselves from the specifics of any singular manifest and thus can only 'explain' a much smaller fraction of reality than a scientific theory which can account for the whole of the manifest. Indeed, anything which might appear true in the present manifest, but could plausibly be false in another manifest can be part of a scientific theory, but cannot be part of the laws of nature, otherwise said laws are susceptible to future falsification. For instance, a simple question such as "do apples always fall downwards" may be quite easy to answer from a scientific standpoint, but be very hard to answer from a 'laws of nature' standpoint. From the scientific standpoint, one may look at ten apples, note that all ten of them fell downward, then a claim by induction that all apples will fall downwards completes the scientific theory, but leaves it susceptible to future falsification. To answer the same question with the laws of nature is much more difficult. One must show that there exists no manifest in which an apple falls up (even ruling out all possible statistical flukes), otherwise one has laws of nature claiming that something cannot happen when in fact some experimental states exist in which *it does happen*, leaving them vulnerable to eventual falsification. The laws of nature are "eternal" over experimental space.

For instance, take Newton's theory of classical gravitation obtained by inspecting a subset of experimental space and note that it was eventually falsified when a larger subset of experimental space was inspected by Einstein to produce general relativity. Here, we have a framework which allows us to look at the whole of experimental space at once, thus giving us the opportunity to identify laws of nature whose only requirement is that they preserve the structure of nature.

3 Main Result

Let us now use these definitions to derive the laws of nature from first principle, and then show the overwhelming similarity to the laws of physics.

3.1 Overview

3.1.1 Halting Probability Ω

Let us start by maximizing the entropy of the random selection of p from Dom(UTM):

$$S = -\sum_{p \in \mathrm{Dom(UTM)}} \rho(p) \log_2 \rho(p) \tag{21}$$

subject to these constraints:

$$\sum_{p \in \mathrm{Dom(UTM)}} \rho(p) = 1 \tag{22}$$

$$\sum_{p \in \mathrm{Dom(UTM)}} \rho(p)|p| = \overline{|p|} \tag{23}$$

Using the method of the Lagrange multipliers, the result is the Gibbs measure (where D is a Lagrange multiplier):

$$\rho(p) = \frac{1}{Z} 2^{-D|p|}, \qquad \text{where } Z = \sum_{p \in \mathrm{Dom(UTM)}} 2^{-D|p|} \tag{24}$$

This is the statistical-physics definition of a sum of programs. Unlike the Halting probability of computer science Ω, here it is $\rho(s)$ (and NOT Z) that is the probability. We note that it is not necessarily all choices of D which causes Z to be non-computable (for instance if $D = 0$ then Z is very much so computable; it is in fact infinite). To recover Ω, the Halting probability[4] of computer science, we would pose the Lagrange multiplier D to 1, then take the encoding of the program to be prefix-free and therefore, via the Kraft-inequality, Z becomes itself a probability measure:

$$\Omega = \sum_{p \in \mathrm{Dom(UTM)}} 2^{-|p|} \tag{25}$$

We further note the work of Tadaki[5] which identifies an 'algorithmic-thermodynamics[6]' definition of Ω by adding D called a 'decompression-term' as follows:

$$\sum_{p \in \mathrm{Dom(UTM)}} = 2^{-D|p|} \tag{26}$$

However, in each of these cases, with the exception of [6], the connection to entropy is lost because the expression of Z is reduced such that it, rather that ρ, acquires the role of the probability measure. However, there is a gain to be had by retaining the connection to an entropy maximum. Indeed, knowing a message from a set of possible messages according to a probability measure that maximizes the entropy, makes knowing said message maximally informative. Likewise, in the case of the statistical physics version of a sum of programs, the probability measure that maximizes the entropy for this system makes our knowledge of a program that halts maximally informative.

3.1.2 Quantum Computing

Let us now investigate the basics of quantum computation. One starts with a state vector:

$$|\psi_a\rangle = \begin{pmatrix} 0 \\ \vdots \\ n \end{pmatrix} \tag{27}$$

Which evolves unitarily to a final state:

$$|\psi_b\rangle = U_0 U_1 \ldots U_m |\psi_a\rangle \tag{28}$$

Clever use of the unitary transformations, often arranged as simple 'gates', allows one to execute a program. 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:

$$\begin{aligned}
&\text{step 1} && |\psi_b\rangle = U_0 U_1 \ldots U_p |\psi_a\rangle && (29)\\
&\text{step 2} && |\psi_c\rangle = U'_0 U'_1 \ldots U'_q |\psi_b\rangle && (30)\\
&\vdots && && (31)\\
&\text{step k} && |\psi_{k'}\rangle = U'_0 U'_1 \ldots U'_v |\psi_k\rangle && (32)\\
&\vdots && && (33)
\end{aligned}$$

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 is itself a completed program, and further it is the case that each step can be infinitely divided, yielding an interesting property specific to quantum computations:

3.1.3 Program-Steps and Programs are indistinguishable

The property of interest in a quantum computation (for our purposes), is that all intermediary steps of the quantum computation are computations in and of themselves. This is because a measurement of a state can take place between any unitary steps. Indeed, each program-steps can be understood as a part of a larger program, or as a program itself, or even subdivided further in infinitely small steps. Quantum computing machines are a special design of a Turing machine in which all program-steps, and all inner-states are also entire programs.

Comparatively, the typical design of a Turing machine is that the machine has an inner state, prints an output to a tape and the program either halts or doesn't. The transformations of the inner state of a Turing machine are not considered programs even if such inner states are, obviously, computable. Because of this property, as it fuses the notions of program, program-steps and inner state of the Turing machine (and based on the fact that universe picked this method of computation for its own computing needs...), I would submit that quantum computing is a conceptually cleaner definition of a system of computation than that of the typical Turing machine defined in terms of tape, output and head — neither of which are programs.

3.1.4 Manifest-to-Manifest Computing... ?

Recall that we have defined a manifest as a tuple of experiments, that is, a tuple of pairs (TM, p) that halt, and we have called that set of all such tuples, experimental space. Now, consider a path in experimental space between manifests. Such a path describes an accumulation of programs over the path, and describes a computation that has the same property of interest as the quantum computing case. Due to the definition of the manifest, paths within experimental space recover a generalized and abstract realization of said property. Any path by an observer in experimental space is guaranteed to only encounter steps that are formulated as completed computations.

3.2 Derivation

Distilled to its core, the laws of nature are simply given as the probability measure that maximizes the entropy of a random selection of a tuple from a space of possible tuples. But there is a caveat to deriving such a measure; measure theory deals with subsets of sets, and not with tuples of tuple-spaces.

Nonetheless the two are very similar. The trick will be to 'fool' measure theory into thinking our tuple-space is a set by adding an invariance constraint on the entropy with respect to a tuple re-ordering. Remarkably, this invariance constraint will be quite impactful on the laws of nature as it will provide support for constructing paths in experimental space in terms of transformations applied to tuples.

Finally, requiring that our tuples contain only programs allows us to think of each transformation as the addition of a new program-step to said programs, which is sufficient to make the theory conceptually self-contained. Let us now do it explicitly.

3.2.1 General Linear Ensemble

Let us start with a sum of programs (i.e. manifests that are comprised of a single element). A probability measure would assign a real number (between 0 and 1) to each program of the sum, representing of course the probability associated with the random selection of said program to be an element of the manifest. Extending this sum to manifests of multiple programs will be done in Section 5.1 using the tensor product.

Let us therefore maximize this entropy:

$$S = -\sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) \tag{34}$$

subject to these constraints:

$$\sum_{q \in \mathbb{Q}} \rho(q) = 1 \tag{35}$$

$$\sum_{q \in \mathbb{Q}} \rho(q) \operatorname{tr} \mathbf{M}(q) = \operatorname{tr} \overline{\mathbf{M}} \tag{36}$$

where $\mathbf{M}(q)$ are a matrix-valued maps[1] from $\mathbb{S}$ to $\mathbb{F}^{n \times n}$ representing the linear transformations of the space, where $\overline{\mathbf{M}}$ is a element-by-element average matrix in $\mathbb{F}^{n \times n}$ and where $\mathbb{F}$ is a field. Here, $\mathbb{Q}$ is an arbitrary sample space of programs, either the full-theory if $\mathbb{Q} = \mathbb{S}$ or a toy model if $\mathbb{Q} \subset \mathbb{S}$.

Usage of the trace of a matrix as a constraint, imposes an invariance with respect to a similarity transformation, accounting for all possible linear re-order of the elements of the tuples of the sum, thus allowing the creation of a measure of a tuple or group of tuples form within a space of tuples. Similarity transformation invariance on the trace is the result of this identity:

$$\operatorname{tr} \mathbf{M} = \operatorname{tr} \mathbf{B}\mathbf{M}\mathbf{B}^{-1} \tag{37}$$

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:

$$\mathcal{L} = -k_B \sum_{q \in \mathbb{Q}} \rho(s) \ln(s) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q) \right) + \tau \left(\operatorname{tr} \overline{\mathbf{M}} - \sum_{q \in \mathbb{Q}} \rho(q) \operatorname{tr} \mathbf{M}(q) \right) \tag{38}$$

[1]There is a possibility of greater generality by considering maps between spaces of different dimensions $\mathbb{S} \to \mathbb{F}^{n \times m}$. In quantum mechanics this is part of the subject matter of 'quantum operations' which includes quantum channels. This mapping from differently-size vector spaces would be required in the general case to account for all possible paths of the observer in experimental space, and would likely come out as a general linear equivalent to a quantum channel transmitting information between manifests as their sizes change. This is likely interesting, but, as we will see, we will not be running out of applications for the general linear ensemble as it is, and thus we have elected to limit the scope to maps in $\mathbb{F}^{n \times n}$ in line with the typical formulations of quantum mechanics.

where α and τ are the Lagrange multipliers. The explicit derivation is made available in Annex B. The result of the maximization process is:

$$\rho(q,\tau) = \frac{1}{Z(\tau)} \det\exp -\tau\mathbf{M}(q) \tag{39}$$

where

$$Z(\tau) = \sum_{q\in\mathbb{Q}} \det\exp -\tau\mathbf{M}(q) \tag{40}$$

3.2.2 Prior

No good probability measure is complete without 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 $\mathbf{P} : \mathbb{Q} \to \mathbb{F}^{n\times n}$ and inject it into the probability measure as well as into the partition function:

$$\rho(q) = \frac{1}{Z} \det\exp\big(\mathbf{P}(q)\big) \det\exp\big(-\tau\mathbf{M}(q)\big) \tag{41}$$

where

$$Z = \sum_{q\in\mathbb{Q}} \det\exp\big(\mathbf{P}(q)\big) \det\exp\big(-\tau\mathbf{M}(q)\big) \tag{42}$$

4 Foundation

To study the properties of the probability amplitude of our measure, we will 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, and finally, the n-dimensional case.

4.1 Overview

4.1.1 Matrix-Valued Vector and Transformations

We will use vectors whose elements are matrices. An example of such a vector is:

$$\mathbf{v} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_n \end{pmatrix} \tag{43}$$

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

$$\mathbf{T} = \begin{pmatrix} \mathbf{M}_{00} & \dots & \mathbf{M}_{0m} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{m0} & \dots & \mathbf{M}_{mm} \end{pmatrix} \tag{44}$$

Note: The scalar element of the vector space are in $\mathbb{F}$. For instance:

$$a\mathbf{v} = \begin{pmatrix} a\mathbf{M}_1 \\ \vdots \\ a\mathbf{M}_n \end{pmatrix} \tag{45}$$

4.1.2 Linear Transformations as Computations

We will be looking for the conditions under which linear paths in experimental spaces are computations.

We begin with a rewriting of the probability measure such that it is 'split' 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:

$$\rho(q, \tau) = \frac{1}{Z} \det \psi(q, \tau) \tag{46}$$

where

$$\psi(q, \tau) = \exp\big(\mathbf{P}(q)\big) \exp\big(-\tau \mathbf{M}(q)\big) \tag{47}$$

Here, the determinant is interpreted as a generalization of the Born rule, and reduces to 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, $\psi(q, \tau)$ will be called the *general linear probability amplitude.*

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

$$\boldsymbol{\psi} := |\psi\rangle := \begin{pmatrix} \psi(q_1, \tau) \\ \psi(q_2, \tau) \\ \vdots \\ \psi(q_n, \tau) \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \tag{48}$$

The transformation which leaves the probability measure $\rho(q, \tau)$ invariant will be reversible linear transformations and express the evolution from one statistical sum of programs $|\psi_a\rangle$ to another $|\psi_b\rangle$ corresponding to the path of an observer in experimental space.

Paths are constructed by chaining transformations on those vectors:

$$|\psi_b\rangle = \underbrace{T_1 T_2 \dots T_n}_{\text{computing steps}} |\psi_a\rangle \tag{49}$$

As more steps are piled on, progressively richer manifests are constructed. Paths in experimental space are realized by completing the missing computational steps required for a starting-point manifest to be the end-point manifest.

4.1.3 General Linear Group

The set of all complex $n \times n$ matrices connects, via the exponential map, to the general linear group in $\mathbb{C}$:

$$\exp : \mathbb{M}(n, \mathbb{C}) \to \mathrm{GL}(n, \mathbb{C}) \tag{50}$$

The map is also possible in $\mathbb{R}$, but in this case the general linear group is reduced to the orientation-preserving general linear group, because the left-hand side of the map cannot produce a matrix with a negative determinant and thus is not surjective in the general case:

$$\exp : \mathbb{M}(n, \mathbb{R}) \to \mathrm{GL}^+(n, \mathbb{R}) \tag{51}$$

The entropy maximization procedure we have used produced a probability measure which embeds the exponential map over matrices, thus connects the arbitrary linear transformation of $\mathbb{M}(n, \mathbb{R})$ to the orientation-preserving linear group $\mathrm{GL}^+(n, \mathbb{R})$.

4.2 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.

4.2.1 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:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{B} \tag{52}$$

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 26 (Geometric representation of a matrix (2×2))**.**

$$A + X\hat{\mathbf{x}} + Y\hat{\mathbf{y}} + B\hat{\mathbf{x}} \wedge \hat{\mathbf{y}} \cong \begin{pmatrix} A+X & -B+Y \\ B+Y & A-X \end{pmatrix} \tag{53}$$

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[7].

Definition 27 (Clifford conjugate (of a $\mathbb{G}(2,\mathbb{R})$ multi-vector))**.**

$$\mathbf{u}^\ddagger := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 \tag{54}$$

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

Definition 28 (Geometric representation of the determinant (of a 2×2 matrix))**.**

$$\begin{array}{rcrcl} \det & : & \mathbb{G}(2,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} & \longmapsto & \mathbf{u}^\ddagger \mathbf{u} \end{array} \tag{55}$$

For example:

$$\det \mathbf{u} = (A - \mathbf{X} - \mathbf{B})(A + \mathbf{X} + \mathbf{B}) \tag{56}$$

$$= A^2 - X^2 - Y^2 + B^2 \tag{57}$$

$$= \det \begin{pmatrix} A+X & -B+Y \\ B+Y & A-X \end{pmatrix} \tag{58}$$

Finally, we define the Clifford transpose:

Definition 29 (Clifford transpose (of a matrix of 2×2 matrix elements))**.** *The Clifford transpose of 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:*

$$\begin{pmatrix} \mathbf{u}_{00} & \dots & \mathbf{u}_{0n} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}_{mn} \end{pmatrix}^\ddagger = \begin{pmatrix} \mathbf{u}_{00}^\ddagger & \dots & \mathbf{u}_{m0}^\ddagger \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}_{nm}^\ddagger \end{pmatrix} \tag{59}$$

If applied to a vector, then:

$$\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix}^\ddagger = \begin{pmatrix} \mathbf{v}_1^\ddagger & \dots \mathbf{v}_m^\ddagger \end{pmatrix} \tag{60}$$

4.2.2 Axiomatic Definition of the Algebra, in 2D

Let $\mathbb{V}$ be a 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:

$$\begin{array}{rcrcl} \langle \cdot, \cdot \rangle & : & \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(2,\mathbb{R}) \\ & & \langle \mathbf{u}, \mathbf{v} \rangle & \longmapsto & \mathbf{u}^\ddagger \mathbf{v} \end{array} \tag{61}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{62}$$

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

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \psi(q)^\ddagger \psi(q) \tag{63}$$

   is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{64}$$

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:

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{65}$$

- $\boldsymbol{\psi}$ 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):

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{66}$$

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

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

$$\langle \mathbf{O}\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v} \rangle \tag{67}$$

  is called an observable.

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

$$\langle \mathbf{O} \rangle = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi} \rangle \tag{68}$$

4.2.3 Reduction to Complex Hilbert Spaces

It is fairly easy to see that if we reduce the expression of our multi-vectors $(A+\mathbf{X}+\mathbf{B}|_{\mathbf{X}\to 0} = A+\mathbf{B}$ and further restrict $\langle\boldsymbol{\psi},\boldsymbol{\psi}\rangle \in \mathbb{R}_{>0}$ to $\langle\boldsymbol{\psi},\boldsymbol{\psi}\rangle = 1$, then we recover the unit vectors of the complex Hilbert spaces:

- Reduction to the conjugate transpose:

$$\Big(\langle\mathbf{u},\mathbf{v}\rangle = \mathbf{u}^\ddagger\mathbf{v}\Big|_{\mathbf{x}\to 0} \implies \langle\mathbf{u},\mathbf{v}\rangle = \mathbf{u}^\dagger\mathbf{v} \tag{69}$$

- Reduction to the unitary transformations:

$$\big(\langle\mathbf{Tu},\mathbf{Tv}\rangle = \langle\mathbf{u},\mathbf{v}\rangle\big|_{\mathbf{x}\to 0} \implies \mathbf{T}^\dagger\mathbf{T} = I \tag{70}$$

- Reduction to the Born rule:

$$\left(\rho(q,\boldsymbol{\psi}) = \frac{1}{\langle\boldsymbol{\psi},\boldsymbol{\psi}\rangle}\psi(q)^\ddagger\psi(q)\right|_{\mathbf{X}\to 0} \implies \rho(q,\boldsymbol{\psi}) = \frac{1}{\langle\boldsymbol{\psi},\boldsymbol{\psi}\rangle}\psi(q)^\dagger\psi(q) \tag{71}$$

- Reduction of observables to Hermitian operators:

$$\big(\langle\mathbf{Ou},\mathbf{v}\rangle = \langle\mathbf{u},\mathbf{Ov}\rangle\big|_{\mathbf{X}\to 0} \implies \mathbf{O}^\dagger = \mathbf{O} \tag{72}$$

Under this reduction, the formalism becomes equivalent to the Dirac-Von-Neumann formalism of quantum mechanics.

4.2.4 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:

$$\langle\mathbf{O}\boldsymbol{\phi},\boldsymbol{\psi}\rangle = \langle\boldsymbol{\phi},\mathbf{O}\boldsymbol{\psi}\rangle \tag{73}$$

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

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

$$\begin{aligned}2\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle &= (O_{00}\phi_1 + O_{01}\phi_2)^{\ddagger}\psi_1 + \psi_1^{\ddagger}(O_{00}\phi_1 + O_{01}\phi_2) \\ &\quad + (O_{10}\phi_1 + O_{11}\phi_2)^{\ddagger}\psi_2 + \psi_2^{\ddagger}(O_{10}\phi_1 + O_{11}\phi_2) &(74)\\ &= \phi_1^{\ddagger}O_{00}^{\ddagger}\psi_1 + \phi_2^{\ddagger}O_{01}^{\ddagger}\psi_1 + \psi_1^{\ddagger}O_{00}\phi_1 + \psi_1^{\ddagger}O_{01}\phi_2 \\ &\quad + \phi_1^{\ddagger}O_{10}^{\ddagger}\psi_2 + \phi_2^{\ddagger}O_{11}^{\ddagger}\psi_2 + \psi_2^{\ddagger}O_{10}\phi_1 + \psi_2^{\ddagger}O_{11}\phi_2 &(75)\end{aligned}$$

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

$$\begin{aligned}2\langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle &= \phi_1^{\ddagger}(O_{00}\psi_1 + O_{01}v_2) + (O_{00}\psi_1 + O_{01}\psi_2)^{\ddagger}\phi_1 \\ &\quad + \phi_2^{\ddagger}(O_{10}\psi_1 + O_{11}\psi_2) + (O_{10}\psi_1 + O_{11}\psi_2)^{\ddagger}\phi_1 &(76)\\ &= \phi_1^{\ddagger}O_{00}\psi_1 + \phi_1^{\ddagger}O_{01}\psi_2 + \psi_1^{\ddagger}O_{00}^{\ddagger}\phi_1 + \psi_2^{\ddagger}O_{01}^{\ddagger}\phi_1 \\ &\quad + \phi_2^{\ddagger}O_{10}\psi_1 + \phi_2^{\ddagger}O_{11}\psi_2 + \psi_1^{\ddagger}O_{10}^{\ddagger}\phi_1 + \psi_2^{\ddagger}O_{11}^{\ddagger}\phi_1 &(77)\end{aligned}$$

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:

$$O_{00}^{\ddagger} = O_{00} \tag{78}$$
$$O_{01}^{\ddagger} = O_{10} \tag{79}$$
$$O_{10}^{\ddagger} = O_{01} \tag{80}$$
$$O_{11}^{\ddagger} = O_{11} \tag{81}$$

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:

$$\mathbf{O}^{\ddagger} = \mathbf{O} \tag{82}$$

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

4.2.5 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:

$$\mathbf{O} = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{83}$$

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

$$\mathbf{O}^\ddagger = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{84}$$

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

$$0 = \det(\mathbf{O} - \lambda I) = \det \begin{pmatrix} a_{00} - \lambda & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} - \lambda \end{pmatrix} \tag{85}$$

implies:

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a - xe_1 - ye_2 - be_{12})(a + xe_1 + ye_2 + be_{12} + a_{11}) \tag{86}$$

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a^2 - x^2 - y^2 + b^2) \tag{87}$$

finally:

$$\lambda = \{\frac{1}{2}\left(a_{00} + a_{11} - \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right), \tag{88}$$

$$\frac{1}{2}\left(a_{00} + a_{11} + \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right)\} \tag{89}$$

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 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[2] 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.

4.3 Algebra of Natural States, in 3D (brief)

The 3D case will be a stepping stone for the 4D case. A general multi-vector of $\mathbb{G}(3, \mathbb{R})$ can be written as follows:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{V} + \mathbf{B} \tag{90}$$

where A is a scalar, $\mathbf{X}$ is a vector, $\mathbf{V}$ is a pseudo-vector and $\mathbf{B}$ is a pseudo-scalar. Such multi-vectors form a complete representation of 2×2 complex matrices:

[2] 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.

$$A + X\sigma_1 + Y\sigma_2 + Z\sigma_3 + V_1 i\sigma_1 + V_2 i\sigma_2 + V_3 i\sigma_3 + B\sigma_1 \wedge \sigma_2 \wedge \sigma_3 \tag{91}$$
$$\cong \begin{pmatrix} A + iB + iV_2 + Z & V_1 + iV_3 + X - iY \\ -V_1 + iV_3 + X + iY & A + iB - iV_2 - Z \end{pmatrix} \tag{92}$$

and the determinant of this matrix, connects to the determinant of the multivector as follows:

$$\begin{array}{rcrcl} \det \cdot & : & \mathbb{G}(3,\mathbb{R}) & \longrightarrow & \mathbb{C} \\ & & \mathbf{u} & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{u} \end{array} \tag{93}$$

where $\mathbf{u}^{\ddagger}$ is the Clifford conjugate in 3D:

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 + \langle \mathbf{u} \rangle_3 \tag{94}$$

To produce a real number, a further multiplication by its complex conjugate is required:

$$\begin{array}{rcrcl} |\cdot| & : & \mathbb{G}(3,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} & \longmapsto & (\mathbf{u}^{\ddagger}\mathbf{u})^{\dagger}\mathbf{u}^{\ddagger}\mathbf{u} \end{array} \tag{95}$$

where $\mathbf{u}^{\dagger}$ is defined as:

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 + \langle \mathbf{u} \rangle_1 + \langle \mathbf{u} \rangle_2 - \langle \mathbf{u} \rangle_3 \tag{96}$$

4.3.1 Axiomatic Definition of the Algebra, in 3D

Let $\mathbb{V}$ be a m-dimensional vector space over $\mathbb{G}(3,\mathbb{R})$. 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 quadri-linear map:

$$\begin{array}{rcrcl} \langle \cdot,\cdot,\cdot,\cdot \rangle & : & \mathbb{V}\times\mathbb{V}\times\mathbb{V}\times\mathbb{V} & \longrightarrow & \mathbb{G}(3,\mathbb{R}) \\ & & \langle \mathbf{u},\mathbf{v},\mathbf{w},\mathbf{x} \rangle & \longmapsto & (\mathbf{u}^{\ddagger}\mathbf{v})^{\dagger}\mathbf{w}^{\ddagger}\mathbf{x} \end{array} \tag{97}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{98}$$

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

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} (\psi(q)^{\ddagger}\psi(q))^{\dagger}\psi(q)^{\ddagger}\psi(q) \tag{99}$$

   is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{100}$$

4.3.2 Reduction to Complex Hilbert Spaces

If we consider an algebra of natural states that comprises only those multi-vectors of the form $\mathbf{u}' \propto \mathbf{u}^\ddagger \mathbf{u}$. We also consider, as we obtain an exponential map due to our entropy maximization process, we only take multi-vectors of this form which are exponentiated, then the algebra reduces to the foundation of quantum mechanics on complex Hilbert spaces (with an extra geometric internal structure). For example, a wave-function would be of this form:

$$\boldsymbol{\psi} = \begin{pmatrix} (\exp \frac{1}{2}\mathbf{u}_1)^\ddagger \exp \frac{1}{2}\mathbf{u}_1 \\ \vdots \\ (\exp \frac{1}{2}\mathbf{u}_m)^\ddagger \exp \frac{1}{2}\mathbf{u}_m \end{pmatrix} \tag{101}$$

Each element of $\boldsymbol{\psi}$ are of this form:

$$(\exp \frac{1}{2}\mathbf{u})^\ddagger \exp \frac{1}{2}\mathbf{u} = \exp \frac{1}{2}(A - \mathbf{X} - \mathbf{V} + \mathbf{B}) \exp \frac{1}{2}(A + \mathbf{X} + \mathbf{V} + \mathbf{B}) \tag{102}$$

$$= \exp(A + \mathbf{B}) \exp \frac{1}{2}(-\mathbf{X} - \mathbf{V}) \exp \frac{1}{2}(\mathbf{X} + \mathbf{V}) \tag{103}$$

$$= \exp(A + \mathbf{B}) \tag{104}$$

Restricting the algebra to such states reduces the quadri-linear map to a bilinear form:

$$\begin{array}{cccccc} \langle \cdot, \cdot \rangle & : & \mathcal{A}(\mathbb{V}) \times \mathcal{A}(\mathbb{V}) & \longrightarrow & \mathbb{C} \\ & & \langle \psi, \phi \rangle & \longmapsto & \psi^\dagger \phi \end{array} \tag{105}$$

yielding, when applied to said reduced subset of vectors, the same theory as that of quantum mechanics on complex Hilbert space, but with an extra geometric structure for its observables. The 3D case is a stepping stone for the 4D case, where this extra geometric structure will be revealed to be (in the 4D case) the relativistic wave-function given in the form of a spinor field.

4.4 Algebra of Natural States, in 4D

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

4.4.1 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:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{106}$$

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:

Definition 30 (Geometric representation of a matrix (4×4))**.**

$$\begin{aligned}
&A + T\gamma_0 + X\gamma_1 + Y\gamma_2 + Z\gamma_3 \\
&\quad + F_{01}\gamma_0 \wedge \gamma_1 + F_{02}\gamma_0 \wedge \gamma_2 + F_{03}\gamma_0 \wedge \gamma_3 + F_{23}\gamma_2 \wedge \gamma_3 + F_{13}\gamma_1 \wedge \gamma_3 + F_{12}\gamma_1 \wedge \gamma_2 \\
&\quad + V_t\gamma_1 \wedge \gamma_2 \wedge \gamma_3 + V_x\gamma_0 \wedge \gamma_2 \wedge \gamma_3 + V_y\gamma_0 \wedge \gamma_1 \wedge \gamma_3 + V_z\gamma_0 \wedge \gamma_1 \wedge \gamma_2 \\
&\quad + B\gamma_0 \wedge \gamma_1 \wedge \gamma_2 \wedge \gamma_3 \\
&\qquad \cong \begin{pmatrix}
A + X_0 - iF_{12} - iV_3 & F_{13} - iF_{23} + V_2 - iV_1 & -iB + X_3 + F_{03} - iV_0 & X_1 - iX_2 + F_{01} - iF_{02} \\
-F_{13} - iF_{23} - V_2 - iV_1 & A + X_0 + iF_{12} + iV_3 & X_1 + iX_2 + F_{01} + iF_{02} & -iB - X_3 - F_{03} - iV_0 \\
-iB - X_3 + F_{03} + iV_0 & -X_1 + iX_2 + F_{01} - iF_{02} & A - X_0 - iF_{12} + iV_3 & F_{13} - iF_{23} - V_2 + iV_1 \\
-X_1 - iX_2 + F_{01} + iF_{02} & -iB + X_3 - F_{03} + iV_0 & -F_{13} - iF_{23} + V_2 + iV_1 & A - X_0 + iF_{12} - iV_3
\end{pmatrix}
\end{aligned} \tag{107}$$

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[7].

Definition 31 (Clifford conjugate (of a $\mathbb{G}(4,\mathbb{R})$ multi-vector))**.**

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 + \langle \mathbf{u} \rangle_3 + \langle \mathbf{u} \rangle_4 \tag{108}$$

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):

$$\lfloor \mathbf{u} \rfloor_{\{3,4\}} := \langle \mathbf{u} \rangle_0 + \langle \mathbf{u} \rangle_1 + \langle \mathbf{u} \rangle_2 - \langle \mathbf{u} \rangle_3 - \langle \mathbf{u} \rangle_4 \tag{109}$$

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

Definition 32 (Geometric representation of the determinant (of a 4×4 matrix))**.**

$$\begin{array}{rccc}
\det & : & \mathbb{G}(4,\mathbb{R}) & \longrightarrow & \mathbb{R} \\
 & & \mathbf{u} & \longmapsto & \lfloor \mathbf{u}^{\ddagger}\mathbf{u} \rfloor_{3,4} \mathbf{u}^{\ddagger}\mathbf{u}
\end{array} \tag{110}$$

4.4.2 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:

$$\begin{array}{rcccc}
\langle \cdot, \cdot, \cdot, \cdot \rangle & : & \mathbb{V} \times \mathbb{V} \times \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(4,\mathbb{R}) \\
 & & \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle & \longmapsto & \lfloor \mathbf{u}^{\ddagger}\mathbf{v} \rfloor_{3,4} \mathbf{w}^{\ddagger}\mathbf{x}
\end{array} \tag{111}$$

is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{112}$$

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

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \lfloor \psi(q)^{\ddagger} \psi(q) \rfloor_{3,4} \psi(q)^{\ddagger} \psi(q) \tag{113}$$

is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{114}$$

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):

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{115}$$

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

$$\langle \mathbf{O}\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{O}\mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{O}\mathbf{x} \rangle \tag{116}$$

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

$$\langle \mathbf{O} \rangle = \frac{\langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \tag{117}$$

4.4.3 Reduction to Complex Hilbert Space

David Hestenes[8] proposes a geometric algebra formulation of the relativistic wave-function, given as $\psi = \sqrt{\rho} e^{iB/2} e^{\mathbf{F}/2}$. David Hestenes connects his wave-function to a complex number via the reverse $\tilde{\psi} := \sqrt{\rho} e^{iB/2} e^{-\mathbf{F}/2}$, such that $\psi \tilde{\psi} = \rho e^{iB}$. Then, via the Born rule, $|\psi \tilde{\psi}|^2 = \rho^2$.

David Hestenes' representation constitute an algebra of natural states in our framework — in fact, the largest one in 4D which reduces to the complex Hilbert space. Let us select a subset of multi-vectors. The subset will contain all multi-vectors resulting from the multiplication of an even-multi-vector by its own Clifford conjugate. Consistent with our entropy maximization procedure, the elements will also be exponentiated.

$$\boldsymbol{\psi} = \begin{pmatrix} (\exp \frac{1}{2}\mathbf{u}_1)^\ddagger \exp \frac{1}{2}\mathbf{u}_1 \\ \vdots \\ (\exp \frac{1}{2}\mathbf{u}_m)^\ddagger \exp \frac{1}{2}\mathbf{u}_m \end{pmatrix} \tag{118}$$

The form of the elements of $\boldsymbol{\psi}$ are.

$$\begin{aligned} \psi^\ddagger \psi &= \exp \frac{1}{2}(A - \mathbf{F} + \mathbf{B}) \exp \frac{1}{2}(A + \mathbf{F} + \mathbf{B}) && (119) \\ &= \exp \frac{1}{2} A \exp -\frac{1}{2}\mathbf{F} \exp \frac{1}{2}\mathbf{B} \exp \frac{1}{2} A \exp \frac{1}{2}\mathbf{F} \exp \frac{1}{2}\mathbf{B} && (120) \\ &= \exp A \exp \mathbf{B} && (121) \\ &= \rho e^{iB} && (122) \end{aligned}$$

On such states, the quadri-linear map is reduced to the Born rule (a bilinear map):

$$\begin{array}{rcrcl} \langle \cdot, \cdot \rangle & : & \mathcal{A}(\mathbb{V}) \times \mathcal{A}(\mathbb{V}) & \longrightarrow & \mathbb{C} \\ & & \langle \psi, \phi \rangle & \longmapsto & \psi^\dagger \phi \end{array} \tag{123}$$

In our example, and with this bilinear map, $\langle \psi^\ddagger \psi, \psi^\ddagger \psi \rangle = \rho^2$.

Of course, the most interesting applications will be when we do not reduce the framework to the complex Hilbert space and keep the full expressive power of the algebra...

5 Applications (Physics)

5.1 Paths in $\mathbb{E}$ are Path Integrals

Let us now reprise our main result. So far, all sums of programs we have used were over manifests comprised of a single program each:

$$Z = \sum_{q \in \mathbb{Q}} \det \exp -\tau \mathbf{M}(q) \tag{124}$$

How do we extend this to a sum of manifests containing multiple programs? We have to use a Cartesian product on the sets, and a tensor product on the probability amplitudes. For instance, let us consider the following sets of manifests:

$$\begin{aligned} \mathbb{Q}_1 &= \{(p_{1a}), (p_{1b})\} && (125) \\ \mathbb{Q}_2 &= \{(p_{2a}), (p_{2b})\} && (126) \end{aligned}$$

The Cartesian product produces manifests comprised of two elements:

$$\mathbb{Q} = \mathbb{Q}_1 \times \mathbb{Q}_2 = \{(p_{1a}, p_{2a}), (p_{1a}, p_{2b}), (p_{1b}, p_{2a}), (p_{1b}, p_{2b})\} \tag{127}$$

At the level of the probability amplitude, we must apply the tensor product. For instance, we start with a wave-function of one program;

$$\boldsymbol{\psi}_1 = \begin{pmatrix} \exp \mathbf{P}(q_{1a}) \\ \exp \mathbf{P}(q_{1b}) \end{pmatrix} \tag{128}$$

Adding a program-step via a linear transformation produces:

$$\mathbf{T}\boldsymbol{\psi}_q = \begin{pmatrix} T_{00} \exp \mathbf{P}(q_{1a}) + T_{01} \exp \mathbf{P}(q_{1b}) \\ T_{10} \exp \mathbf{P}(q_{1a}) + T_{11} \exp \mathbf{P}(q_{1b}) \end{pmatrix} \tag{129}$$

If we tensor product this wave-function:

$$\boldsymbol{\psi}_2 = \begin{pmatrix} \exp \mathbf{P}(q_{2a}) \\ \exp \mathbf{P}(q_{2b}) \end{pmatrix} \tag{130}$$

along with a program-step:

$$\mathbf{T}'\boldsymbol{\psi}_2 = \begin{pmatrix} T'_{00} \exp \mathbf{P}(q_{2a}) + T'_{01} \exp \mathbf{P}(q_{2b}) \\ T'_{10} \exp \mathbf{P}(q_{2a}) + T'_{11} \exp \mathbf{P}(q_{2b}) \end{pmatrix} \tag{131}$$

Then the tensor product of these states produces manifests as follows:

$$\mathbf{T}\boldsymbol{\psi}_1 \otimes \mathbf{T}'\boldsymbol{\psi}_2 = \begin{pmatrix} (T_{00} \exp \mathbf{P}(q_{1a}) + T_{01} \exp \mathbf{P}(q_{1b}))(T'_{00} \exp \mathbf{P}(q_{2a}) + T'_{01} \exp \mathbf{P}(q_{2b})) \\ (T_{00} \exp \mathbf{P}(q_{1a}) + T_{01} \exp \mathbf{P}(q_{1b}))(T'_{10} \exp \mathbf{P}(q_{2a}) + T'_{11} \exp \mathbf{P}(q_{2b})) \\ (T_{10} \exp \mathbf{P}(q_{1a}) + T_{11} \exp \mathbf{P}(q_{1b}))(T'_{00} \exp \mathbf{P}(q_{2a}) + T'_{01} \exp \mathbf{P}(q_{2b})) \\ (T_{10} \exp \mathbf{P}(q_{1a}) + T_{11} \exp \mathbf{P}(q_{1b}))(T'_{10} \exp \mathbf{P}(q_{2a}) + T'_{11} \exp \mathbf{P}(q_{2b})) \end{pmatrix} \tag{132}$$

Now, each element of the resulting vector is a manifest of multiple programs, but its probability is a sum over a path. Finally, under an appropriate limiting process, both the number of programs in a vector and the number of tensor products grows to infinity resulting in the conventional path integral used in physics.

The probability measure associated to manifests of multiple programs are sum of paths, and recover the path integral in the limiting case.

5.2 Gauge Completeness

The typical gauge theory in quantum mechanics is obtained by the production of a gauge covariant derivative over a $U(1)$ invariance:

$$\psi' = e^{i\theta}\psi \tag{133}$$

Localizing the invariance group $\theta \to \theta(x)$ yields the corresponding covariant derivative:

$$D_\mu = \partial_\mu - i\frac{e}{\hbar}A_\mu(x) \tag{134}$$

Where $A_\mu(x)$ is the gauge field, in this case associated with electromagnetism. 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 specifically constructing said probability measure as a Lagrangian.

However, there exists some gauge for which quantum mechanics using the complex norm is unable to support: for instance, the gauge of the general linear group cannot be supported without breaking the linear evolution of the complex probability amplitude. This is were the increased expressive power of our framework shows its value; in fact, all possible matrix-group gauge can be supported without breaking said linearity. This is because the wave-function connects to the probability measure via a quadri-linear form and can thus remain linear even under the application of the general linear group (in 4D). Extensions to more dimensions are possible.

The fundamental invariance group of our measure is the orientation-preserving general linear group $\mathrm{GL}^+(n, \mathbb{R})$, if the algebra is even, or the complex general linear group $\mathrm{GL}(n, \mathbb{C})$ if the algebra is odd, rather than $U(1)$. Gauging the $\mathrm{GL}^+(n, \mathbb{R})$ group is known to substantially connect to general relativity, as the resulting $\mathrm{GL}(4, \mathbb{R})$-valued field can be viewed as the Christoffel symbols Γ^μ.

Let us first show the general linear invariance.

5.2.1 General Linear Invariance

Consider an arbitrary probability measure from our algebra of natural states in 4D:

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \lfloor \psi(q)^\ddagger \psi(q) \rfloor_{3,4} \psi(q)^\ddagger \psi(q) \tag{135}$$

$$= \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \det \psi(q) \tag{136}$$

A global orientation-preserving general linear transformation $\mathbf{G}$ is applied as follows:

$$\mathbf{G}\boldsymbol{\psi} = \boldsymbol{\psi}' \tag{137}$$

where $\mathbf{G}$ is a 4×4 matrix, and where $\boldsymbol{\psi}$ is a m-dimensional vector whose elements are 4×4 matrices. Thus, $\mathbf{G}\boldsymbol{\psi}$ entails an element-wise multiplication.

Let us now show that invariance with respect to said transformation. The determinant of $\mathbf{G}$ factors out:

$$\rho(\mathbf{G}\psi(q), \mathbf{G}\boldsymbol{\psi}) = \frac{1}{\langle \mathbf{G}\boldsymbol{\psi}, \mathbf{G}\boldsymbol{\psi}, \mathbf{G}\boldsymbol{\psi}, \mathbf{G}\boldsymbol{\psi}\rangle} \det \mathbf{G} \det \psi(q) \tag{138}$$

Finally, since the partition function is simply a sum of determinant, then $\det \mathbf{G}$ can be factored out on the denominator, and the terms cancel:

$$\begin{aligned}
\rho(\mathbf{G}\psi(q), \mathbf{G}\boldsymbol{\psi}) &= \frac{1}{\det \mathbf{G}} \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}\rangle} \det \mathbf{G} \det \psi(q) && (139)\\
&= \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}\rangle} \det \psi(q) && (140)\\
&= \rho(\psi(q), \boldsymbol{\psi}) && (141)
\end{aligned}$$

5.2.2 Lagrangian

This general linear invariance finds itself in any probability measure we might construct from this algebra. Let us now consider the case of a Lagrangian.

A typical Lagrangian density in quantum mechanics relies upon the existence of a measure of the momentum:

$$\overline{P} = \frac{1}{Z} \int_M P(\mathbf{x})\psi(\mathbf{x})^* \psi(\mathbf{x}) \,\mathrm{d}\mathbf{x} \tag{142}$$

and a measure of the potential energy:

$$\overline{T} = \frac{1}{Z} \int_M T(\mathbf{x})\psi(\mathbf{x})^* \psi(\mathbf{x}) \,\mathrm{d}\mathbf{x} \tag{143}$$

Now, if and only if there exists a $\hat{p}$ and a basis of $\psi(\mathbf{x})$ such that $P(\mathbf{x})\psi(\mathbf{x}) = \hat{p}\psi(\mathbf{x})$, then $\hat{p}$ is called the momentum operator. In relativistic quantum mechanics $\hat{p} = \gamma^0 \hbar i \not{\partial}$. Then, the Lagrangian density is a measure of the energy:

$$\mathcal{L}(\mathbf{x}) = \psi^*(\mathbf{x})\hat{p}\psi(\mathbf{x}) - \psi^*(\mathbf{x})\gamma_0 mc^2 \psi(\mathbf{x}) \tag{144}$$

A similar probability measure can be constructed for our framework, using the determinant instead of the complex norm:

$$\overline{P} = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \int_M P(\mathbf{x}) \det \psi(\mathbf{x}, \tau) \, \mathrm{d}\mathbf{x} \tag{145}$$

And the potential energy as:

$$\overline{V} = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \int_M V(\mathbf{x}) \det \psi(\mathbf{x}, \tau) \, \mathrm{d}\mathbf{x} \tag{146}$$

resulting, for the general case, in the Lagrangian density:

$$\mathcal{L}(\mathbf{x}, \tau) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} (P(\mathbf{x}) - V(\mathbf{x})) \det \psi(\mathbf{x}, \tau) \tag{147}$$

Since we have a determinant on top, and a partition function as a sum of determinants at the bottom, the measure is invariant with respect to the general linear group as $\det \mathbf{G}$ will factor out and cancel.

5.2.3 Example Gauge

Since all finite dimensional groups have matrix representations, it then follows that our framework —as it works with any matrices— is able to create a linear probability amplitude for said group; including, of course, those groups resulting from the direct product of groups such as the affine group, producing the metric-affine theory of gravity[9] when gauged, or the Poincaré group producing the Einstein–Cartan gravity theory[10] when gauged. Furthermore, all groups resulting from the direct product of groups are also supported by the framework. Here is a list of examples of groups that are supported by a linear probability amplitude in our framework, but are not with the complex norm only:

Gauge group	Name	Theory
$\mathrm{GL}^+(n, \mathbb{R})$	general linear group	general relativity
$T(n, \mathbb{R}) \times \mathrm{GL}(n, \mathbb{R})$	affine group	metric-affine gravity
$T(n, \mathbb{R}) \times \mathrm{SO}(n, \mathbb{R})$	Poincaré group	Einstein–Cartan gravity theory
$T(n, \mathbb{R}) \times \mathrm{SO}(n, \mathbb{R}) \times U(1) \times SU(2) \times SU(3)$	sm+gr group	"toy unification"
G	any matrix group	general case

(148)

We are not necessarily claiming that these gauges all lead to physically realized theories. Our primary goal here is simply to show that our framework supports any gauge.

In the general case, let $\mathfrak{g}$ be the matrix representation lie algebra of any group G which may include any of the above groups or direct products of groups. Then, consider the following constraints:

$$\sum_{q\in\mathbb{Q}} \rho(q) = 1 \tag{149}$$

$$\sum_{q\in\mathbb{Q}} \rho(q) \operatorname{tr} \mathfrak{g}(q) = \operatorname{tr} \overline{\mathfrak{g}(q)} \tag{150}$$

Solving the Lagrange equation, we obtain:

$$\rho(q) = \frac{1}{Z} \det \exp -\tau(\mathfrak{g}(q)) \tag{151}$$

and the wave-function as:

$$\psi(q) = \exp -\tau(\mathfrak{g}(q)) \tag{152}$$

Here, as before, the exponential map generates the group associated with the algebra:

$$\exp : \tau\mathfrak{g} \to G \tag{153}$$

and contains an evolution term as the Lagrangian multiplier τ which forms a one-parameter sub group of G. Multi-parameter constructions are also possible, as shown in the Annexes.

Regardless of the group used, the probability amplitude will be linear, and said amplitude connects to the probability via the determinant, here acting as a generalization of the Born rule. Consequently, the map from $\psi(q)$ to $\rho(q)$ is invariant with respect to a global transformation of said group. Then, producing a gauge-invariant derivative for the local action of the group $G \to G(q)$ induces a number of compensating gauge fields associated to these groups.

6 Testable Prediction

Certain linear transformations of the wave-function, under the general linear group and its subgroups, would produce richer interference patterns that what is possible merely with complex interference. The possibility of richer interference patterns have been proposed before; specifically, I note the work of B. I. Lev.[11] which suggest (theoretically) the possibility of an extended interference pattern associated with the David Hestenes form of the relativistic wave-function and for the subset of rotors.

We note that interference experiments have paid off substantial dividends in the history of physics and are somewhat easy to construct and more affordable that many alternative experiments.

6.1 Geometric Interference

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

6.1.1 Geometric Algebra Dot Product

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

$$\begin{array}{rcrcl} \cdot & : & \mathbb{G}(2n,\mathbb{R}) \times \mathbb{G}(2n,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} \cdot \mathbf{v} & \longmapsto & \frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) \end{array} \tag{154}$$

For example,

$$\mathbf{u} = A_1 + X_1 e_1 + Y_1 e_2 + B_1 e_{12} \tag{155}$$

$$\mathbf{v} = A_2 + X_2 e_1 + Y_2 e_2 + B_2 e_{12} \tag{156}$$

$$\implies \mathbf{u} \cdot \mathbf{v} = A_1 A_2 + B_1 B_2 - X_1 X_2 - Y_1 Y_2 \tag{157}$$

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 det 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:

$$\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) = \frac{1}{2}\left((\mathbf{u}+\mathbf{v})^\ddagger(\mathbf{u}+\mathbf{v}) - \mathbf{u}^\ddagger\mathbf{u} - \mathbf{v}^\ddagger\mathbf{v}\right) \tag{158}$$

$$= \mathbf{u}^\ddagger\mathbf{u} + \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} + \mathbf{v}^\ddagger\mathbf{v} - \mathbf{u}^\ddagger\mathbf{u} - \mathbf{v}^\ddagger\mathbf{v} \tag{159}$$

$$= \mathbf{u}^\ddagger\mathbf{v} + \mathbf{v}^\ddagger\mathbf{u} \tag{160}$$

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

$$
\begin{aligned}
&\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det\mathbf{u} - \det\mathbf{v}) && (161)\\
&= \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) && (162)\\
&= \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) && (163)\\
&= \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 && (164)\\
&= \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} && (165)
\end{aligned}
$$

6.1.2 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 this case the exponential of $\exp\mathbf{u} = \exp a \exp\mathbf{s}$ because it commutes $\exp a$ commutes with $\exp\mathbf{s}$.

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

$$\psi = \psi_1 + \psi_2 = e^{\frac{2}{n}A_1}e^{\frac{2}{n}\mathbf{S}_1} + e^{\frac{2}{n}A_2}e^{\frac{2}{n}\mathbf{S}_2} \tag{166}$$

$$\tag{167}$$

The general interference pattern will be of the following form:

$$
\begin{aligned}
\det\psi_1 + \psi_2 &= \det\psi_1 + \det\psi_2 + \psi_1\cdot\psi_2 && (168)\\
&= e^{2A_1} + e^{2A_2} + \psi_1\cdot\psi_2 && (169)
\end{aligned}
$$

where $\det\psi_1 + \det\psi_2$ is a sum of probabilities, and where $\psi_1\cdot\psi_2$ is the interference pattern.

6.1.3 Complex Interference (Recall)

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} \tag{170}$$

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

$$\psi^\dagger\psi = \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 \tag{171}$$

$$= e^{A_1}e^{-\mathbf{B}_1}e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2}e^{A_2}e^{\mathbf{B}_2} + e^{A_1}e^{-\mathbf{B}_1}e^{A_2}e^{\mathbf{B}_2} + e^{A_2}e^{-\mathbf{B}_2}e^{A_1}e^{\mathbf{B}_1} \tag{172}$$

$$= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}(e^{-\mathbf{B}_1+\mathbf{B}_2} + e^{-(-\mathbf{B}_1+\mathbf{B}_2)}) \tag{173}$$

$$= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2}\cos(B_1 - B_2)}_{\text{interference}} \tag{174}$$

6.1.4 Geometric Interference in 2D

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{X}_1+\mathbf{B}_1} + e^{A_2}e^{\mathbf{X}_2+\mathbf{B}_2} \tag{175}$$

To lighten the notation we will write it as follows:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{S}_1} + e^{A_2}e^{\mathbf{S}_2} \tag{176}$$

where

$$\mathbf{S} = \mathbf{X} + \mathbf{B} \tag{177}$$

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

$$\psi^\dagger\psi = \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 \tag{178}$$

$$= e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_1}e^{\mathbf{S}_1} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_1}e^{\mathbf{S}_1} \tag{179}$$

$$= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}((e^{\mathbf{S}_1})^\ddagger e^{\mathbf{S}_2} + (e^{\mathbf{S}_2})^\ddagger e^{\mathbf{S}_1}) \tag{180}$$

$$= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{e^{A_1+A_2}(e^{-\mathbf{X}_1-\mathbf{B}_1}e^{\mathbf{X}_2+\mathbf{B}_2} + e^{-\mathbf{X}_2-\mathbf{B}_2}e^{\mathbf{X}_1+\mathbf{B}_1})}_{\text{interference}} \tag{181}$$

6.1.5 Geometric Interference in 4D

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}(\mathbf{X}_1+\mathbf{F}_1+\mathbf{V}_1+\mathbf{B}_1)} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}(\mathbf{X}_2+\mathbf{F}_2+\mathbf{V}_2+\mathbf{B}_2)} \tag{182}$$

To lighten the notation we will write it as follows:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{S}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{S}_2} \tag{183}$$

where

$$\mathbf{S} = \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{184}$$

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

$$\begin{aligned} \lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi &= \lfloor \psi_1^\ddagger \psi_1 \rfloor_{3,4} \psi_1^\ddagger \psi_1 + \lfloor \psi_2^\ddagger \psi_2 \rfloor_{3,4} \psi_2^\ddagger \psi_2 + \psi_1 \cdot \psi_2 && (185) \\ &= e^{2A_1} + e^{2A_2} + \left(e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{S}_1} \right) \cdot \left(e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{S}_2} \right) && (186) \end{aligned}$$

In many cases of interested, the pattern simplifies nicely.

6.1.6 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:

$$\psi = \psi_1 + \psi_2 \tag{187}$$

where

$$\begin{aligned} \psi_1 &= (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1})^\ddagger (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1}) = e^{A_1} e^{\mathbf{B}_1} && (188) \\ \psi_2 &= (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1})^\ddagger (e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1}) = e^{A_2} e^{\mathbf{B}_2} && (189) \end{aligned}$$

Thus

$$\psi = e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} \tag{190}$$

The quadri-linear map, becomes a bilinear map:

$$\begin{aligned} \psi^\dagger \psi &= (e^{A_1} e^{-\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2})(e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2}) && (191) \\ &= e^{A_1} e^{-\mathbf{B}_1} e^{A_1} e^{\mathbf{B}_1} + e^{A_1} e^{-\mathbf{B}_1} e^{A_2} e^{\mathbf{B}_2} + e^{A_2} e^{-\mathbf{B}_2} e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2} e^{A_2} e^{\mathbf{B}_2} && (192) \\ &= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1 + A_2} \cos(B_1 - B_2)}_{\text{complex interference}} && (193) \end{aligned}$$

6.1.7 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:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{B}_2} \tag{194}$$

The sub-product part is:

$$\psi^{\ddagger}\psi = (e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2})(e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}) \tag{195}$$

$$= e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2} \tag{196}$$

$$= e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} + 2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \tag{197}$$

The final product is:

$$\begin{aligned}
\lfloor\psi^{\ddagger}\psi\rfloor_{3,4}\psi^{\ddagger}\psi &= (e^{A_1}e^{-\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2} + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}) \\
&\quad \times (e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} + 2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}) &(198)\\
&= e^{A_1}e^{-\mathbf{B}_1}e^{A_1}e^{\mathbf{B}_1} + e^{A_1}e^{-\mathbf{B}_1}e^{A_2}e^{\mathbf{B}_2} + e^{A_1}e^{-\mathbf{B}_1}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\
&\quad + e^{A_2}e^{-\mathbf{B}_2}e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2}e^{A_2}e^{\mathbf{B}_2} + e^{A_2}e^{-\mathbf{B}_2}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\
&\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_1}e^{\mathbf{B}_1} \\
&\qquad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_2}e^{\mathbf{B}_2} \\
&\qquad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} &(199)\\
&= e^{2A_1} + e^{2A_2} + 2e^{A_1+A_2}\cos(B_1 - B_2) \\
&\quad + e^{A_1}e^{-\mathbf{B}_1}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\
&\quad + e^{A_2}e^{-\mathbf{B}_2}2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} \\
&\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_1}e^{\mathbf{B}_1} \\
&\quad + 2e^{\frac{1}{2}(A_1+A_2)}e^{-\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}e^{A_2}e^{\mathbf{B}_2} \\
&\quad + 4e^{A_1+A_2} &(200)\\
&= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2}\cos(B_1 - B_2)}_{\text{complex interference}} \\
&\quad \underbrace{+2e^{\frac{1}{2}(A_1+A_2)}(e^{A_1} + e^{A_2})\cos\frac{1}{2}(B_1 - B_2) + 4e^{A_1+A_2}}_{\text{deep phase interference}} &(201)
\end{aligned}$$

6.1.8 Geometric Interference in 4D (Deep Spinor Rotation)

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

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2} \tag{202}$$

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

$$\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi \tag{203}$$

Let us start with the sub-product:

$$\psi^\ddagger \psi = (e^{\frac{1}{2}A_1} e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2})(e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2}) \tag{204}$$

$$\begin{aligned} &= e^{\frac{1}{2}A_1} e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_1} e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} \\ &\quad + e^{\frac{1}{2}A_2} e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} \end{aligned} \tag{205}$$

$$= e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (e^{-\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{F}_2}) + e^{-\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{F}_1}) \tag{206}$$

$$= e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \tag{207}$$

where $R = e^{\frac{1}{2}\mathbf{F}}$, and where $\tilde{R} = e^{-\frac{1}{2}\mathbf{F}}$.
The full product is:

$$\begin{aligned} \lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi = &\Big(e^{A_1} e^{-\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \Big) \\ &\times \Big(e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \Big) \end{aligned} \tag{208}$$

$$\begin{aligned} &= e^{A_1} e^{-\mathbf{B}_1} e^{A_1} e^{\mathbf{B}_1} + e^{A_1} e^{-\mathbf{B}_1} e^{A_2} e^{\mathbf{B}_2} + e^{A_1} e^{-\mathbf{B}_1} e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \\ &\quad + e^{A_2} e^{-\mathbf{B}_2} e^{A_1} e^{\mathbf{B}_1} + e^{A_2} e^{-\mathbf{B}_2} e^{A_2} e^{\mathbf{B}_2} + e^{A_2} e^{-\mathbf{B}_2} e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \\ &\quad + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{A_1} e^{\mathbf{B}_1} \\ &\qquad + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{A_2} e^{\mathbf{B}_2} \\ &\qquad + e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) e^{\frac{1}{2}(A_1+A_2)} e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) \end{aligned} \tag{209}$$

$$= e^{2A_1} + e^{2A_2} + 2e^{A_1+A_2} \cos(B_1 - B_2) \tag{210}$$

$$+ e^{\frac{1}{2}(A_1+A_2)} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)(\tag{211}$$

$$e^{A_1} (e^{\frac{1}{2}(-\mathbf{B}_1+\mathbf{B}_2)} + e^{\frac{1}{2}(\mathbf{B}_1-\mathbf{B}_2)}) \tag{212}$$

$$+ e^{A_2} (e^{\frac{1}{2}(\mathbf{B}_1-\mathbf{B}_2)} + e^{\frac{1}{2}(-\mathbf{B}_1+\mathbf{B}_2)})) \tag{213}$$

$$+ e^{A_1+A_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)^2 \tag{214}$$

$$\begin{aligned} &= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2} \cos(B_1 - B_2)}_{\text{complex interference}} \\ &\quad + \underbrace{2e^{\frac{1}{2}(A_1+A_2)} (e^{A_1} + e^{A_2}) (\tilde{R}_1 R_2 + \tilde{R}_2 R_1) (\cos \frac{1}{2}(B_1 - B_2)) + e^{A_1+A_2} (\tilde{R}_1 R_2 + \tilde{R}_2 R_1)^2}_{\text{deep spinor interference}} \end{aligned} \tag{215}$$

6.1.9 Geometric Interference Experiment (Sketch)

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 produce a series of interference patterns of decreasing complexity, and as such they provide a method to experimentally identify which group of geometric transformations the world obeys, 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 gauge-theoretical theory of gravity; whereas identification of a reduced interference pattern (produced by $A, \mathbf{F}, \mathbf{B}$) and subsequent showing a failure to observe the full general linear interference ($\mathbf{X} \to 0, \mathbf{V} \to 0$), would suggest at most spinor-level interference.

In any such case, a general experimental setup would send a particle into two distinct paths. Then, either: a) one of the paths undergoes a general linear transformation, while the other doesn't or b) both paths undergo a different general linear transformation. Then, the paths are recombined to produce an interference pattern on a screen. Depending on the nature of the transformation, a deformation of the interference pattern based on the geometry of the setup should be observed.

One would can further utilize the non-commutativity of the general linear transformations to identify only the difference between complex-interference and general linear interference. One would apply the same general linear transformations to each path, but would reverse the order in which the transformations are applied. The resulting interference pattern would then be compared to a case where both paths are transformed in the same order. Then, complex-interference, as it is fully commutative, would predict the same interference pattern irrespective of the order the transformations are applied in — whereas, with general linear interference, as it is non-commutative, would predict different interference patterns.

To achieve this, it may be necessary to use a three-dimensional detector, whose idealized construction is a homogeneous bath of impurities (allowing photons to 'click' anywhere within the volume of the detector), instead of a two-dimensional screen, since the opportunity for non-commutative behaviour often kicks in at three dimensions or higher. In a real experiment, it is probably easier to use a 2d x-y screen, and stepping it along an orthogonal z-axis, capturing the 2d interference pattern at each step, then numerically reconstructing the volumetric interference pattern out of the steps.

7 Discussion (Science + Physics)

An interesting consequence of integrating both science and physics into a singular mathematical theory is that one can speak, with some amount of mathematical rigour, about subject matters which exceeds the normal scope of physics, but not necessarily that of science + physics. For instance, it is in this increased scope that the interpretation of quantum mechanics including the collapse of the wave-function resides... further explaining why physics, by itself, is and was unable to settle these questions. In this section, we attack problems that specifically are unresolvable in physics alone, but where the union of physics + science may provide direction. Please note that this section is a bit more "relaxed" and "speculative" that the other sections, and is meant to make the reader think, pounder and to encourage discussion, rather than provide final answers.

7.1 Observer Bias — An Origin Story

We have defined the laws of nature as the solution to an entropy maximization problem. In contrast, the definition of the observer references only a probability measure in general and as such is not necessarily entropy-maximized. The laws of nature are consequently the preferred solution only to a subset of all possible observers; called the natural, or unbiased observers. Other so-called *biased observers* may find that alternative explanations are 'better' or 'preferred' than the laws of nature, for a myriad of reasons. This can open up quite the can of worm, but here we will restrict our investigation to a small subset of the biased observers; specifically those that remarkably manage to improve upon the laws of nature as an explanatory tool. This subset of observers still maximizes the entropy, and thus do recover the laws of nature, but also induces a bias on experimental space, such that said bias is 'felt' as an entropic force. The modification is an improvement because a specific type of bias, called the *survivorship bias*, <u>must</u> be accounted for, so as to produce a complete description of reality (from the perspective of a random observer).

For an example of bias, consider that I certainly think of myself as an observer. Then, it is perhaps the case that I consider certain experiments to be of special interest to me, such as possibly those experiments that involve my own body. For instance, I may wish to know if I have an allergy to pollen. To know that, I might go for a summer in a rural area, versus a summer in an urban area, and see if I develop any allergies in the rural area but none in the urban area, thereby making pollen the prime suspect. In contrast, the question "can I out-swim a shark?" can be legitimately interesting as a scientific question, but it goes against my observer bias to try it — the well known maxim being that if one dies in experimental space, then one dies in real life. A natural observer who might find the experiment "can I out-swim a shark?" to be equally desirable to grab as a unit of knowledge as any other experiments, could be expected to terminate faster on average than one who is biased, thus overall producing a less informative path in experimental space. Since the length of an observer's path in experimental space is subject to the choice of experiments, an entropy maximization procedure which attempts to maximize knowledge must also maximize survivorship bias to achieve the best possible result.

At the statistical-physics level, this biased selection on the level of program-steps, is expressed as an entropic force on the path of the observer pointing towards regions of experimental space preferential to said bias. Consider a function $\mathrm{BIAS}(\mathbf{q})$ which assigns a real value to each program-step. Maximizing the entropy with a constraint in the direction of BIAS:

$$\sum_{\mathbf{q}\in\mathbb{Q}^n} \nabla(\rho(\mathbf{q})\,\mathrm{BIAS}(\mathbf{q})) = \nabla\overline{\mathrm{BIAS}} \tag{216}$$

produces, in the probability measure, an exponential suppression term for each path running counter to the direction of said bias. One might get the impression that an observer must be able to exert control on a physical system to steer its path willfully along the direction of increased bias, but fear not, one does not actually need to postulate such an ability. The symmetry of the situation does not require more than a mere correlation. Indeed, one can simply take as a starting point the set of all observers, and since the length of a path of each observer is in $[0, \infty]$, where 0 could mean a Boltzmann brain spontaneously forming then disintegrating, and ∞ means an observer lucky enough to experience immortality, then a survivorship bias, as an entropic force, is necessarily present for such set, iff the average path length is greater than 0. This is certainly always the case if the number of observers is finite, and if it is infinite, we may select a limiting condition which preserves the average length given by the entropy for finite cases. A correlation, —thankfully, no causation is necessary—, between high survivorship bias and a random selection of a typical observer is thus statistically expected.

In the general computing case, knowing the length of ones' path in advance is either undecidable, or at least highly unfeasible. If the complexity may grow unbounded, a plausibly optimal general strategy is to maximize future freedom of action, defined as an entropic force $F = T\nabla S\tau$ and previously proposed as a general definition of intelligence[12]. In this case, the intelligence bias would be a specific implementation of a survivorship bias.

The entropic force resulting from survivor bias gives us an opportunity to re-interpret the anthropic principle (a mere truism) as an emergent entropic force (thus improving its explanatory power), and to further rethink the odds of the Boltzmann brain (thus attributing a minimalistic origin to the observer). First, let us recall that the anthropic principle, in its minimal version, attributes a binary [YES] or [NO] value to possible universes regarding whether they support life or not, and concludes self-evidently that only [YES] universes admit observers — and in a slightly less minimal formulation, the anthropic principle is a claim on the numerical value of the constants of nature. Second, rather than thinking of a minimal brain randomly generated in a thermal soup (Boltzmann brain), we will switch the space from the thermal soup to a random path in experimental space. A postulated *algorithmic-Boltzmann brain* is overwhelmingly favoured, statistically, to be situated in a region of experimental space of high *survivorship bias*, rather than in a region of very little. For example, a 'society' of X billion members, each member being the culmination of Y billion years of natural selection pressure, would represent a state of experimental space where the cumulative entropic force of *survivorship bias* outclasses, by many orders of magnitude, the force generated by a singular Boltzmann brain in a thermal soup, rendering the former overwhelmingly likelier to contain observers than the latter. Observers are favoured by entropy to occupy the 'algorithmic niche' of highest survivorship bias. Comparatively, a thermal soup, the preferred environment of the Boltzmann brain, is an environment of no (or even negative) survivor bias. Although it may be possible, in principle, for a Boltzmann brain to be spontaneously produced, it cannot do so without the system countering a massive negative survivorship bias penalty on the entropy, thus making the occurrence of such incredibly unlikely overall.

Since the laws of nature are obtained under the assumption of an unbiased observer, they will invariably make the universe appear "unnaturally" fit, with respect to the random expectation, for a spontaneously emerging algorithmic-Boltzmann brain which has said bias.

7.2 Interpretation of Quantum Mechanics

In statistical physics, one considers a number of constraint on the entropy, such as:

$$\sum_{q\in\mathbb{Q}} \rho(q)) = 1 \tag{217}$$

$$\sum_{q\in\mathbb{Q}} \rho(q)E(q) = \overline{E} \tag{218}$$

$$\sum_{q\in\mathbb{Q}} \rho(q)V(q) = \overline{V} \tag{219}$$

Maximizing the entropy produces the Gibbs measure:

$$\rho(q) = \frac{1}{Z}\exp\bigl(-\beta(E(q) + pV(q))\bigr) \tag{220}$$

Our framework admits an origin in the same entropy maximization procedure as that of statistical physics. The contraints are:

$$\sum_{q\in\mathbb{Q}} \rho(q) = 1 \tag{221}$$

$$\sum_{q\in\mathbb{Q}} \rho(q) \operatorname{tr} \begin{pmatrix} 0 & -h(q) \\ h(q) & 0 \end{pmatrix} = \operatorname{tr} \begin{pmatrix} 0 & -\overline{h} \\ \overline{h} & 0 \end{pmatrix} \tag{222}$$

In the statistical physics interpretation, the constraints simply refer to various instruments that are acting on the system. For instance, an energy constrain implies an energy measure tool, and a volume constrain likewise implies a volume measure tool, each yielding a sequence of measurements on the system.

In the case of the trace, and since the trace remains invariant to a similarity transformations, the interpretation of statistical physics remain valid but must be adjusted to account for an 'instrument' which is unable to measure *the phase*. With said constraint, consider how simple and how free of clutter, this foundation to quantum mechanics is: *A statistical physics ensemble whose entropy is constrained by an unmeasured phase, is mathematically identical to quantum mechanics.* Everything we need for quantum mechanics follows from this; complex Hilbert space, measurements, states as rays of an Hilbert space, observables, unitary transformation, time evolution, etc. are all a consequence of maximizing the entropy under this simple constraint. Finally, under this constraint, the wave-function is simply a special case of the Gibbs measure.

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \det \psi(q) \tag{223}$$

7.2.1 Interpretation of the Collapse

In our interpretation, the collapse of the wave-function is a behaviour inherited from how the laws of nature are derived, and results from a clash between two scientifically valid, but different, definition of the observer.

Let us first define two observers: one called the *trivial* observer, and the other called the *natural* observer. We are already familiar with the natural observer: it is the observer which maximizes the entropy of the random selection of the reference manifest, thus identifying the laws of nature as a result. Now, for the trivial observer: it is one who assigns a probability measure of one to the reference manifest on the grounds that it happened, and zero to all others on the grounds that they did not happen.

Let us now start with simple examples, then we will ramp up the complexity as we go along. The goal is to identify a pattern.

Example 1 - Coin toss: Let us take the example of the following sequence of HEAD (H) or TAILS (T):

$$\text{HHT} \tag{224}$$

How can one associate a probability measure to this sequence? The possibilities boil down to two mutually exclusive ways. The first is simply to claim the sequence HHT is the one that has happened in reality and thus its probability is post-facto given the value of 1. Its main disadvantage is that it is quite trivial. The second way has the advantage that it much less trivial: we assume that there could have been alternatives to what has happened, then solve for a probability measure over these alternatives:

1. We start by listing all possible sequences we assume could have been alternatives to the actual sequence (and also add the actual sequence to the set):

$$\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT} \tag{225}$$

2. Then, we create a probability measure which assumes, all other things being equal, that each sequence is equally likely (there are other plausible assumptions, unfair coin toss, maximizing entropy, etc.):

$$\rho(q) = \frac{1}{Z} \tag{226}$$

Note the difference of 'model-depth' between the first and second approach. The first approach, although a completely correct description of what has occurred is nonetheless very trivial and quite uninformative, but the second one which includes a plurality of sequences that never happened, is much more interesting!

Example 2 - Statistical physics: This logic is repeated for any system that are attributed a probabilistic measure, including those of statistical physics. For instance, if one measures the volume of a gas in a box, and its energy, one would get two finite sequences of measurements:

$$V_1 V_2 V_3 V_4 V_5 V_6 \tag{227}$$

$$E_1 E_2 E_3 E_4 E_5 E_6 \tag{228}$$

One can then claim that the probability of each of these sequences is 1 because they are the ones that have actually happened, but here again it feels quite trivial to do so. A much less trivial approach is to list all possible alternatives to this sequence, then solve for a probability measure that maximizes the entropy of this ensemble of possibilities. In this case, the alternative sequences are presented in the form of two constrains on the entropy:

$$\overline{V} = \sum_{q \in \mathbb{Q}} \rho(q) V(q) \tag{229}$$

$$\overline{E} = \sum_{q \in \mathbb{Q}} \rho(q) E(q) \tag{230}$$

The probability measure, obtained by maximizing the entropy under these constraints is the familiar Gibbs measure:

$$\rho(q) = \frac{1}{Z} \exp -\beta(E(q) + pV(q)) \tag{231}$$

Example 3 - Quantum mechanics: The same occurs if we use the constrains of a unmeasured phase of a system of statistical physics, as we have used in this work to recover quantum mechanics from entropy maximization. In nature, we would be presented with a sequence of measurements, each having an unmeasured phase:

$$(h_1^* h_1), (h_2^* h_2), (h_3^* h_3), (h_4^* h_4^*) \tag{232}$$

One could normalize the probability of this sequence to 1, after-all it is indeed the one that has happened! However, here also, it is much more interesting to assume there could have alternatives to this sequence. Then, maximizing the entropy for the set of all possible sequences, expressed as a constraint:

$$\operatorname{tr} \begin{pmatrix} 0 & \overline{h} \\ \overline{h} & 0 \end{pmatrix} = \sum_{q \in \mathbb{Q}} \rho(q) \operatorname{tr} \begin{pmatrix} 0 & h(q) \\ h(q) & 0 \end{pmatrix} \tag{233}$$

produces the wave-function has the probability measure. Then, playing with a wave-function is much less trivial that playing with a singular fixed sequence, and thus we find intellectual satisfaction.

Example 4 - Relativistic wave-function Even in the case of the full-blown relativistic wave-function, the logic is the same. One experimentally measures a sequence of numbers:

$$\operatorname{tr} \mathbf{M}_1, \operatorname{tr} \mathbf{M}_2, \operatorname{tr} \mathbf{M}_3 \tag{234}$$

But, to avoid triviality, one assumes that, rather that being normalized to one, those numbers ought to be the result of a random selection between alternatives, in this case given as:

$$\operatorname{tr} \overline{\mathbf{M}} = \sum_{q \in \mathbb{Q}} \rho(q) \operatorname{tr} \mathbf{M}(q) \tag{235}$$

and obtains the general linear wave-function as a result of maximizing the entropy:

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \det \psi(q) \tag{236}$$

Discussion - Statistical interpretation: The statistical interpretation[13] of quantum mechanics is, in essence, an attempt to justify why and when it is acceptable for us to assume alternative sequences (even if only one happened), by restricting the domain of applicability to *similarly prepared ensembles*. Then, if similarly prepared ensembles produces different sequences, it implies there could have been alternatives. This is an improvement, but it has the loop-hole that one can always claim that any group of *similarly prepared ensembles* is itself a larger uniquely prepared ensemble, and thus sending us directly back to square one. What is then the correct way to think about this?

Discussion - Manifests: The largest sequence associated with any physical system is the manifest of the present reality. We have access to only one such sequence and we are unable to produce similarly prepared ensembles on that scale. Yet, the same entropy maximization process is required to get the laws of nature. Here as well, we could claim the path to the actual manifest is normalized to a probability of one, because it is the one that has occurred. However, if one assumes as alternatives all possible paths between all possible manifests, then maximizes of entropy over the selection of one path from this set, then one obtains the laws of nature. As we have, by definition, only one "similarly prepared ensemble", the statistical interpretation requiring a plurality of similarly prepared ensembles cannot help us make sense of this.

The question then is: why do we need alternative possibilities to define the laws of physics? Do we need not just reference the present manifest — are they not the laws that pertain to the present reality? An answer is available and palatable but only if the totality of the laws of nature are obtained by entropy maximization (as in the present case) — otherwise it is a near impossible pill to swallow. Remarkably, the laws of nature are not the laws of the reality that has occurred (the probability of such normalizes to one) —a more appropriate construction to understand what has occurred would be to use a scientific theory which is an axiomatic representation of the manifest, and subject to future falsification—, but instead are the laws regarding what *could have been*, or what could be. It is the boundaries of what we assume 'could have been', post entropy maximization, that constitutes the laws of nature. Finally, our assumption of what 'could have been' is made unique by assuming all possible manifests, and those boundaries are also made unique by entropy maximization.

For instance, if I claim to be a physicist, then it is plausible to claim that I could have instead been a doctor, even if the later did not occur because, presumably, 'being a doctor' does not in principle violate the laws of nature. However, it is not plausible to claim I could have instead been superman, because in this case the laws of nature would presumably be violated. In this example, the laws of nature rules out superman but not the doctor, even if reality rules out both the doctor and superman. The laws of nature rules out less than what reality rules out. The laws of nature, since they are the same for all possible manifests, cannot not rule out what could have plausibly happened.

The appearance of a wave-function collapse is the result of a clash between two valid definitions of observers: the trivial (yielding at most a scientific theory of the manifest) vs the natural observer (yielding the laws of nature). There exists no link between the two from a physics stand-point —because one *is* physics, and the other isn't—, but the link exists from a science+physics stand-point —because they are both sicence—.

References

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

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

[3] Stephen Wolfram. *A new kind of science*, volume 5. Wolfram media Champaign, IL, 2002.

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

[5] Kohtaro Tadaki. A generalization of chaitin's halting probability omega and halting self-similar sets. *Hokkaido Math. J.*, 31(1):219–253, 02 2002.

[6] John Baez and Mike Stay. Algorithmic thermodynamics. *Mathematical. Structures in Comp. Sci.*, 22(5):771–787, September 2012.

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

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

[9] Friedrich W Hehl, J Dermott McCrea, Eckehard W Mielke, and Yuval Ne'eman. Metric-affine gauge theory of gravity: field equations, noether identities, world spinors, and breaking of dilation invariance. *Physics Reports*, 258(1-2):1–171, 1995.

[10] K Hayashi and T Shirafuji. Gravity from the poincaré gauge theory of fundamental interactions, i. general formulation, ii. equations of motion for test bodies and various limits, iii. weak field approximation, iv. mass and energy of particle spectrum. *Prog. Theor. Phys*, 64:866–883, 1980.

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

[12] Alex Wissner-Gross. Ted speaker: Alex wissner-gross: A new equation for intelligence, 2016.

[13] Leslie E Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42(4):358, 1970.

A Notation

S will denote the entropy, $\mathcal{A}$ the action, L the Lagrangian, and $\mathcal{L}$ the Lagrangian density. 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 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 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 $a_0a_1 + x_0x_1\hat{\mathbf{x}} + y_0y_1\hat{\mathbf{y}} + b_0b_1\hat{\mathbf{x}} \wedge \hat{\mathbf{y}}$.

B Lagrange equation

The Lagrangian equation to maximize is:

$$\mathcal{L}(\rho, \alpha, \tau) = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q)\right) + \tau \operatorname{tr}\left(\overline{\mathbf{M}} - \sum_{q \in \mathbb{Q}} \rho(q)\mathbf{M}(q)\right) \tag{237}$$

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:

$$\frac{\partial \mathcal{L}}{\partial \rho(p)} = -k_B \ln \rho(p) - k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p) \tag{238}$$

$$0 = k_B \ln \rho(p) + k_B + \alpha + \tau \operatorname{tr} \mathbf{M}(p) \tag{239}$$

$$\implies \ln \rho(p) = \frac{1}{k_B} \left(-k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p)\right) \tag{240}$$

$$\implies \rho(p) = \exp\left(\frac{-k_B - \alpha}{k_B}\right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(p)\right) \tag{241}$$

$$= \frac{1}{Z} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(p)\right) \tag{242}$$

where Z is obtained as follows:

$$1 = \sum_{q \in \mathbb{Q}} \exp\left(\frac{-k_B - \alpha}{k_B}\right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q)\right) \tag{243}$$

$$\implies \left(\exp\left(\frac{-k_B - \alpha}{k_B}\right)\right)^{-1} = \sum_{q \in \mathbb{Q}} \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q)\right) \tag{244}$$

$$Z := \sum_{q \in \mathbb{Q}} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(q)\right) \tag{245}$$

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:

$$\overline{\mathbf{M}}_1 = \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_1(q) \tag{246}$$

$$\vdots \tag{247}$$

$$\overline{\mathbf{M}}_n = \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_n(q) \tag{248}$$

Then the Lagrange equation becomes:

$$\begin{aligned} \mathcal{L} = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q)\right) + \tau_1 \operatorname{tr} \left(\overline{\mathbf{M}}_1 - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_1(q)\right) + \dots \\ + \tau_n \operatorname{tr} \left(\overline{\mathbf{M}}_n - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_n(q)\right) \end{aligned} \tag{249}$$

and the measure references all n constraints:

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{\tau_1}{k_B}\mathbf{M}_1(q) - \cdots - \frac{\tau_n}{k_B}\mathbf{M}_n(q)\right) \tag{250}$$

B.2 Multiple constraints - General Case

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

$$\overline{M}_{00}\begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{00}(q) \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \tag{251}$$

$$\vdots \tag{252}$$

$$\overline{M}_{01}\begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{01}(q) \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \tag{253}$$

$$\vdots \tag{254}$$

$$\overline{M}_{nn}\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{nn}(q) \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} \tag{255}$$

For a $n \times n$ matrix, there are n^2 constraints.

The probability measure which maximizes the entropy is as follows:

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{1}{k_B}\boldsymbol{\tau} \odot \mathbf{M}(q)\right) \tag{256}$$

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