Can we change the past by forgetting?

Saibal Mitra

November 30, 2008

Abstract

Assuming the validity of the Many Worlds Interpretation of quantum mechanics (MWI), we examine the possibility of changing the past by memory erasure. We find that while observers cannot change the probability of experiencing certain outcomes, they can nevertheless randomize past events and thus end up in a different sector of the multiverse. We also discuss Tegmark's quantum suicide thought experiment and show that, contrary to conventional wisdom, it *cannot* be used by an observer to test the validity of the MWI for herself.

1 Introduction

Everyone makes bad decisions or receives bad news from time to time. Wouldn't it be nice if we could just travel back in time and set things right? Time travel has been contemplated by many philosophers, science fiction writers, and theoretical physicists. Paradoxes associated when attempting to change the past have been widely discussed.

So, it seems that changing the past is a futile dream precisely because you would create a conflict with the facts that have already happened. However, the Many Worlds Interpretation of quantum mechanics (MWI) [2] may offer a way out. David Deutsch has argued [1] that a time traveller will end up in a different sector of the multiverse if he were to make different decisions in the past than he made in the sector he started out in.

In this essay, we'll not speculate about time travel. Instead, assuming the validity of the MWI, we'll explore if we can jump to another sector of the multiverse by wiping out information about certain past events and then filling the gaps in the memory.

2 The formalism

For the purpose of this essay, it is only necessary to make the following assumptions. We assume that the whole universe, including the observer, always evolves in time according to the Schrödinger equation:

$$i\hbar\frac{\partial\left|\psi\right\rangle}{\partial t} = H\left|\psi\right\rangle . \tag{1}$$

We assume that the states the observer can subjectively find herself in, can be described classically, similarly to specifying the states of the bits of a computer to be zero or one. To describe the full quantum mechanical state vector of the observer will, of course, require a huge number of additional information. In this essay we will take the view that whatever the exact quantum mechanical state vector is, the observer can only be aware of some classically describable macrostates of itself. We can then consider the additional information needed to specify the exact quantum state of the observer as part of the rest of the universe.

The set of all macrostates of an observer then defines an orthonormal basis. We will denote these basis vectors formally as $|O_n\rangle$. The generic form of a quantum state of the universe can then be formally written as:

$$|\psi\rangle = \sum_{n,m} a_{n,m} |O_n\rangle |\phi_m\rangle \tag{2}$$

where the $|\phi_m\rangle$ form a complete set of orthonormal states describing the rest of the universe. If we sum (2) over m, we can write it formally as

$$|\psi\rangle = \sum_{n} |O_n\rangle |U_n\rangle \tag{3}$$

The $|U_n\rangle$ can then be arbitrary states describing the rest of the universe.

Suppose an initial state

$$|\psi_{\text{initial}}\rangle = |O_1\rangle |U_1\rangle \tag{4}$$

evolves in time to become a superposition of the form

$$|\psi_{\text{a while later}}\rangle = |O_2\rangle |U_2\rangle + |O_3\rangle |U_3\rangle \tag{5}$$

We then interpret this as two parallel universes, one containing the observer in state $|O_2\rangle$, the other containing the observer in the state $|O_3\rangle$. We assume the validity of the Born rule, i.e. that the norms of the $|U_j\rangle$ give the probabilities for the observer to find herself in state $|O_j\rangle$

3 Memory erasure

In the future our descendants will probably be machines [3]. Just like our PCs, there will be a need to back up the content of the memory. If a future machine observer saves a copy of its memory every day then, in case of corrupted memory files, it will erase its entire memory and load the saved memory of the previous day. Can this feature also be used to undo real world facts? Suppose that the observer learns about a comet that is about to impact with the Earth in a few days time. The previous day, the observer and rest of the universe was in a state of the form:

$$|\psi_{\text{initial}}\rangle = |O_{\text{initial}}\rangle |U_{\text{initial}}\rangle \tag{6}$$

The state of the rest of the universe can be expanded in the form:

$$|U_{\text{initial}}\rangle = a |\text{Comet}\rangle + b |\text{Comet}\rangle \tag{7}$$

Where $|\text{(No) Comet}\rangle$ indicates that (no) comet is about to impact Earth. These states are orthonormal. The initial state is properly normalized, therefore

$|a|^2 + |b|^2 = 1$. One can argue that a superposition consisting of widely different macroscopic states should decohere very fast and that therefore the observer must be entangled with that superposition. However, as explained in the previous section, the ket vectors describing the observer only describes whatever the observer is aware of. If we expand the entangled state describing the observer and the rest of the universe in the tensor products of the states $|O_n\rangle$ and the states describing the rest of the universe (which will include parts of the observer that the observer has no direct knowledge of), then the coefficient of each $|O_n\rangle$ will be a superposition of states describing the rest of universe. In case of the state in (8), the observer really cannot know about the next day being either just another boring day or about a comet on its way to hit Earth. Assuming that there exist no such thing as psychic abilities, we should assume that the above initial state is the correct physical representation of a situation where the observer has not received any information about a possible significant event. No amount of meditation should allow the observer to be able to somehow sense that it is part of the second term of the superposition and not the first term.

The next day, the wavefunction will have changed to

$$|\psi_{\text{next day}}\rangle = a\,|O_{\text{We are doomed}}\rangle\,|\text{Comet}\rangle + \sum_j |O_j\rangle\,|U_j\rangle\,. \tag{8}$$

Here the states $|U_j\rangle$ are orthogonal to $|\text{Comet}\rangle$. The observer finds itself with probability $|a|^2$ in the "doomed" state. Suppose that the observer in this state and perhaps also in some other states among the $|O_j\rangle$'s, decide to reset their memories to the previous day's state. The state then changes to the following general form:

$$|\psi_{\text{reset}}\rangle = |O_{\text{initial}}\rangle \left[a\,|\text{Comet;erased memory}\rangle + \sum_j \left|U'_j\right\rangle\right] + \sum_k |O_k\rangle\,|U_k\rangle \tag{9}$$

We note that the state of the rest of the universe will contain the information about the state of the memory that was erased. Memory erasure from the entire universe is not possible due to the fact that time evolution must be unitary. The summation over j is over values for which observer $|O_j\rangle$ would decide to reset its memory. The prime in $\left|U'_j\right\rangle$ indicates that the state has been modified due to the dumping of the memory. The summation over k is over all values for which $|O_k\rangle$ would decide not to reset its memory.

Suppose that after reloading the previous day's memory, the observer is first notified that its memory was reset. The sector of the multiverse this observer is in, is then described by a state vector of the form:

$$|\psi_{\text{notified}}\rangle = N\,|O_{\text{initial}}\rangle \left[a\,|\text{Comet;erased memory}\rangle + \sum_j \left|U'_j\right\rangle\right] \tag{10}$$

where N is a normalization constant. We see that while the observer $|O_{\text{notified}}\rangle$ now knows that its memory was reset, it doesn't know anything about the events of the day. It has to perform new observations to find out what happened. The outcome of that is not pre-determined if some of the $\left|U'_j\right\rangle$ are nonzero. Suppose the observer resets its memory pseudo-randomly with a probability of p when it is not facing a freak disaster and also in case of a predicted freak disaster with extremely low prior probability q. Then N in (10) is given by

$$N = [q + (1-q)\, p]^{-\frac{1}{2}} \tag{11}$$

where we have used that $|a|^2 = q$[1], that due to the unitarity of the memory dumping process the states $\left|U'_j\right\rangle$ are orthogonal to the state $|\,\text{Comet;erased memory}\rangle$ and equal in norm to the corresponding states $|\,U_j\rangle$ in (8), and that which $|\,O_j\rangle$ decides to reset its memory does not correlate with the norm of the $|\,U_j\rangle$.

The probability for the observer $|\,O_{\text{notified}}\rangle$ to find that a freak disaster is coming its way will thus be given by:

$$P_{\text{disaster}} = \frac{q}{q(1-p)+p}. \tag{12}$$

This is approximately q/p if $q \ll p$. It is a simple exercise to show that despite the memory erasure, the observer in the initial state $|\,O_{\text{initial}}\rangle$ has the same probability q of ultimately end up in the sector where the comet is about to hit Earth. This means that we cannot use memory erasure to test the validity of the MWI. Tegmark's quantum suicide experiment [4] purportedly would be able to affect probabilities. It could therefore provide machine civilizations with a better method to escape disasters than memory erasure. However, I disagree with the usual conclusion of this thought experiment. Let's take a closer look at that though experiment.

4 Quantum suicide

In Tegmark's quantum suicide experiment one starts with an initial state containing an observer who is about the measure the z-component of a spin polarized in the x-direction. A suicide machine will kill the observer if the spin points in the negative z-direction. Tegmark argues that since the superposition that represents the final state contains only one conscious observer in the parallel universe where the spin is up, the observer will find herself there with 100

So, we see that the notion of conditional probabilities like: "what is the probability to find yourself in state Y, given that you were in state X a minute earlier" is, in general, inherently problematic. The sum of all the probabilities for all possible Y will not sum to 1 if during the experiment the observer can lose consciousness. If we discard unconscious states and renormalize the probabilities, then we're in trouble if any unconscious state would ever regain consciousness.

Instead, we should assume that the laws of physics only yield prior probabilities for finding us in certain states. So, in the above case there is a prior probability $P(X)$ for an observer to be in state X. And for all possible states Y that correspond to the observer remembering being in state X some time earlier, there is an a priori probability $P(Y)$. The conditional probability which we measure in experiments is then the ratio $P(Y)/P(X)$. When we consider suicide experiments, this quantity does not define a bona fide conditional probability. The sum over all Y of $P(Y)$ won't equal $P(X)$ anymore, indicating that the (non fundamental) notion of a conditional probability has broken down.

[1] Strictly speaking, we should have generalized to more than one possible disaster, but this generalization is straightforward, yielding the same result

In the case of the observer facing the comet impact, quantum suicide by e.g. resetting the memory to previous day's state first and then only restarting the central processing unit if the news is good, won't improve the probability of experiencing a good outcome.

5 Discussion

Given that the probability of experiencing a disaster does not change, we may ask what the point of the memory resetting is. One can argue that it has a big psychological effect: when you are confronted with a random freak disaster, you know that you will reset your memory and then you will very likely find yourself in another sector of the multiverse. Also, you can imagine that after memory resetting and the notification about the fact that the memory has been reset, you will first get some psychological preparation in the event you happen to find yourself in a problematic sector of the multiverse. After this preparation, you will be informed about the situation you are in.

Nevertheless, memory resetting does still intuitively look like a very nonsensical thing to do. After the memory resetting and finding out that the comet is going to hit, you could just check that you went into the memory resetting procedure when you already knew that the comet was going to hit. So you could argue that you just stay in the sector of the multiverse where you are. Now, while it is of course true that any observer will find a consistent history upon investigation, it still the case that after memory resetting and being notified about that, you are in the state (10). In this state the outcome of making the observation of what happened is not a priori fixed.

The observer in this state has an approximate probability of finding that the comet is about to hit with probability q/p if $q \ll p$. Then the fact that the total probability of ending up in the disastrous situation is still q is explained by the fact that you have a prior probability of p of engaging in a memory resetting exercise which, multiplied by the conditional probability of q/p of getting the disastrous outcome when the memory has been reset, yields the probability q. So, if $q \ll p$, it is the fraction of p of the observer states who engage in the routine memory resettings who are the dominant source of the observers who end up experiencing a disaster. The fact that q is the same as the a priori probability of experiencing disaster if the memory had not been reset, then implies that the probability of escaping disaster by resetting the memory is almost equal to 1.

Therefore, if the MWI is true, the observer who after memory resetting ends up having to face a disaster, should really believe that he originated from a different sector of the multiverse where the disaster would not have happened. It is then, of course, true that when he investigates further he'll find that he appears to have faced the same disaster before memory resetting. But the observer who faced that disaster was not him, that was a (former) copy of him who branched off from him in the multiverse since the start of the day. This copy has now very likely escaped to a safe place in the multiverse and he has taken his place.

References

[1] http://en.wikipedia.org/wiki/The_Fabric_of_Reality#Time_travel

[2] H. Everett, Rev. Mod. Phys. **29**, 454 (1957).

[3] Hans Moravec, http://www.frc.ri.cmu.edu/ hpm/

[4] M. Tegmark, Fortsch. Phys. **46**, 855-862 (1998).