10. The Dynamics by Itself (Albert Chap 6)

Topics
I. Many Worlds
II. The Bare Theory
III. Many Minds

Consider a system composed of a human observer h, a *Color* measuring device m, and an initially *hard* electron e. Before measurement, state is represented by $|ready\rangle_h|ready\rangle_m|hard\rangle_e$. After measurement, according to the Schrödinger dynamics, state is represented by the superposition:

$$\sqrt{1/2}|believes\ e\ black\rangle_h|\text{“}black\text{”}\rangle_m|black\rangle_e + \sqrt{1/2}|believes\ e\ white\rangle_h|\text{“}white\text{”}\rangle_m|white\rangle_e \qquad (*)$$

Option (A1) says this is a state in which
(a) electron e has no definite color
(b) measuring device m has no definite reading
(c) human observer h has no definite belief about measurement outcome

** Recall Option (A1): Literal interpretation of superpositions: Properties are indeterminate in a superposed state.*

BUT: ***Measurements are supposed to have unique outcomes!***
(This is the motivation for the *Projection Postulate*.)

Recall: ***The Measurement Problem*** = How to reconcile the *Schrödinger dynamics* with the *Projection Postulate*; *i.e.*, how to reconcile states like (*) with our experience that *measurements have unique outcomes.*

***GRW's Solution*:** Keep the Projection Postulate and modify the Schrödinger dynamics so that superpositions like (*) will *not* occur after measurements.

***MW Solution*:** Keep the Schrödinger dynamics and give up the Projection Postulate! Attempt to explain how measurements do not *really* have unique outcomes (even though we think they do).

I. The Many Worlds (*MW*) Interpretation DeWitt (1970), Everett (1957)

MW Claims:

(A) States evolve ***only*** *via* the Schrödinger dynamics (no Projection Postulate).

(B) Each term in a superposition represents a state in a *distinct physical world*.

$$\underbrace{\sqrt{1/2}\,|\textit{believes e black}\rangle_h|\,\text{“}\textit{black}\text{”}\rangle_m|\textit{black}\rangle_e}_{\substack{\textit{state of h-m-e}\\ \textit{system in World A}}} + \underbrace{\sqrt{1/2}\,|\textit{believes e white}\rangle_h|\,\text{“}\textit{white}\text{”}\rangle_m|\textit{white}\rangle_e}_{\substack{\textit{state of h-m-e}\\ \textit{system in World B}}}$$

According to *Eigenvector/Eigenvalue Rule*, in *both* Worlds A and B,

(a) electron e has definite color

(b) measuring device m has definite reading

(c) human observer h has definite belief about measurement outcome

Eigenvector/Eigenvalue Rule: A state possesses the value λ of a property represented by operator O *if and only if* that state is an eigenvector of O with eigenvalue λ.

Consequences:

(1) Any given measurement does not have one unique outcome! Rather, when a measurement occurs, *all* its possible outcomes are realized, one per world.

(2) ***SO***: Each time a measurement occurs, the world splits, in general into an infinite number of worlds.

(3) There is no interaction between worlds (we don't experience these splits -- we *think* measurements have unique outcomes).

One way to think of this:

$$|\textit{universe}\rangle \xrightarrow[\textit{evolution}]{\textit{Schrödinger}} \sqrt{1/2}\,|\textit{universe}\rangle_A + \sqrt{1/2}\,|\textit{universe}\rangle_B$$

universal state vector at any given instant

universal state vectors for Worlds A and B after Color measurement

Important Note: According to *MW*, a "measurement" isn't necessarily confined to a special type of interaction between measuring devices, human observers and measured things. In general, *any* interaction between two or more physical systems may precipitate a splitting of worlds. (Any interaction that is governed by the Schrödinger dynamics may result in a superposition.)

Three Problems with MW

1. ***The "Preferred Basis" Problem***

   The terms in a superposition depend on the basis in which we expand it. And what basis we expand it in is, in general, completely arbitrary (one basis is just as good, mathematically, as another). So if the terms in a superposition represent different worlds, what worlds there are will arbitrarily depend on the basis we choose to expand the superposition in! (And this seems strange: Shouldn't there be a fact of the matter as to what the "many worlds" are after an interaction?)

   Ex: ***Recall***: $|black\rangle_e = \sqrt{1/2}|hard\rangle_e + \sqrt{1/2}|soft\rangle_e$

   $|white\rangle_e = \sqrt{1/2}|hard\rangle_e - \sqrt{1/2}|soft\rangle_e$

   SO: We can rewrite:

$$\sqrt{1/2}\underbrace{|believes\ e\ black\rangle_h|\text{"black"}\rangle_m|black\rangle_e}_{\substack{\textit{World A with}\\ \textit{black electron}}} + \sqrt{1/2}\underbrace{|believes\ e\ white\rangle_h|\text{"white"}\rangle_m|white\rangle_e}_{\substack{\textit{World B with}\\ \textit{white electron}}}$$

   as:

$$\sqrt{1/2}\underbrace{|Q+\rangle_{h\&m}|hard\rangle_e}_{\substack{\textit{World C with}\\ \textit{hard electron}}} + \sqrt{1/2}\underbrace{|Q-\rangle_{h\&m}|soft\rangle_e}_{\substack{\textit{World D with}\\ \textit{soft electron}}}$$

   where: $|Q+\rangle_{h\&m} = \sqrt{1/2}|believes\ e\ black\rangle_h|\text{"black"}\rangle_m + \sqrt{1/2}|believes\ e\ white\rangle_h|\text{"white"}\rangle_m$

   $|Q-\rangle_{h\&m} = \sqrt{1/2}|believes\ e\ black\rangle_h|\text{"black"}\rangle_m - \sqrt{1/2}|believes\ e\ white\rangle_h|\text{"white"}\rangle_m$

SO: If we initially had a *hard* electron and we let it interact with a *Color* measuring device, what does *MW* say about how the world splits? Does it split into Worlds A and B, or does it split into Worlds C and D?

Task: Find a "fundamental" basis in terms of which all superpositions should be expanded.

2. ***<u>The Problem of Probabilities</u>***

When applied to (*), the *Born Rule* says: When *h* measures the *Color* of *e*, there is a probability of ½ that the outcome will be *black*, and a probability of ½ that the outcome will be *white*.

<u>BUT</u>: *MW* says: When *h* measures the *color* of *e*, the world *splits* into a world in which the outcome is *black* with ***certainty***, and a world in which the outcome is *white* with ***certainty***! Where did the probabilities go?

<u>MW</u>: "All outcomes occur."

<u>Born Rule</u>: "Each outcome has a distinct probability of occurring."

<u>Possible Responses</u>:

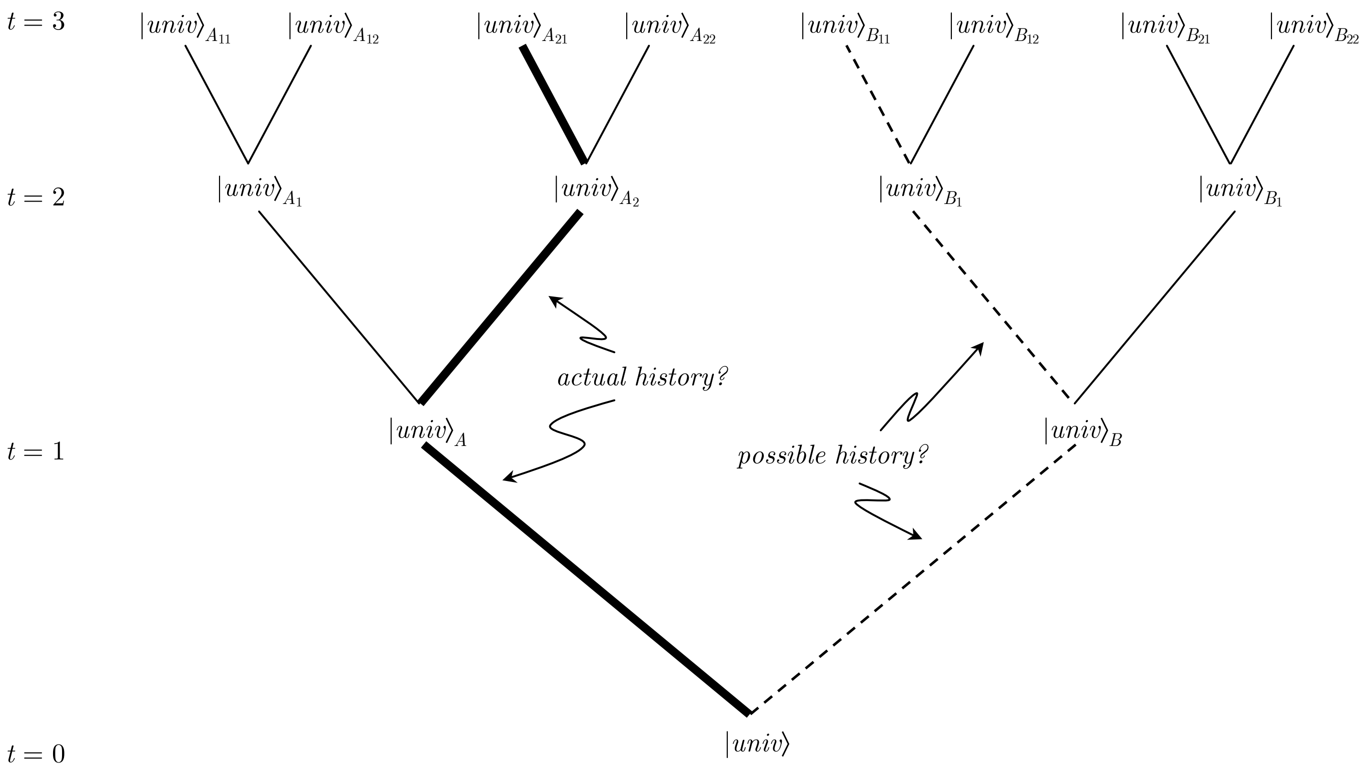
(i) ***<u>MW probabilities are objective</u>***: They are defined over possible worlds. So when a *Color* measurement is conducted, the world splits with probability ½ into World *A*, and probability ½ into World *B*.

<u>Task</u>: To justify this talk, we need to specify an indeterministic dynamics of worlds: a dynamical law that tells us how a given world evolves over time indeterministically into others. (At the least, this requires some notion of the identity of a world over time).

(ii) ***<u>MW probabilities are subjective</u>***: They reflect the state of knowledge of the human observer in the act of measurement. So when a *Color* measurement is conducted, *h* doesn't know which world (*A* or *B*) she will end up in; she only knows the probability of which world she will end up in.

<u>Tasks</u>: Again, this requires some notion of the identity of *h* over time to make sense. (On the surface, there seem to be *two h*'s after the measurement, one in each world with certainty!) Moreover, this way of talking seems to be falling back on a distinction between *measurements* (interactions involving human observers) and other types of interactions, which is what *MW* is trying to get away from in the first place.

There's also a deeper problem: Even if we agree on how to interpret *MW* probabilities (either in terms of worlds or measures of ignorance), why do they replicate the probabilities in quantum mechanics? Why, after conducting a *color* measurement, is there a probability of ½ for a *black* electron world, and a probability of ½ for a *white* electron world? This follows from the *Born Rule* in quantum mechanics, but why should it follow for *MW* probabilities? (The problem here is that *MW* doesn't really *need* probabilities: it's completely deterministic (it's based solely on the *Schrödinger dynamics*). So how do we justify introducing probabilities into it, and why are the probabilities we introduce into it the same ones that occur in quantum mechanics?)



To make sense of probabilities, we need to know how worlds relate over time. We need to be able to pick out *possible* "histories" of worlds, and then be able to distinguish them from the *actual* "history". This indeterministic world dynamics would then let us say things like: "The probability of World B_1 occurring, given we're in World *B,* is so-and-so." (Again, even if we could construct such a dynamics, we'd still have to be able to explain why it reproduces the same probabilities that the Born Rule requires.)

3. ***The Problem of Conservation Laws***

When the universe splits, aren't conservation laws violated?

$$\left(\begin{array}{c}\textit{universe}\\ \textit{splits}\end{array}\right) \Rightarrow \left(\begin{array}{c}\textit{number of physical}\\ \textit{objects increases}\end{array}\right) \Rightarrow \left(\begin{array}{c}\textit{violation of conservation}\\ \textit{of mass/energy?}\end{array}\right)$$

Possible Response:

A world includes *spacetime* as well as physical objects. And worlds split "outside" of spacetime (each "new" world contains its own spacetime and its own physical objects). So no violations of mass/energy conservation.

Problem: Does it make sense to say splits occur outside of time, in particular? Don't we want to say something like: "At time t_1, there is one world. At time $t_2 > t_1$, after a *Color* measurement, there are two worlds."

II. The "Bare Theory"

MW says: Keep the Schrödinger dynamics and give up the Projection Postulate! Attempt to explain how measurements do not *really* have unique outcomes (even though we think they do).

The *Bare Theory* says the same thing, but differs on the explanation of why measurements don't really have unique outcomes. It tries to offer an explanation without all the "world-talk".

> ***Bare Theory Claims***:
> (A) States evolve ***only*** *via* the Schrodinger dynamics (no Projection Postulate).
> (B) We are deluded in thinking measurements have unique outcomes. (We never have definite beliefs about measurement outcomes; at best, we have "effective knowledge" about outcomes.)

Motivation: What would it feel like to be in a superposition? Suppose h conducts a *Color* measurement on e and ends up in standard state (*):

$$\sqrt{1/2}\,|\textit{believes e black}\rangle_h|\textit{“black”}\rangle_m|\textit{black}\rangle_e + \sqrt{1/2}\,|\textit{believes e white}\rangle_h|\textit{“white”}\rangle_m|\textit{white}\rangle_e \qquad (*)$$

> ***Now ask h***: "Do you have *any* definite belief about the value of the *Color* of e?

Note first:

Suppose O is a linear operator representing some property and let $|A\rangle$ and $|B\rangle$ be eigenvectors of O with the same eigenvalue, call it λ:

$$O|A\rangle = \lambda|A\rangle$$
$$O|B\rangle = \lambda|B\rangle$$

Then any linear superposition $\alpha|A\rangle + \beta|B\rangle$ of these eigenvectors will also be an eigenvector of O with eigenvalue λ (where α and β are numbers).

Proof:

$$\begin{aligned} O(\alpha|A\rangle + \beta|B\rangle) &= \alpha(O|A\rangle) + \beta(O|B\rangle) \\ &= \alpha\lambda|A\rangle + \beta\lambda|B\rangle \\ &= \lambda(\alpha|A\rangle + \beta|B\rangle) \end{aligned}$$

Now Note:

(i) If after measurement, *h-m-e* state is given by the first term of (*), then h will respond to the question with "*Yes*".

(ii) If *h-m-e* state is given by the second term of (*), then h will respond with "*Yes*".

(iii) Think of the question as a property of the *h-m-e* system and "*Yes*" as a value of this property. By the proof above, the superposition (*) will also have this property (if we adopt the *Eigenvector/Eigenvalue Rule*), and its value in (*) will also be "*Yes*"!

SO: By the *Eigenvector/Eigenvalue Rule*, when (*) obtains, h doesn't have a definite belief about the *Color* of e. But h "effectively knows" what the *Color* of e is: h will answer "Yes" if asked if she knows what the *Color* of e is.

Consequence:

According to the Bare Theory, we are ***always*** mistaken about the values of the properties of things. The ***only*** beliefs that we are never mistaken about are beliefs about whether or not some definite measurement result was observed.

Problem: Is this a coherent position to take?

(1) Consider *introspection*: If we're certain about anything, shouldn't it be about our own beliefs? Suppose you are h. Ask yourself: "Did I just see a definite *Color* result for e?" According to the Bare Theory, you will answer "Yes", but this is mistaken! You really do not have a definite belief about what the *Color* of e is! Furthermore, if you now ask yourself: "Did I just say 'Yes'?", you will answer "Yes", but this is mistaken, too! You really have no definite belief about whether or not you just said "Yes". Only if you ask yourself: "Did I just give a definite reply?", and you answer "Yes", will you not be mistaken!

(2) Is the Bare Theory self-defeating? Any belief we might have for evidence for it (or for quantum mechanics in general) would be mistaken, according to it!

III. The Many Minds (*MM*) Interpretation Albert and Loewer 1988

Bare Theory: Tries to tell a story about how belief states in superpositions can still be said to have "effective" collapses, even if they really don't.

Many Minds: Distinguishes between physical states and mental states and says, physical states can be in superpositions, but mental states never are.

MM Claims:

(A) *Physical states* evolve ***only*** *via* the Schrodinger dynamics (no Projection Postulate).

(B) *Mental states* ("minds") evolve *via* an indeterministic dynamics in such a way that they are *never* in superpositions.

Motivation: Suppose h conducts a *Color* measurement on e and ends up in standard state (*):

$$\sqrt{½}|\textit{believes e black}\rangle_h|\text{“}\textit{black}\text{”}\rangle_m|\textit{black}\rangle_e + \sqrt{½}|\textit{believes e white}\rangle_h|\text{“}\textit{white}\text{”}\rangle_m|\textit{white}\rangle_e \qquad (*)$$

MM says:

(i) This is a *physical* state. In particular, the h-states are physical "brain" states of h.

(ii) Corresponding to these *brain states* of h are *mental states* (which aren't represented in (*)).

(iii) (*), as a physical superposition, means: h is in the mental state associated with the brain state $|\textit{believes e black}\rangle_h$ with probability ½, and h is in the mental state associated with the brain state $|\textit{believes e white}\rangle_h$ with probability ½.

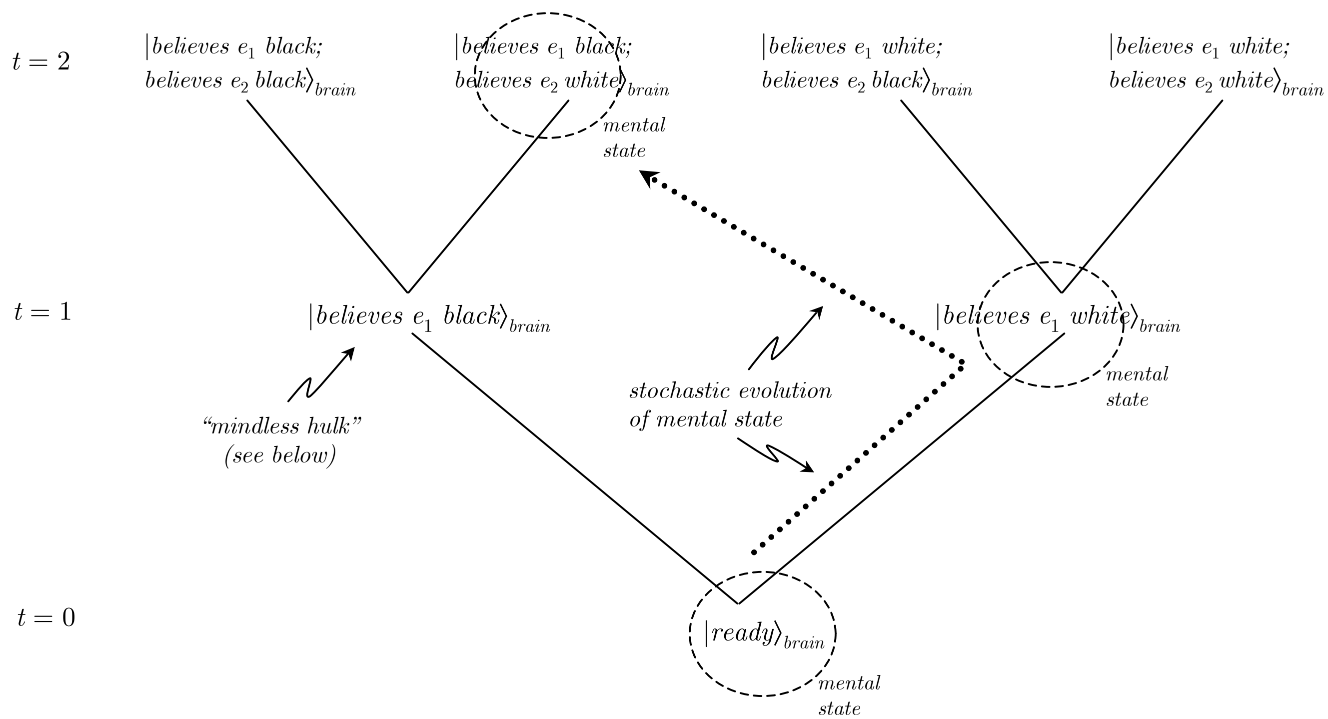
Why this is supposed to help:

Suppose h is in the physical state (*) and has a definite belief about what the *Color* of e is. According to the *Eigenvector/Eigenvalue Rule*, this is a false belief. But h's mental state (according to *MM*) evolves in a way that is consistent with supposing h's belief were true. For instance, if she thinks e is *black*, then her mental state evolves to the mental state corresponding to the brain state $|\textit{believes e black}\rangle_h$.

> This is indeed better than the Bare Theory: What explains our "effective knowledge" of measurement outcomes are our mental states. So we aren't completely deceived by measurements: While our *brains* may be in superpositions, our *minds* are not! In particular, we will be correct about what we *report* we believe (even though our beliefs themselves will be incorrect).

How do mental states evolve?

MM just stipulates that a mental state evolves stochastically in such a way that it always corresponds to the appropriate brain state with the appropriate associated quantum mechanical probability.



Suppose h conducts a series of *Color* measurements on a number of electrons e_1, e_2, ... The above tree represents the evolution of h's physical brain state (*via* the *Schrödinger dynamics*) after the first two measurements. *MM* says h's mental state evolves stochastically, "jumping" about in the tree of physical states in such a way that the probability of it being associated with any brain state is given by the *Born Rule*. The above tree represents one possible trajectory h's mental state could take. At $t = 2$, h's brain state is in a superposition of 4 terms, but her mental state is associated with the belief that e_1 is black and e_2 is white.

Note that this random jumping entails that h is deluded about both her *current* and *past* beliefs about measurement outcomes, but not her beliefs about *what* her beliefs are! For instance, at $t = 1$, she thinks she measured e_1 white, whereas at $t = 2$, she thinks she measured e_1 to be black at $t = 1$. But she will never be able to tell that she's been inconsistent! At any given time, her mental state will correspond to a brain state with definite measurement outcomes for the electrons. So her own introspective beliefs about her current and past beliefs will be reliable (unlike in the Bare Theory).

Initial Problem:

Without further ado, only one of the terms in (*) will actually be associated with an actual mental state. (Albert: Most people we meet will be "mindless hulks"!)

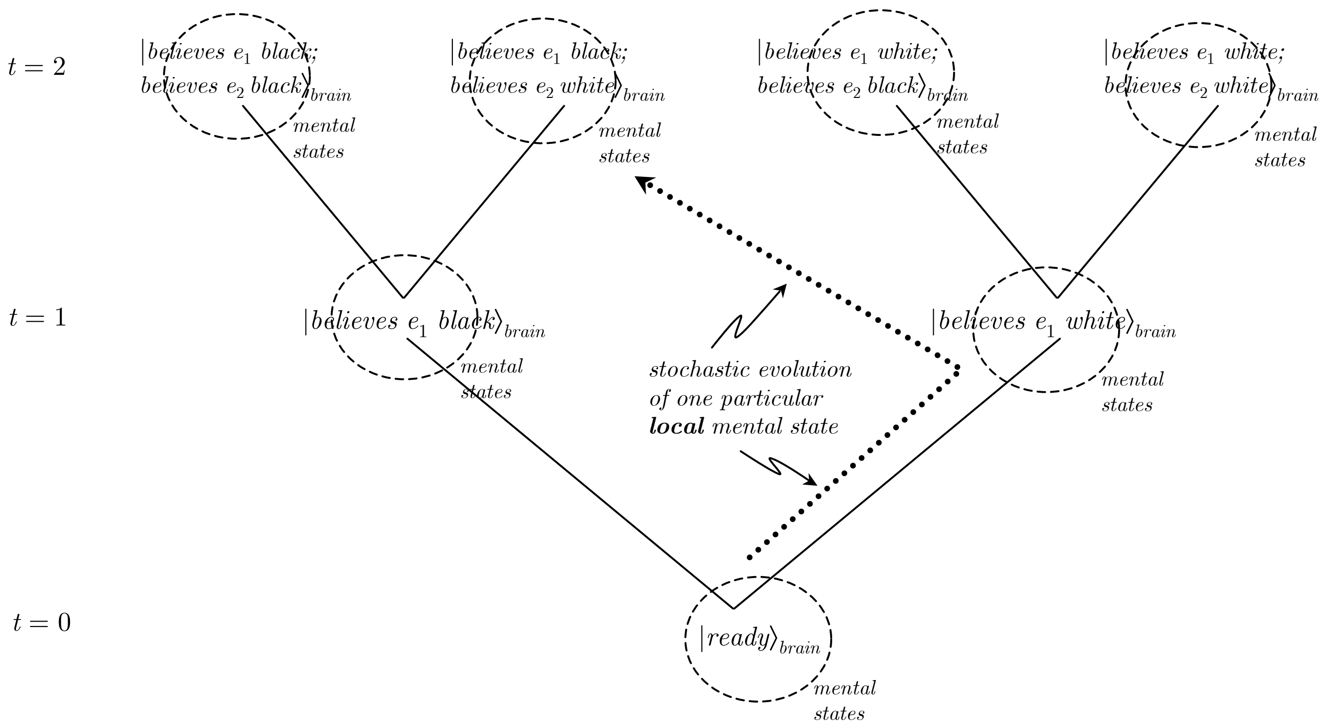
Remedy:

(1) Claim that every sentient physical system has a *continuous infinity* of minds. (!)

(2) The proportion of mental states (minds) associated with a given brain state is equal to the square of the absolute value of the expansion coefficient of that brain state in the universal state vector.

(3) Each mind evolves stochastically as before.

SO: For (*), half of h's minds will be associated with the brain state $|believes\ e\ black\rangle_h$ and the other half will be associated with the brain state $|believes\ e\ white\rangle_h$.



All branches here are associated with some (or all) of the minds of h. Each of h's minds (call it a local mental state) evolves stochastically. Each of her minds, as before, has consistent introspective beliefs, but possibly delusional beliefs about measurement outcomes. The complete collection of all of h's minds, call it her "global mental state", evolves deterministically according to the Schrodinger dynamics (it's what get's divided up horizontally among the branches at any given time according to the *Born Rule*).

General Problems:

(1) ***What's a mind***?

If we can agree on an explicit distinction between mental states and physical states, why go to all the trouble of *MM*? Why not just use this distinction as a means of implementing the *Projection Postulate*? (*Recall*: The problem with reconciling the Projection Postulate with the Schrödinger dynamics was, in one form, determining just when the Projection Postulate applies. If we had a distinction between minds and bodies, we could simply say: Apply the Projection Postulate whenever a mind interacts with a body.)

(2) ***How do probabilities appear in MM?***

MM, arguably, avoids the *MW* problems of preferred bases and conservation laws (how?), but what about the problem with probabilities? Again, what seems to be needed is an indeterministic dynamics of minds (as opposed to worlds) that agrees with the probabilities that quantum mechanics prescribes. (Albert says these probabilities can simply be put in by fiat, stipulating that they agree with *QM* prescriptions. But is this an adequate response? Why can't *MW* respond in a similar way?)