Incompatibility of the Many Worlds Theory with the Independence Postulate

Samuel Epstein*

October 12, 2022

Abstract

The Many Worlds Theory and the Independence Postulate are in conflict, as shown through the existence of a finite experiment that measures the spin of a large number of electrons. After the experiment there are branches of positive probability which contain forbidden sequences that break the Independence Postulate.

1 Introduction

The Many Worlds Theory (**MWT**) was formulated by Hugh Everett [Everett(1957)] as a solution to the measurement problem of Quantum Mechanics. Branching (a.k.a splitting of worlds) occurs during any process that magnifies microscopic superpositions to the macro-scale. This occurs in events including human experiments such as the double slit experiments, or natural processes such as radiation resulting in cell mutations.

One question is if **MWT** causes issues with the foundations of computer science. The physical Church Turing Thesis (**PCTT**) states that any functions computed by a physical system can be simulated by a Turing machine. A straw man argument for showing **MWT** and **PCTT** are in conflict is an experiment that measures the spin of an unending number of electrons, with each measurement bifurcating the current branch into two sub-branches. This results in a single branch in which the halting sequence is outputted. However this branch has Born probability converging to 0, and can be seen as a deviant, atypical branch, as described in Section 2.3.

In fact, conflicts do emerge between **MWT** and Algorithmic Information Theory (**AIT**). Whereas classical information theory deals with in part the entropy of random variables, **AIT** deals with the entropy of individual sequences of 1's and 0's. This entropy measure, known as Kolmogorov complexity, is equal to the level of compressibility of finite and infinite sequences with respect to a reference Turing machine. Another notion of **AIT** is the amount of algorithmic mutual information of sequences, which represents their shared algorithmic information content. In **AIT**, the Independence Postulate (**IP**), [Levin(1984), Levin(2013)], is an unprovable inequality on the information measure of two sequences. **IP** is a finitary Church Turing Thesis, postulating that certain infinite and *finite* sequences cannot be found in nature, a.k.a. have high "addresses". One such example is finite prefixes of the halting problem. If a forbidden sequence can be found with a low address, then an "information leak" occurs. L. A. Levin states [Levin(2013)]

> *The toolkit of our models may change (e.g., quantum amplitudes work somewhat differently than probabilities) but it is hard to expect new realistic primitives allowing such "information leaks".*

*JP Theory Group, Boston MA 02130. samepst@jptheorygroup.org

However **MWT** represents a theory in which such information leaks can occur. This article introduces the MeasurePrefix experiment which produces branches of positive Born Probability with memory leaks. The Born probability is assumed to be derived by decision theory, as introduced in [Deutsch(1999)]. The MeasurePrefix experiment measures the spin of a large number, N, of electrons, creating a superposition of branches (worlds) each containing a unique sequence of N measurements. For each branch, a prefix of the sequence is taken, via an encoding scheme, and is assigned a small address (a.k.a ip-address). For a significant (Born) probability of branches, a forbidden sequence is posted to a small address, thus easily found in nature. This small address breaks **IP**, representing a information leak. The forbidden sequence used is not prefixes of the halting sequence, but prefixes of Chaitin's Omega, an **AIT** definition introduced in Section 4.2.

2 Many Worlds Theory

There is an inherent problem in quantum mechanics. On one hand, the dynamics of quantum states is prescribed by unitary evolution. This evolution is deterministic and linear. On the other hand, measurements result in the collapse of the wavefunction. This evolution is non-linear and nondeterministic. This conflict is called the measurement problem of quantum mechanics.

The time of the collapse is undefined and the criteria for the kind of collapse are strange. The Born rule assigns probabilities to macroscopic outcomes. The projection postulate assigns new microscopic states to the system measured, depending on the the macroscopic outcome. The apparatus itself should be modeled in quantum mechanics. However it's dynamics is deterministic. Probabilities only enter the conventional theory with the measurement postulates. For **MWT**, the collapse of the wave function is the change in dynamical influence of one part of the wavefunction over another, the decoherence of one part from the other. The result is a branching structure of the wavefunction and a collapse only in the phenomenological sense.

MWT was proposed by Everett as a way to remove the measurement postulate from quantum mechanics. The theory consists of unitary evolutions of quantum states without measurement collapses. For **MWT**, the collapse of the wave function is the change in dynamical influence of one part of the wavefunction over another, the decoherence of one part from the other. The result is a branching structure of the wavefunction and a collapse only in the phenomenological sense.

2.1 Deterministic Measurements

An example of a deterministic measuring of an electron can be seen in an idealized Stern-Gerlarch experiment with a single electron with spin $|\phi_\uparrow\rangle$ and $|\phi_\downarrow\rangle$. This description can be found in [Saunders et al.(2010)]. There is a measuring apparatus $\mathcal{A}$, which is in an initial state of $|\psi^{\mathcal{A}}_{\text{ready}}\rangle$. After $\mathcal{A}$ reads spin-up or spin-down then it is in state $|\psi^{\mathcal{A}}_{\text{reads spin}\uparrow}\rangle$ or $|\psi^{\mathcal{A}}_{\text{reads spin}\downarrow}\rangle$, respectively. The evolution for when the electron is solely spin-up or spin-down is

$$|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \overset{\text{unitary}}{\longrightarrow} |\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\uparrow}\rangle$$
$$|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \overset{\text{unitary}}{\longrightarrow} |\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\downarrow}\rangle.$$

Furthermore, one can model the entire quantum state of an observer $\mathcal{O}$ of the apparatus, with

$$
\begin{aligned}
&|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\uparrow}\rangle
\end{aligned}
$$

$$
\begin{aligned}
&|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}} &|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\downarrow}\rangle.
\end{aligned}
$$

For the general case, the electron is in a state $|\phi\rangle = a|\phi_\uparrow\rangle + b|\phi_\downarrow\rangle$, where $|a|^2 + |b|^2 = 1$. In this case, the final superposition would be of the form:

$$
\begin{aligned}
&a|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\uparrow}\rangle \\
+&b|\phi_{\text{absorbed}}\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\downarrow}\rangle.
\end{aligned}
$$

This is a superposition of two branches, each of which describes a perfectly reasonable physical story. This bifurcation is one method on how the quantum state of universe bifurcates into two branches.

2.2 Relative States

Everett described **MWT** in terms of relative states. Any substate of the (pure) quantum state of the universe $|\Psi\rangle$ is itself a quantum state, which can be a mixed quantum state. Any state of a subsystem of $|\Psi\rangle$, uniquely defines a relative state of the rest of $|\Psi\rangle$. How relative states relate to branching worlds can be seen in the following example.

Everett gave a model of a measuring apparatus $\mathcal{B}$ that performs multiple independent non-disturbing measurements. This same model will be used in Section 5. The ready state of $\mathcal{B}$ is $|\psi^{\mathcal{B}}[\dots]\rangle$. Let the system to be measured be spanned by an orthogonal set of state $\{|\phi_0\rangle, |\phi_1\rangle\}$. This description can also be found in [Saunders et al.(2010)]. The measurement of each ϕ_i has the unitary evolution

$$
|\phi_i\rangle \otimes |\psi^{\mathcal{B}}[\dots]\rangle \xrightarrow{\text{unitary}} |\phi_i\rangle \otimes |\psi^{\mathcal{B}}[\dots\alpha_i]\rangle,
$$

where each α_i is a recording of $|\phi_i\rangle$. For a general state in $|\phi\rangle = c_0|\phi_0\rangle + c_1|\phi_1\rangle$, then the evolution of the measurement dynamics is

$$
|\phi\rangle \otimes |\psi^{\mathcal{B}}[\dots]\rangle \xrightarrow{\text{unitary}} \sum_{i\in\{0,1\}} c_i|\phi_i\rangle \otimes |\psi^{\mathcal{B}}[\dots\alpha_i]\rangle \tag{1}
$$

If there are n systems in the identical state $|\phi\rangle$, each of which is measured independently by $\mathcal{B}$, then the measurement dynamics would be

$$
\begin{aligned}
&\bigotimes_{i=1}^{n} |\phi\rangle \otimes |\psi^{\mathcal{B}}[\dots]\rangle \xrightarrow{\text{unitary}} \\
&\sum_{\{a_1,\dots,a_n\}\in\{0,1\}^n} c_{a_1}c_{a_2}\cdot\dots\cdot c_{a_n}|\phi_{a_1}\rangle \otimes |\phi_{a_2}\rangle \otimes \dots \otimes |\phi_{a_n}\rangle |\psi^{\mathcal{B}}[a_n\dots a_2 a_1]\rangle
\end{aligned} \tag{2}
$$

Everett formulated the notion of relative quantum state. In the above example each state $|\psi^{\mathcal{B}}[a_n \dots]\rangle$, there exists the relative state of the other subsystem $|a_1\rangle \otimes |a_2\rangle \otimes \cdots \otimes |a_n\rangle$. There is no actual subsystem state, only the state of a subsystem relative to the state of the other subsystem.

2.3 Probability from Decision Theory

In standard quantum mechanics, measurements are probabilistic operations. Measurements on a state vector $|\psi\rangle$, which is a unit vector over Hilbert space $\mathcal{H}$, are self-adjoint operators $\mathcal{O}$ on $\mathcal{H}$. Observables are real numbers that are the spectrum $\mathrm{Sp}(\mathcal{O})$ of $\mathcal{O}$. A measurement outcome is a subset $E \subseteq \mathrm{Sp}(\mathcal{O})$ with associated projector P_E on $\mathcal{H}$. Outcome E is observed on measurement of $\mathcal{O}$ on $|\psi\rangle$ with probability $P(E) = \langle\psi|P_E|\psi\rangle$. This is known as the Born rule. After this measurement, the new state becomes $P_E|\psi\rangle/\sqrt{\langle\psi|P_E|\psi\rangle}$. This is known as the projection postulate.

However, the Born rule and the projection postulate are not assumed by **MWT**. The dynamics are totally deterministic. Each branch is equally real to the observers in it. To address these issues, Everett first derived a typicality-measure that weights each branch of a state's superposition. Assuming a set of desirable constraints, Everett derived the typicality-measure to be equal to the norm-squared of the coefficients of each branch, i.e. the Born probability of each branch. Everett then drew a distinction between typical branches that have high typicality-measure and exotic atypical branches of decreasing typicality-measure. For the repeated measurements of the spin of an electron $|\phi\rangle = a|\phi_\uparrow\rangle + b|\phi_\downarrow\rangle$, the relative frequencies of up and down spin measurements in a typical branch converge to $|a|^2$ and $|b|^2$, respectively. The notion of typicality can be extended to measurements with many observables.

In a more recent resolution to the relation between **MWT** and probability, Deutsch introduced a decision theoretic interpretation [Deutsch(1999)] which we assume in this paper to be the basis for deriving the probability of a branch. Deutsch obtains the Born rule from the non-probabilistic axioms of quantum theory and non-probabilistic axioms of decision theory. Deutsch proved that rational actors are compelled to adopt the Born rule as the probability measure associated with their available actions.

The setup is game where the agent receives a payoff for each measurement of an observable $\mathcal{O}$ of a quantum state $|\psi\rangle$. The game is of the form $(|\psi\rangle, \mathcal{O}, \Pi)$, where $|\psi\rangle$ and $\mathcal{O}$ are defined as before, and Π is the payoff function, mapping measurement outcomes to reals. The value of the game is $V((|\psi\rangle, \mathcal{O}, \Pi))$. The agent prefers game G_1 to game game G_2 if the $V(G_1) > V(G_2)$. It is assumed that V follows the non-probabilistic decision-theoretic axioms of substitutivity, (weak)-additivity, and the zero-sum-rule. It is also assumed if the state $|\psi\rangle$ is an eigenvector of $\mathcal{O}$, then with probability one, the measurement is the corresponding eigenvalue of $|\psi\rangle$. Under these assumptions, it was proved in [Deutsch(1999)] that

$$V((|\psi\rangle, \mathcal{O}, \Pi) = \sum_i |\langle\psi|\phi_i\rangle|^2 \Pi(x_i). \tag{3}$$

The eigenstates of $\mathcal{O}$ are $|\phi_i\rangle$, with corresponding eigenvalues x_i. A rational agent who knows that the Born-rule weight of an outcome is p is rationally compelled to act as if that outcome had probability p. According to [Deutsch(1999)],

> *A rational decision maker behaves as if he believed that each possible outcome x_i had a probability, given by the conventional formula $|\langle\psi|\phi_i\rangle|^2$, and as if he were maximizing the probabilistic expectation value of the payoff.*

Thus one should act as if the Born rule is true if one is a rational agent. While Deutsch does not explicitly mention **MWT**, in [Wallace(2003)], a version of the results of [Deutsch(1999)] explicitly connecting to **MWT** is given. Thus with respect to **MWT**, the probabilities of the two branches of Equation 1 are $|c_0|^2$ and $|c_1|^2$, respectively. The probability of each branch in Equation 2 is $|c_{a_1} c_{a_2} \dots_{a_n}|^2$.

3 Complexity and Information

IP detailed in Section 4, is a statement that uses complexity and information terms in Algorithmic Information Theory. In this section the concepts needed for **IP** are detailed.

3.1 Self Delimiting Codes

When it is clear from the context, we will use whole numbers interchangeably with their binary representations. For example, each whole number $n \in \mathbb{W}$ can be associated with the $(n+1)$th sequence of a length increasing lexicographical ordering $\{\xi_n\}_{n=1}^{\infty}$. Each ξ_n is a finite sequence, with

$$(0,0),(1,1),(2,00),(3,01),(4,10),(5,11),(6,000)\dots$$

Thus $\xi_6 = 000$. A prefix free set of sequences S is a set of finite sequences such that there does not exist two distinct sequences x, y in S where one sequence is a prefix of the other. We say such S is a *self-delimiting* code because there exists a method to determine where each code word $x \in S$ ends without reading past its last symbol. Let $\|x\|$ be the length of the sequence x. One such self-delimiting code is $\langle x\rangle' = 1^{\|x\|}0x$, where the decoding algorithm would first count the number of $1's$ before the first 0 to determine the length of x and then output the $\|x\|$ remaining bits in the input, (corresponding to x). Thus $\|\langle x\rangle'\| = 2\|x\|+1$. For a finite sequence x, we use the sequence $\langle x\rangle$ to denote a more efficient self-delimiting code which will be used in Section 5, with

$$\langle x\rangle = \langle \xi_{\|x\|}\rangle' x. \tag{4}$$

For example, $\langle 01111\rangle = 1101101111$. Thus

$$\|\langle x\rangle\| \leq \|x\| + 2\lceil \log \|x\|\rceil + 1. \tag{5}$$

3.2 Algorithms

The set of finite sequences is $\{0,1\}^*$. The set of infinite sequences is $\{0,1\}^\infty$. The set of finite and infinite sequences is $\{0,1\}^{*\infty}$. Our paper uses Turing machines M which have four tapes: a main input tape, an auxiliary input tape, a work tape, and an output tape. The alphabet for all tapes is $\{0,1,\$\}$. We give M a (partial) functional representation $M : \{0,1\}^* \times \{0,1\}^{*\infty} \to \{0,1\}^*$, defined by $y = M_\alpha(x)$ when

1. M starts with all its heads in the leftmost square. The main input tape starts with $x\$^\infty$. The auxiliary tape is set to α if it is an infinite sequence, otherwise it starts with $\alpha\$^\infty$. The work and output tape start with $\$^\infty$.
2. During its operation, M reads exactly $\|x\|$ bits from the main input tape.
3. The output tape is $y\$^\infty$ when M halts.

when this does not occur for inputs x and α then $M_\alpha(x) = \perp$ is undefined. The domain of such M is prefix free, where for all $x, y \in \{0,1\}^*$, $\alpha \in \{0,1\}^{*\infty}$, with $y \neq \emptyset$, it must be that $T_\alpha(x) = \perp$ or $T_\alpha(xy) = \perp$.

We use a fixed universal Turing machine U, where for each Turing machine T, there exists $t \in \{0,1\}^*$, where for all $x \in \{0,1\}^*$ and $\alpha \in \{0,1\}^{*\infty}$, $U_\alpha(tx) = T_\alpha(x)$. One example is for such t to be equal to $\langle i\rangle$, where i is the first index of T in an enumeration of Turing machines.

3.3 Complexity and Information

The Kolmogorov complexity of sequence $x \in \{0,1\}^*$ relative to sequence $\alpha \in \{0,1\}^{*\infty}$ is defined to be the shortest U-program which produces x.

Definition 1 (Kolmogorov Complexity) $\mathbf{K}(x|\alpha) = \min\{\|p\| : U_\alpha(p) = x\}$.

$\mathbf{K}$ is a measure of the information content of a sequence. If a string x is random, $\mathbf{K}(x) \approx \|x\|$, in that there is no algorithmic means to compress it. Otherwise, if $\mathbf{K}(x) \ll \|x\|$, then x does not have much information content. For numbers $n \in \mathbb{N}$, their Kolmogorov complexity is of logarithmic magnitude, with $\mathbf{K}(n) = \mathbf{K}(\xi_n) = O(\log n)$. For sequences $x \in \{0,1\}^*$, their Kolmogorov complexity is bounded by $\mathbf{K}(x) < \|x\| + \mathbf{K}(\|x\|) + O(1)$, as there is a Turing machine that when given a program for a number n, computes this number n and then reads n bits on the input tape, and copies the contents to the output. The Kolmogorov complexity of a pair of strings $x, y \in \{0,1\}^*$, is $\mathbf{K}(x,y) = \mathbf{K}(\langle x\rangle y)$. We define information with respect to two finite or infinite sequences $\alpha, \beta \in \{0,1\}^{*\infty}$. This definition was first introduced in [Levin(1974)].

Definition 2 (Information) $\mathbf{I}(\alpha : \beta) = \log \sum_{x,y\in\{0,1\}^*} 2^{\mathbf{K}(x)+\mathbf{K}(y)-\mathbf{K}(x,y)-\mathbf{K}(x|\alpha)-\mathbf{K}(y|\beta)}$.

We derive the following inequality for use in Section 4. Let $z \in \{0,1\}^*$ be a finite string. Then

$$\begin{aligned}
\mathbf{I}(z : z) &= \log \sum_{x,y\in\{0,1\}^*} 2^{\mathbf{K}(x)+\mathbf{K}(y)-\mathbf{K}(x,y)-\mathbf{K}(x|z)-\mathbf{K}(y|z)} \\
&> 2\mathbf{K}(z) - \mathbf{K}(z,z) - 2\mathbf{K}(z|z) \\
&> \mathbf{K}(z) - c_{\mathrm{Inf}},
\end{aligned}$$

where c_{Inf} is a small constant. If $\mathbf{I}(\alpha : \beta)$ is high, then the two sequences α and β share a lot of mutual algorithmic information. One example is an infinite sequence $\alpha \in \{0,1\}^\infty$ that is random, where there is some constant $c \in \mathbb{N}$ where for all prefixes $\alpha_n \in \{0,1\}^n$ of α of size n, we have that $n - c < \mathbf{K}(a_n)$. Then

$$\begin{aligned}
\mathbf{I}(\alpha : \alpha) &> \log \sum_{n\in\mathbb{N}} 2^{2\mathbf{K}(\alpha_n)-\mathbf{K}(\alpha_n,\alpha_n)-2\mathbf{K}(\alpha_n|\alpha)} \\
&> \log \sum_{n\in\mathbb{N}} 2^{2\mathbf{K}(\alpha_n)-\mathbf{K}(\alpha_n)-2\mathbf{K}(\alpha_n|\alpha)} - O(1) \\
&> \log \sum_{n\in\mathbb{N}} 2^{2\mathbf{K}(\alpha_n)-\mathbf{K}(\alpha_n)-2\mathbf{K}(n)} - O(1) \\
&> \log \sum_{n\in\mathbb{N}} 2^{\mathbf{K}(\alpha_n)-2\mathbf{K}(n)} - O(1) \\
&> \log \sum_{n\in\mathbb{N}} 2^{n-O(\log n)} - O(1) \\
&= \infty.
\end{aligned}$$

4 The Independence Postulate

In [Levin(1984), Levin(2013)], **IP** was introduced. The statement of **IP** is as follows.

> **IP**: *Let $\alpha \in \{0,1\}^{*\infty}$ be a sequence defined with an n-bit mathematical statement (e.g., in Peano Arithmetic or Set Theory), and a sequence $\beta \in \{0,1\}^{*\infty}$ can be located in the physical world with a k-bit instruction set (e.g., ip-address). Then $\mathbf{I}(\alpha : \beta) < k+n+c_{\mathrm{IP}}$, for some small absolute constant c_{IP}.*

One consequence of **IP** is a finite version of the (physical) Church-Turing Thesis. **IP** says that the only finite sequences that can be found in nature (i.e. have short physical addresses) will have non-recursive descriptions that are equal in length to their recursive descriptions. This can be seen when **IP** is applied to the case when $\alpha = \beta$ is a finite sequence which has a non-recursive description of length n that is much shorter than its recursive description $\mathbf{K}(\alpha)$, with $n \ll \mathbf{K}(\alpha)$. Let k be the shortest physical address of α. Then by **IP**,

$$\mathbf{K}(\alpha) < \mathbf{I}(\alpha : \alpha) + c_{\mathrm{Inf}} < k + n + c_{\mathrm{Inf}}$$
$$\mathbf{K}(\alpha) - n - c_{\mathrm{Inf}} < k.$$

Thus k is large and α cannot be easily located in the physical world. The only sequences α with short physical addresses must have $n \approx \mathbf{K}(\alpha)$.

4.1 Halting Sequence

One primary example of **IP** is prefixes of the halting sequence. The halting sequence $\mathcal{H}$ is the infinite sequence where $\mathcal{H}[i] = 0$ if the universal Turing machine halts on the input of the ith encoded Turing and $\mathcal{H}[i] = 1$, otherwise. Let H_n be the finite sequence that is the prefix of size 2^n of $\mathcal{H}$. It is well known that there is some some small constant c_{HK} independent of n such that $\mathbf{K}(H_n) \in (n - c_{\mathrm{HK}}, n + \mathbf{K}(n) + c_{\mathrm{HK}})$. The entire halting sequence $\mathcal{H}$ can be described in a mathematical statement of size equal to some small constant c_{HM}. Each n can be described using a program of size $\mathbf{K}(n)$. Therefore each H_n can be defined by a mathematical statement of size $c_{\mathrm{HM}} + \mathbf{K}(n)$. So by **IP** applied to $\alpha = \beta = H_n$ where k_n is the smallest physical address of H_n

$$n - c_{\mathrm{HK}} < \mathbf{K}(H_n) < \mathbf{I}(H_n : H_n) + c_{\mathrm{Inf}} < c_{\mathrm{HM}} + \mathbf{K}(n) + c_{\mathrm{Inf}} + k_n$$
$$n - \mathbf{K}(n) - (c_{\mathrm{HK}} + c_{\mathrm{HM}} + c_{\mathrm{Inf}}) < k_n.$$

Therefore the prefixes of $\mathcal{H}$ do not exist in reality because, up to a small additive constant ($c_{\mathrm{HK}} + c_{\mathrm{HM}} + c_{\mathrm{Inf}}$), the information content of H_n is in the range $(n, n + \mathbf{K}(n))$ but by **IP** the smallest physical address to reach H_n is more than $n - \mathbf{K}(n)$. As is usual for Algorithmic Information Theory, the statements have logarithmic precision.

4.2 Chaitin's Omega

We show another set of forbidden sequences, and the inequalities derived here will be used in Section 5. Let Ω be equal to the probability that the universal Turing machine U halts. More formally $\Omega = \sum\{2^{-\|p\|} : U(p) \text{ halts}, p \text{ is a finite sequence}\}$. Let O be the infinite sequence equal to the binary expansion of Ω. It is well known that O is a random sequence, where each prefix $O_n \in \{0,2\}^n$ of size n of O has $n - c_{\mathrm{OK}} < \mathbf{K}(O_n) < n + \mathbf{K}(n) + c_{\mathrm{OK}}$, for some small constant c_{OK}.

The reasoning from this point follows similarly to that of Section 4.1. The binary expansion O of Ω can be described by a short mathematical statement of length c_{OM}. Each n can be described using a program of size $\mathbf{K}(n)$. So each O_n can be defined by a mathematical statement of size $c_{\mathrm{OM}} + \mathbf{K}(n)$. So by **IP** applied to $\alpha = \beta = O_n$, where k_n is the smallest physical address of O_n:

$$\begin{aligned} n - c_{\mathrm{OK}} < \mathbf{K}(O_n) < \mathbf{I}(O_n : O_n) + c_{\mathrm{Inf}} < c_{\mathrm{OM}} + \mathbf{K}(n) + c_{\mathrm{Inf}} + k_n \\ n - \mathbf{K}(n) - (c_{\mathrm{OK}} + c_{\mathrm{OM}} + c_{\mathrm{Inf}}) < k_n \\ n - \mathbf{K}(n) - c_{\mathrm{O}} < k_n, \end{aligned} \tag{6}$$

where $c_{\mathrm{O}} = c_{\mathrm{OK}} + c_{\mathrm{OM}} + c_{\mathrm{Inf}}$ is a small constant. So the sequences O_n are "forbidden" in that they have information content between n and $n + \mathbf{K}(n)$ and a large physical address no smaller than $n - \mathbf{K}(n)$, up to an additive constant.

4.3 Peano Arithmetic

IP can also be used in instances where $\alpha \neq \beta$, and one canonical example is to logic, and in particular Peano Arithmetic (PA). PA is a logic system that encodes statements of arithmetic through a set of initial axioms and a deduction system. Gödel proved that PA is incomplete, in that there are well formed formulas in the language of PA which are true but are unprovable in PA. Suppose we order every well formed formula of PA and let the infinite sequence L be defined such that its ith bit is 1 iff the ith formula of PA is true. Then L is undecidable, in that there is no algorithm that can compute it. However Gödel himself thought that there can be other means to produce true axioms of mathematics [Gödel(1961)]:

> *Namely, it turns out that in the systematic establishment of the axioms of mathematics, new axioms, which do not follow by formal logic from those previously established, again and again become evident. It is not at all excluded by the negative results mentioned earlier that nevertheless every clearly posed mathematical yes-or-no question is solvable in this way. For it is just this becoming evident of more and more new axioms on the basis of the meaning of the primitive notions that a machine cannot imitate.*

However, as detailed in [Levin(2013)], **IP** forbids such information leaks. The sequence L can be defined by a small mathematical formula of size n. Let β be any source of information with a reasonably small physical address of size k, such as the contents of an entire mathematical library. Then by **IP**, with $\alpha = L$, this information source will have negligible shared information with L (which encodes PA):

$$\mathbf{I}(\beta : L) < k + n + c_{\mathrm{IP}}.$$

5 MeasurePrefix Experiment

Informally, the MeasurePrefix experiment in this section will measure the spin of a very large number of isolated electrons. The measuring apparatus will contain a very large sequence $x \in \{0,1\}^*$. From this sequence, a smaller sequence y is taken from the larger sequence and posted to a unique address. With non-zero probability this posted sequence will be a prefix of the binary expansion of Chaitin's Omega, as discussed in Section 4.2.

We define the function *Prefix* : $\{0,1\}^* \to \{0,1\}^*$ as follows. Given an input $x \in \{0,1\}^*$, *Prefix*, returns the unique sequence $y \in \{0,1\}^*$, such that $\langle y \rangle$ is a prefix of x or is equal to x. The code $\langle \cdot \rangle$ was defined by Equation 4. If no such y exists, for example in the case where $x = 1^{(n)}$, then *Prefix* returns 0. Let $N \in \mathbb{N}$ be a large number, say $N \approx 10^9$. Let a be an ip-address, which can be encoded by 32 bits. Let $|\phi\rangle$ be the state of the spin of an electron where $|\psi\rangle = \frac{1}{\sqrt{2}}|\psi_0\rangle + \frac{1}{\sqrt{2}}|\psi_1\rangle$. There is a measuring apparatus $\mathcal{A}$ with initial state of $|\psi^{\mathcal{A}}\rangle$, and after reading N spins of N electrons, it is in the state $|\psi^{\mathcal{A}}[x]\rangle$, where $x \in \{0,1\}^N$, whose ith bit is 1 iff the ith measurement returns $|\phi_1\rangle$. Let the apparatus to publicly display a finite sequence at an ip-address be $\mathcal{B}$. Its initial state without a sequence to display is $|\psi^{\mathcal{B}}\rangle$. Its state displaying a sequence $y \in \{0,1\}^*$ is $|\psi^{\mathcal{B}}[y]\rangle$. The MeasurePrefix experiment proceeds with the following unitary evolution:

$$
\begin{aligned}
&\bigotimes_{i=1}^{N} |\phi\rangle \otimes |\psi^{\mathcal{A}}\rangle \otimes |\psi^{\mathcal{B}}\rangle \\
\xrightarrow{\text{unitary}} & \sum_{a_1,\dots,a_N \in \{0,1\}^N} 2^{-N/2} \bigotimes_{i=1}^{N} |\phi_{a_i}\rangle \otimes |\psi^{\mathcal{A}}[a_1 a_2 \dots a_n]\rangle \otimes |\psi^{\mathcal{B}}\rangle \\
\xrightarrow{\text{unitary}} & \sum_{a_1,\dots,a_N \in \{0,1\}^N} 2^{-N/2} \bigotimes_{i=1}^{N} |\phi_{a_i}\rangle \otimes |\psi^{\mathcal{A}}[a_1 a_2 \dots a_n]\rangle \otimes |\psi^{\mathcal{B}}[\mathit{Prefix}(a_1 a_2 \dots a_n)]\rangle.
\end{aligned}
$$

Let $M \in \mathbb{N}$ be the largest natural number such that $M + 2\lceil \log M \rceil + 1 \leq N$. For $n \in [1, M]$, let $\mathcal{O}_n$ be the prefix of size n of the binary expansion of Chaitin's Omega, defined in Section 4.2. Using probability of a branch defined in Section 2.3, the probability that a branch has $\mathcal{B}$ displaying $\mathcal{O}_n$ is $\geq 2^{-n-2\lceil \log n \rceil - 1}$. On such branches, the smallest physical address that $\mathcal{O}_n$ can be located by is ≤ 32 (bits), the size of an ip-address. However by **IP**, the smallest physical address of $\mathcal{O}_n$, is $> n - \mathbf{K}(n) - c_{\mathrm{O}}$, where c_{O} is a small constant, (as seen in Equation 6).

A branch has a information leak of size $m - l$ if it has a physical address for a sequence of size l but **IP** dictates that the smallest physical for that sequence is greater than m. So $2^{-n-2\lceil \log n \rceil - 1} \leq$ Prob(Branch where $\mathcal{B}$ displays $\mathcal{O}_n$) $\leq$ Prob(Branch with a information leak of size $n - \mathbf{K}(n) - c_{\mathrm{O}} - 32$). Thus for some small constant $c \in \mathbb{N}$, for $n \in [1, M]$,

$$
\text{Prob(Branch with an information leak of size } n - 2\log n - c) \geq 2^{-n-2\log n - c}. \tag{7}
$$

6 Conclusion

Thus any branch that performs the MeasurePrefix experiment bifurcates in sub-branches. A subset of them have positive Born probability and have "information leaks", i.e. break the inequalities of **IP**. Thus assuming **MWT** to be valid, **IP** cannot be formulated by an observer because there exist successor observers where information leaks occur. Thus if taken both to be valid, **MWT** and **IP** are in conflict. The author cannot envision a change to **MWT** to reconcile this disparity. One solution is to just accept that there exists "information leaks" which breaks **IP**.

In general, there are other avenues of research in the intersection of **MWT** and **AIT**. One area is to revise the definition of typical branches described by Everett. Everett showed that in the limit of the number of measurements, almost all relative sequences of measurements, with respect to the Born probability, will exhibit the standard quantum statistics. Thus typical branches will have such statistics. Let B be a branch displaying the repeated measurement of spin of independent electrons in the state $|\phi\rangle = \frac{1}{\sqrt{2}}|\phi_0\rangle + \frac{1}{\sqrt{2}}|\phi_1\rangle$. If B displays the records $01010101010\dots$ it exhibits the quantum statistics, but it is highly atypical. To account for this, one can define typical branches to be algorithmically incompressible. This represents an interesting open area of study.

References

[Deutsch(1999)] David Deutsch. Quantum theory of probability and decisions. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455 (1988), 1999.

[Everett(1957)] Hugh Everett. "relative state" formulation of quantum mechanics. *Rev. Mod. Phys.*, 29, 1957.

[Gödel(1961)] Kurt Gödel. The modern development of the foundations of mathematics in the light of philosophy. *In: Kurt Gödel. Collected Works. Volume III. Oxford University Press.*, 1961.

[Levin(1974)] L. A. Levin. Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory. *Problemy Peredachi Informatsii*, 10(3):206–210, 1974.

[Levin(1984)] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Levin(2013)] L. A. Levin. Forbidden information. *J. ACM*, 60(2), 2013. ISSN 0004-5411.

[Saunders et al.(2010)] S. Saunders, J. Barrett, A. Kent, and D. Wallace. *Many Worlds?: Everett, Quantum Theory, & Reality.* OUP Oxford, 2010.

[Wallace(2003)] D. Wallace. Everettian rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 34(3), 2003.