Quantum Mechanics and Algorithmic Information Theory

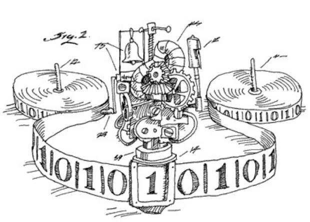
Samuel Epstein
samepst@jptheorygroup.org

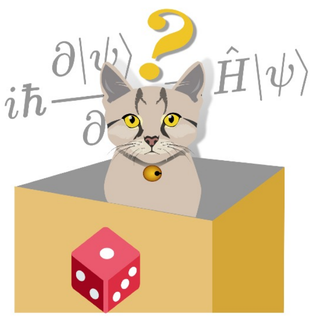
May 16, 2023

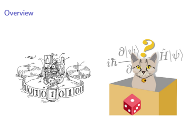
Quantum Mechanics and Algorithmic Information Theory
Samuel Epstein
samepst@jptheorygroup.org
May 16, 2023

Hello my name is Sam Epstein and this is my talk on the applications of Algorithmic Information Theory to Quantum Mechanics.

Overview







Concepts in AIT such as Kolmogorov complexity and algorithmic mutual information can give insights into aspects of quantum mechanics.

Closely related to this talk is quantum information theory, the intersection of classical information theory with quantum mechanics. QIT, in part, deals with bandwidth and information between the inputs and outputs of noisy quantum channels that realize quantum mechanics.

AIT enhances our understanding of quantum mechanics as it gives a formal measure of the information contents of individual quantum states and their measurements.

Source

Samuel Epstein. An Introduction to Algorithmic Information Theory and Quantum Mechanics. 2023. (hal-04072076v1)

1. Quantum Complexity
2. Quantum Outliers
3. Information
4. Measurements
5. Infinite Quantum Spin Chains
6. Many World Theory

Source
Samuel Epstein. An Introduction to Algorithmic Information Theory and Quantum Mechanics. 2023. (hal-04072076v1)
1. Quantum Complexity
2. Quantum Outliers
3. Information
4. Measurements
5. Infinite Quantum Spin Chains
6. Many World Theory

Source

Samuel Epstein. An Introduction to Algorithmic Information Theory and Quantum Mechanics. 2023. (hal-04072076v1)

1. Quantum Complexity
2. Quantum Outliers
3. Information
4. **Measurements**
5. Infinite Quantum Spin Chains
6. **Many World Theory**

Source
Samuel Epstein. An Introduction to Algorithmic Information Theory and Quantum Mechanics. 2023. (hal-04072076v1)
1. Quantum Complexity
2. Quantum Outliers
3. Information
4. Measurements
5. Infinite Quantum Spin Chains
6. Many World Theory

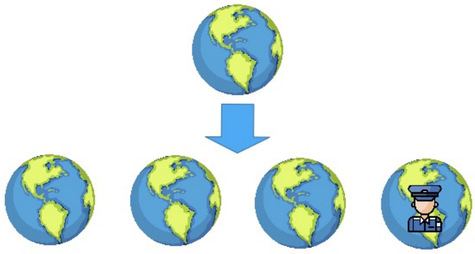
Outline

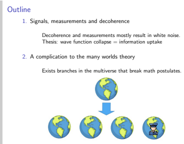
1. Signals, measurements and decoherence

   Decoherence and measurements mostly result in white noise.
   Thesis: wave function collapse = information uptake

2. A complication to the many worlds theory

   Exists branches in the multiverse that break math postulates.





Algorithmic Information Theory can be used to show that for an overwhelming majority of pure or mixed states, when a POVM measurement is applied, then the result is algorithmic white noise. This white noise cannot be treated with randomized processing to create a more coherent signal.
Decoherence is the process when a quantum state interacts with the environment for a period of time, losing coherence, i.e. the off diagonal terms, and becoming a virtually classical mixture. In this talk, we show that for an overwhelming majority of pure and mixed states, white noise if the result of decoherence.
This part of the talk is the prelude to the thesis that the wavefunction collapse caused by measurements results in uptake of algorithmic information. Though not covered in the talk, one can prove that the collapse enables non-negligible information from further measurements and partial information cloning. A quantum operation or a unitary transform does not have this property.
The Many Worlds Theory was formulated by Hugh Everret in 195? as an answer to the measurment problem. Superpositions of particles at the substomic level are propagated to superpositions of measuring devices, and then to their environments and so on. Thus the multiverse is constantly splitting branches, causing multiple parallel worlds. Some of these worlds are quite exotic, in that typical laws of physics, such as the second law of thermodynamics do not hold. In this talk we show that there are worlds where certain mathematical postulates do not hold. Namely the Independence Postulate of Algorithmic Information Theory must break in a significant measure of worlds in the multiverse. The diagram shows the splitting of the worlds, where one branch contains information that breaks the independence postulate.

Signals from Classical and Quantum Sources

Classical conservation of information, for randomized processing g,

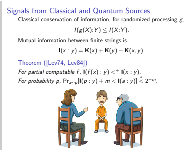
$$I(g(X){:}Y) \leq I(X{:}Y).$$

Mutual information between finite strings is

$$\mathbf{I}(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y).$$

Theorem ([Lev74, Lev84])

For partial computable f, $\mathbf{I}(f(x):y) <^{+} \mathbf{I}(x:y)$.
For probability p, $\Pr_{a\sim p}[\mathbf{I}(p:y) + m < \mathbf{I}(a:y)] \stackrel{*}{<} 2^{-m}$.



Information conservation inequalities were originally proven for classical information theory. Randomized processing cannot increase the mutual information between two random variables.
Conservation inequalities have been proven for algorithmic mutual information.

Information between Signals

Definition
For probabilities p and q, $\mathbf{I}_{\mathrm{Prob}}(p:q) = \log \sum_{x,y} 2^{\mathbf{I}(x:y)} p(x)q(y)$.

Example
Low $\mathbf{I}_{\mathrm{Prob}}(p:p) \Rightarrow p$ has
- large measure on simple strings,
- or low measure on a large set of complex strings.

Singleton p and q over x and y, $\mathbf{I}_{\mathrm{Prob}}(p:q) = \mathbf{I}(x:y)$.

$\mathbf{I}(\mathbf{m}:\mathbf{m}) = O(1)$.

Uniform measure over $\{0,1\}^n$ has $\mathbf{I}_{\mathrm{Prob}}(U_n:U_n) =^+ \mathbf{K}(n)$.

For ML random sequence α, $p(\alpha_n) = n^{-2}$ has $\mathbf{I}_{\mathrm{Prob}}(p:p) = \infty$.

One very important definition is the self information of a probability. Whereas in strings, the self information is equal to the Kolmogorov complexity, in probabilites, self information can be much less than its complexity.

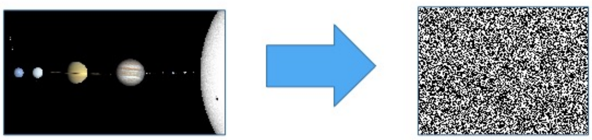
Channels

Definition
Channel $f : \{0,1\}^* \times \{0,1\}^* \to \mathbb{R}_{\geq 0}$,
$f(\cdot|x)$ is a probability.

For probability p, $fp(x) = \sum_z f(x|z)p(z)$.

Theorem
For probabilities p, q, channel f, $\mathbf{I}_{\mathrm{Prob}}(fp:q) <^+ \mathbf{I}_{\mathrm{Prob}}(p:q)$.

Corollary
$\mathbf{I}_{\mathrm{Prob}}(fp:fp) <^+ \mathbf{I}_{\mathrm{Prob}}(p:p)$

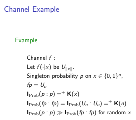


There are information non-growth laws with respect to random channels. When a channel is applied to a probability p, it creates a new probability, where every element is processed by the channel.
The information non-growth theorem is pessimistic, in that ultimatly all probabilities loss coherence.

Channel Example

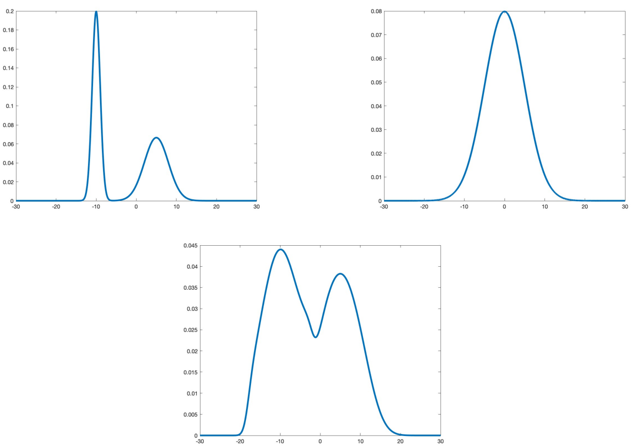
Example

Channel f :
Let $f(\cdot|x)$ be $U_{\|x\|}$.
Singleton probability p on $x \in \{0,1\}^n$,
$fp = U_n$
$\mathbf{I}_{\mathrm{Prob}}(p:p) =^+ \mathbf{K}(x)$
$\mathbf{I}_{\mathrm{Prob}}(fp:fp) = \mathbf{I}_{\mathrm{Prob}}(U_n:U_n) =^+ \mathbf{K}(n)$.
$\mathbf{I}_{\mathrm{Prob}}(p:p) \gg \mathbf{I}_{\mathrm{Prob}}(fp:fp)$ for random x.

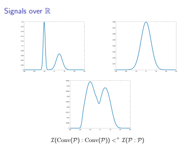


An example of this if the channel takes in a string of length n and outputs uniformly the strings of length n. Probabilities that are singletons are translated to uniform probabilities, which is a loss of signal.

Signals over $\mathbb{R}$



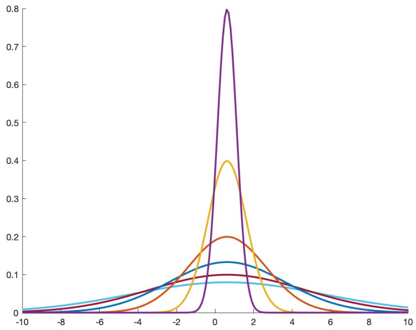
$$\mathcal{I}(\mathrm{Conv}(\mathcal{P}) : \mathrm{Conv}(\mathcal{P})) <^+ \mathcal{I}(\mathcal{P} : \mathcal{P})$$

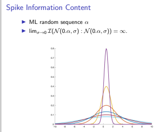


Information between probabilities can be extended probabilities over the real line. I'm not going to go over the definition, but it is sufficient to say that it involves a mapping from points in $\mathbb{R}$ to infinite sequences and using a definition of information of probabilities over infinite sequences.
There are several properties of information over $\mathbb{R}$ which are illuminating. The first is that the convolution of any probability with a probability kernel will result in a loss of self information. The example here shows that the probability when convoluted with the normal distribution will result in loss of signal coherence.

Spike Information Content

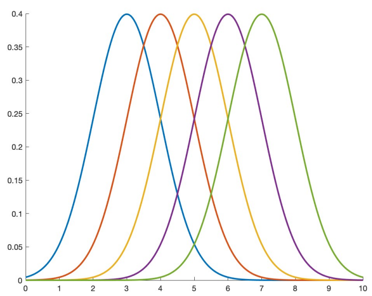
- ML random sequence α
- $\lim_{\sigma \to 0} \mathcal{I}(\mathcal{N}(0.\alpha, \sigma) : \mathcal{N}(0.\alpha, \sigma)) = \infty.$



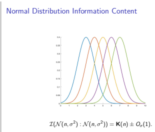


Probabilities over $\mathbb{R}$ can have infinite self-information. Take a ML random sequence α. The limit of the self information of normal distributions at 0 point α as the variance goes to 0 is infinity. Thus a random spike has infinite self information.

Normal Distribution Information Content



$$\mathcal{I}(\mathcal{N}(n, \sigma^2) : \mathcal{N}(n, \sigma^2)) = \mathbf{K}(n) \pm O_\sigma(1).$$



As distributions move away from the origin, the more self information that it has. For example normal distributions that are centered around n will have $\mathbf{K}(n)$ self information, with the additive constant dependent on the variance.

Measurements

A PVM Π is a collection of Projectors $\{\Pi_i\}$.

$\sum_i \Pi_i = I$, $\mathrm{Tr}\Pi_i\Pi_j = 0$ for $i \neq j$.

Given mixed state σ measures i with probability $\mathrm{Tr}\Pi_i\sigma$,

The probability over measurements is denotes $\Pi\sigma$, where $\Pi\sigma(i) = \mathrm{Tr}\Pi_i\sigma$.

The state collapses to $\sigma' = \Pi_i\sigma\Pi_i/\sqrt{\mathrm{Tr}\Pi_i\sigma}$.

Results also work for POVMs,

Measurements
A PVM Π is a collection of Projectors $\{\Pi_i\}$.
$\sum_i \Pi_i = I$, $\mathrm{Tr}\Pi_i\Pi_j = 0$ for $i \neq j$.
Given mixed state σ measures i with probability $\mathrm{Tr}\Pi_i\sigma$,
The probability over measurements is denotes $\Pi\sigma$, where $\Pi\sigma(i) = \mathrm{Tr}\Pi_i\sigma$.
The state collapses to $\sigma' = \Pi_i\sigma\Pi_i/\sqrt{\mathrm{Tr}\Pi_i\sigma}$.
Results also work for POVMs,

That concludes our discussion of signal information with respect to Algorithmic Information Theory. In the next section, we describe how quantum information theory can shed insight into how signals evolve and are created. Here is a recap of measurements. The results also work for POVMs.

Quantum Prepare-And-Measure Channels

Alice: random variable X, values 1 to n with prob. $\{p_i\}_{i=1}^n$.

Alice prepares density matrix σ_X where $\{\sigma_i\}_{i=1}^n$ is chosen according to X.

Bob: measurements on σ_X with PVM Π getting Y.

Still a classical channel,

$$\mathbf{I}_{\mathrm{Prob}}(Y:Y) <^+ \mathbf{I}_{\mathrm{Prob}}(X:X).$$

Quantum Prepare-And-Measure Channels
Alice: random variable X, values 1 to n with prob. $\{p_i\}_{i=1}^n$.
Alice prepares density matrix σ_X where $\{\sigma_i\}_{i=1}^n$ is chosen according to X.
Bob: measurements on σ_X with PVM Π getting Y.
Still a classical channel,
$\mathbf{I}_{\mathrm{Prob}}(Y:Y) <^+ \mathbf{I}_{\mathrm{Prob}}(X:X)$.

- Prepare and measure channel $\Rightarrow$ new insights.

Constructing Signals from Quantum Sources

Computable Density matrix σ

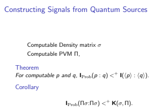
Computable PVM Π,

Theorem

For computable p and q, $\mathbf{I}_{\mathrm{Prob}}(p:q) <^{+} \mathbf{I}(\langle p\rangle : \langle q\rangle)$.

Corollary

$$\mathbf{I}_{\mathrm{Prob}}(\Pi\sigma{:}\Pi\sigma) <^{+} \mathbf{K}(\sigma,\Pi).$$



What about explicitly creating a signal (i.e. probability) from a quantum mixed state and a measurement. However due to following theorem the information between probabilities is less than the information of their encodings. Thus simple quantum states and measurement apparatuses will produce simple signals.

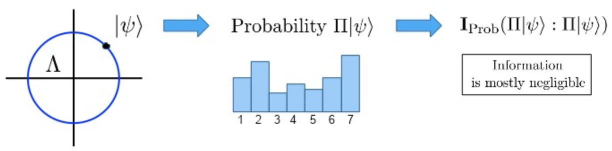
Signals from Measurements

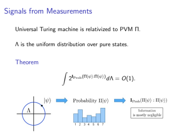
Universal Turing machine is relativized to PVM Π.

Λ is the uniform distribution over pure states.

Theorem

$$\int 2^{\mathbf{I}_{\mathrm{Prob}}(\Pi|\psi\rangle:\Pi|\psi\rangle)} d\Lambda = O(1).$$





It could be that a large majority of uncomputable quantum pure states can produce a signal with a simple PVM. However the equation says otherwise: An overwhelming number of pure quantum states, when applied to a measurement, will result in probabilities that have negligible self-information. In fact, there is a form of the equation which a uniform distribution over mixed states.

Signals from Decoherence

Decoherence:

Quantum state + environment $\Rightarrow$ lose coherence.

Off-diagonal elements of density matrix $\sigma \Rightarrow 0$.

Idealized result $\Rightarrow$ classical probability $p_\sigma(i) = \sigma_{ii}$.

Theorem

$$\int 2^{\mathbf{I}_{\mathrm{Prob}}(p_{|\psi\rangle}:p_{|\psi\rangle})} d\Lambda = O(1).$$

Signals from Decoherence
Decoherence:
Quantum state + environment ⇒ lose coherence.
Off-diagonal elements of density matrix $\sigma \Rightarrow 0$.
Idealized result ⇒ classical probability $p_\sigma(i) = \sigma_{ii}$.
Theorem
$\int 2^{\mathbf{I}_{\mathrm{Prob}}(p_{|\psi\rangle}:p_{|\psi\rangle})} d\Lambda = O(1)$.

In decoherence, the quantum state interacts with environment, and the off diagonal terms decrease at an exponential rate to the time of the interaction. The result is a classical mixture, a density matrix with the only non-negligible terms being on the diagonal.

The equation states, that for almost all quantum pure states, the resultant classical mixture is white noise. Again, this is also true for a uniform distribution over mixed states.

Wave Function Collapse = Signal Uptake

PVM Π, state $|\psi\rangle$ creates distribution Γ.

Example

$\Pi = \{|i\rangle\langle i|\}_{i=1}^{2^n} \Rightarrow \Gamma(|i\rangle) = |\langle\psi|i\rangle|^2$.

Λ_Π is dist. when Π is applied to uniform prior Λ.

Example

$\Lambda_\Pi(|i\rangle) = 2^{-n} \Rightarrow$ Measurements, Partial Cloning.

Theorem

If PVM Π has 2^{n-c} projectors,

$$n - 2c <^{\log} \log \int 2^{\mathbf{I}_{\mathrm{Prob}}(\Pi|\psi\rangle:\Pi|\psi\rangle)} d\Lambda_\Pi$$

Wave Function Collapse = Signal Uptake
PVM Π, state $|\psi\rangle$ creates distribution Γ.
Example
$\Pi = \{|i\rangle\langle i|\}_{i=1}^{2^n} \Rightarrow \Gamma(|i\rangle) = |\langle\psi|i\rangle|^2$.
Λ_Π is dist. when Π is applied to uniform prior Λ.
Example
$\Lambda_\Pi(|i\rangle) = 2^{-n} \Rightarrow$ Measurements, Partial Cloning.
Theorem
If PVM Π has 2^{n-c} projectors,
$n - 2c <^{\log} \log \int 2^{\mathbf{I}_{\mathrm{Prob}}(\Pi|\psi\rangle:\Pi|\psi\rangle)} d\Lambda_\Pi$

When an arbitrary state is applied to a PVM, the result is a distribution over the resultant collapsed states. The distribution Λ_Π is the application of the collapse to each $|\psi\rangle$, distributed according to the prior Λ.

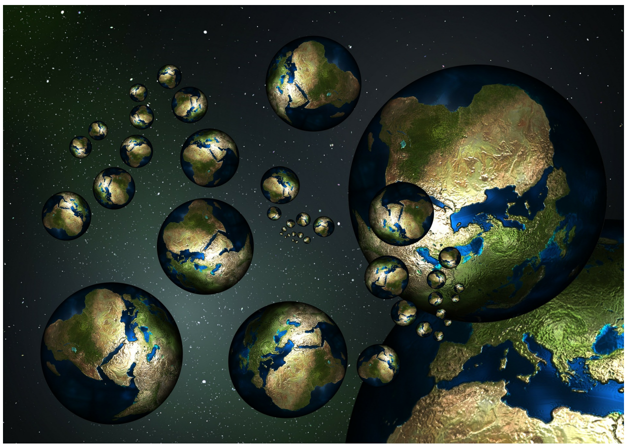
Take a PVM with a lot of projectors. Then when states are distributed accoring to Λ_Π, then most will have a signal with respect to measurement Π. Thus given a uniform prior, if you collapse the states with a measurement, and then perform the same measurement, you will get probabilities with large self-information.

Conclusion

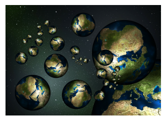
1. Simple computable quantum states and measurements.
2. Given a measurement, most quantum states produce white noise.
3. Most quantum states decohere into white noise.
4. Conservation inequalities prevents processing of this noise.
5. **The wave function collapse from measurements cause information uptake.**

Conclusion
1. Simple computable quantum states and measurements.
2. Given a measurement, most quantum states produce white noise.
3. Most quantum states decohere into white noise.
4. Conservation inequalities prevents processing of this noise.
5. The wave function collapse from measurements cause information uptake.

Many Worlds Theory and the Independence Postulate



Many Worlds Theory and the Independence Postulate



Many Worlds Theory and the Independence Postulate

1. Many Worlds Theory.

    Measurement Problem.
    Incompleteness of Copenhagen Interpretation

2. Independence Postulate.

    Forbidden strings such as prefix of halting sequence.

3. Measure spin of many electrons $\Rightarrow$ exists a branch with forbidden sequence.

Many Worlds Theory and the Independence Postulate
1. Many Worlds Theory. Measurement Problem. Incompleteness of Copenhagen Interpretation
2. Independence Postulate. Forbidden strings such as prefix of halting sequence.
3. Measure spin of many electrons $\Rightarrow$ exists a branch with forbidden sequence.

The Many Worlds Theory was formulated as solution to the measurement problem. Whereas the evolution of the quantum state is deterministic, the measurement postulate is probabilistic. Furthermore, the exact time of the measurement is undefined. The copenhagen interpretation views quantum mechanics as an incomplete theory, segregating the micro and macro scopic realms. The setup is as follows.

Many Worlds Theory Example Branching of World

Measure the electron spin: $|\phi_\uparrow\rangle$ or $|\phi_\downarrow\rangle$.
Measuring apparatus $\mathcal{A}$ in initial state $|\psi^{\mathcal{A}}_{\text{ready}}\rangle$

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

Many Worlds Theory Example Branching of World
Measure the electron spin: $|\phi_\uparrow\rangle$ or $|\phi_\downarrow\rangle$.
Measuring apparatus $\mathcal{A}$ in initial state $|\psi^{\mathcal{A}}_{\text{ready}}\rangle$
$|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \xrightarrow{\text{unitary}} |\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle$
$|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \xrightarrow{\text{unitary}} |\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle.$

Many Worlds Theory Example Branching of World

Can model observer $\mathcal{O}$ of the apparatus,

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

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

Many Worlds Theory Example Branching of World

If electron is $|\phi\rangle = a|\phi_\uparrow\rangle + b|\phi_\downarrow\rangle$, with $|a|^2 + |b|^2 = 1$,

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

The state of the apparatus and environment is undefined. It only has a defined relative state to the electron. Hence this is why Evveret described the Many Worlds Theory as the relative state formulation of quantum mechanics. Thus the apparatus and environment are in two branching worlds, one measuring spin up and another measuring spin down. There are also other phenomina that cause branching, such as radiation of cells.

Deriving Born Rule

The Born Rule states:

If the electron is in the state

$$a|\phi_{\uparrow}\rangle + b|\phi_{\downarrow}\rangle$$

Then the spin measurement will measure

$$\begin{array}{ll} |\phi_{\uparrow}\rangle & \text{with probability } |a|^2 \\ |\phi_{\downarrow}\rangle & \text{with probability } |b|^2 \end{array}$$

Compatible with Many Worlds Theory?

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.

Deriving the Born Rule

1. Everett's Typicality Measure
2. Vaidman's Subjective Probability

Probability Postulate. *An observer should set his subjective probability of the outcome of a quantum experiment in proportion to the total measure of existence of all worlds with that outcome.*

- "Total measure of existence" follows Born rule

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.

Another attempt uses subjective probability [**?**]. The experimenter puts on a blindfold before he finishes performing the experiment. After he finishes the experiment, he has uncertainty about what world he is in. This uncertainty is the foundation of a probability measure over the measurements. The probability measure are postulated to be proportion to that equal to the born rule. We will assume this intrepretation of probability for the talk.

The Independence Postulate

Each singleton language $\{\text{Halting}[1 \ldots m]\}$ is computable.

> **IP:**([Lev84, Lev13]) *Let α be a sequence defined with an n-bit mathematical statement (e.g., in Peano Arithmetic), and a sequence β can be located in the physical world with a k-bit instruction set (e.g., ip-address). Then* $\mathbf{I}(\alpha : \beta) < k + n + c$*, for some small absolute constant* c.

1. Logic
2. Statistics
3. Induction
4. Games
5. Physics

The Physical Church Turing Thesis states that the only functions that exist in the world are computable ones. However this leaves a nagging problem. Take a system that produces a single sequence of the first 2^m bits of the halting sequence, $\mathcal{H}$, for very large m.
All singletons strings are computable, so the Physical Church Turing Thesis says nothing about them. To address this, among other applications, the independence postulate was formulated.
Gödel had hopes that there could exist a means to derive new statements about arithmetic. The Independence Postulate dowses this hope since due to Levin13, any consistent completion of Peano Arithmetic has infinite mutual information with the halting sequence.
It has recently been proven that all randomized algorithmic sampling method will produce ever greater outliers. However this leaves open the possibility of systems that are too complex to be algorithmic. Take, for example your local weather forecast. IP extends the existence of outliers to such systems.
The goal of induction is to guess the hidden part of the environment from an incomplete observation. The question is, do you use Occam's razor, or a probabilistic method, involving priors and posteriors. Without getting into too much detail, IP suggests that Occam's razor is the preffered strategy.
In the cybernetic agent model, the agent interacts with an environment exchanging actions. IP implies that if there is a simple randomized player that wins with good probability, then there exists a simple deterministic player who will win against the environment.

Forbidden Strings

Set $x = \alpha = \beta \in \{0,1\}^*$, and $\mathbf{I}(\alpha : \beta) = \mathbf{K}(\alpha) + \mathbf{K}(\beta) - \mathbf{K}(\alpha, \beta)$.
$\mathbf{I}(x : x) = \mathbf{K}(x)$.

> **IP':** *Let x be defined with an n-bit mathematical statement and can be located in the physical world with a k-bit address. Then* $\mathbf{K}(x) <^+ k + n$.

$$\Omega = \sum \{2^{-\|p\|} : U(p) \text{ halts.}\} \in (0,1)$$

$$\Omega_m = \text{The first } m \text{ bits of } \Omega.$$

$$\mathbf{K}(\Omega_m) =^{\log} m.$$

$$\Omega_m \text{ is defined by a formula of size } O(\log m).$$

Ω_m is forbidden:

$$m <^{\log} \text{ length of any physical address of } \Omega_m.$$

The Experiment

Measure the spin $|\phi_0\rangle$ and $|\phi_1\rangle$ of N isolated electrons:

$$|\phi\rangle = \frac{1}{\sqrt{2}}|\phi_0\rangle + \frac{1}{\sqrt{2}}|\phi_1\rangle.$$

Experiment located at LHC, has address size $O(\log N)$.
Apparatus $|\psi^{\mathcal{A}}\rangle$, after measuring, is $|\psi^{\mathcal{A}}[x]\rangle$, $x \in \{0,1\}^N$.

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

The Experiment

If in state $|\psi^{\mathcal{A}}[\Omega_N]\rangle$, then a memory leak of size $N - O(\log N)$ has occurred.

Probability(A memory leak of size $N - O(\log N)) \geq 2^{-N}$.

Blindfolded mathematican declaring **IP**:
wrong with probability 2^{-N}.

Rebuttal

> *How is this different from flipping N coins? There is a 2^{-N} chance of getting Ω_N.*

Protected by

1. Algorithmic Conservation Inequalities
2. In physics one can postulate away events with extremely small probabilities.

   More generally, how to protect against a probability P learning information about a target sequence with good probability?

Rebuttal
How is this different from flipping N coins? There is a 2^{-N} chance of getting Ω_N.
Protected by
1. Algorithmic Conservation Inequalities
2. In physics one can postulate away events with extremely small probabilities.
More generally, how to protect against a probability P learning information about a target sequence with good probability?

Rebuttal

Let P be a probability over sequences.

Then $\Pr_{\alpha\sim P}[\mathbf{I}(\langle P\rangle:\beta)+m<\mathbf{I}(\alpha:\beta)]\overset{*}{<}2^{-m}$.

  Memory leak probability is very small.

Repeat event many times: some leaks could occur.

  Most events $\Rightarrow$ large address.

Pr[Leak at small address] is small$\Rightarrow$ postulate away.
Like increase of thermodynamic entropy.

Rebuttal
Let P be a probability over sequences.
Then Pr_{α∼P}[I(⟨P⟩ : β) + m < I(α : β)] <* 2^{-m}.
Memory leak probability is very small.
Repeat event many times: some leaks could occur.
Most events ⇒ large address.
Pr[Leak at small address] is small⇒ postulate away.
Like increase of thermodynamic entropy.

Conclusion

Memory leaks inevitable in Many Worlds Theory

No probabilitic memory leaks in IP

Critics: Another problem with MWT.

Advocates: The multiverse has an interesting relation to Algorithmic Information Theory.

Conclusion
Memory leaks inevitable in Many Worlds Theory
No probabilitic memory leaks in IP
Critics: Another problem with MWT.
Advocates: The multiverse has an interesting relation to Algorithmic Information Theory.

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.

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

L. A. Levin.
Forbidden information.
J. ACM, 60(2), 2013.