Quantum Probability as an Application of Data Compression Principles

Allan F. Randall
Dept. of Information & Communication Eng. Tech.
Centennial College
Toronto, Ontario, Canada
research@allanrandall.ca

Realist, no-collapse interpretations of quantum mechanics, such as Everett's, face the probability problem: how to justify the norm-squared (Born) rule from the wavefunction alone. While any basis-independent measure can *only* be norm-squared (due to the Gleason-Busch Theorem) this fact conflicts with various popular, non-wavefunction-based phenomenological measures—such as observer, outcome or world counting—that are frequently demanded of Everettians. These alternatives conflict, however, with the wavefunction realism upon which Everett's approach rests, which seems to call for an objective, basis-independent measure based only on wavefunction amplitudes. The ability of quantum probabilities to destructively interfere with each other, however, makes it difficult to see how probabilities can be derived solely from amplitudes in an intuitively appealing way. I argue that the use of algorithmic probability can solve this problem, since the objective, single-case probability measure that wavefunction realism demands is exactly what algorithmic information theory was designed to provide. The result is an intuitive account of complex-valued amplitudes, as coefficients in an optimal lossy data compression, such that changes in algorithmic information content (entropy deltas) are associated with phenomenal transitions.

1 Introduction

Approaches to quantum foundations based on Everett's no-collapse and/or wavefunction realism postulates [6] inevitably hit the probability problem [1, 5, 7, 11–14, 16, 25–27]: how to justify the norm-amplitude squared (Born) rule [2] from the wavefunction alone. While any probability based on an objective, basis-independent measure can *only* be norm-squared, due to the Gleason-Busch Theorem [4, 8], this is often seen as being in conflict with various intuitive and popular ideas and conceptions of probability. Most particularly, it is felt that some kind of phenomenological counting measure, rather than a wavefunction-based measure, ought to be employed, such as observer, outcome or world counting [6, 11]. While little rigorous argument has been presented for this feeling, it appears to be quite strong, to the extent that such a non-wavefunction-based measure is widely cited as a requirement for any pure wavefunction-based (no-collapse) interpretation of quantum probability.

This would seem, on the face of it, a somewhat contradictory demand: since wavefunction realism is the hall-mark of no-collapse interpretations, why are we requiring anything but a purely wavefunction-based measure of them? The answer may lie in the lack of an intuitive explanation for any pure wavefunction measure—how can wavefunction amplitudes actually "count" towards a probability measure, in the way marbles count when we pick one at random from a bag? Instead of something straightforward, we have counterintuitive puzzles like complex counts and negative probabilities. What is needed, then, is *not* a Born rule proof, which we already have (actually we have lots of them!). What is needed is an intuitive story to tell about why complex-valued counts can make sense, connecting the ontological

A.A. Abbott and D.C. Horsman (Eds.): 7th International Workshop on Physics and Computation (PC 2016).
EPTCS 214, 2016, pp. 29–40, doi:10.4204/EPTCS.214.6

© Allan F. Randall
This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works License.

wavefunction to the phenomena without appealing to any reality beyond the wavefunction and without making phenomenal entities the basis of the probability counts.

I will explore the feasibility of an algorithmic probability measure, by making the minimal assumptions about such a measure as are consistent with *(1)* wavefunction realism and *(2)* the assumption that quantum probability should be algorithmic and information-theoretic in nature. The result is a derivation of the Born rule that provides an intuitive rationale for complex counts, conceiving them as changes in algorithmic information content (entropy deltas) associated with phenomenal transitions. It remains true to the realist, objectivist intent of no-collapse interpretations, and gives an intuitive story for the role of probability interference under such interpretations. Although only the impact on the Born rule is discussed here, this is intended as a first step towards a complete and consistent algorithmic interpretation of quantum mechanics (see also [19]).

2 The Probability Objection

We assume with Everett the doctrine of wavefunction realism [6]:

> "A wave function that obeys a linear wave equation everywhere and at all times supplies a complete mathematical model for every isolated physical system without exception."

It might seem that this makes probabilities meaningless since all possible outcomes are realized. However, one can model observers within the wavefunction as machines or robots, with sensors to collect information, circuitry to process that information, and memory to store the information so that it may affect future behaviour. We then seek to calculate, given a phenomenon being experienced by one of these observers, what the transition probabilities are for all subsequent possible "continuer" experiences for that observer, as described by the unitary evolution of the wavefunction.

Unfortunately, it is not at all obvious how one is supposed to go about performing such a calculation. One problem is that the very idea of multiple different mutually incompatible experiences is alien to our common experience. We do not wake up one morning to find five other copies of ourselves running around the house. However, a well-known and purely classical thought experiment can yield just this scenario without requiring any quantum mechanics. Imagine an ingenious replicator machine that can make a complete (or complete-enough) scan of the physical state of a human being (including the brain) and then build any number of perfect working replicas or copies of that human being. Assume for convenience that the original body and brain are destroyed in the scanning process, so there is no temptation to confer a privileged status on any one resulting copy.

Imagine you are about to be so copied. You are told that the scanner produces two identical copies, each of which exits through a separate exit door. Each door goes to a small room with a closed box on the floor. The rooms are identical, except that in one room, the closed box has a live cat inside, while in the other room, it contains a dead cat. You are told that when your copies exit the machine, they will be asked to open the box that appears in front of them. What, you are asked, is the probability, from *your* perspective, of finding a live cat in the box? Obviously, the answer is $1/2$, since there are two copies or continuers made, and you could find yourself equally well as either continuer after replication.

Probability objectors to Everett reason that wavefunction probabilities should work the same way. In the original cat experiment [20], the wavefunction computed two and only two continuers, and hence an *a priori* probability of $1/2$ for each. Or, more generally, for n continuers/copies, we have

$$p\,(m_i \mid m) = \frac{1}{n}, \tag{1}$$

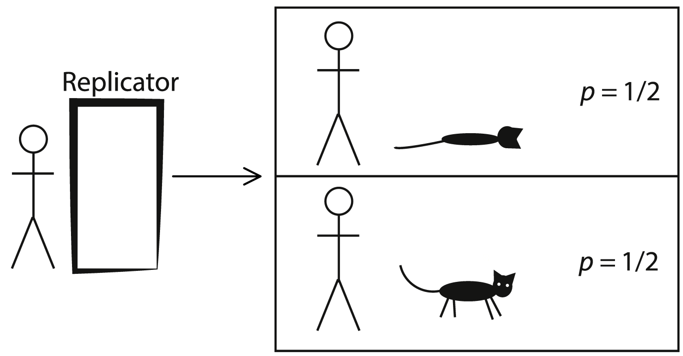

Figure 1: Replicator copies: *two* copies yields *two* counts of equal measure.

resulting in a simple, flat probability distribution across all the Everettian "branches" (or "worlds" or "copies" or "continuers"). Of course, we know this is *not* the actual probability rule, which is instead Born's norm-amplitude-squared rule.

To avoid the reification of subsystems, which goes against wavefunction realism, we will express the wavefunction in Everettian terms, by decomposing the entire wavefunction of the universe according to an observation basis that is defined relative to an observer experience m and its corresponding continuers:

$$|\psi\rangle = a_1 |\psi_1\rangle + \cdots + a_n |\psi_n\rangle , \tag{2}$$

so that, according to the Born rule,

$$p(m_i \mid m) = p(\psi_i \mid \psi) \propto |\langle \psi_i \mid \psi \rangle|^2 = |a_i|^2 . \tag{3}$$

This formulation makes a couple of assumptions:

1. It is possible to treat each continuer as an independent, orthogonal basis element.
2. It is possible to treat the wavefunction $|\psi\rangle$ as a linear combination of these continuers.

Neither of these assumptions is ultimately problematic, however. If we suspend assumption 1, it means there are at least two continuers that are not linearly independent, and so would have to be included as multiple copies *within one Everettian branch*. Since we, in fact, do not generally experience multiple continuers within our branch or "world"—since nothing like replicator machines seem to exist—assumption 1 does not present a serious problem. We could easily suspend it, by using the Born rule to assign a probability to each $|\psi_i\rangle$, and then split $|\psi_i\rangle$'s measure equally amongst its internal continuers or copies.

We could also suspend assumption 2. If the linearly-independent continuers of m cannot be linearly combined to reproduce $|\psi\rangle$, then we simply add a unit-norm remainder state $|\psi_D\rangle$, in which no internal continuer of m can be found, but which allows the expansion in terms of continuers to add up to the wavefunction:

$$|\psi\rangle = a_1 |\psi_1\rangle + \cdots + a_n |\psi_n\rangle + a_D |\psi_D\rangle , \tag{4}$$

so that $|\psi_D\rangle$ might be called the "dead branch", since it is the world(s) or branch(es) without any continuer (the original observer having apparently "died" in that branch).

In most laboratory applications of Born's rule, however, both assumptions 1 and 2 clearly hold, so we will continue to talk in terms of these two assumptions, without loss of generality.

Experiment upholds the Born rule, which calculates probabilities based on amplitudes, so that equal amplitude means equal probability, higher amplitudes mean higher probability, and lower amplitudes mean lower probability. In fact, Everett [6] proves that so long as one assumes "amplitude dependence"—that the measure of an outcome is a function solely of amplitude—then the Born rules follows.

The point of the objectors to Everett is that there is no *a priori* justification for amplitude-dependence under the assumptions of wavefunction realism. A lower-amplitude branch is not somehow "less computed" by the wavefunction than the high-amplitude branches. So how do we justify giving it a lower measure, in the way we justified counting copies in the replicator experiment? Thus far, it seems that the world or outcome counters have the upper hand, since counting outcomes (or worlds or continuers) would seem to be the equivalent of counting replicator copies. However, as we will soon see, the situation is not as simple as that.

3 Response to the Probability Objection

The most mathematically straightforward, and most formalistic, response to this objection (although not one very commonly used) is simply to point to the Gleason-Busch theorem [4, 8], which tells us that Born's rule is mathematically the *only* possible probability rule, under an assumption of additivity for the probabilities (demanded by the nature of probabilities) along with the assumption of "noncontextuality" or "basis-independence", which states simply that the measure used to compute probability is independent of the basis used to expand $|\psi\rangle$. Thus *any* basis $\{\ldots, |\psi_i\rangle, \ldots\}$ that contains continuer state $|\psi_i\rangle$, regardless of whether the other basis elements correspond to continuers, can use the *same* probability rule to calculate *the same* probability for m_i.

In Everettian terms, noncontextuality is often thought of in terms of branch-dependence, since this is a consequence of noncontextuality, and one that can be intuitively applied to branches. Mathematically, branch-dependence states that there exists a function f such that

$$p(m_i \mid m) \propto f(|\psi_i\rangle), \tag{5}$$

so that the measure of a particular continuer is a function only of that continuer's branch of the wavefunction (other than for normalization). From this, one can prove amplitude-dependence, and from there the Born rule, via either Everett's proof [6] or (for dimensions greater than two) via certain versions of Gleason's original proof [15].

But *is* this assumption at all intuitive, within wavefunction realism? The probability objectors say no, since the wavefunction computes all the continuers, and probability is something that can only be calculated from the perspective of the observer. Hence, we have no reason to assume that the measure we use must be noncontextual, since it is intended to yield only the probability from the perspective of this observer (m). This is arguably a subjectivist conception of the probability measure: all that matters is how many subjectively indistinguishable states there are. If there are n such states, then the probability—*from the perspective of* m—of finding oneself in *one* of those n continuer states or copies *must* be $1/n$. Or so goes the reasoning.

There is, unfortunately, little in the way of rigorous argumentation for this "world-counting" or "observer-counting" assumption. When it is used, it is usually simply assumed. Note that in many very straightforward applications of probability theory, the categories used are subjective indistinguishables, while the countables are still assumed to represent actually existing, objective entities. For instance, if I pick a marble at random from a bag with 7 blue and 3 red marbles, there is a 30% chance of picking "red". Now "red" is a subjective category, true. But the actual objective entities being picked are just marbles. I could distinguish them differently if I chose. I choose "red" versus "blue" because that makes sense for me. However, the objects counted are still the actual marbles, with the objective count of things falling within my subjective category placed in the numerator, and the total objective count of things in the denominator. In neither case am I "counting" subjective indistinguishables, but rather objectively existing things. So we still call this an "objective" probability rule, not a subjective one (so long as we consider the marbles themselves to objectively exist).

It might seem that counting distinct observers is like counting copies in the replicator experiment, but a little reflection reveals otherwise. Imagine that, instead of *two* copies, there are now *three* copies, three exit doors and three rooms with closed boxes. One room has a box with a dead cat, while the other two have a box with a live cat. What now are the probabilities of seeing a live cat, given the method of counting copies? It would seem now that the probability of seeing a live cat is clearly $2/3$, given that there are three copies and two of them see a living cat.

But are there *really* two live-cat copies, in other words *two distinct observers* that see a live cat? Probably in a realistic scenario, the answer is clearly yes, since there will be at least subtle but real differences in the two rooms with the live cats. However, if we really can make the rooms *absolutely identical* in every respect, other than the cat inside the box, then it is arguable that there is really only one observer that sees a live cat, since there is no formal difference between the two observers, and under wavefunction realism, observers are considered to be merely emergent phenomenon arising from a purely formal computational system. If the *experiences* of the two live-cat copies are absolutely identical, then surely—from the perspective of this observer—there is only *one* indistinguishable experiential continuer-state, and thus these two live-cat rooms actually constitute a *single* observer.

And clearly, the probability of seeing a live cat remains $2/3$, in either case, for if we did the same experiment on someone repeatedly, they would measure a 2:1 ratio of live:dead outcomes, regardless of how precisely (or imprecisely) we made the two live-cat rooms match.

In other words, there is *one* continuer/observer present while there are *two* physical copies/rooms. Thus, we are *not*, after all, counting observers. One of the two observers actually consists of two *copies*, and *this* is what we count. The number of copies, then, is comparable to wavefunction amplitude, while the observer-count (or world-count) is itself more or less irrelevant to the measure.

Of course, this by no means proves that amplitudes are the appropriate countable for wavefunctions. However, if amplitudes could be counted in a classical manner to produce probabilities, then there would be a much stronger argument for equating copy-counts to amplitudes. But amplitudes are complex-valued, which means they can exhibit destructive interference. This makes the comparison with the replicator rooms obscure. What would it mean to send someone through a replicator machine, that does everything required to produce both live-cat copies, but somehow, because these two copies have opposite phase, they cancel each other out, and now we have *no one* observing a live cat? How can copies cancel each other out, or interfere with each other?

We can perhaps make things less obscure by replacing the replicator copies with multiple copies simulated in a computer, in a virtual reality. For here we have a situation (a computer program) whose functioning we can more easily consider purely on its formal properties, as a Turing machine, since the "observers" and "rooms" are simulated and not really constituted from actual physical objects. This is actually a benefit within the Everettian context, since we are *supposed* to be calculating probabilities, in the first place, from the formal structure of the wavefunction alone.

In this scenario, the program is actually computing the live-cat observer *twice*. So twice the computing cycles, computer memory and other computing resources are being used to simulate the two observers. Perhaps, then, they should count as "two copies", even though they are subjectively indistinguishable as observers. However, once again, the fact that the amplitudes can interfere call this inter-

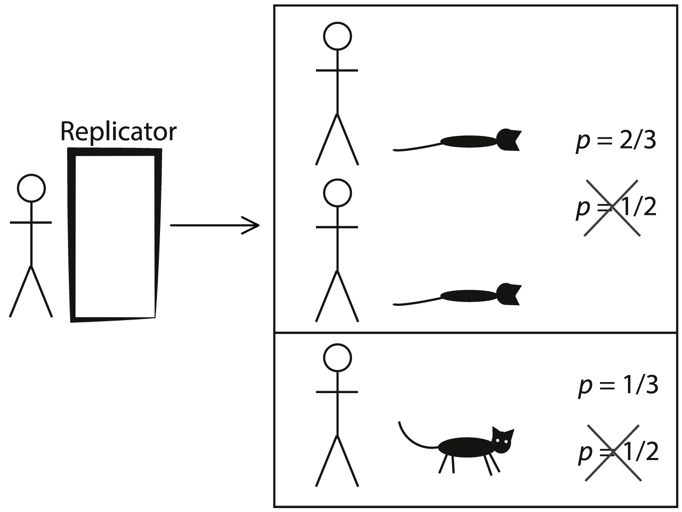

Figure 2: Replicator copies: *two* observers, *three* copies yields *three* objective counts of equal measure in *two* subjective categories.

pretation into question. For the resources dedicated to the calculation of the wavefunction are, it would seem, the same whether the amplitude is high or low. And two interfering high amplitudes can result in *no* amplitude for a particular outcome. Yet, it is not clear that "no resources" went into the computation of that outcome. In a sense, it is obvious that no resources went into its computation, since the wavefunction does not even compute that outcome at all! Yet, if we change the phases just *slightly* so that there is a *tiny* remaining amplitude for that outcome, then it is suddenly again very unclear how we could argue that there is somehow "very little" computing resources dedicated to that outcome, just because the amplitude is low. The fact that the amplitude is low does not, surely, mean that it takes less resources to compute it, provided that it *is* in fact computed. So, while a straight simulation of the replicator example as a computer simulation would seem to strengthen the copy-count/amplitude analogy, any incorporation of probability interference seems to call the analogy into question.

4 The Algorithmic Assumption

In order to find anything to count within a "purely formal" system like a pure wavefunction, we have to have some idea just *what* a pure formalism is. While there are competing conceptions of formality, the idea of computation as the core of all formal systems carries a great deal of weight. Gödel [9] claimed, for instance, that Turing's machines provided "a precise and unquestionably adequate definition of the general concept of formal system" and that "the term 'formal system' or 'formalism' should never be used for anything but this notion."

I will therefore adopt the minimal *algorithmic assumption* needed to justify the application of algorithmic probabilities to emergent observers:

> We can, if we wish, model an observer's *experience* as a set of *finitistic algorithms* (Turing machines) each of which corresponds to an abstract computer program that *sufficiently and uniquely describes* that observer experience.

I will refer to any such sufficient and unique algorithmic description as "outputting" that observer/

experience. Note that some convention must be adopted to interpret a program's state in terms of "output", such that a program cannot be said to describe/output an experience unless it is distinguished from any other experiences the same program might also incidentally describe.

This terminology is not meant to rule out the possibility that an experience may only be output "in the limit" of an infinite number of computational steps of a nonhalting machine (see [3, 10, 17, 18]). Note that in spite of the finitistic nature of computational models, the continuum can be modelled in such systems. Also, no assumption is made here that there is either a finite or infinite number of continuers for any given experience.

Under algorithmic probability, two different continuers might both be computed, but one may still have higher probability, since it is output by many more machines than the other—just as the replicator might produce more physical copies that correspond to one continuer than another. But instead of counting physical copies, as in the replicator example, in this purely formal ontology, we count *programs*.

To calculate a probability for a given experience m, we must find all the valid continuer experiences $\{m_1, \ldots, m_n\}$, $0 \leq n \leq \infty$. We then need to find a way of determining, from the given continued experience, m, what the probability $p(m_i \mid m)$ is for some particular valid continuer m_i. In the quantum context, the continued experience must be output by the pre-measurement wavefunction of the universe (under some appropriate interpretation of "output").

If we accept the view that formal systems are essentially Turing machines, or abstract computer programs, then what are the "countables" of a purely formal system, in which observers are merely computed emergent properties? If we take the wavefunction of the universe as the only existing thing, and treat it as a purely formal system, then does this not mean that we have just exactly *one* thing to count, namely the wavefunction of the universe considered as a program?

The answer to this seems to be no, since the wavefunction can, in fact, be computed by an infinite number of distinct programs. If we consider each such program code to be a distinct encoding of the wavefunction, then the wavefunction "counts" as far more than a single object—we must count every distinct program code that generates the wavefunction. The distinct encodings are all different, but they all encode the same algorithm for generating our wavefunction. In a sense, they are different machines that perform the same calculation, just as I can add 2+2 on an abacus or a pocket calculator.

Solomonoff [21–23] initiated the field of algorithmic probability with the discovery of a well-defined algorithmic measure of information. The "information" or "entropy" $H(x)$ of a code or symbol sequence x can be defined more or less as the size or length in bits of $\mathscr{C}(x)$, the "optimal compression" of x, meaning the shortest self-delimiting program that generates x:

$$H(x) = |\mathscr{C}(x)|. \tag{6}$$

A program is "self-delimiting" if one can tell from the encoding itself where the end of the program is. The probability of a randomly chosen Turing machine (or program) generating x is then simply

$$p(x) = 2^{-H(x)}. \tag{7}$$

A program x is considered to be "algorithmically random" if and only if it is incompressible:

$$x = \mathscr{C}(x). \tag{8}$$

So long as you can further compress x, it must contain some kind of pattern or symmetry, else you would not have been able to shorten your description of it.

The strict expression of $p(x)$ is more complicated than indicated above, and involves *summing* probabilities over all the infinity of programs that generate x, rather than simply taking the shortest program.

However Solomonoff [24] proved that summing over *all* programs (even an infinite number of them) was *almost* the same thing as taking the shortest program, since the shorter programs contribute exponentially more to the probability than the longer programs. The result is a semi-measure rather than a measure, due to the fact that some encoding are nonhalting and produce no output, but Solomonoff converts this to a proper measure by normalizing.

One potential problem with the algorithmic measure is the question of its objectivity. I have thus far assumed that we have adopted a particular computational language in which to encode our programs, yielding a particular bit-count for any program. However, choice of a different language will yield different bit counts and hence a different probability measure. Nonetheless, Solomonoff's invariance theorem [24] proves that the measure is invariant between languages, up to an additive constant given by the size of the translation manual between the languages. Thus, for simple languages, it does seem that the measure should be very close to an objective measure.

I cannot *prove* that algorithmic probability is the correct *a priori* measure for continuers, of course, since we cannot (at least not currently) completely formalize observers and their method of interfacing with the rest of the wavefunction. Nonetheless, we will take the algorithmic information-theoretic perspective as a fundamental assumption of this paper. If we can show that such a measure is consistent with the wavefunction and with the Born rule, with an intuitive *a priori* justification akin to our count of replicator copies, then algorithmic probability will make for a very intuitive choice for an *a priori* measure. At the very least, we can expect it to give world-counting a run for its money!

5 Transition Probabilities

How do we calculate an algorithmic probability for continuer m_i given pre-measurement observer m? What we cannot do is to pick m apart, in order to count the number of bits in some subsystem of m associated with the observer's experience, as this would reify subsystems, and we are looking for a purely formal probability rule. This leaves us with m in its entirety for which to calculate an entropy.

Presumably, a continuer state m_i and its continued state m may or may not share a lot of mutual information. If they do, then m_i can be described in very few bits once we already have a description of m. Otherwise, m_i will require many bits to describe, even given a description of m. Those continuers that share very little information will have very low probability, since it will take a large number of bits to describe them (we will call these "maverick continuers"). Continuers m_i that have significant probability will be those with high mutual information content with the continued state m.

Thus, the unshared residual information required to describe a continuer m_i given m is the change in entropy (or information) in transition from m to m_i, meaning the size of the optimal compression of m_i given that we already have m:

$$\triangle H_i = H(m_i|m) = H(m_i) - H(m_i : m) = H(m_i) - H(m), \tag{9}$$

where $H(x : y)$ is defined as the "mutual information" between x and y, or the optimal number of bits it would take to generate x if already given y "for free", so that $H(x : y) = H(x) - H(x|y)$.

Thus, the probability of continuer m_i is based on the (typically small) ΔH_i associated with the transition $m \to m_i$. We will call ΔH_i the "alteration" bits for the transition, as it is the number of bits that would have to be added to the optimal compression of m to generate the optimal compression of m_i.

The Solomonoff probability $p\,(m_i \mid m)$ of the transition $m \longrightarrow m_i$ is:

$$p(m_i|m) = \frac{p\,(m_i)}{p\,(m)} \propto 2^{-\triangle H_i} = 2^{-(H(m_i)-H(m))} = \frac{2^{-H(m_i)}}{\sum_{k=1}^{n} 2^{-H(m_k)}}. \tag{10}$$

This probability rule has *nothing to do with quantum mechanics*, being a general rule for calculating probabilities for multiple continuers in a formal system, using algorithmic probability. The rule is clearly different from counting branches or observers to compute probabilities (which would set $p(m_i) = \frac{1}{n}$), since the entropies of the individual branches are not necessarily equal ($\triangle H_1 \neq \triangle H_2$). Since there exists a function f such that $p(m_i|m) \propto f(m_i)$, it follows that the unnormalized measure of a branch depends only on the information content of *that* branch, as the content of the prior state (or of the other continuers) serves only to normalize the measure. This is, of course, simply a form of branch-dependence, so if it were possible to apply this framework to quantum mechanics, the Born rule would necessarily hold.

6 Quantum Transition Probabilities

To apply our algorithmic measure to continuer quantum states, we need to interpret such states as optimal compressions. As it turns out, it is very natural to interpret a quantum state as an optimal compression. Established real-world data compression algorithms intended for synthetic decompression (into something experienced by an observer), are typically highly lossy, and are usually some form of "transform" compressor, using some variation on the discrete Fourier transform (or DFT):

$$\mathscr{F}\left(\{\langle x \mid \psi\rangle : x = 1\cdots N\}, R\right) = \{\langle k \mid \psi\rangle : k = 1\cdots R\}, \tag{11}$$

quantizing the original N data points $\langle x \mid \psi\rangle$ into R frequency amplitudes $\langle k \mid \psi\rangle$, where $R \ll N$, so that the inverse transform

$$\mathscr{F}^{-1}\left(\{\langle k \mid \psi\rangle : k = 1\cdots R\}, N\right) = \{\langle x \mid \psi\rangle : x = 1\cdots N\} \tag{12}$$

decompresses back to the original data set at resolution (dimensionality) N, with

$$\left\{|x\rangle = \left[e^{i\frac{1}{R}x\tau}\cdots e^{i\frac{k}{R}x\tau}\cdots 1\right] : x = 1\cdots N\right\} \text{ and } \left\{|k\rangle = \left[e^{-ik\frac{1}{N}\tau}\cdots e^{-ik\frac{x}{N}\tau}\cdots 1\right] : k = 1\cdots R\right\} \tag{13}$$

being the orthonormal "spatial" and "frequency" bases, respectively, and where τ is angular unity, the circumference of the unit circle (more precisely, τ is some appropriate minimum-bit-length finite-precision approximation to angular unity, since no actual data compression takes place on a machine with infinite precision). Note that the Fourier transform is a unitary transformation and is the solved form of Schrödinger's equation.

We thus interpret $|\psi\rangle$, when expressed in a pure frequency basis of dimension R, to be the optimal compression of pre-measurement observer state m, so that $H(m) = bR$, where b is the number of bits used by the compression algorithm to represent a single complex amplitude.

The compressed amplitude list itself, being an optimal compression, is algorithmically random, containing no further structure or symmetry. Once transformed to the spatial basis, where the dimensionality N is larger, it is no longer an optimal compression, but nonetheless encodes the same wavefunction, and so has the same entropy or information content. Since Fourier transforms are unitary, the unitary evolution of the wavefunction is in the same category: it does not increase entropy. However, when an observer splits (at least from that observer's perspective) then bits have been added to the system, and the optimal compression will be greater than R by an amount we may call $\triangle H_i$.

Thus, the algorithmic transition rule from the previous section can be readily applied to the inner product vector space of Fourier compressions. Note that the Solomonoff probability for m_i is based on a *different* DFT than the one for m, with a higher dimensionality and entropy. The probability of m_i is based on ΔH_i, with the entropy of m serving only to normalize the measure (as in non-quantum version):

$$p(m_i|m) = \frac{p(m_i)}{p(m)} \propto 2^{-\Delta H_i} = 2^{-(|\mathscr{F}_{min}(m_i)| - |\mathscr{F}_{min}(m)|)} = \frac{2^{-|\mathscr{F}_{min}(m_i)|}}{\sum_{k=1}^{n} 2^{-|\mathscr{F}_{min}(m_k)|}}. \tag{14}$$

Thus, as with the non-quantum version, there exists a function f such that $p(m_i|m) \propto f(m_i)$, so that the measure of a continuer depends only on *that* continuer branch, not on the context of the other possible branches (other than for normalization). This is branch-dependence translated to a periodic (DFT) context, and it rules out world-counting as the basis for a probability rule.

Thus, if we apply a measure on the discrete Fourier spectrum, it must, for each frequency, depend only on the individual amplitude for that frequency (other than for normalization). This makes intuitive sense, since the overall list of amplitudes is an optimal compression, and hence algorithmically random, so there can be no structure or symmetry found within it, with which to extract anything further that could be informative for our measure.

A corollary to branch-dependence is basis-independence (noncontextuality). If our measure depends *only* on the given branch, then it cannot depend on the complement of the branch (the remainder of the wavefunction) yielding "complement-independence". And if the complement does not matter, then it cannot matter what basis we consider $|\psi_i\rangle$ to be a part of. Given this noncontextuality, Gleason-Busch [4, 8] assures us that all measures of projections onto subspaces of the vector space are norm-squared measures, and the probability of observer state m in pre-measurement universe $|\psi\rangle$ becoming continuer m_i in post-measurement universe $|\psi_i\rangle$ follows the Born rule:

$$p(m_i|m) \propto |\langle\psi_i|\psi\rangle|^2 = |a_i|^2. \tag{15}$$

7 Discussion

We have seen that algorithmic probabilities based on wavefunction realism lead very naturally to the Born measure. It might seem odd that norm-amplitude alone can tell us the measure of a continuer, when algorithmically we are supposed to be using a bit-counting measure, not an "amplitude-counting" measure. Since each amplitude occupies the same number of bits in a DFT, it might appear that amplitude-counting cannot possibly be a form of bit-counting. But recall that it is the $\triangle H_i$ alteration bits for continuer m_i that are counted, not bits *within* a DFT representation, which would violate wavefunction realism.

Imagine the extreme case of just two continuers, m_1 and m_2 (with no non-continuers) where $\mathscr{C}(m_1)$ is virtually identical to $\mathscr{C}(m)$, but for a few minor alterations, while $\mathscr{C}(m_2)$ is radically different (perhaps a maverick continuer). This could be the case even if the *apparent* difference between the two continuers, to the observer, is only slight. The measures for m_1 and m_2 can be directly bit-counted by independently compressing each continuer. But assuming we already have the observation-basis expansion $|\psi_m\rangle = a_1|\psi_1\rangle + a_2|\psi_2\rangle$, is there no way to compute the probabilities from *this* representation, simply in terms of $a_1|\psi_1\rangle$ and $a_2|\psi_2\rangle$, without counting bits or performing optimal compressions? Considering that $\mathscr{C}(m_1)$ is already almost identical to $\mathscr{C}(m)$, clearly a_1 must be very high. This follows straightforwardly from the definition of the DFT. A *low* $\triangle H_i$ alteration-bit-count means a *large* amplitude in the preferred expansion. By the same reasoning, a *high* alteration-bit-count corresponds to a *small* amplitude, and since $\mathscr{C}(m_2)$ contributes almost nothing to $\mathscr{C}(m)$, a_2 will be tiny.

It thus makes sense for the alteration bits (or change in entropy) to be reflected in the observation-basis expansion as norm-amplitudes. It also makes sense that the formula for a norm-amplitude-based measure will be norm-amplitude-squared, since this is a conserved measure across unitary transformations that obeys additivity—and due to Gleason-Busch, we know it is the *only* such measure. And under this alternative view of probability, the idea of probability interference is no longer mysterious, as it is an inherent and expected feature of the DFT frequency-based representation scheme.

Returning to the replicator example, simply replace the three replicator copies with three different programs, and we end up with the same basic justification for our new algorithmic measure, just applied to the domain of programs rather than physical subsystems.

Some readers may fear I have artificially chosen an optimal compression scheme that just happens to be the discrete form of the solved Schrödinger equation. However, keep in mind that we are permitted in this exercise to assume the wavefunction under wavefunction realism. I would, however, challenge anyone concerned about the choice of the DFT to propose an alternative *a priori* guess for an optimal compression scheme that they consider to be at least as good, and our discussion can proceed from there. Considering that, in terms of real world practical technology, the DFT or something very like it, is at the heart of nearly all practical synthetic (lossy) compression schemes, it seems a very plausible choice.

It is worth noting, in addition, that while real-world lossy compression algorithms can take more specific forms than the relatively straightforward DFT compression used in this paper, the resulting algorithms are still usually describable as "transform compression" or "modified DFT" algorithms, being based on essentially the same scheme. Given that branch-dependence is a general property of algorithmic continuers, independent of which compression algorithm is used, basis-independence should arise given just about any transform-based compression algorithm, and the Born rule will still follow.

References

[1] David Z. Albert (2010): *Probability in the Everett picture*. In Simon Saunders, Jonathan Barrett, Adrian Kent & David Wallace, editors: *Many Worlds? Everett, quantum theory, and reality*, Oxford University Press, Oxford, doi:10.1093/acprof:oso/9780199560561.003.0013.

[2] Max Born (1926): *Zur quantenmechanik der stoßvorgänge*. *Zeitschrift für Physik* 37(12), pp. 863–867, doi:10.1007/BF01397477.

[3] Mark S. Burgin (1983): *Inductive Turing machines*. *Notices of the Academy of Sciences of the USSR* 270(6), pp. 1289–1293.

[4] Paul Busch (2003): *Quantum states and generalized observables: a simple proof of Gleason's theorem*. *Physics Review Letters* 91(120403), doi:10.1103/PhysRevLett.91.120403.

[5] David Deutsch (1999): *Quantum Theory of Probability and Decisions*. *Proceedings of the Royal Society of London A* 455(1988), pp. 3129–3137, doi:10.1098/rspa.1999.0443.

[6] Hugh Everett III (1957): *'Relative state' formulation of quantum mechanics*. *Reviews of Modern Physics* 29, pp. 454–462, doi:10.1103/RevModPhys.29.454.

[7] E. Farhi, J. Goldstone & S. Gutmann (1989): *How probability arises in quantum mechanics*. *Annals of Physics* 192, pp. 368–382, doi:10.1016/0003-4916(89)90141-3.

[8] Andrew M. Gleason (1957): *Measures on the closed subspaces of a Hilbert space*. *Journal of Mathematics and Mechanics* 6(6), pp. 885–893, doi:10.1512/iumj.1957.6.56050.

[9] Kurt Gödel (1967): *On formally undecidable propositions of Principia Mathematica and related systems I.* In J van Heijenoort, editor: *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Harvard University Press, pp. 592–617, doi:10.1016/j.hm.2003.07.003.

[10] E. Mark Gold (1965): *Limiting recursion*. *Journal of Symbolic Logic* 30(1), pp. 28–48, doi:10.2307/2270580.

[11] Neill Graham (1973): *The measure of relative frequency*. In: *The Many Worlds Interpretation of Quantum Mechanics*, Princeton University Press, Princeton, pp. 229–253, doi:10.1515/9781400868056.

[12] Sam Gutmann (1995): *Using classical probability to guarantee properties of infinite quantum sequences*. Physical Review A 52(5), pp. 3560–3562, doi:10.1103/PhysRevA.52.3560.

[13] James B. Hartle (1968): *Quantum Mechanics of Individual Systems*. American Journal of Physics 36(8), pp. 704–712, doi:10.1119/1.1975096.

[14] Adrian Kent (1990): *Against many worlds interpretations*. International Journal of Modern Physics A 5(9), pp. 1745–1762, doi:10.1142/S0217751X90000805.

[15] Fabrizio Logiurato & Augusto Smerzi (2012): *Born Rule and Noncontextual Probability* 1202.2728. doi:10.4236/jmp.2012.311225.

[16] Arnold Neumaier (1999): *On the many-worlds interpretation*. Available at `http://www.mat.univie.ac.at/~neum/manyworlds.txt`.

[17] Hilary Putnam (1965): *Trial and error predicates and the solution to a problem of Mostowski*. Journal of Symbolic Logic 30(1), pp. 49–57, doi:10.2307/2270581.

[18] Allan F. Randall (2006): *Limit recursion and Gödel's incompleteness theorem*. MA thesis, York University. Available at `http://allanrandall.ca/Godel.pdf`.

[19] Allan F. Randall (2014): *An Algorithmic Interpretation of Quantum Probability*. Ph.D. thesis, York University, Toronto, doi:10315/27640.

[20] Erwin Schrödinger (1935): *Die gegenwärtige situation in der Quantenmechanik*. Naturwissenschaften 23, pp. 807–812,823–828,844–849, doi:10.1007/BF01491914.

[21] Ray Solomonoff (1960): *A Preliminary Report on a General Theory of Inductive Inference*. Zator Company and United States Air Force Office of Scientific Research, ZTB-138, Cambridge.

[22] Ray Solomonoff (1964): *A formal theory of inductive inference. Part I*. Information and Control 7, pp. 1–22, doi:10.1016/S0019-9958(64)90223-2.

[23] Ray Solomonoff (1964): *A formal theory of inductive inference. Part II*. Information and Control 7, pp. 224–254, doi:10.1016/S0019-9958(64)90131-7.

[24] Ray Solomonoff (1978): *Complexity-based induction systems: comparisons and convergence theorems*. IEEE Transactions on Information Theory 24(4), pp. 422–432, doi:10.1109/TIT.1978.1055913.

[25] Euan J. Squires (1990): *On an alleged 'proof' of the quantum probability law*. Physics Letters A 145(2,3), pp. 67–68, doi:10.1016/0375-9601(90)90192-Q.

[26] David Wallace (2010): *How to prove the Born rule*. In S. Saunders, J. Barrett, A. Kent & D. Wallace, editors: Many Worlds? Everett, Quantum Theory, and Reality, Oxford University Press, Oxford, pp. 227–263, doi:10.1093/acprof:oso/9780199560561.003.0010.

[27] Wojciech Hubert Zurek (2005): *Probabilities from Entanglement, Born's Rule from Envariance*. Physics Review A 71, p. 052105, doi:10.1103/PhysRevA.71.052105.