The Design of a Formal System of Science: A Compelling Foundation for Physics

Alexandre Harvey-Tremblay[1]

[1]Independent scientist, aht@protonmail.ch

October 30, 2021

Abstract

A formal system of science is presented as a candidate for a more powerful foundation of physics. Rather than starting with an axiomatic theory of physics as is typically done (quantum mechanics, general relativity, classical mechanics, etc), we instead had the idea to start at the level of science by designing a formal system of science. Our goal was then to attempt, as a challenge, to derive the laws of physics as a theorem of the scientific method. This exercise turned out surprisingly productive; correcting definitions and resolving open problems spawning philosophy, mathematics, science (philosophy of) and physics. Some of our notable results include a definition of experimental data (done purely mathematically), a definition of the observer (again done purely mathematically), and a derivation of the laws of physics as the product of the scientific method (and thus, we allege, done uncontroversially). This group of results constitutes the primary items of modern theoretical physics currently missing mathematical formalisation. The first part of the paper consists of constructing an experimental basis that is purely mathematical. For this we employ the set of all halting programs and we leverage modern notions of undecidability to produce a system that never runs out of new knowledge to discover, supporting a perpetual application of the scientific method. The system culminates in a definition for the observer in terms of a measure space over halting programs, such that maximizing its entropy produces a quantum-computation model of physics able to support, remarkably, up to quantum gravity. Finally, testable predictions are proposed.

Contents

1 Introduction

In classical philosophy an axiom is a statement which is (believed to be) 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 demoted to simply be 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 loopholes, 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 loophole: 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 now 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. What used to be "justified true belief" would now become "infallible true belief". This eliminates the loophole, but it is an unpopular solution because adding it is assumed to reduce epistemology to radical skepticism in which almost nothing is knowledge, thus rendering epistemology non-comprehensive.

Here, we will adopt the insight of Kirkham regarding the requirement of infallibility, whilst also resolving the non-comprehensiveness objection. To do so, we will structure our statements such that they are individually infallible, yet as a group form a Turing complete language. Turing completeness will guarantee comprehensiveness. Our tool of choice will be halting programs. As we will see, halting programs carry all desired features to make this possible. Using them, we will be able to tackle epistemological knowledge using infallible statements, as put forward by Kirkham. That may sound impressive, but there is a catch. Although we can describe the knowledge once it is acquired, acquiring it in the first place will be difficult, in some case even arbitrarily difficult. After-all, halting programs are of course subject to the halting problem. Indeed, we still have to identify those program that halts from those that do not, and because of the halting problem, as there exists no general algorithm able to do so, the system will be irreducibly experimental (more on that later).

With these two features, the sum-total of epistemological knowledge becomes well-defined and is in fact recursively enumerable. The implications of such a construction should not be underestimated as this was a long standing goal of the ancients of philosophy, even if attempts to crack it have generally fallen out of favour in modern times. In any case, our formal system of science will thoroughly exploit this construction.

Finally, let us state that attempts to find a logical basis for epistemology have been made ad nauseam in the past but they failed for primarily two reasons. First, they were attempted before Gödel-type theorems were known or appreciated, and attempts were directed at constructing *decidable* logical bases. Secondly, instead of directing efforts to recursively enumerable bases following the discovery of said theorems, efforts simply felt out of favour. However, it is possible to construct a logical basis for epistemology provided that such basis is recursively enumerable (and not decidable), and further the limitations induced by recursive enumeration ought to instead be seen as an *opportunity*; in this case, to create a formal system to map out epistemology, such that it may serve as the foundation to a formalization of science.

For more general information regarding the connection between mathematics, science and programs, we recommend the works of Gregory Chaitin[3, 4, 5]. A familiarity with his work is assumed. Let us now continue.

1.1 Halting Programs as Knowledge

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

Let us take the example of a statement that may appear as an obvious true statement such as "$1+1=2$", but is in fact not infallible. Here, we will provide the 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.

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

If we use the tools of theoretical computer science we can produce statements free of all loopholes, thus ensuring they are infallible. Those statements are programs:

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

$$\textit{iff } \mathrm{UTM}(\mathrm{TM}, p) \textit{ halts, then } f = (\mathrm{TM}, p) \textit{ is a halting program} \tag{1}$$

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 an halting program because it is verifiable on all universal Turing machines.

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

1.2 Halting Images

We will now use halting programs to redefine the foundations of mathematics in terms of *halting images*, replacing *formal axiomatic systems*.

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

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

Definition 2 (Halting Image)**.** *Let* $\mathbb{D} = \mathrm{Dom}(\mathrm{UTM})$ *be the set of all halting programs for a given universal Turing machine. A halting image, or simply a image,* $\mathbf{m}$ *of* n *halting program is an element of the* n*-fold Cartesian product of* $\mathbb{D}$*:*

$$\mathbf{m} \in \mathbb{D}^n, \quad \mathbf{m} := \big((\mathrm{TM}_1, p_1), \ldots, (\mathrm{TM}_n, p_n)\big) \tag{2}$$

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

- *A halting image contains some, but not necessarily all, programs of the domain of the universal Turing machine.*

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

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

Formal axiomatic systems do not excel at pure description because they are be more akin to an *interpretation* of mathematical knowledge based on a preference of some patterns or tools (we like sets, thus ZFC!, or we prefer categories, thus category theory!). New knowledge and new problems will eventually force one to challenge this preference. Not so with halting images! Halting images are the true starting point of the logical inquiry as they represent an infallible and non-preferential description of mathematical knowledge.

We will now explore the concept more rigorously.

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

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

The theorem implies that we will never run out of new knowledge to discover.

Definition 3 (Premises)**.** *Let* $\mathbf{m}$ *be a halting image. The premises in* $\mathbf{m}$ *are defined as the set of all* TM *in* $\mathbf{m}$*:*

$$\mathbb{P} := \mathrm{proj}_1(\mathbf{m}) \tag{3}$$

Definition 4 (Theorems)**.** *Let* $\mathbf{m}$ *be a halting image. The theorems of* $\mathbf{m}$ *are defined as the set of all* p *in* $\mathbf{m}$*:*

$$\mathbb{T} := \mathrm{proj}_2(\mathbf{m}) \tag{4}$$

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

Definition 6 (Scope (of a premise))**.** *The set of all theorems in* $\mathbf{m}$ *in which a premise is repeated is called the scope of the premise. For instance if* $\mathbf{m} = \{(\mathrm{TM}_1, p_1), (\mathrm{TM}_1, p_2)\}$*, 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 halting image formulation to the standard formal axiomatic system (FAS) formulation:

Definition 7 (Enumerator (of a FAS))**.** *A function* $\mathrm{enumerator}_{\mathrm{FAS}}$ *is an enumerator for* FAS *if it recursively enumerates the theorems of* FAS*. For instance:*

$$\mathrm{enumerator}_{\mathrm{FAS}}(s) = \begin{cases} 1 & \mathrm{FAS} \vdash s \\ \nexists / \text{ does-not-halt} & \text{otherwise} \end{cases} \tag{5}$$

Definition 8 (Domain (of FAS))**.** *Let* FAS *be a formal axiomatic system, let* $\mathbf{m}$ *be a halting image and let* $\mathrm{enumerator}_{\mathrm{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then the domain of* FAS*, denoted as* $\mathrm{Dom(FAS)}$*, is the set of all sentences* $s \in \mathbb{L}$ *which halts for* $\mathrm{enumerator}_{\mathrm{FAS}}$*.*

Definition 9 (Formal Axiomatic Representation)**.** *Let* FAS *be a formal axiomatic system, let* $\mathbf{m}$ *be a halting image and let* $\mathrm{enumerator}_{\mathrm{FAS}}$ *be a function which recursively enumerates the theorems of* FAS*. Then* FAS *is a formal axiomatic representation of* $\mathbf{m}$ *iff:*

$$\mathrm{Dom(FAS)} = \mathrm{proj}_2(\mathbf{m}) \tag{6}$$

Definition 10 (Factual Isomorphism)**.** *Two formal axiomatic systems* FAS_1 *and* FAS_2 *are factually-isomorphic if and only if* $\mathrm{Dom}(\mathrm{FAS}_1) = \mathrm{Dom}(\mathrm{FAS}_2)$.

1.3 Discussion — The Mathematics of Knowledge

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

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

In all cases, knowledge is an algorithm along with an input, such that the algorithm halts for it, lest it is not knowledge. The set of all known pairs forms a halting image.

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

$$\mathbf{m} := \big((\mathrm{TM}_1, A), (\mathrm{TM}_1, \neg A)\big) \tag{7}$$

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

```
fn main(input: String){
 if p=="A" {
```

```
  return;
 }
 if p=="not A"{
  return;
 }
 loop();
}
```

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

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

Now, suppose one has a sizeable halting image which may contain a plurality of pairs:

$$\mathbf{m} := \big((\mathrm{TM}_1, p_1), (\mathrm{TM}_2, \neg p_1), (\mathrm{TM}_1, p_2), (\mathrm{TM}_2, p_1), (\mathrm{TM}_2, \neg p_3)\big) \quad (8)$$

Here, the negation of some, but not all, is also present across the pairs: in this instance, the theorems p_1 and p_3 are negated but for different premises. What interpretation can we give to such elements of a halting image? 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 halting programs are pairs which contain a premise and a theorem. The premise contains the 'definition' under which the flavour qualifies as real ice cream. A flavour with a large spread is considered real ice cream by most definitions (i.e. vanilla or chocolate ice cream), and one with a tiny spread would be considered real ice cream by only very few definitions (i.e. tofu-based ice cream). Then, within this example, the presence of p_1 and its negation 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 halting image ought to be sufficiently flexible to handle this, without itself exploding into a contradiction.

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

1.4 Discussion — Idempotency, and Epistemic Regress

Let us now take a well known axiomatic theory of physics to use as an example. Specifically, let us pick non-relativistic quantum mechanics. Typical formulations admit a number of axioms, and specifically the Von-Neumann formulation is as follows: the 1) physical states are the unit vectors $\mathbf{u}$ of a complex Hilbert space, 2) observables are self-adjoint operators ($O = O^{\dagger}$), 3) the expectation value of an observable is the 'Born rule' applied to a statistical sum weighted by the value of the observable (equivalent to the inner product : $\langle O \rangle = \langle O\mathbf{u}|\mathbf{u}\rangle$).

Since people are naturally curious they may wonder: "is there a more fundamental explanation for these axioms? — can we derive, say, the Born rule more fundamentally?". A possible direction, currently under some amount of investigation, is towards a non-local hidden variable theory. Now suppose one successfully invents an axiomatic non-local hidden variable theory which does derive those rules, allegedly, from *"first principles"*, and further assumes that such a reformulation would gain sufficient traction to replace the current formulation. This may provide some of us with a small dopamine hit, but surely the next generation will eventually begin to wonder the same as we did but applied to this new formulation: "is there a more fundamental explanation for these axiomatic hidden-variables? — why this specific set of hidden-variables and not others?". The cycle of questioning is expected to simply repeated itself. The desire for the ultimate fundamental theory has not been satisfied.

In general all formal axiomatic systems are subject to a perpetual realization of epistemic regress: "a belief is justified because it is based on another belief that is also justified...". For an axiomatic system of physics, epistemic regress is simply imported from its mathematical origin as a formal axiomatic system; any FAS is logically entailed by infinitely-many other FAS, thus one is free to perpetually replace a given FAS by another, believed to be more fundamental, FAS while keeping the epistemology of system equivalent. Even if one were to invent an axiomatic basis able to derive the Born rule as theorem, and somehow show it to be an equivalent representation of quantum mechanics, then this new construction will depend upon a new set of axioms itself subject to the epistemic regress problem. One may perhaps fool himself or herself into believing a more fundamental theory has been achieved, however one has merely identified a new element out the set of all equivalent FAS for the given physical theory.

Now, let us try the same with halting images and see if they also are subject to the epistemic regress problem. Can one, as a challenge, produce a "more fundamental" halting image for a given halting image? Before we investigate, let us note that a halting image uniquely represents a specific instance of mathematical knowledge. To change the halting image with another, entails to represent a different instance of mathematical knowledge. For instance, say one has $\mathbf{m}_a := (f_1, f_2)$. Now suppose one wishes to produce a 'more fundamental' halting image of $\mathbf{m}_a$, calling it $\mathbf{m}_b$. Well either $\mathbf{m}_a = \mathbf{m}_b$ at which point neither are more or less fundamental (they are the same), or $\mathbf{m}_a \neq \mathbf{m}_b$ at which point they describe different systems. One cannot improve the fundamentality of the halting image in the typical sense of replacing it with an alternative, as all alternatives describe different systems.

What if we try to "improve the fundamentality" of a halting image by picking better formal axiomatic representations for it? In this case we are free to pick any of its formal axiomatic representations, but any such choice simply link back to the same halting image — the knowledge remains the same. We thus say that halting images are *idempotent* representations of knowledge. Here we use the word in the following sense: idem + potence = (same + power). Changing the formal axiomatic representation does not change the halting image.

The relationships and differences can be visualized by noting that a halting image admits infinitely-many possible formal axiomatic representations, whereas a formal axiomatic system admits a single halting image as a representation of its domain. Idempotency, with respects to a change of formal axiomatic representation, occurs at the level of the halting image representation. Finally, epistemic regress immediately terminates for halting images due to this idempotency.

1.5 Axiomatic Information

The first result of interest will be the introduction of axiomatic information. Since halting images are idempotent (and thus admit no more-fundamental representations), any account for the elements of any particular halting image is relegated to having been 'randomly picked', according to a probability measure ρ, from the set of all possible halting programs. We can quantify the information of the pick using the entropy.

Definition 11 (Axiomatic Information)**.** *Let* $\mathbb{D} = \mathrm{Dom}(\mathrm{UTM})$ *be the domain of a universal Turing machine. Then, let* $\rho : \mathbb{D} \to [0,1]$ *be a probability measure over* $\mathbb{D}$*. Finally, let* $\mathbf{m}$ *be a halting image subset of* $\mathbb{D}^n$*. The axiomatic information of a single element of* $\mathbf{m}$ *is quantified as the entropy of* ρ*:*

$$S = -\sum_{q \in \mathbb{D}} \rho(q) \ln \rho(q) \tag{9}$$

For instance, a well-known (non-computable) probability measure regarding a sum of prefix-free programs is the Halting probability[6] of computer science:

$$\Omega = \sum_{p \in \mathrm{Dom(UTM)}} 2^{-|p|} \implies \rho(p) = 2^{-|p|} \tag{10}$$

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

2 The Formal System of Science

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

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

Definition 12 (Scientific method)**.** *A function which recursively enumerates epistemological knowledge, is called a scientific method.*

Mathematically speaking, this is a very simple definition. First, it is indeed purely mathematical, and formal, and in fact coincide with the definition of a universal function — which is a non-controversial mathematical concept. We have previously defined epistemological knowledge as halting programs (this made it comprehensible) and it's domain as that of a universal Turing machine (this made is comprehensive). Now we simply define a recursive enumeration function for said domain and we give it a name. The notion of the scientific method, a previously informal (naive[1]) construction, is now imported into pure mathematics and as such we have produced a net gain for science, compared to not having it.

The features of the scientific method are found implicitly in the definition. Indeed, implicit in said definition lies a requirement for the function to verify the input to be knowledge by running its corresponding program to completion, and reporting success once proven to halt. That it may or may not halt is the hypothesis, and the execution of the function is the 'experiment' which verifies the hypothesis. If an input runs for an abnormally long time, one may try a different hypothesis hoping to reach the conclusion differently. Since knowledge is element-wise infallible, each terminating experiments are formally reproducible

[1]We refer to the word naive in the mathematical sense; i.e. as a theory which is not formalized. No negative connotation are implied.

as many times as one needs to, to be satisfied with its validity. All of the tenets of the scientific method are implicit in the definition, and its domain is that of knowledge itself, just as we would expect from the scientific method. Finally, the domain of the function is arbitrarily complex and countably infinite, therefore we never run out of new knowledge allowing for a perpetual and never ending application of the scientific method. Mathematically, it is a remarkably simple definition for such an otherwise rich concept.

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

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

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

3. The elimination of all naive concepts or notions (no more "magic" or "handwaving") is now required. If one has a worldview that relies upon a plurality of non-formalizable concepts, then one's worldview will not survive this formalization. For many, this is interpreted as killing the "fun" or the "imagination" from reality. Since this is the first time a fully formalizable model of reality has been presented, then no one's pre-existing worldview is expected to survive (ouch!).

Does one even stand a chance at maintaining his or her informal (naive) worldview, when facing such definitions? Many of our base definitions were carefully chosen to merely *match* and *rebrand* pre-existing and well respected mathematical definitions; this was a strategic choice to make it incredibly difficult (not to say impossible) to find fatal flaws. In our experience the battery of aversion we typically receive boils down to an equivalent formulation of "I can't find the error, but it *must* be wrong because [my worldview] requires it to be different" or variations of "I just don't see it, bye!". Of course, no actual pinpointing of a fatal error is ever produced (otherwise we would either correct it, or immediately abandon the project altogether depending on the nature of the error presented).

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

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

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

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

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

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

```
  }
  return false;
}
```

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

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

Let us return to our discussion on aversion. At the other end of the aversion spectrum, we find some readers (it would be overly optimistic to expect it from all readers, but hopefully some) that accept and understand that the proposed system induces what amounts to a checkmate position for informal (naive) worldview. Of those readers, most will then condition themselves to accept a re-adjustment of their worldview such that it becomes conductive to complete formalisation. This will be no easy task, because many concepts central to mainstream science and physics are *not* formalizable absent of importing what people like Max Tegmark calls physical baggage (which, unless such baggage is re-expressible as programs, we consider it to be non-formalizable). For these readers, their desire for formalization is greater than their attachment to their informal (naive) worldview, and they are willing to make the necessary sacrifices to work completely formally. Just like the beginnings of mainstream science quickly displaced (most of) the charlatans and purveyors of quackery physical or medical products, formal science via its definition of epistemology displaces the charlatans and purveyors of informal "intellectual products". Resistance from those purveyors is of course inevitable.

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

2.1 Terminating Protocols (as Knowledge about Nature)

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

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

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

> [Protocol]: In natural and social science research, a protocol is most commonly a predefined procedural method in the design and

> implementation of an experiment. Protocols are written whenever it is desirable to standardize a laboratory method to ensure successful replication of results by others in the same laboratory or by other laboratories. Additionally, and by extension, protocols have the advantage of facilitating the assessment of experimental results through peer review.

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

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

Hinkelmann, Klaus and Kempthorne, Oscar in 'Design and Analysis of Experiments, Introduction to Experimental Design'[7] note the following:

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

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

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

Let us now produce a thought experiment to help us understand how the concepts connect, and how they are organized in a formal system of science.

2.1.1 The Universal Verifier (Thought experiment)

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

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

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

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

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

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

How would a theoretical physicist work with such a machine?

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

The mathematical definition of the Physical Verifier is as follows:

Definition 13 (Physical Verifier)**.** *Let p be a sentence of a language $\mathbb{L}$, which we call a protocol. A physical verifier is defined as a function:*

$$\begin{array}{rcccl} \mathrm{PV} & : & \mathbb{L} & \longrightarrow & \{\mathbb{L}, \nexists\} \\ & & p & \longmapsto & r \end{array} \tag{11}$$

- *The domain of the verifier* Dom(PV) *includes the set of all protocols which terminates for it.*

Definition 14 (Experimental Image)**.** *Let* Dom(PV) *be the domain of a Physical Verifier* PV*. An experimental image is a tuple of n elements of* Dom(PV)*:*

$$\mathbf{m} := (\mathrm{Dom}(\mathrm{PV}))^n \tag{12}$$

- *The experimental image only contains protocols that have terminated.*
- *The experimental image corresponds, intuitively, to a sequence of related or unrelated experiments, that have been verified by the machine.*
- *The experimental image corresponds to an instance of natural knowledge (knowledge about nature). It represents knowledge in the epistemological sense because the protocols maps to halting programs, and knowledge about nature specifically, because the machine performs the requested experiment in nature (just like an experimental physicist would).*
- *Finally, as epistemological knowledge is comprehensive, then all systems which admits knowledge, physical or otherwise, can be represented in the form of a specific experimental image associated to a specific verifier, and said image constitutes a complete representation of the knowledge said system has produced thus far for its operator.*

Let us now define the universal version of the Physical Verifier. A Universal Verifier is a Physical Verifier that is able to construct any possible Physical Verifiers and have it run a protocol.

Definition 15 (Universal Verifier)**.** *Let* PV *be a Physical Verifier, and* p *be a sentence of a language* $\mathbb{L}$ *which we call a protocol. Then a Universal Verifier is defined as:*

$$\mathrm{UPV}(\mathrm{PV}, p) = \mathrm{PV}(p) \tag{13}$$

The relation must of course hold for all possible Physical Verifiers, and all possible protocols.

Let us make note of the following comments:

1. The Universal Verifier takes a pair of input (PV, p), whereas the Physical Verifier takes a single input (p).
2. The Universal Verifier must first construct the provided PV, or otherwise simulate its behaviour on a computer, before it can instruct it to run the second input which is the protocol p. Comparatively, the Physical Verifier can begin to execute the protocol immediately.
3. A protocol executed by a Physical Verifier can reference a physical context, if such context is exposed by the machine in its instructions. For instance, the instruction **capture** can refer to the EV taking a picture with a camera and outputting it in jpeg format. But what is a camera? This is not specified in the protocol, but in the instructions, and thus represent physical baggage of a similar nature as the concept of a 'fellow expert' does in peer-review.

4. The Universal Verifier, in contrast, must include a complete description of the PV, enough to perfectly construct or simulate it. Physical baggage such as a camera cannot be referenced informally in the specifications of the PV, otherwise the UPV cannot construct the PV. If the protocol calls for the usage of a camera, then the behaviour of the camera must be completely specified in formal terms within the instructions to build the PV. Consequently, all rules and/or physical laws which are required to be known, including any initial conditions, must be precisely provided in the PV description, so that the UPV can construct it. For some highly convoluted experiments, such as what is the effect on the price of selling 100 shares of AMCE on the NYSE on Jan 1 2021 at 12pm? or is this a good recipe for apple pie?... the aphorism from Carl Sagan "If you wish to make an apple pie from scratch, you must first invent the universe" is adopted quite literally by the universal verifier. The Universal Verifier must create (or at least simulate) the universe, let interstellar matter accretes into stars, let biological evolution run its course, then finally conduct the experiment once the required actors are in play, by feeding them apple pie. For a universal verifier, certain protocols, due to their arbitrary physical contexts requirements and general complexity, cannot be created more efficiently than from literal scratch and by going through the full sequence of events until the end of the experiment.

5. Since it is free of physical context, or baggage, the universal verifier meets the definition of the Universal Turing Machine. However, it can 'construct' other machines too, whereas a universal Turing machine can only simulate other machines.

The differences between Physical Verifiers and Universal Verifiers is sufficiently distinctive to allow us to split explanatory models, or theories, into two classes. The first class comprise those whose formulation references the physical context, and the other class comprise those that do not. Dependance upon the physical context means that the first class is formulated in reference to a specific Physical Verifier. Those theories will be called the 'scientific theories' and they are subject to a larger class of scientific theorems, such as falsifiability, eventual revisions, valid for one possible world as opposed to all possible worlds, admits non-formalizable physical baggage, etc. as more knowledge is produced by the PV. The formulation of such theories can in many case be extremely "efficient" because they can directly refer to exposed physical baggage such as that of an electric wire, without having to first specify the creation of the universe in which the wire can exist. Or they can feed apple pie to subjects, by referencing apples and flour directly, without having to first create the universe. They include all scientific theories which references physical baggage directly, for instance biology, sociology and many many others. We will classify them in a moment.

The second class of theories include those that are formulated in reference to a universal verifier, and as such are context-free. In this class, all objects (such as electrical wire) must be formally specified (or derived from simpler formally specified concepts) and cannot be referenced brutely by manufacturer ID numbers or as physical baggage. This class comprises the most fundamental models. Said models will not be dependant on the physical context, and as such they are valid for all possible worlds. Those theories are by their nature and scope completely inviolable. They refer to the possibilities and limits of constructing any constructible objects and/or to the realization of any knowledge-bearing protocol. This class of theories will be presented starting from section 2.2 onwards, and it comprises the main result of this paper. They are called the (fundamental) 'laws of physics'. In our classification, and quite remarkably, they are distinct from the scientific theories.

2.1.2 Classification of Scientific Theories

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

$$\mathrm{proj}_2(\mathbf{m}) \cap \mathrm{Dom}(\mathrm{ST}) \neq \varnothing \tag{14}$$

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

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

$$\mathrm{proj}_2(\mathbf{m}) = \mathrm{Dom}(\mathrm{ST}) \tag{15}$$

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

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

$$\mathrm{Dom}(\mathrm{ST}) \subset \mathrm{proj}_2(\mathbf{m}) \tag{16}$$

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

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

$$\mathrm{proj}_2(\mathbf{m}) \subset \mathrm{Dom}(\mathrm{ST}) \tag{17}$$

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

$$\mathbb{S} := \mathrm{Dom}(\mathrm{ST}) \setminus \mathbf{m} \tag{18}$$

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

2.1.3 The Fundamental Theorem of Science

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

Theorem 2 (The Fundamental Theorem of Science)**.** *Let* $\mathbf{m}_1$ *and* $\mathbf{m}_2$ *be two halting images, such that the later includes and is larger than the former:* $\mathbf{m}_1 \subset \mathbf{m}_2$*. If* ET_2 *is an empirical theory of* $\mathbf{m}_2$*, then it follows that* ET_2 *is a predictive theory of* $\mathbf{m}_1$*. Finally, up to factual-isomorphism,* $\mathrm{Dom}(\mathrm{ET}_2)$ *has measure* 0 *over the set of all distinct domains spawned by the predictive theories of* $\mathbf{m}_2$*.*

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

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

2.1.4 Completing the Model (Final Details)

We now introduce a few more definitions to complete the formal system of science.

Definition 20 (Domain of science)**.** *We note* $\mathbb{D}$ *as the domain* (Dom) *of science. We can define* $\mathbb{D}$ *in reference to a universal Turing machine* UTM *as follows:*

$$\mathbb{D} := \mathrm{Dom}(\mathrm{UTM}) \tag{19}$$

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

Definition 21 (Halting Space)**.** *Let* $\mathbf{m}$ *be a halting image comprised of* n *halting programs, and let* $\mathbb{M} = \bigcup_{i=1}^{n} \mathrm{proj}_i(\mathbf{m})$ *be the set comprised of the halting programs of* $\mathbf{m}$*. The halting space* $\mathbb{E}$ *of* $\mathbf{m}$ *is the "powertuple" of* $\mathbf{m}$*:*

$$\mathbb{E} := \bigcup_{i=0}^{n} (\mathbb{M})^i \tag{20}$$

- *Put simply, halting space is the set of all possible halting images (including the empty halting image).*
- *Conceptually, a powertuple is similar to a powerset where the notion of the set is replaced by that of the tuple.*
- *All elements of a halting space are halting images, and all "sub-tuples" of a halting image are elements of its halting space.*

Theorem 3 (Scientific method (Constructive proof of existence))**.** *Existence of the scientific method spawning the entire domain of epistemology.*

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. $\square$

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.

2.2 Axiomatic Foundation of Physics

We have used a dovetailing algorithm in Theorem 3 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 one? Perhaps we would find it appealing to optimize for an efficient algorithm, or 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 property of the algorithm which implements the strategy. We must optimize for axiomatic information.

Axiomatic information is manifest under the assumption that the halting image, as it is an idempotent formulation of knowledge admitting no more-fundamental representation, cannot have an earlier deterministic reason (as that would of course be a more-fundamental representation). Axiomatic information does not represent knowledge itself, rather it encodes the *state* a knowledge-bearing system is in, under the imposition that the state cannot be compressed deterministically (or at least that any attempted compression of such would misguide us).

The strategy to construct the 'correct' model of the scientific method will be to *maximize the axiomatic information* gained from the scientific method as it produces an halting image, and this means, in the technical sense, to maximize the entropy of a probability measure on halting space. Maximizing the entropy of the *state* of the system, will give us, via the resulting probability measure, a physical theory detailing and limiting the possible transformations of such *states*, which are then used to limit, bound and define the actions or steps permissible for knowledge creation and verification. Maximum entropy will further guarantee the idempotence of knowledge under every transformations of the probability measure.

The play here between information and entropy can also be interpreted in the sense of Claude Shannon's theory of information, a 'solution' to the model (a specific halting image) is a 'message' selected from the set of all possible elements of the halting space, and the model is chosen to make this message maximally informative for the observer.

To have information (in the information-theoretical sense) one must work with a probability measure. With this we introduce the definition of the observer. This is the system's only axiom.

Axiom 1 (Observer)**.** *An observer, denoted as $\mathcal{O}$, is a measure space over halting space:*

$$\mathcal{O} := (\mathbf{m}, \mathbb{E}, \rho : \mathbb{E} \to [0, 1]) \tag{21}$$

where ρ is a probability measure, $\mathbf{m}$ is a halting image, and $\mathbb{E}$ is the halting space of $\mathbf{m}$. The definition is reminiscent of a measure space in measure theory, with the important difference that the elements of $\mathbb{E}$ are tuples and not sets. The consequences of this difference will be investigated in the main result section, where a prescription to derive a measure from this triplet is provided.

An observer is a specialization of the definition of the halting image in the sense that an halting image is an element of halting space, and the observer is a *measure space* over halting space. 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, inherited from the requirement of maximizing axiomatic information. Just like we did earlier with a "one-liner" definition of the scientific method, and then showed that the richness of the concept was implicit in the relatively simple definition, here a similar richness will be recovered but for the domain of physics. Although it is still obscure at this point, this definition of the observer will perfectly coincide with what we understand an observer to be in quantum physics, and also to what we understand it to be in general relativity, and will near effortlessly resolve many still open problems.

Thesis 1 (Model of Physics)**.** *The probability measure that maximizes the entropy of $\mathcal{O}$ constitutes the laws of physics.*

The laws of physics are a specialization of the definition of the observer, in the sense that an observer admits a measure over experimental space, and the laws of physics are the *entropy-maximized measure* of the same measure space.

The subject matter of the laws of physics derived here-in coincide with a coalescence of three mathematically related but philosophically distinct concepts: the Universal Turing Machine, the Universal Constructor and the Universal Verifier, and thus are able to account for all construction and verification rules whether physical, simulated or mathematical and over all possible knowledge-bearing states. This setup will allow us to derive the laws of physics under incredibly strong claims of comprehensiveness. This level of proof exceeds anything else available in the present arsenal of physics that the author is aware of.

3 Main Result

Let us now use these definitions to derive the laws of physics from first principle, show that they at least recover the familiar laws of physics, and then investigate extended results and predictions.

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

$$\mathcal{O}_i := (\mathbf{m}_i, \mathbb{E}, \rho : \mathbb{E} \to [0,1]) \tag{22}$$

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

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

$$M := (X, \Sigma, \mu(X)) \tag{23}$$

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

1. We assign a non-negative number to each element of $\mathbb{E}$.
2. We equip said numbers with the addition operation, converting the construction to a vector space.
3. We maximize the entropy of a single halting program under the effect of constraints, by using the method of the Lagrange multipliers.

4. We prescribe that any and all constraints on said entropy must remain invariant with respect to a change of basis of said vector space.
5. We use the tensor product n-times over said vector space to construct a probability measure of n-tuples of halting programs.
6. We use the direct sum to complete the measure over the whole of halting space by combining the measures of different sizes as a single measure.

Explicitly, we maximize the entropy:

$$S = -\sum_{f \in \mathbb{M}} \rho(f) \ln \rho(f) \tag{24}$$

subject to these constraints:

$$\sum_{f \in \mathbb{M}} \rho(f) = 1 \tag{25}$$

$$\sum_{f \in \mathbb{M}} \rho(f) \operatorname{tr} \mathbf{M}(f) = \operatorname{tr} \overline{\mathbf{M}} \tag{26}$$

where $\mathbb{M}$ is the set of halting programs associated with the halting image $\mathbf{m}$, where $\mathbf{M}(f)$ are a matrix-valued maps from $\mathbb{M}$ to $\mathbb{C}^{n\times n}$ representing the linear transformations of the vector space and where $\overline{\mathbf{M}}$ is a element-by-element average matrix.

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

$$\operatorname{tr} \mathbf{M} = \operatorname{tr} \mathbf{B}\mathbf{M}\mathbf{B}^{-1} \tag{27}$$

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:

$$\mathcal{L} = -k_B \sum_{f \in \mathbb{M}} \rho(f) \ln(f) + \alpha \left(1 - \sum_{f \in \mathbb{M}} \rho(f)\right) + \tau \left(\operatorname{tr} \overline{\mathbf{M}} - \sum_{f \in \mathbb{M}} \rho(f) \operatorname{tr} \mathbf{M}(f)\right) \tag{28}$$

where α and τ are the Lagrange multipliers. The explicit derivation is made available in Annex B. This derivation (minus the trace) is standard and shown in most introductory textbooks of statistical physics to derive the Gibbs measure. The result of the maximization process is:

$$\rho(f,\tau) = \frac{1}{Z(\tau)} \det \exp -\tau \mathbf{M}(f) \tag{29}$$

where

$$Z(\tau) = \sum_{f \in \mathbb{M}} \det \exp -\tau \mathbf{M}(f) \tag{30}$$

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

$$\rho(f) = \frac{1}{Z} \det \exp \big(\mathbf{P}(f)\big) \det \exp \big(-\tau \mathbf{M}(f)\big) \tag{31}$$

where

$$Z = \sum_{f \in \mathbb{M}} \det \exp \big(\mathbf{P}(f)\big) \det \exp \big(-\tau \mathbf{M}(f)\big) \tag{32}$$

3.1 Completing the Measure

From the previous entropy maximization procedure, we have produced a measure over a sum of singular halting programs. Whereas the measure we are after is a sum over the whole of the halting space of a given halting image, which contains all sub-tuples of the halting image. Completing the measure over said space will require us to sum over differently-sized tuples. To do so, first, we will use the tensor product to produce measures summing over multiple elements, and second, we will use the direct sum to combine the differently-sized measures into a single final measure.

3.1.1 Split to Amplitude / Probability Rule

Before we are able to proceed with both the tensor product and the direct sum, we must introduce a split over the mathematical operators of the measure. I was hoping to avoid doing this until deeper into the section on the physics, because it implies an element of physics that is too delicious not to discuss right away, but I do not see an easy way to to perform the upcoming operation (tensor product/direct sum) without first doing the split.

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

$$\rho(f,\tau) = \frac{1}{Z} \det \psi(f,\tau) \tag{33}$$

where

$$\psi(f,\tau) = \exp\big(\mathbf{P}(f)\big) \exp\big(-\tau \mathbf{M}(f)\big) \tag{34}$$

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

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

$$\boldsymbol{\psi} := |\psi\rangle := \begin{pmatrix} \psi(f_1,\tau) \\ \psi(f_2,\tau) \\ \vdots \\ \psi(f_n,\tau) \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \tag{35}$$

3.1.2 Tensor Product

So far, the sums we have used were over halting images comprised of a single program each. How do we extend this to halting images containing multiple programs? We have to use a Cartesian product on the sets of halting images and a tensor product on the probability amplitudes. For instance, let us consider the following sets of halting programs:

$$\mathbb{M}_1 = \{(f_{1a}), (f_{1b})\} \tag{36}$$

$$\mathbb{M}_2 = \{(f_{2a}), (f_{2b})\} \tag{37}$$

The Cartesian product produces halting images comprised of two elements:

$$\mathbf{m} \in \mathbb{M}_1 \times \mathbb{M}_2 = \{(f_{1a}, f_{2a}), (f_{1a}, f_{2b}), (f_{1b}, f_{2a}), (f_{1b}, f_{2b})\} \tag{38}$$

At the level of the probability amplitude, the Cartesian product of sets translates to the tensor product. For instance, we start with a wave-function of one program;

$$\boldsymbol{\psi}_1 = \begin{pmatrix} \exp \mathbf{P}(f_{1a}) \\ \exp \mathbf{P}(f_{1b}) \end{pmatrix} \tag{39}$$

Adding a program-step via a linear transformation produces:

$$\mathbf{T}\boldsymbol{\psi}_1 = \begin{pmatrix} T_{00} \exp \mathbf{P}(f_{1a}) + T_{01} \exp \mathbf{P}(f_{1b}) \\ T_{10} \exp \mathbf{P}(f_{1a}) + T_{11} \exp \mathbf{P}(f_{1b}) \end{pmatrix} \tag{40}$$

If we tensor product this wave-function:

$$\boldsymbol{\psi}_2 = \begin{pmatrix} \exp \mathbf{P}(f_{2a}) \\ \exp \mathbf{P}(f_{2b}) \end{pmatrix} \tag{41}$$

along with a program-step:

$$\mathbf{T}'\boldsymbol{\psi}_2 = \begin{pmatrix} T'_{00} \exp \mathbf{P}(f_{2a}) + T'_{01} \exp \mathbf{P}(f_{2b}) \\ T'_{10} \exp \mathbf{P}(f_{2a}) + T'_{11} \exp \mathbf{P}(f_{2b}) \end{pmatrix} \tag{42}$$

Then the tensor product of these states produces the probability measure of a halting image as follows:

$$\mathbf{T}\boldsymbol{\psi}_1 \otimes \mathbf{T}'\boldsymbol{\psi}_2 = \begin{pmatrix} (T_{00} \exp \mathbf{P}(f_{1a}) + T_{01} \exp \mathbf{P}(f_{1b}))(T'_{00} \exp \mathbf{P}(f_{2a}) + T'_{01} \exp \mathbf{P}(f_{2b})) \\ (T_{00} \exp \mathbf{P}(f_{1a}) + T_{01} \exp \mathbf{P}(f_{1b}))(T'_{10} \exp \mathbf{P}(f_{2a}) + T'_{11} \exp \mathbf{P}(f_{2b})) \\ (T_{10} \exp \mathbf{P}(f_{1a}) + T_{11} \exp \mathbf{P}(f_{1b}))(T'_{00} \exp \mathbf{P}(f_{2a}) + T'_{01} \exp \mathbf{P}(f_{2b})) \\ (T_{10} \exp \mathbf{P}(f_{1a}) + T_{11} \exp \mathbf{P}(f_{1b}))(T'_{10} \exp \mathbf{P}(f_{2a}) + T'_{11} \exp \mathbf{P}(f_{2b})) \end{pmatrix} \tag{43}$$

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

3.1.3 Direct Sum

In the previous section, we have introduce a way to produce measure summing over halting images having multiple element but of fixed sizes. Here, we wish to produce a measure that sums over halting images of different sizes. Taking the direct sum of the measures of different sizes (where each individual size is produced from the tensor product), accomplishes the goal and yields an amplitude given has follows:

$$|\psi\rangle = |\psi_1\rangle \oplus (|\psi'_1\rangle \otimes |\psi'_2\rangle) \otimes (|\psi''_1\rangle \otimes |\psi''_2\rangle \otimes |\psi''_3\rangle) \oplus ... \tag{44}$$

In quantum field theory, and when $\mathbf{M}(f)$ is reduced to the complex field, these are the states of a Fock Space, which we have obtained here simply my maximizing the entropy of the measure associated with our simple definition of the observer (Axiom 1).

3.2 Discussion - Fock Space, Measures over Tuples

We may consider it even more fundamental to interpret our result from the angle of measure theory in the sense that an entropy-maximized measure over the tuples of a tuple-space (as an extension to typical measure theory defined for the subsets of a set) induces a Fock Space, along with the appropriate probability rule (Born rule) for use in quantum mechanics. Rather can being of "obscure" origins, the measures used in quantum mechanics would thus result quite intuitively from this simple extension of measure theory defined for sets, to tuples.

We should mention that, although tuples can represent anything, in our system Axiom 1 requires the tuples to represent halted program states. But this is a very unrestrictive constraints; it simply enforces, at the most general level possible, that all states of the wave-function (or Fock Space) are comprehensible to the scientific method.

3.3 Overview

We will first provide a small overview of the main result and how the physics manifest itself, then we will provide a thorough investigation in two parts:

Part 1: We will show that a special case of this result is an equivalent representation of standard quantum mechanics, and discuss the implications of this formulation.

Part 2: We will show that this result, in the general case, is a quantum theory of gravity which adheres to both 1) the axioms of quantum mechanics and 2) to the theory of general relativity.

3.3.1 Connection to Computation

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

$$|\psi_a\rangle = \begin{pmatrix} 0 \\ \vdots \\ n \end{pmatrix} \tag{45}$$

Which evolves unitarily to a final state:

$$|\psi_b\rangle = U_0 U_1 \dots U_m |\psi_a\rangle \tag{46}$$

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:

$$\begin{aligned}
&\text{step 1} && |\psi_b\rangle = U_0 U_1 \dots U_p |\psi_a\rangle && (47)\\
&\text{step 2} && |\psi_c\rangle = U'_0 U'_1 \dots U'_q |\psi_b\rangle && (48)\\
&\vdots && && (49)\\
&\text{step k} && |\psi_{k'}\rangle = U'_0 U'_1 \dots U'_v |\psi_k\rangle && (50)\\
&\vdots && && (51)
\end{aligned}$$

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.

Comparatively, the linear transformations of our main result are here interpreted in the same manner as those used in quantum computations, but extended to the general linear group.

As discussed we can split our main result into a first step which is linear with respect to a 'probability amplitude', and a second which connects the amplitude to the probability:

$$\rho(f,\tau) = \frac{1}{Z} \det \psi(f,\tau) \tag{52}$$

where

$$\psi(f,\tau) = \exp\big(\mathbf{P}(f)\big) \exp\big(-\tau \mathbf{M}(f)\big) \tag{53}$$

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

$$\boldsymbol{\psi} := |\psi\rangle := \begin{pmatrix} \psi(f_1,\tau) \\ \psi(f_2,\tau) \\ \vdots \\ \psi(f_n,\tau) \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix} \tag{54}$$

Paths will be constructed by chaining transformations on those vectors:

$$|\psi_b\rangle = \underbrace{\mathbf{T}_1 \mathbf{T}_2 \dots \mathbf{T}_n}_{\text{computing steps}} |\psi_a\rangle \tag{55}$$

As more transformations are chained, progressively richer halting images are constructed. Paths in halting space are realized by completing the missing computational steps required for a starting-point halting image to be the end-point halting image.

Comparatively, quantum mechanical computations are simply a special cases when he transformations are unitary:

$$|\psi_b\rangle = \underbrace{\mathbf{U}_1\mathbf{U}_2\ldots\mathbf{U}_n}_{\text{computing steps}}|\psi_a\rangle \tag{56}$$

3.3.2 Matrix-Valued Vector and Transformations

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

$$\mathbf{v} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_m \end{pmatrix} \tag{57}$$

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

$$\mathbf{T} = \begin{pmatrix} \mathbf{M}_{00} & \ldots & \mathbf{M}_{0m} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{m0} & \ldots & \mathbf{M}_{mm} \end{pmatrix} \tag{58}$$

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

$$a\mathbf{v} = \begin{pmatrix} a\mathbf{M}_1 \\ \vdots \\ a\mathbf{M}_m \end{pmatrix} \tag{59}$$

We are now ready to begin investigating the main result as a physical theory.

4 Foundation of Physics (Standard QM)

Remarkably, and as we will now see our main result is able to produce, as theorems, the standard axioms of quantum mechanics –including the famously elusive wave-function collapse mechanism, and the origin of the Born rule– thus it constitute, necessarily, a more fundamental formalism of QM that all other alternatives requiring axiomatic definitions. Finally, it necessitate, implies and guarantees not only ensemble interpretation of quantum mechanics, but also derives it from maximum entropy principles. It thus offers a "total solution" to the foundations, origins and interpretation of QM.

The study of the main result will be in two parts. In the first part, to recover QM, we will reduce the expressivity of the constraint of the main result to that of the matrix representation of the complex number, as opposed to a general matrix. In this special case, we will recover standard QM without gravity. In the section following this one, we will investigate the unadulterated version of the main result and show that it produces a complete theory of quantum gravity adhering both to 1) all axioms of QM, and also 2) is a complete formulation of general relativity, thus producing what we interpret as a theory of quantum gravity.

4.1 Born Rule

Definition 22 (Born Rule)**.** *The standard definition of the Born rule connects the probability to the complex norm of the wave-function:*

$$\rho = |\psi|^2 \tag{60}$$

The Born rule was postulated by Born in 1926, but attempts to derive it from first principles has been elusive. Notable proposals are those of Andrew M. Gleason[8], Kastner, R. E[9], and Lluis Masanes and Thomas Galley[10], and others.

In our formalism the determinant rule is the result of maximizing the entropy of a measure over a tuple. In the case of a matrix representation of the complex number, the determinant rule becomes the Born rule. Indeed, we first note that a complex number is represented by a matrix as follows:

$$a + ib \cong \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{61}$$

Then, we note that its determinant is the same as the complex norm:

$$\det \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = a^2 + b^2 \tag{62}$$

Finally, the determinant rule reduces to the Born rule forthe complex case:

$$\rho(f, \tau) = \frac{1}{Z} \det \left(\exp -\beta \begin{pmatrix} a(f) & 0 \\ 0 & a(f) \end{pmatrix} \exp -\tau \begin{pmatrix} 0 & -b(f) \\ b(f) & 0 \end{pmatrix} \right) \tag{63}$$

$$= \frac{1}{Z} \exp -\beta 2a(f) \tag{64}$$

where $\exp -\beta 2a(f)$ is the preparation of the ensemble (the prior) and Z is the normalization constants. In the case where $\beta = 1/k_B T$, then it is a thermal preparation. The Born rule is thus revealed to be the probability measure derived from maximizing the entropy of the selection of a tuple from a set of tuples, and in this sense is analogous to the Gibbs measure in statistical physics.

4.2 Axioms of QM

4.2.1 Dirac-Von Neumann Axioms (State vector)

The standard Dirac-Von Neumann Axioms are recovered as theorems. This becomes apparent if we split the probability measure into a two-step process:

$$\rho(f,\tau) = \det \psi(f,\tau)|_{\psi\in\mathbb{C}} = |\psi(f,\tau)|^2 \tag{65}$$

where

$$\psi(f,\tau) = \frac{1}{\sqrt{Z}}\left(\exp -\beta \begin{pmatrix} a(f) & 0 \\ 0 & a(f) \end{pmatrix} \exp -\tau \begin{pmatrix} 0 & -b(f) \\ b(f) & 0 \end{pmatrix}\right) \tag{66}$$

The formalism in terms of Hilbert space is obtained simply by taking ψ rather than ρ as the object of study. To show this is almost trivial; $\psi \in \mathbb{C}$ and ρ is given as the complex norm, therefore the ψ are the unit vectors of a complete complex vector space.

4.2.2 Dirac-Von Neumann Axioms (Expectation Value)

In statistical physics, an observable is simply a real value tied to each element of the probability measure:

$$\overline{O} = \sum_{q\in\mathbb{Q}} \rho(q)O(q) \tag{67}$$

Applied to our matrix-based constraints, this definition becomes that of a self-adjoint operator. And the expectation is a sum as follows:

$$\overline{O} = \sum_{q\in\mathbb{Q}} \det \psi(q)O(q)|_{\psi\in\mathbb{C}} = \sum_{q\in\mathbb{Q}} \psi(q)^*\psi(q)O(q) \tag{68}$$

and is thus the same definition as that of the expectation value of a self-adjoint operator acting on the unit vectors of a complex Hilbert space:

$$\sum_{q\in\mathbb{Q}} \psi(q)^*\psi(q)O(q) = \langle\psi|\, O\, |\psi\rangle \tag{69}$$

4.2.3 Time-Evolution (Schrödinger equation)

The last axiom of QM is usually a statement that the time-evolution of a state vector ψ is given by the Schrödinger equation. To derive it from our framework, that we can write:

$$\psi(f,\tau) = \frac{1}{\sqrt{Z}}\left(\exp -\beta \begin{pmatrix} a(f) & 0 \\ 0 & a(f) \end{pmatrix} \exp -\tau \begin{pmatrix} 0 & -b(f) \\ b(f) & 0 \end{pmatrix}\right) \tag{70}$$

as:

$$\psi(f,\tau) = \frac{1}{\sqrt{Z}} \exp -\tau i b(f) \tag{71}$$

where, to simplify, we have taking the prior to be unity ($a(f) = 0$). This is the familiar form of the quantum mechanical unitary evolution operator, where we rewrite $\tau \to t$ and $b \to H$. The Taylor expansion to the first linear term: $U(\delta t) \approx 1 - i\delta t H$. Then:

$$\left|\psi(t+\delta t)\right\rangle = U(\delta t)\left|\psi(t)\right\rangle \approx (1 - i\delta t H)\left|\psi\right\rangle \tag{72}$$

$$\implies \left|\psi(t+\delta t)\right\rangle - \left|\psi\right\rangle \approx -i\delta t H \left|\psi\right\rangle \tag{73}$$

$$\implies i\frac{\left|\psi(t+\delta t)\right\rangle - \left|\psi\right\rangle}{\delta t} \approx H\left|\psi\right\rangle \tag{74}$$

$$\implies i\frac{\partial\left|\psi(t)\right\rangle}{\partial t} = H\left|\psi\right\rangle \tag{75}$$

which is the time-dependant Schrödinger equation.

5 Foundation of Physics (Quantum Gravity)

In this section we will show that the probability measure here-in derived adheres to 1) the axioms of quantum mechanics, and 2) is simultaneously a gauge-theoretical theory of gravitation.

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. In all cases, the probability amplitude transforms linearly with respect to general linear transformations and the probability measure, obtained from the determinant, is positive-definite.

5.1 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.

5.1.1 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:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{B} \tag{76}$$

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

$$A + X\hat{\mathbf{x}} + Y\hat{\mathbf{y}} + B\hat{\mathbf{x}} \wedge \hat{\mathbf{y}} \cong \begin{pmatrix} A+X & -B+Y \\ B+Y & A-X \end{pmatrix} \tag{77}$$

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[11].

Definition 24 (Clifford conjugate (of a $\mathbb{G}(2,\mathbb{R})$ multi-vector))**.**

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 \tag{78}$$

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

Definition 25 (Geometric representation of the determinant (of a 2×2 matrix))**.**

$$\begin{array}{rcccl} \det & : & \mathbb{G}(2,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{u} \end{array} \tag{79}$$

For example:

$$\begin{aligned} \det \mathbf{u} &= (A - \mathbf{X} - \mathbf{B})(A + \mathbf{X} + \mathbf{B}) && (80) \\ &= A^2 - X^2 - Y^2 + B^2 && (81) \\ &= \det \begin{pmatrix} A+X & -B+Y \\ B+Y & A-X \end{pmatrix} && (82) \end{aligned}$$

Finally, we define the Clifford transpose:

Definition 26 (Clifford transpose (of a matrix of 2×2 matrix elements))**.** *The Clifford transpose is the geometric analogue to the conjugate transpose. Like the conjugate transpose can be interpreted as a transpose followed by an element-by-element application of the complex conjugate, here the Clifford transpose is a transpose, followed by an element-by-element application of the Clifford conjugate:*

$$\begin{pmatrix} \mathbf{u}_{00} & \dots & \mathbf{u}_{0n} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}_{mn} \end{pmatrix}^{\ddagger} = \begin{pmatrix} \mathbf{u}^{\ddagger}_{00} & \dots & \mathbf{u}^{\ddagger}_{m0} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{m0} & \dots & \mathbf{u}^{\ddagger}_{nm} \end{pmatrix} \tag{83}$$

If applied to a vector, then:

$$\begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{pmatrix}^{\ddagger} = \begin{pmatrix} \mathbf{v}_1^{\ddagger} & \dots \mathbf{v}_m^{\ddagger} \end{pmatrix} \tag{84}$$

5.1.2 Axiomatic Definition of the Algebra, in 2D

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

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

$$\begin{array}{rcrcl} \langle \cdot, \cdot \rangle & : & \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(2,\mathbb{R}) \\ & & \langle \mathbf{u}, \mathbf{v} \rangle & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{v} \end{array} \tag{85}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{86}$$

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

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \psi(q)^{\ddagger} \psi(q) \tag{87}$$

   is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{88}$$

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:

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{89}$$

- $\boldsymbol{\psi}$ 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):

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{90}$$

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

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

$$\langle \mathbf{O}\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v}\rangle \tag{91}$$

  is called an observable.

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

$$\langle \mathbf{O}\rangle = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}\rangle}\langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi}\rangle \tag{92}$$

5.1.3 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:

$$\Big(\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^{\ddagger}\mathbf{v}\Big|_{\mathbf{x}\to 0} \implies \langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^{\dagger}\mathbf{v} \tag{93}$$

- Reduction to the unitary transformations:

$$\big(\langle \mathbf{T}\mathbf{u}, \mathbf{T}\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle\big|_{\mathbf{x}\to 0} \implies \mathbf{T}^{\dagger}\mathbf{T} = I \tag{94}$$

- Reduction to the Born rule:

$$\Bigg(\rho(q, \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}\rangle}\psi(q)^{\ddagger}\psi(q)\Bigg|_{\mathbf{X}\to 0} \implies \rho(q, \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}\rangle}\psi(q)^{\dagger}\psi(q) \tag{95}$$

- Reduction of observables to Hermitian operators:

$$\big(\langle \mathbf{O}\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v}\rangle\big|_{\mathbf{X}\to 0} \implies \mathbf{O}^{\dagger} = \mathbf{O} \tag{96}$$

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

5.1.4 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:

$$\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi}\rangle = \langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi}\rangle \tag{97}$$

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

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

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

$$\begin{aligned} 2\langle \mathbf{O}\boldsymbol{\phi}, \boldsymbol{\psi} \rangle &= (O_{00}\phi_1 + O_{01}\phi_2)^{\ddagger}\psi_1 + \psi_1^{\ddagger}(O_{00}\phi_1 + O_{01}\phi_2) \\ &\quad + (O_{10}\phi_1 + O_{11}\phi_2)^{\ddagger}\psi_2 + \psi_2^{\ddagger}(O_{10}\phi_1 + O_{11}\phi_2) \qquad (98) \\ &= \phi_1^{\ddagger}O_{00}^{\ddagger}\psi_1 + \phi_2^{\ddagger}O_{01}^{\ddagger}\psi_1 + \psi_1^{\ddagger}O_{00}\phi_1 + \psi_1^{\ddagger}O_{01}\phi_2 \\ &\quad + \phi_1^{\ddagger}O_{10}^{\ddagger}\psi_2 + \phi_2^{\ddagger}O_{11}^{\ddagger}\psi_2 + \psi_2^{\ddagger}O_{10}\phi_1 + \psi_2^{\ddagger}O_{11}\phi_2 \qquad (99) \end{aligned}$$

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

$$\begin{aligned} 2\langle \boldsymbol{\phi}, \mathbf{O}\boldsymbol{\psi} \rangle &= \phi_1^{\ddagger}(O_{00}\psi_1 + O_{01}v_2) + (O_{00}\psi_1 + O_{01}\psi_2)^{\ddagger}\phi_1 \\ &\quad + \phi_2^{\ddagger}(O_{10}\psi_1 + O_{11}\psi_2) + (O_{10}\psi_1 + O_{11}\psi_2)^{\ddagger}\phi_1 \qquad (100) \\ &= \phi_1^{\ddagger}O_{00}\psi_1 + \phi_1^{\ddagger}O_{01}\psi_2 + \psi_1^{\ddagger}O_{00}^{\ddagger}\phi_1 + \psi_2^{\ddagger}O_{01}^{\ddagger}\phi_1 \\ &\quad + \phi_2^{\ddagger}O_{10}\psi_1 + \phi_2^{\ddagger}O_{11}\psi_2 + \psi_1^{\ddagger}O_{10}^{\ddagger}\phi_1 + \psi_2^{\ddagger}O_{11}^{\ddagger}\phi_1 \qquad (101) \end{aligned}$$

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:

$$O_{00}^{\ddagger} = O_{00} \qquad (102)$$

$$O_{01}^{\ddagger} = O_{10} \qquad (103)$$

$$O_{10}^{\ddagger} = O_{01} \qquad (104)$$

$$O_{11}^{\ddagger} = O_{11} \qquad (105)$$

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:

$$\mathbf{O}^{\ddagger} = \mathbf{O} \qquad (106)$$

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

5.1.5 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:

$$\mathbf{O} = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{107}$$

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

$$\mathbf{O}^\ddagger = \begin{pmatrix} a_{00} & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} \end{pmatrix} \tag{108}$$

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

$$0 = \det(\mathbf{O} - \lambda I) = \det \begin{pmatrix} a_{00} - \lambda & a - xe_1 - ye_2 - be_{12} \\ a + xe_1 + ye_2 + be_{12} & a_{11} - \lambda \end{pmatrix} \tag{109}$$

implies:

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a - xe_1 - ye_2 - be_{12})(a + xe_1 + ye_2 + be_{12} + a_{11}) \tag{110}$$

$$0 = (a_{00} - \lambda)(a_{11} - \lambda) - (a^2 - x^2 - y^2 + b^2) \tag{111}$$

finally:

$$\lambda = \{\frac{1}{2}\left(a_{00} + a_{11} - \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right), \tag{112}$$

$$\frac{1}{2}\left(a_{00} + a_{11} + \sqrt{(a_{00} - a_{11})^2 + 4(a^2 - x^2 - y^2 + b^2)}\right)\} \tag{113}$$

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

5.2 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:

$$\mathbf{u} = A + \mathbf{X} + \mathbf{V} + \mathbf{B} \tag{114}$$

[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.

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:

$$A + X\sigma_1 + Y\sigma_2 + Z\sigma_3 + V_1 i\sigma_1 + V_2 i\sigma_2 + V_3 i\sigma_3 + B\sigma_1 \wedge \sigma_2 \wedge \sigma_3 \tag{115}$$

$$\cong \begin{pmatrix} A + iB + iV_2 + Z & V_1 + iV_3 + X - iY \\ -V_1 + iV_3 + X + iY & A + iB - iV_2 - Z \end{pmatrix} \tag{116}$$

and the determinant of this matrix connects to the determinant of the multi-vector as follows:

$$\begin{array}{rcrcl} \det \cdot & : & \mathbb{G}(3,\mathbb{R}) & \longrightarrow & \mathbb{C} \\ & & \mathbf{u} & \longmapsto & \mathbf{u}^{\ddagger}\mathbf{u} \end{array} \tag{117}$$

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

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 - \langle \mathbf{u} \rangle_1 - \langle \mathbf{u} \rangle_2 + \langle \mathbf{u} \rangle_3 \tag{118}$$

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

$$\begin{array}{rcrcl} |\cdot| & : & \mathbb{G}(3,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & & \mathbf{u} & \longmapsto & (\mathbf{u}^{\ddagger}\mathbf{u})^{\dagger}\mathbf{u}^{\ddagger}\mathbf{u} \end{array} \tag{119}$$

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

$$\mathbf{u}^{\ddagger} := \langle \mathbf{u} \rangle_0 + \langle \mathbf{u} \rangle_1 + \langle \mathbf{u} \rangle_2 - \langle \mathbf{u} \rangle_3 \tag{120}$$

5.2.1 Axiomatic Definition of the Algebra, in 3D

Let $\mathbb{V}$ be an 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:

$$\begin{array}{rcrcl} \langle \cdot,\cdot,\cdot,\cdot \rangle & : & \mathbb{V} \times \mathbb{V} \times \mathbb{V} \times \mathbb{V} & \longrightarrow & \mathbb{G}(3,\mathbb{R}) \\ & & \langle \mathbf{u},\mathbf{v},\mathbf{w},\mathbf{x} \rangle & \longmapsto & (\mathbf{u}^{\ddagger}\mathbf{v})^{\dagger}\mathbf{w}^{\ddagger}\mathbf{x} \end{array} \tag{121}$$

   is positive-definite:

$$\langle \boldsymbol{\psi}, \boldsymbol{\psi} \rangle \in \mathbb{R}_{>0} \tag{122}$$

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

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle}(\psi(q)^{\ddagger}\psi(q))^{\dagger}\psi(q)^{\ddagger}\psi(q) \tag{123}$$

   is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{124}$$

5.2.2 Reduction to Complex Hilbert Spaces

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

$$\boldsymbol{\psi} = \begin{pmatrix} (\exp \frac{1}{2}\mathbf{u}_1)^\ddagger \exp \frac{1}{2}\mathbf{u}_1 \\ \vdots \\ (\exp \frac{1}{2}\mathbf{u}_m)^\ddagger \exp \frac{1}{2}\mathbf{u}_m \end{pmatrix} \tag{125}$$

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

$$(\exp \frac{1}{2}\mathbf{u})^\ddagger \exp \frac{1}{2}\mathbf{u} = \exp \frac{1}{2}(A - \mathbf{X} - \mathbf{V} + \mathbf{B}) \exp \frac{1}{2}(A + \mathbf{X} + \mathbf{V} + \mathbf{B}) \tag{126}$$

$$= \exp(A + \mathbf{B}) \exp \frac{1}{2}(-\mathbf{X} - \mathbf{V}) \exp \frac{1}{2}(\mathbf{X} + \mathbf{V}) \tag{127}$$

$$= \exp(A + \mathbf{B}) \tag{128}$$

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

$$\begin{array}{cccc} \langle\cdot,\cdot\rangle & : & \mathcal{A}(\mathbb{V}) \times \mathcal{A}(\mathbb{V}) & \longrightarrow & \mathbb{C} \\ & & \langle\psi,\phi\rangle & \longmapsto & \psi^\dagger\phi \end{array} \tag{129}$$

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.

5.3 Algebra of Natural States, in 4D

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

5.3.1 Geometric Representation (in 4D)

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

$$\mathbf{u} = A + \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{130}$$

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 27 (Geometric representation of a matrix (4×4))**.**

$$
\begin{aligned}
&A + T\gamma_0 + X\gamma_1 + Y\gamma_2 + Z\gamma_3\\
&\quad + F_{01}\gamma_0\wedge\gamma_1 + F_{02}\gamma_0\wedge\gamma_2 + F_{03}\gamma_0\wedge\gamma_3 + F_{23}\gamma_2\wedge\gamma_3 + F_{13}\gamma_1\wedge\gamma_3 + F_{12}\gamma_1\wedge\gamma_2\\
&\quad + V_t\gamma_1\wedge\gamma_2\wedge\gamma_3 + V_x\gamma_0\wedge\gamma_2\wedge\gamma_3 + V_y\gamma_0\wedge\gamma_1\wedge\gamma_3 + V_z\gamma_0\wedge\gamma_1\wedge\gamma_2\\
&\quad + B\gamma_0\wedge\gamma_1\wedge\gamma_2\wedge\gamma_3\\
&\qquad\cong \begin{pmatrix}
A+X_0-iF_{12}-iV_3 & F_{13}-iF_{23}+V_2-iV_1 & -iB+X_3+F_{03}-iV_0 & X_1-iX_2+F_{01}-iF_{02}\\
-F_{13}-iF_{23}-V_2-iV_1 & A+X_0+iF_{12}+iV_3 & X_1+iX_2+F_{01}+iF_{02} & -iB-X_3-F_{03}-iV_0\\
-iB-X_3+F_{03}+iV_0 & -X_1+iX_2+F_{01}-iF_{02} & A-X_0-iF_{12}+iV_3 & F_{13}-iF_{23}-V_2+iV_1\\
-X_1-iX_2+F_{01}+iF_{02} & -iB+X_3-F_{03}+iV_0 & -F_{13}-iF_{23}+V_2+iV_1 & A-X_0+iF_{12}-iV_3
\end{pmatrix}
\end{aligned}
\tag{131}
$$

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[11].

Definition 28 (Clifford conjugate (of a $\mathbb{G}(4,\mathbb{R})$ multi-vector))**.**

$$\mathbf{u}^{\ddagger} := \langle\mathbf{u}\rangle_0 - \langle\mathbf{u}\rangle_1 - \langle\mathbf{u}\rangle_2 + \langle\mathbf{u}\rangle_3 + \langle\mathbf{u}\rangle_4 \tag{132}$$

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):

$$\lfloor\mathbf{u}\rfloor_{\{3,4\}} := \langle\mathbf{u}\rangle_0 + \langle\mathbf{u}\rangle_1 + \langle\mathbf{u}\rangle_2 - \langle\mathbf{u}\rangle_3 - \langle\mathbf{u}\rangle_4 \tag{133}$$

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

Definition 29 (Geometric representation of the determinant (of a 4×4 matrix))**.**

$$
\begin{array}{rccc}
\det & : & \mathbb{G}(4,\mathbb{R}) & \longrightarrow & \mathbb{R}\\
 & & \mathbf{u} & \longmapsto & \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u}
\end{array}
\tag{134}
$$

5.3.2 Axiomatic Definition of the Algebra, in 4D

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

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

$$
\begin{array}{rccc}
\langle\cdot,\cdot,\cdot,\cdot\rangle & : & \mathbb{V}\times\mathbb{V}\times\mathbb{V}\times\mathbb{V} & \longrightarrow & \mathbb{G}(4,\mathbb{R})\\
 & & \langle\mathbf{u},\mathbf{v},\mathbf{w},\mathbf{x}\rangle & \longmapsto & \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{w}^{\ddagger}\mathbf{x}
\end{array}
\tag{135}
$$

   is positive-definite:

$$\langle\boldsymbol{\psi},\boldsymbol{\psi},\boldsymbol{\psi},\boldsymbol{\psi}\rangle \in \mathbb{R}_{>0} \tag{136}$$

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

$$\rho(\psi(q), \boldsymbol{\psi}) = \frac{1}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \lfloor \psi(q)^{\ddagger} \psi(q) \rfloor_{3,4} \psi(q)^{\ddagger} \psi(q) \tag{137}$$

is positive-definite:

$$\rho(\psi(q), \boldsymbol{\psi}) \in \mathbb{R}_{>0} \tag{138}$$

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):

$$\sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \mathbf{T}\boldsymbol{\psi}) = \sum_{\psi(q) \in \boldsymbol{\psi}} \rho(\psi(q), \boldsymbol{\psi}) = 1 \tag{139}$$

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

$$\langle \mathbf{O}\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{O}\mathbf{v}, \mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{O}\mathbf{w}, \mathbf{x} \rangle = \langle \mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{O}\mathbf{x} \rangle \tag{140}$$

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

$$\langle \mathbf{O} \rangle = \frac{\langle \mathbf{O}\boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle}{\langle \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi}, \boldsymbol{\psi} \rangle} \tag{141}$$

5.3.3 Reduction to 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.

$$\boldsymbol{\psi} = \begin{pmatrix} (\exp \frac{1}{2}\mathbf{u}_1)^{\ddagger} \exp \frac{1}{2}\mathbf{u}_1 \\ \vdots \\ (\exp \frac{1}{2}\mathbf{u}_m)^{\ddagger} \exp \frac{1}{2}\mathbf{u}_m \end{pmatrix} \tag{142}$$

The form of the elements of $\boldsymbol{\psi}$ is:

$$\psi^{\ddagger}\psi = \exp\frac{1}{2}(A - \mathbf{F} + \mathbf{B})\exp\frac{1}{2}(A + \mathbf{F} + \mathbf{B}) \tag{143}$$

$$= \exp\frac{1}{2}A\exp-\frac{1}{2}\mathbf{F}\exp\frac{1}{2}\mathbf{B}\exp\frac{1}{2}A\exp\frac{1}{2}\mathbf{F}\exp\frac{1}{2}\mathbf{B} \tag{144}$$

$$= \exp A\exp\mathbf{B} \tag{145}$$

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

$$\begin{array}{rcccl} \langle\cdot,\cdot\rangle & : & \mathcal{A}(\mathbb{V})\times\mathcal{A}(\mathbb{V}) & \longrightarrow & \mathbb{C} \\ & & \langle\psi,\phi\rangle & \longmapsto & \psi^{\dagger}\phi \end{array} \tag{146}$$

In our example, and with this bilinear map, $\langle\psi^{\ddagger}\psi, \psi^{\ddagger}\psi\rangle = \exp 2A$.

We note the similarity of this sub-algebra to David Hestenes[12]'s 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 $\psi\tilde{\psi} = \rho e^{iB}$.

5.4 Law of Motion

5.4.1 Probability-Preserving Evolution (Left Action in 2D)

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

$$\left\langle\psi\right|\mathbf{G}^{\ddagger}\mathbf{G}\left|\psi\right\rangle = \langle\psi|\psi\rangle \tag{147}$$

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

$$\mathbf{G}^{\ddagger}\mathbf{G} = I \tag{148}$$

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

$$\mathbf{G} = \begin{pmatrix} u & v \\ w & x \end{pmatrix} \tag{149}$$

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

$$\mathbf{G}^{\ddagger}\mathbf{G} = \begin{pmatrix} v^{\ddagger} & u^{\ddagger} \\ w^{\ddagger} & x^{\ddagger} \end{pmatrix}\begin{pmatrix} v & w \\ u & x \end{pmatrix} = \begin{pmatrix} v^{\ddagger}v + u^{\ddagger}u & v^{\ddagger}w + u^{\ddagger}x \\ w^{\ddagger}v + x^{\ddagger}u & w^{\ddagger}w + x^{\ddagger}x \end{pmatrix} \tag{150}$$

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

$$v^\ddagger v + u^\ddagger u = 1 \tag{151}$$
$$v^\ddagger w + u^\ddagger x = 0 \tag{152}$$
$$w^\ddagger v + x^\ddagger u = 0 \tag{153}$$
$$w^\ddagger w + x^\ddagger x = 1 \tag{154}$$

This is the case if

$$\mathbf{G} = \frac{1}{\sqrt{v^\ddagger v + u^\ddagger u}} \begin{pmatrix} v & u \\ -e^{\varphi} u^\ddagger & e^{\varphi} v^\ddagger \end{pmatrix} \tag{155}$$

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

$$\mathbf{U} = \frac{1}{\sqrt{|a|^2 + |b|^2}} \begin{pmatrix} a & b \\ -e^{i\theta} b^\dagger & e^{i\theta} a^\dagger \end{pmatrix} \tag{156}$$

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

$$\implies \mathbf{G}^\ddagger \mathbf{G} = \frac{1}{v^\ddagger v + u^\ddagger u} \begin{pmatrix} v^\ddagger & -e^{-\varphi} u \\ u^\ddagger & e^{-\varphi} v \end{pmatrix} \begin{pmatrix} v & u \\ -e^{\varphi} u^\ddagger & e^{\varphi} v^\ddagger \end{pmatrix} \tag{157}$$
$$= \frac{1}{v^\ddagger v + u^\ddagger u} \begin{pmatrix} v^\ddagger v + u^\ddagger u & v^\ddagger u - v^\ddagger u \\ u^\ddagger v - u^\ddagger v & u^\ddagger u + v^\ddagger v \end{pmatrix} \tag{158}$$
$$= I \tag{159}$$

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

$$\mathbf{D} = \begin{pmatrix} e^{x_1 \hat{\mathbf{x}} + y_1 \hat{\mathbf{y}} + i b_1} & 0 \\ 0 & e^{x_2 \hat{\mathbf{x}} + y_2 \hat{\mathbf{y}} + i b_2} \end{pmatrix} \tag{160}$$

where $\mathbf{G} = P\mathbf{D}P^{-1}$. It follows quite easily that $D^\ddagger D = I$, because each diagonal entry produces unity: $e^{-x_1 \hat{\mathbf{x}} - y_1 \hat{\mathbf{y}} - i b_1} e^{x_1 \hat{\mathbf{x}} + y_1 \hat{\mathbf{y}} + i b_1} = 1$.

5.4.2 Probability-Preserving Evolution (Adjoint Action in 2D)

Since the elements of $|\psi\rangle$ are matrices, in the general case, the transformation is given by adjoint action:

$$\mathbf{G} \, |\psi\rangle \, \mathbf{G}^{-1} \tag{161}$$

The bilinear form is:

$$(\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1})^{\ddagger}(\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1}) = (\mathbf{G}^{-1})^{\ddagger}\,\langle\psi|\,\mathbf{G}^{\ddagger}\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1} \tag{162}$$

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

$$(\mathbf{G}^{-1})^{\ddagger}\,\langle\psi|\,\mathbf{G}^{\ddagger}\mathbf{G}\,|\psi\rangle\,\mathbf{G}^{-1} = \langle\psi|\psi\rangle \tag{163}$$

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

$$\mathbf{D} = \begin{pmatrix} e^{a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1} & 0 & 0 \\ 0 & e^{a_2+x_2\hat{\mathbf{x}}+y_2\hat{\mathbf{y}}+ib_2} & 0 \\ 0 & 0 & \ddots \end{pmatrix} \tag{164}$$

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

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

$$e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1^{\ddagger}e^{a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1}e^{a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1 e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{165}$$

$$= e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}\psi_1^{\ddagger}e^{2a_1}\psi_1 e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{166}$$

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

$$= \psi_1^{\ddagger}\psi_1 e^{2a_1}e^{-a_1+x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+ib_1}e^{-a_1-x_1\hat{\mathbf{x}}-y_1\hat{\mathbf{y}}-ib_1} \tag{167}$$

$$= \psi_1^{\ddagger}\psi_1 e^{2a_1}e^{-a_1}e^{-a_1} = \psi_1^{\ddagger}\psi_1 \tag{168}$$

5.4.3 General Linear Schrödinger Equation (Left Action)

The standard Schrödinger equation can be derived as follows. First, assume $U(t) = e^{-itH}$, and its Taylor expansion to the first linear term: $U(\delta t) \approx 1 - i\delta t H$. Then:

$$\left|\psi(t+\delta t)\right\rangle = U(\delta t)\left|\psi(t)\right\rangle \approx (1 - i\delta t H)\,|\psi\rangle \tag{169}$$

$$\implies \left|\psi(t+\delta t)\right\rangle - |\psi\rangle \approx -i\delta t H\,|\psi\rangle \tag{170}$$

$$\implies i\frac{\left|\psi(t+\delta t)\right\rangle - |\psi\rangle}{\delta t} \approx H\,|\psi\rangle \tag{171}$$

$$\implies i\frac{\partial\left|\psi(t)\right\rangle}{\partial t} = H\,|\psi\rangle \tag{172}$$

Now, we wish to use the same derivation, but apply it to the 2D general linear version of the unitary group:

$$U^\dagger U = I \to (\mathbf{G}^{-1})^\ddagger \langle\psi| \mathbf{G}^\ddagger \mathbf{G} |\psi\rangle \mathbf{G}^{-1} = \langle\psi|\psi\rangle \tag{173}$$

In the general linear case, the imaginary number i is replaced with an arbitrary matrix $\mathbf{M}$, via the relation:

$$\mathbf{G} = e^{-\mathbf{M}\tau H} \tag{174}$$

where H is self-adjoint : $H^\ddagger = H$.

Then, the general linear Schrödinger equation for the one-parameter group of the general linear group $\mathbf{G}(\tau) = e^{-\mathbf{M}\tau H}$, for the left action is:

$$\begin{aligned}
\big|\psi(\tau+\delta\tau)\big\rangle &= G(\delta\tau)\big|\psi(\tau)\big\rangle && (175)\\
&\approx (1-\mathbf{M}\delta\tau H)\big|\psi(\tau)\big\rangle && (176)\\
&= \big|\psi(\tau)\big\rangle - \mathbf{M}\delta\tau H\big|\psi(\tau)\big\rangle && (177)\\
\implies -\frac{\partial\big|\psi(\tau)\big\rangle}{\partial\tau} &= \mathbf{M}H\big|\psi(\tau)\big\rangle && (178)
\end{aligned}$$

and iff $\exists\mathbf{M}^{-1}$, then

$$-\mathbf{M}^{-1}\frac{\partial\big|\psi(\tau)\big\rangle}{\partial\tau} = H\big|\psi(\tau)\big\rangle \tag{179}$$

5.4.4 General Linear Schrödinger Equation (Adjoint Action)

And for the adjoint action, it is:

$$\begin{aligned}
\big|\psi(\tau+\delta\tau)\big\rangle &= G(\delta\tau)\big|\psi(\tau)\big\rangle G(\delta\tau)^{-1} && (180)\\
&\approx (1-\mathbf{M}\delta\tau H)\big|\psi(\tau)\big\rangle(1+\mathbf{M}\delta\tau H) && (181)\\
&= \big|\psi(\tau)\big\rangle(1+\mathbf{M}\delta\tau H) - \mathbf{M}\delta\tau H\big|\psi(\tau)\big\rangle(1+\mathbf{M}\delta\tau H) && (182)\\
&= \big|\psi(\tau)\big\rangle + \big|\psi(\tau)\big\rangle\mathbf{M}\delta\tau H - \mathbf{M}\delta\tau H\big|\psi(\tau)\big\rangle - \mathbf{M}\delta\tau H\big|\psi(\tau)\big\rangle\mathbf{M}\delta\tau H && (183)\\
&\approx \big|\psi(\tau)\big\rangle + \big|\psi(\tau)\big\rangle\mathbf{M}\delta\tau H - \mathbf{M}\delta\tau H\big|\psi(\tau)\big\rangle && (184)\\
&= \big|\psi(\tau)\big\rangle + \delta\tau\Big[\mathbf{M}H, \big|\psi(\tau)\big\rangle\Big] && (185)\\
\implies \frac{\partial\big|\psi(\tau)\big\rangle}{\partial\tau} &= \Big[\mathbf{M}H, \big|\psi(\tau)\big\rangle\Big] && (186)
\end{aligned}$$

5.4.5 Conservation of Probability (Left Action in 2D)

For a parametrization of ψ, the probability must normalize. For instance, a x parametrization would yield:

$$\int \psi(\tau,x)^\ddagger \psi(\tau,x)\,\mathrm{d}x = N(\tau) \tag{187}$$

To lighten the notation we will not explicitly write the dependance of ψ in (τ, x).

$$\frac{dN(\tau)}{d\tau} = 0 = \int \frac{\partial \psi^\ddagger \psi}{\partial \tau}\,\mathrm{d}x \tag{188}$$

$$= \int \frac{\partial \psi^\ddagger}{\partial \tau}\psi\,\mathrm{d}x + \int \psi^\ddagger \frac{\partial \psi}{\partial \tau}\,\mathrm{d}x \tag{189}$$

We now inject the following relation (derived from the general linear Schrödinger equation):

$$\frac{\partial \psi}{\partial \tau} = \mathbf{M} H \psi \tag{190}$$

$$\frac{\partial \psi^\ddagger}{\partial \tau} = (\mathbf{M} H \psi)^\ddagger = \psi^\ddagger H^\ddagger \mathbf{M}^\ddagger \tag{191}$$

Injecting them, we get:

$$\frac{dN(\tau)}{d\tau} = \int \frac{\partial \psi^\ddagger}{\partial \tau}\psi\,\mathrm{d}x + \int \psi^\ddagger \frac{\partial \psi}{\partial \tau}\,\mathrm{d}x \tag{192}$$

$$= \int \psi^\ddagger \mathbf{M}^\ddagger H^\ddagger \psi\,\mathrm{d}x + \int \psi^\ddagger \mathbf{M} H \psi\,\mathrm{d}x \tag{193}$$

We note that $\mathbf{M}^\ddagger = -\mathbf{M}$ (this requires that $\langle \mathbf{M} \rangle_0 = 0$, which is the case for the left action) and that $H^\ddagger = H$ (which it is iff it is self-adjoint), therefore:

$$= -\int \psi^\ddagger \mathbf{M} H \psi\,\mathrm{d}x + \int \psi^\ddagger \mathbf{M} H \psi\,\mathrm{d}x \tag{194}$$

Finally, adding the requirement that $[\mathbf{M}, H] = 0$, we get the conservation of probability:

$$= 0 \tag{195}$$

The general linear form of the Schrödinger equation is a conservation of probability law of the general linear case.

5.4.6 Conservation of Probability (Adjoint Action in 2D)

For a parametrization of ψ, the probability must normalize. For instance, a x parametrization would yield:

$$\int \psi(\tau,x)^{\ddagger}\psi(\tau,x)\,\mathrm{d}x = N(\tau) \tag{196}$$

To lighten the notation we will not explicitly write the dependance of ψ in (τ, x).

$$\frac{dN(\tau)}{d\tau} = 0 = \int \frac{\partial\psi^{\ddagger}\psi}{\partial\tau}\,\mathrm{d}x \tag{197}$$

$$= \int \frac{\partial\psi^{\ddagger}}{\partial\tau}\psi\,\mathrm{d}x + \int \psi^{\ddagger}\frac{\partial\psi}{\partial\tau}\,\mathrm{d}x \tag{198}$$

We now inject the following relation (derived from the general linear Schrödinger equation):

$$\frac{\partial\psi}{\partial\tau} = \mathbf{M}H\psi - \psi\mathbf{M}H \tag{199}$$

$$\frac{\partial\psi^{\ddagger}}{\partial\tau} = (\mathbf{M}H\psi - \psi\mathbf{M}H)^{\ddagger} = \psi^{\ddagger}H^{\ddagger}\mathbf{M}^{\ddagger} - H^{\ddagger}\mathbf{M}^{\ddagger}\psi^{\ddagger} \tag{200}$$

Injecting them, we get:

$$\frac{dN(\tau)}{d\tau} = \int \frac{\partial\psi^{\ddagger}}{\partial\tau}\psi\,\mathrm{d}x + \int \psi^{\ddagger}\frac{\partial\psi}{\partial\tau}\,\mathrm{d}x \tag{201}$$

$$= \int \psi^{\ddagger}H^{\ddagger}\mathbf{M}^{\ddagger}\psi - H^{\ddagger}\mathbf{M}^{\ddagger}\psi^{\ddagger}\psi\,\mathrm{d}x + \int \psi^{\ddagger}\mathbf{M}H\psi - \psi^{\ddagger}\psi\mathbf{M}H\,\mathrm{d}x \tag{202}$$

$$= \int \psi^{\ddagger}H^{\ddagger}\mathbf{M}^{\ddagger}\psi - H^{\ddagger}\mathbf{M}^{\ddagger}\psi^{\ddagger}\psi + \psi^{\ddagger}\mathbf{M}H\psi - \psi^{\ddagger}\psi\mathbf{M}H\,\mathrm{d}x \tag{203}$$

We note that $\mathbf{M}^{\ddagger} = A - \mathbf{m}$ and that $H^{\ddagger} = H$, therefore (we also pose $\mathbf{M} = A + \mathbf{m}$):

$$= \int \psi^{\ddagger}H(A-\mathbf{m})\psi - H(A-\mathbf{m})\psi^{\ddagger}\psi + \psi^{\ddagger}(A+\mathbf{m})H\psi - \psi^{\ddagger}\psi(A+\mathbf{m})H\,\mathrm{d}x \tag{204}$$

$$= \int -\psi^{\ddagger}H\mathbf{m}\psi + H\mathbf{m}\psi^{\ddagger}\psi + \psi^{\ddagger}\mathbf{m}H\psi - \psi^{\ddagger}\psi\mathbf{m}H\,\mathrm{d}x \tag{205}$$

Finally, adding the requirement that $[\mathbf{m}, H] = 0$, we get the conservation of probability:

$$= 0 \tag{206}$$

The general linear form of the Schrödinger equation is a conservation of probability law of the general linear case.

The 4D case is omitted due to excessive verbosity.

5.5 Gravity as a Gauge Theory

5.5.1 Unitary Gauge (Recap)

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

$$D_\mu = \partial_\mu + iqA_\mu(x) \tag{207}$$

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

$$\psi \to e^{-iq\theta(x)}\psi \quad \text{and} \quad A_\mu \to A_\mu + \partial_\mu\theta(x) \tag{208}$$

Then, applies the covariant derivation, one gets:

$$D_\mu\psi = \partial_\mu\psi + iqA_\mu\psi \tag{209}$$

$$\to \partial_\mu(e^{-iq\theta(x)}\psi) + iq(A_\mu + \partial_\mu\theta(x))(e^{-iq\theta(x)}\psi) \tag{210}$$

$$= e^{-iq\theta(x)}D_\mu\psi \tag{211}$$

Finally, the field is given as follows:

$$F_{\mu\nu} = [\mathcal{D}_\mu, \mathcal{D}_\nu] \tag{212}$$

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

5.5.2 General Linear Gauge

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, and the commutator of the covariant derivatives as the Riemann tensor..

A general linear transformation of ψ:

$$\psi'(x) \to g\psi(x)g^{-1} \tag{213}$$

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

$$D_\mu \psi = \partial_\mu \psi - [iqA_\mu, \psi] \tag{214}$$

Finally, the field is given as follows:

$$R_{\mu\nu} = [D_\mu, D_\nu] \tag{215}$$

where $R_{\mu\nu}$ is the Riemann tensor.

Since this is the result of the GL gauge invariance, then gravity is fundamentally integrated throughout the present quantum mechanical framework because of GL invariance, for the same reason that electromagnetism is fundamentally integrated within quantum theory over the complex norm because of U(1) invariance.

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 has been proposed before; specifically, I note the work of B. I. Lev.[13] which suggests (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:

$$\begin{array}{rrcl} \cdot & : \quad \mathbb{G}(2n,\mathbb{R}) \times \mathbb{G}(2n,\mathbb{R}) & \longrightarrow & \mathbb{R} \\ & \mathbf{u} \cdot \mathbf{v} & \longmapsto & \frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) \end{array} \tag{216}$$

For example,

$$\mathbf{u} = A_1 + X_1 e_1 + Y_1 e_2 + B_1 e_{12} \tag{217}$$
$$\mathbf{v} = A_2 + X_2 e_1 + Y_2 e_2 + B_2 e_{12} \tag{218}$$
$$\implies \mathbf{u} \cdot \mathbf{v} = A_1 A_2 + B_1 B_2 - X_1 X_2 - Y_1 Y_2 \tag{219}$$

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:

$$\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) = \frac{1}{2}\left((\mathbf{u}+\mathbf{v})^{\ddagger}(\mathbf{u}+\mathbf{v}) - \mathbf{u}^{\ddagger}\mathbf{u} - \mathbf{v}^{\ddagger}\mathbf{v}\right) \tag{220}$$
$$= \mathbf{u}^{\ddagger}\mathbf{u} + \mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u} + \mathbf{v}^{\ddagger}\mathbf{v} - \mathbf{u}^{\ddagger}\mathbf{u} - \mathbf{v}^{\ddagger}\mathbf{v} \tag{221}$$
$$= \mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u} \tag{222}$$

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

$$\frac{1}{2}(\det(\mathbf{u}+\mathbf{v}) - \det \mathbf{u} - \det \mathbf{v}) \tag{223}$$
$$= \frac{1}{2}\left(\lfloor(\mathbf{u}+\mathbf{v})^{\ddagger}(\mathbf{u}+\mathbf{v})\rfloor_{3,4}(\mathbf{u}+\mathbf{v})^{\ddagger}(\mathbf{u}+\mathbf{v}) - \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} - \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v}\right) \tag{224}$$
$$= \frac{1}{2}\left(\lfloor\mathbf{u}^{\ddagger}\mathbf{u} + \mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u} + \mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}(\mathbf{u}^{\ddagger}\mathbf{u} + \mathbf{u}^{\ddagger}\mathbf{v} + \mathbf{v}^{\ddagger}\mathbf{u} + \mathbf{v}^{\ddagger}\mathbf{v}) - \dots\right) \tag{225}$$
$$\begin{aligned}
= {} & \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} \\
& + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} \\
& + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} \\
& + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} - \dots
\end{aligned} \tag{226}$$
$$\begin{aligned}
= {} & \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} \\
& + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{u}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} \\
& + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{u}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{v} \\
& + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{u} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{u}^{\ddagger}\mathbf{v} + \lfloor\mathbf{v}^{\ddagger}\mathbf{v}\rfloor_{3,4}\mathbf{v}^{\ddagger}\mathbf{u}
\end{aligned} \tag{227}$$

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 general, the exponential $\exp \mathbf{u}$ equals $\exp a \exp \mathbf{s}$ because a commutes with $\mathbf{s}$.

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

$$\psi = \psi_1 + \psi_2 = e^{\frac{2}{n}A_1}e^{\frac{2}{n}\mathbf{S}_1} + e^{\frac{2}{n}A_2}e^{\frac{2}{n}\mathbf{S}_2} \tag{228}$$
$$\tag{229}$$

The general interference pattern will be of the following form:

$$\begin{aligned} \det\psi_1 + \psi_2 &= \det\psi_1 + \det\psi_2 + \psi_1\cdot\psi_2 &(230)\\ &= e^{2A_1} + e^{2A_2} + \psi_1\cdot\psi_2 &(231)\end{aligned}$$

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:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} \tag{232}$$

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

$$\begin{aligned} \psi^\dagger\psi &= \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 &(233)\\ &= e^{A_1}e^{-\mathbf{B}_1}e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2}e^{A_2}e^{\mathbf{B}_2} + e^{A_1}e^{-\mathbf{B}_1}e^{A_2}e^{\mathbf{B}_2} + e^{A_2}e^{-\mathbf{B}_2}e^{A_1}e^{\mathbf{B}_1} &(234)\\ &= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}(e^{-\mathbf{B}_1+\mathbf{B}_2} + e^{-(-\mathbf{B}_1+\mathbf{B}_2)}) &(235)\\ &= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2}\cos(B_1-B_2)}_{\text{interference}} &(236)\end{aligned}$$

6.1.4 Geometric Interference in 2D

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{X}_1+\mathbf{B}_1} + e^{A_2}e^{\mathbf{X}_2+\mathbf{B}_2} \tag{237}$$

To lighten the notation we will write it as follows:

$$\psi = \psi_1 + \psi_2 = e^{A_1}e^{\mathbf{S}_1} + e^{A_2}e^{\mathbf{S}_2} \tag{238}$$

where

$$\mathbf{S} = \mathbf{X} + \mathbf{B} \tag{239}$$

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

$$
\begin{aligned}
\psi^\dagger\psi &= \psi_1^\dagger\psi_1 + \psi_2^\dagger\psi_2 + \psi_1^\dagger\psi_2 + \psi_2^\dagger\psi_1 && (240)\\
&= e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_1}e^{\mathbf{S}_1} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_1}(e^{\mathbf{S}_1})^\ddagger e^{A_2}e^{\mathbf{S}_2} + e^{A_2}(e^{\mathbf{S}_2})^\ddagger e^{A_1}e^{\mathbf{S}_1} && (241)\\
&= e^{2A_1} + e^{2A_2} + e^{A_1+A_2}((e^{\mathbf{S}_1})^\ddagger e^{\mathbf{S}_2} + (e^{\mathbf{S}_2})^\ddagger e^{\mathbf{S}_1}) && (242)\\
&= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{e^{A_1+A_2}(e^{-\mathbf{X}_1-\mathbf{B}_1}e^{\mathbf{X}_2+\mathbf{B}_2} + e^{-\mathbf{X}_2-\mathbf{B}_2}e^{\mathbf{X}_1+\mathbf{B}_1})}_{\text{interference}} && (243)
\end{aligned}
$$

6.1.5 Geometric Interference in 4D

Consider a two-state wave-function:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}(\mathbf{X}_1+\mathbf{F}_1+\mathbf{V}_1+\mathbf{B}_1)} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}(\mathbf{X}_2+\mathbf{F}_2+\mathbf{V}_2+\mathbf{B}_2)} \tag{244}$$

To lighten the notation we will write it as follows:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{S}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{S}_2} \tag{245}$$

where

$$\mathbf{S} = \mathbf{X} + \mathbf{F} + \mathbf{V} + \mathbf{B} \tag{246}$$

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

$$
\begin{aligned}
\lfloor\psi^\ddagger\psi\rfloor_{3,4}\psi^\ddagger\psi &= \lfloor\psi_1^\ddagger\psi_1\rfloor_{3,4}\psi_1^\ddagger\psi_1 + \lfloor\psi_2^\ddagger\psi_2\rfloor_{3,4}\psi_2^\ddagger\psi_2 + \psi_1\cdot\psi_2 && (247)\\
&= e^{2A_1} + e^{2A_2} + \left(e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{S}_1}\right)\cdot\left(e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{S}_2}\right) && (248)
\end{aligned}
$$

In many cases of interest, the pattern simplifies:

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:

$$\psi = \psi_1 + \psi_2 \tag{249}$$

where

$$
\begin{aligned}
\psi_1 &= (e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1})^\ddagger(e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1}) = e^{A_1}e^{\mathbf{B}_1} && (250)\\
\psi_2 &= (e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1})^\ddagger(e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1}) = e^{A_2}e^{\mathbf{B}_2} && (251)
\end{aligned}
$$

Thus

$$\psi = e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} \tag{252}$$

The quadri-linear map becomes a bilinear map:

$$
\begin{aligned}
\psi^\dagger\psi &= (e^{A_1}e^{-\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2})(e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2}) && (253)\\
&= e^{A_1}e^{-\mathbf{B}_1}e^{A_1}e^{\mathbf{B}_1} + e^{A_1}e^{-\mathbf{B}_1}e^{A_2}e^{\mathbf{B}_2} + e^{A_2}e^{-\mathbf{B}_2}e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{-\mathbf{B}_2}e^{A_2}e^{\mathbf{B}_2} && (254)\\
&= \underbrace{e^{2A_1} + e^{2A_2}}_{\text{sum}} + \underbrace{2e^{A_1+A_2}\cos(B_1 - B_2)}_{\text{complex interference}} && (255)
\end{aligned}
$$

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:

$$\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2} \tag{256}$$

The sub-product part is:

$$
\begin{aligned}
\psi^\ddagger\psi &= (e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2})(e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}) && (257)\\
&= e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{B}_2} && (258)\\
&= e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} + 2e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)} && (259)
\end{aligned}
$$

The final product is:

$$
\begin{aligned}
\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi &= \left(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)}\right) \\
&\quad \times \left(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)}\right) \qquad (260) \\
&= 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} \\
&\quad\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\quad + 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)} \qquad (261) \\
&= 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} \qquad (262) \\
&= \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}} \qquad (263)
\end{aligned}
$$

6.1.8 Geometric Interference in 4D (Deep Spinor Rotation)

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

$$
\psi = \psi_1 + \psi_2 = e^{\frac{1}{2}A_1} e^{\frac{1}{2}\mathbf{F}_1} e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2} e^{\frac{1}{2}\mathbf{F}_2} e^{\frac{1}{2}\mathbf{B}_2} \qquad (264)
$$

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

$$
\lfloor \psi^\ddagger \psi \rfloor_{3,4} \psi^\ddagger \psi \qquad (265)
$$

Let us start with the sub-product:

$$
\begin{aligned}
\psi^\ddagger\psi &= (e^{\frac{1}{2}A_1}e^{-\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{-\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2})(e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2}) \quad (266)\\
&= e^{\frac{1}{2}A_1}e^{-\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_1}e^{-\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2}\\
&\quad + e^{\frac{1}{2}A_2}e^{-\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_1}e^{\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{B}_1} + e^{\frac{1}{2}A_2}e^{-\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2}e^{\frac{1}{2}A_2}e^{\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{B}_2} \quad (267)\\
&= e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}(e^{-\frac{1}{2}\mathbf{F}_1}e^{\frac{1}{2}\mathbf{F}_2}) + e^{-\frac{1}{2}\mathbf{F}_2}e^{\frac{1}{2}\mathbf{F}_1}) \quad (268)\\
&= e^{A_1}e^{\mathbf{B}_1} + e^{A_2}e^{\mathbf{B}_2} + e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}(\tilde{R}_1R_2 + \tilde{R}_2R_1) \quad (269)
\end{aligned}
$$

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}_1R_2 + \tilde{R}_2R_1)\Big)\\
&\quad\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}_1R_2 + \tilde{R}_2R_1)\Big) \quad (270)\\
&= 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}_1R_2 + \tilde{R}_2R_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}_1R_2 + \tilde{R}_2R_1)\\
&\quad + e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(-\mathbf{B}_1-\mathbf{B}_2)}(\tilde{R}_1R_2 + \tilde{R}_2R_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}_1R_2 + \tilde{R}_2R_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}_1R_2 + \tilde{R}_2R_1)e^{\frac{1}{2}(A_1+A_2)}e^{\frac{1}{2}(\mathbf{B}_1+\mathbf{B}_2)}(\tilde{R}_1R_2 + \tilde{R}_2R_1) \quad (271)\\
&= e^{2A_1} + e^{2A_2} + 2e^{A_1+A_2}\cos(B_1 - B_2) \quad (272)\\
&\quad + e^{\frac{1}{2}(A_1+A_2)}(\tilde{R}_1R_2 + \tilde{R}_2R_1)(\quad (273)\\
&\quad\ e^{A_1}(e^{\frac{1}{2}(-\mathbf{B}_1+\mathbf{B}_2)} + e^{\frac{1}{2}(\mathbf{B}_1-\mathbf{B}_2)}) \quad (274)\\
&\qquad + e^{A_2}(e^{\frac{1}{2}(\mathbf{B}_1-\mathbf{B}_2)} + e^{\frac{1}{2}(-\mathbf{B}_1+\mathbf{B}_2)})) \quad (275)\\
&\quad + e^{A_1+A_2}(\tilde{R}_1R_2 + \tilde{R}_2R_1)^2 \quad (276)\\
&= \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}_1R_2 + \tilde{R}_2R_1)(\cos\frac{1}{2}(B_1 - B_2)) + e^{A_1+A_2}(\tilde{R}_1R_2 + \tilde{R}_2R_1)^2}_{\text{deep spinor interference}} \quad (277)
\end{aligned}
$$

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 subsequently 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 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.

7 Discussion

The final piece of the puzzle is to provide an interpretation of quantum mechanics as well as a mechanism for the wave-function collapse. Previously, we have derived from first principles the probability measure (Born rule) as the fundamental measure and the wave-function as a convenient construction to work with linear transformations. In fact, we have found the Born rule as a special case of the Gibbs measure. Consequently we will discuss an interpretation imported from statistical physics, but extended to account for our usage of the trace operator; that is, that the entropy is maximized under the constraint of *measurement-events* collected by *phase-invariant instruments* distributed in space-time... Let us see in more details.

First, let us note that our recovery of Fock Spaces follows the same mathematical process as that as that used in statistical physics to recover the Gibbs measure (Lagrange multipliers maximization), and thus our proposed interpretation will admit significant parallels with that of statistical physics.

In statistical physics, constraints on the entropy are interpreted as instruments acting on the system. For instance, an energy constraint on the entropy:

$$\overline{E} = \sum_{q \in \mathbb{Q}} \rho(q) E(q) \tag{278}$$

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

$$\overline{V} = \sum_{q \in \mathbb{Q}} \rho(q) V(q) \tag{279}$$

associated to a volume-meter acting on the system and converging towards an average volume value $\overline{V}$, also by producing a sequence of measurements of the volume $V_1, V_2, \ldots$. With these two constraints, the typical system of statistical physics is obtained, and its Gibbs measure is:

$$\rho(q, \beta, p) = \frac{1}{Z} \exp\big(-\beta(E(q) + pV(q))\big) \tag{280}$$

Comparatively, in our main result, the statistical physics interpretation can also be adopted. Instead of an energy-meter or a volume-meter, we have a phase-invariant meter, and the constraint is given as follows:

$$\operatorname{tr} \begin{pmatrix} 0 & -\overline{b} \\ \overline{b} & 0 \end{pmatrix} = \sum_{q \in \mathbb{Q}} \rho(q) \operatorname{tr} \begin{pmatrix} 0 & -b(q) \\ b(q) & 0 \end{pmatrix} \tag{281}$$

The usage of the trace enforces the phase-invariance of the instrument. Yet, and quite simply, maximizing the entropy under this constraints produces the probability measure of the wave-function including the Born rule. This is the true origin of the Born rule, here reported for the first time. The interpretation simply becomes that of an instrument performing a sequence of measurement on the system such that an average value is obtained, but instead of the simpler scalar instruments used in statistical physics, here we have a phase-invariant instrument. This instrument is responsible for the quantum mechanical behaviour associated with the wave-function. What is an example of such a detector; quite simply a photo-counter would be one. Such an instrument produces a sequence of incidences ('clicks') as photons are detected and "advanced features" such as an interference pattern is a consequence of this phase-invariance.

What is then the interpretation of quantum mechanics? Consider this comparative table between standard quantum mechanics, and our interpretation:

$$\text{QM:} \quad \text{Measurement}(\underbrace{\text{wave-function}}_{\text{Axiomatic}}) \implies \underbrace{\text{Result}}_{\text{collapse problem?}} \tag{282}$$

$$\text{Ours:} \quad \text{Max-Entropy}(\underbrace{\text{Result + Preparation}}_{\text{Axiomatic}}) \implies \underbrace{\text{wave-function}}_{\text{derived}} \tag{283}$$

Our interpretation places the ontology of the system at the level of 'knowledge' (of the preparation) and of axiomatic information (the result of the measurements describing the state the system is in). The probability measure consistent which this knowledge is *derived* as the wave-function. This is the reverse of the standard quantum mechanical interpretation. We maximize the entropy of the result+preparation to get the wave-function, rather than to "minimize" the wave-function via a measurement to the get the result. The measurement operation is philosophically problematic, and introduces the problem of the wave-function collapse; whereas the maximization of entropy is a non-problematic operation. The ontology of this interpretation matches that which natures gives us as the given (halting or experimental image), thus never encounters a wave-function collapse problem.

The interpretation of quantum mechanics appears to us, as that which John A. Wheeler had in mind when he wrote "Information, Physics, Quantum; The Search for Links.". where he makes statements of this nature:

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

Raw fact; a count — in our scheme this is encoded in the form of axiomatic information. The other part, not identified by Wheeler but nonetheless necessary to complete the description, is knowledge — in our scheme this is encoded in the form of protocols (or halting programs). The 'knowledge' corresponds to the steps required to construct an experiment in which photons are sent according to a repeatable and well-defined preparation. When a 'click' is registered, it yield more than just a bit; it also associated to a unit of 'knowledge' given in the form of a protocol able to produce similar preparations.

Our wave-function, as it is a sum over knowledge, refers to both the possible results of a measurement, and to the protocol used to prepare the system. For instance, a sum $Z = \det M(f_1) + \det M(f_2)$, would upon a mathematical projection yield a pick of either f_1 or f_2, which includes axiomatic information in the form of the pick result, as well as arbitrary preparation in the form of the protocol f_1 or f_2.

In the quantum gravity generalization to general linear transformations, the interpretation of quantum mechanics takes its most simple and neat interpretation. The phase-invariant instruments are upgraded from a complex phase to a general linear phase. The probability is now associated with a sequence of 'clicks' recorded in space-time as events. Thus, the framework describes reality as a sequence of space-time 'clicks' (or events) which, under entropy maximization, are associated to a general linear wave-function in lieu of the Gibbs ensemble, which describes quantum gravity. As you would recall, general relativity is primarily a theory of events in space-time, and quantum gravity as we have recovered it assigns a probability and an entropy to said events, such that the measure over said events is a wave-function able to support the transformations required by general relativity.

8 Conclusion

We believe the formal system of science here-in presented to be a more powerful formulation of the laws of physics, because it can provide mathematical definitions for experimental data, the observer, it can derive the laws of physics from the scientific method, resolve the problem of the wave-function collapse, identify the wave-function as a special form of the Gibbs ensemble, account for the origin of the Born rule and finally provide a (reversed) interpretation of quantum physics matching that of statistical physics, where the ontology of system measurements takes precedence over the derived measure.

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

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

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

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

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

[8] Andrew M Gleason. Measures on the closed subspaces of a hilbert space. In *The Logico-Algebraic Approach to Quantum Mechanics*, pages 123–133. Springer, 1975.

[9] Ruth E Kastner. *The transactional interpretation of quantum mechanics: the reality of possibility*. Cambridge University Press, 2013.

[10] Lluís Masanes, Thomas D Galley, and Markus P Müller. The measurement postulates of quantum mechanics are operationally redundant. *Nature communications*, 10(1):1–6, 2019.

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

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

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

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:

$$\mathcal{L}(\rho, \alpha, \tau) = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q)\right) + \tau \operatorname{tr} \left(\overline{\mathbf{M}} - \sum_{q \in \mathbb{Q}} \rho(q)\mathbf{M}(q)\right) \tag{284}$$

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:

$$\frac{\partial \mathcal{L}}{\partial \rho(p)} = -k_B \ln \rho(p) - k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p) \tag{285}$$

$$0 = k_B \ln \rho(p) + k_B + \alpha + \tau \operatorname{tr} \mathbf{M}(p) \tag{286}$$

$$\implies \ln \rho(p) = \frac{1}{k_B} \left(-k_B - \alpha - \tau \operatorname{tr} \mathbf{M}(p)\right) \tag{287}$$

$$\implies \rho(p) = \exp\left(\frac{-k_B - \alpha}{k_B}\right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(p)\right) \tag{288}$$

$$= \frac{1}{Z} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(p)\right) \tag{289}$$

where Z is obtained as follows:

$$1 = \sum_{q \in \mathbb{Q}} \exp\left(\frac{-k_B - \alpha}{k_B}\right) \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q)\right) \quad (290)$$

$$\implies \left(\exp\left(\frac{-k_B - \alpha}{k_B}\right)\right)^{-1} = \sum_{q \in \mathbb{Q}} \exp\left(-\frac{\tau}{k_B} \operatorname{tr} \mathbf{M}(q)\right) \quad (291)$$

$$Z := \sum_{q \in \mathbb{Q}} \det \exp\left(-\frac{\tau}{k_B} \mathbf{M}(q)\right) \quad (292)$$

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:

$$\overline{\mathbf{M}}_1 = \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_1(q) \quad (293)$$

$$\vdots \quad (294)$$

$$\overline{\mathbf{M}}_n = \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_n(q) \quad (295)$$

Then the Lagrange equation becomes:

$$\mathcal{L} = -k_B \sum_{q \in \mathbb{Q}} \rho(q) \ln \rho(q) + \alpha \left(1 - \sum_{q \in \mathbb{Q}} \rho(q)\right) + \tau_1 \operatorname{tr} \left(\overline{\mathbf{M}}_1 - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_1(q)\right) + \dots + \tau_n \operatorname{tr} \left(\overline{\mathbf{M}}_n - \sum_{q \in \mathbb{Q}} \rho(q) \mathbf{M}_n(q)\right) \quad (296)$$

and the measure references all n constraints:

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{\tau_1}{k_B} \mathbf{M}_1(q) - \dots - \frac{\tau_n}{k_B} \mathbf{M}_n(q)\right) \quad (297)$$

B.2 Multiple constraints - General Case

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

$$\overline{M}_{00}\begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{00}(q) \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \tag{298}$$

$$\vdots \tag{299}$$

$$\overline{M}_{01}\begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{01}(q) \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} \tag{300}$$

$$\vdots \tag{301}$$

$$\overline{M}_{nn}\begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = \sum_{q\in\mathbb{Q}} \rho(q) M_{nn}(q) \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} \tag{302}$$

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

$$\rho(q) = \frac{1}{Z} \det \exp\left(-\frac{1}{k_B} \boldsymbol{\tau} \odot \mathbf{M}(q)\right) \tag{303}$$

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