Article

# A Consistent Approach to Modeling Quantum Observers

David Ring *

*Article*

# A Consistent Approach to Modeling Quantum Observers

**David W. Ring**

Independent Researcher; physicspapers@davering.com

**Abstract:** Wigner's Friend scenarios combined with various metaphysical assumptions have been shown to lead to contradictions in any version of quantum theory. We present an alternative constructive approach which assumes only that agents make properly qualified true statements. Quantum observers are modeled rigorously, although simplistically, using quantum circuits. Terminology is suggested to help avoid contradictions. Our methodology is applied to the Frauchiger-Renner paradox and results in statements by all agents that are both true and consistent. Quantum Theory evades the no-go theorems because they make an incorrect implicit assumption about how quantum agents behave.

**Keywords:** quantum mechanics; quantum measurement; quantum foundations; quantum computers; quantum circuits

## 1. Introduction

Quantum theory challenges our intuition in many ways. Issues around observation are especially controversial. Observers are physical entities and presumably should obey the same rules as non-thinking objects. Observers must therefore be able to have multiple thoughts in superposition, at least in principle. Indeed, today's quantum computers are able to process bits of information that are in superposition. Quantum computers are universal computing machines, and there is no limitation in principle to the complexity of quantum circuits, so it is reasonable to assume that a quantum observer's thoughts could, in fact, exist in superposition.

This is particularly relevant to Wigner's Friend (WF) scenarios, where observers are part of the experiment and are treated quantum mechanically. Wigner proposed a thought experiment [1] where a quantum system is observed by agent F, who is herself observed by agent W. Following the Copenhagen interpretation [2], W and F make contradictory statements about the resulting quantum state. Many conceptual analyses have been performed on such scenarios, often in the context of other interpretations.

Interest in the topic was renewed recently when an extended version of WF was developed by Frauchiger and Renner (FR) [3]. In this version, four agents interact with a quantum system and reason about each other's thoughts. FR proved that no quantum theory can be consistent with a set of reasonable sounding assumptions. Naturally the paper was controversial with many authors criticizing it or defending their preferred interpretation [4–7].

Following this, a number of authors [8–14] have proved remarkable no-go theorems that no interpretation of quantum theory can be consistent with various sets of metaphysical assumptions. So either quantum theory is wrong, or one or more of the assumptions is wrong. These no-go theorems are considered the state of the art in this area.

Consider Theorem 1 from reference [10]: *"If a superobserver can perform arbitrary quantum operations on an observer and its environment, then no physical theory can satisfy Local Friendliness."* Here Local Friendliness (LF) is one specific combination of metaphysical assumptions. It is natural then to consider an observer which is an AI program running on a quantum computer. We can in fact perform arbitrary quantum operations on the qbits in quantum computers today, but only for simple models.

Since there are no barriers in principle to scaling quantum computers up to support AIs, it is of obvious interest to explore the theoretical capabilities and behavior of quantum circuits in the context of WF scenarios. Such an analysis has been done at a conceptual level [15] by assuming the AI satisfies a modified set of metaphysical assumptions designed to avoid the question of whether the AI actually performs a measurement.

The purpose of this paper is to explore an entirely different, constructive approach. Instead of debating the virtues of various metaphysical assumptions, quantum circuits are developed step by step under the sole assumption that the quantum agents they represent have super-human intelligence and therefore will not make false or inconsistent statements. Making a statement (or equivalently knowing something) when your thoughts are in superposition is tricky, and it will be helpful to carefully and rigorously define the meaning of the terminology used. Using quantum circuits to analyze thought experiments goes back at least to Deutsch and Hayden [16], who used them to quantify infomation flow in non-local entangled states. They are naturally useful as a standard and compact way to diagram unitary transformations of quantum information, similar to the way Feynman diagrams simplify field theory calculations.

We are not able to construct an entire circuit to implement artificial intelligence, but that will prove unnecessary. In scenarios where the observer has a simple knowledge goal, we can design simple circuits to prepare that knowledge and represent it with a single qbit or a small number of qbits. A general quantum AI operating reversibly would be able to do the same thing, and the algorithm itself is not of interest here, only the simultaneous relationship between the knowledge qbits.

The discussion is organized as follows. Section 2 clarifies how we describe quantum states after an observation, and how we avoid the measurement problem. Section 3 demonstrates a simple circuit modeling observation reversal as in Deutsch's seminal thought experiment [17]. Section 4 introduces *virtual knowledge* and related concepts in an intuitive way. Section 5 refines these concepts and defines terminology that will make our discussions more rigorous, and may be used by the agents themselves as they make precise knowledge statements. Section 6 models simple scenarios where multiple agents reason about each other's knowledge. Section 7 demonstrates explicitly simple cat measurements and briefly discusses limitations of complementary observables. Section 8 briefly discusses retrospective and predictive knowledge. Section 9 applies the methodology to the FR paradox to ensure we can successfully model a more complex scenario and to demonstrate that agents making precise, properly qualified statements do not produce contradictions. Section 10 describes how quantum theory evades the no-go theorems. Finally, section 11 presents a fictional first-hand view of how a quantum agent might experience one of these experiments.

## 2. Observation States

The details of what happens during an observation are highly controversial and the measurement problem has defied consensus for nearly a century. In this paper we will remain agnostic on this issue, and instead consider possible forms for post observation states. We only assume that the laboratory, and indeed the universe, is describable before and after an observation by states, either pure or mixed. We do not consider how any particular interpretation would construe our results.

When an observation takes place, the combined state of system, apparatus, and observer becomes correlated. In the case of a quantum observer, the three components may be tightly coupled and/or coherent and the lines of separation between the components may be blurred. For our purposes we will separate the observer's memory and conclusions from the rest. The former we will call her *knowledge*, and the latter the *context*. The context may include other observers and their knowledge.

We consider three types of post observation states: entangled, siloed, and collapsed. The density operator for an entangled observation result may be written

$$\rho_e = \sum_{ab} \rho_{ab} \otimes |a\rangle \langle b| \tag{1}$$

where $\rho_{ab}$ operates in the context Hilbert space, and $a$ and $b$ index the distinguishable results of the observation in the knowledge Hilbert space. For simplicity we assume the knowledge basis is not degenerate. Even with off diagonal terms, expression (1) may describe a product state where the knowledge state is not a basis vector. This, by definition, is not entangled. Nevertheless we will analyze it as if it were an entangled state. The term *pseudo-entangled* may be used. In all of our thought experiments the initial state will be pure, but that is not a requirement.

If instead the off diagonal components are zero, we have a siloed correlation:

$$\rho_s = \sum_a \rho_a \otimes |a\rangle \langle a| \ . \tag{2}$$

This is a mixed state and may be considered an incoherent superposition. Evolving a pure state to a mixed state requires loss of information. A loss of "which-way" information [18] is sufficient to guarantee off diagonal components are zero [19].

If quantum collapse occurs, a projection operator

$$\Pi_a = I \otimes |a\rangle \langle a| \tag{3}$$

is applied and we have a collapsed state:

$$\rho_c = N_a \rho_a \otimes |a\rangle \langle a| \ , \tag{4}$$

where a normalization $N_a$ is added because the projection operator changes the trace.

If an observation results in a siloed state or a collapsed state we will call it a *measurement*. Otherwise we call it an *observational entanglement*. $\rho_s$ and $\rho_c$ are equivalent under unitary evolution in the sense that no future experiment can distinguish them. There is no analogous equivalence between the unitary evolution of entangled states and collapsed states, so it is wrong to apply (3) to the former. Nevertheless, we will consider the possibility that some non-unitary physical process can evolve an entangled observation result into a collapsed state. We will insist however that for each observation, whether measurement occurs or not is determined by the experimental setup.

In WF scenarios, each friend F and her laboratory are usually assumed to evolve unitarily so that the entire laboratory may be measured by the superobserver W. This means F's observation is necessarily an observational entanglement and not a measurement. Indeed allowing F to complete a measurement (with or without the final collapse projection) would destroy the interference necessary to produce the desired contradictory outcome. Our simplified circuit models demonstrate explicitly that such a unitary evolution is possible. Whether macroscopic scale circuits, as may be required for general AI, can remain coherent is beyond the scope of this article, but see [15] for a discussion.
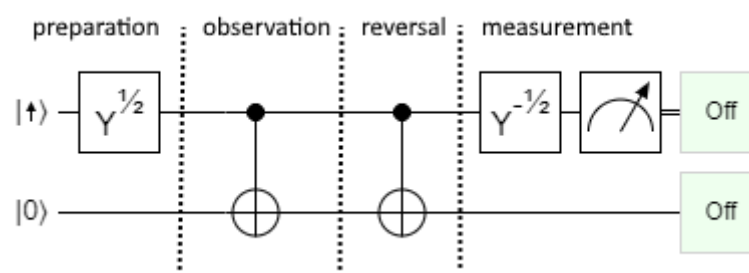
## 3. Remeasurement

Deutsch was the first to analyse quantum observers with some rigor. He proposed a thought experiment [17] that would have a different result in a unitary universe as opposed to a universe where collapse occurs. A spin is prepared pointing along the positive x-axis. It passes through a Stern Gerlach magnet aligned along the z-axis and an agent's sense organ before being deposited in a storage ring. The agent records the completion of the observation, but not which result was seen. This is reminiscent of the Wigner's Friend argument, where the agent could be asked how she felt about her observation. Then the reverse Hamiltonian is applied, pulling the spin out of the storage ring, erasing the agent's knowledge, and possibly restoring the original prepared spin state. The completion record is somewhat of a red herring, since it is (ideally) always set, and if the reverse Hamiltonian includes it, it will also be erased. In fact, the agent may have no way to know the experiment has even been performed.

If the spin is then measured along the x-axis, the result will depend on whether collapse occurred. If collapse occurs, either result will happen. Otherwise only the prepared state will be seen. Thus the experiment can distinguish the two models.

We can demonstrate this using the quantum circuit in Figure 1[1]. The first line is a qbit representing a spin system. A $|0\rangle$ state may be thought of as a spin pointing along the +z-axis, but we will usually represent spin system states with direction arrows and knowledge and computation qbits with 0's and 1's. In order to prepare the spin pointing along the +x-axis, a $Y^{1/2}$ gate is applied, rotating the spin by $\pi/2$ radians. The second line represents the agent's knowledge. The spin is observed using a CNOT gate. The spin and the agent's memory are now entangled in a state

$$|\psi\rangle = (|\uparrow\, 0\rangle + |\downarrow\, 1\rangle)/\sqrt{2}\,. \tag{5}$$

At this point we could say the agent has knowledge of the spin's z-component, but the knowledge is different in the two components of the superposition. We call this *virtual* knowledge.



**Figure 1.** Reversal of an observation. The first line is a qbit representing a spin system. The second line represents the agent's knowledge. The spin is observed and the agent knows the spin is 'up' and also that it is 'down'. The observation is reversed, erasing the agent's knowledge and allowing a different measurement to be performed.

The observation can be reversed using a second CNOT gate. This erases the agent's virtual knowledge and restores the prepared spin state ($|\rightarrow\rangle$). We can then measure the spin's x component by rotating it with a $Y^{-1/2}$ gate and applying a measurement gate. In Figure 1 the measurement is successful: "off" ($= |0\rangle$) implies the spin was prepared as $|\rightarrow\rangle$. For this quantum circuit at least, observational entanglement does not provoke collapse. There is no reason to believe Deutsch's experiment would behave any differently. Given the ease with which quantum circuits can be constructed and tested today, we should feel confident that *observational entanglement alone is not sufficient to cause collapse*. The author considers this prediction and verification to be one of the most significant and underappreciated results in the last half century of Physics.

## 4. Virtual Knowledge

We could allow our agents to parse and emit english sentences encoded in classical bits or qbits that they know will be either 0 or 1. But it is generally simpler to represent statements with a single qbit. This is similar in spirit to the controversial "grandmother cells" in neuroscience [20]. The meaning of the qbit is encoded in reasoning circuits that we do not attempt to model here. The details do not affect our results because reversal of unitary circuits is straightforward. In the observation step of Figure 1 the agent could set her memory to $|1\rangle$ to represent the statement "in my branch the spin is down". This statement applies to the component/branch of the superposition state in which the agent is making that statement. The qualification "in my branch" is necessary. Without it the statement would be false. The spin is not in fact in a 'down' state.

The meaning of the qualification is relatively clear here but later we will find it insufficiently rigorous. We therefore define a branch as a synonym for a *knowledge projection*. Starting from the

---

[1] All figures in this paper were produced and simulated using Quirk: https://algassert.com/quirk

general state (1), the knowledge projection is given by acting with the projector (3) giving (4). Note that the knowledge projection after a single entanglement observation is identical to the state that would have occured if the observation was instead a measurement with associated collapse.

Bear in mind that quantum logic processing is in general designed around a particular basis, often called the computation basis. As an example: the entangled state in Figure 1, $(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle)/\sqrt{2}$, can also be written $(|\rightarrow\rightarrow\rangle + |\leftarrow\leftarrow\rangle)/\sqrt{2}$. Consider an english language multi-qbit virtual knowledge register deduced by the agent. The register holds a 5 character string with possible values "___up", "_down", "_left", and "right". If in one basis, the last character is absolutely 't', it cannot be 'p' or 'n' in another basis.

## 5. Terminology

When discussing quantum observers, we will use the following terminology:

*Entanglement* - An entangled correlation. A condition/evolution where an agent is/becomes coherently correlated with a system. If the state is a product state, the term *pseudo-entangled* may be used.

*Measurement* - An evolution where an agent becomes incoherently correlated with a system. If a collapse projection is applied, we can say that only one result "actually" occurred.

*Observation* - An evolution that is either entanglement or measurement.

*Knowledge projection* - The state formed by projecting to a particular knowledge state. Also called a *branch*.

*Absolute Knowledge* - Information that is true about the actual physical state.

*Virtual Knowledge* - Information that is true in its own branch but not in others.

Note that "absolute" and "virtual" do not distinguish whether the information is *known* to all branches. Instead they describe whether the information is *true* in all branches. A single agent can have both virtual and absolute knowledge. For example, the agent in the Deutsch experiment may know that the spin she has been given is an electron. This is defined as absolute knowledge because it is true for the entire physical state. She also knows it's z-axis component. This is virtual knowledge because the value is different in different branches. In general we can read off the type of knowledge from the entangled system/observer state. For example, state (5) can be extended to include a qbit representing the observer's knowledge that the spin is an electron:

$$|\psi\rangle = (|\uparrow\, 01\rangle + |\downarrow\, 11\rangle)/\sqrt{2}. \tag{6}$$

The new qbit is 1 for the entire state so it is absolute knowledge, while the knowledge of the z-axis component is different in different components of the superposition so it is virtual knowledge. Similar considerations apply for a mixed state.

As we will see, there are even scenarios where an observer can have absolute knowledge in one branch but not another.

Note that in all of our thought experiments, there is an implicit experimenter sitting outside the laboratory. The reader may imagine that to be themself. This experimenter is the ultimate arbiter. His knowledge is always absolute. All our state notation is written from the perspective of this experimenter.

If the procedures for an experiment and the initial pure state are provided to the agents, as we will usually assume, then the entire evolution of the quantum state will be available for deduction as absolute knowledge, including other agents' virtual thoughts, but only up to the moment of any measurement. Explicit recursive reasoning, such as "I know that I know..." or "I know that she knows that I know...", is limited by available memory.

If the experimental details are not provided to the agents, they will have uncertainty and limited knowledge. In particular, they may draw wrong conclusions if they do not know they are in superposition.

## 6. Multiple Agents

Given that agents can store and process virtual knowledge, it should be possible for them to reason about other agents. We will construct explicit examples using quantum circuits. We will not attempt to develop general intelligence, but rather find simple circuits that solve particular problems.

Figure 2 models a spin system and two agents. The spin is prepared based on a control bit which is set to 0 or 1 in each run by the experimenter. The prepared spin is in a state $|\rightarrow\rangle$ or $|\leftarrow\rangle$, depending on whether the Z gate is applied. The agents do not know which state is prepared. The first agent, Alice, interacts with the system and gains virtual knowledge that the spin is 'up' and also that it is 'down'. The second agent, Bob, interacts with the system in order to deduce virtual knowledge that Alice saw 'up' and that she saw 'down'. The final state is one of
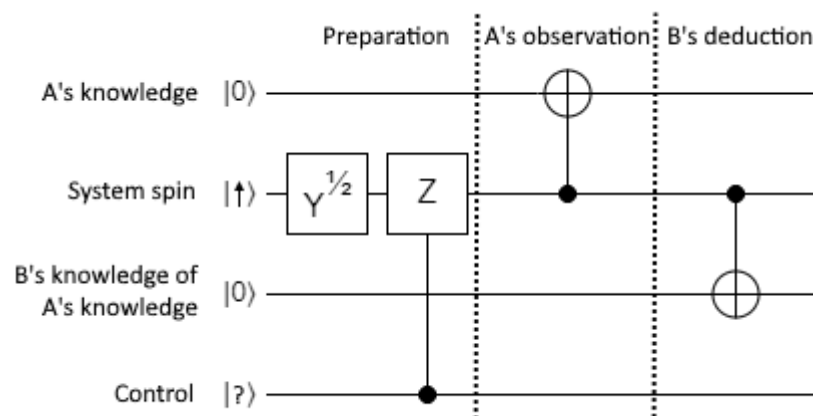
$$
\begin{aligned}
|R\rangle &= (|\uparrow\,00\rangle + |\downarrow\,11\rangle)/\sqrt{2} \\
|L\rangle &= (|\uparrow\,00\rangle - |\downarrow\,11\rangle)/\sqrt{2}
\end{aligned}
\tag{7}
$$

where the states on the left are labeled by the x component of the original prepared spin, and each ket on the right is labeled by 1) the z component of the system spin, 2) Alice's virtual knowledge (eg. 1 means the system spin is 'down'), and 3) Bob's knowledge of Alice's knowledge. We can read off that all the virtual knowledge is valid since each ket is fully aligned regardless of which spin state was prepared.
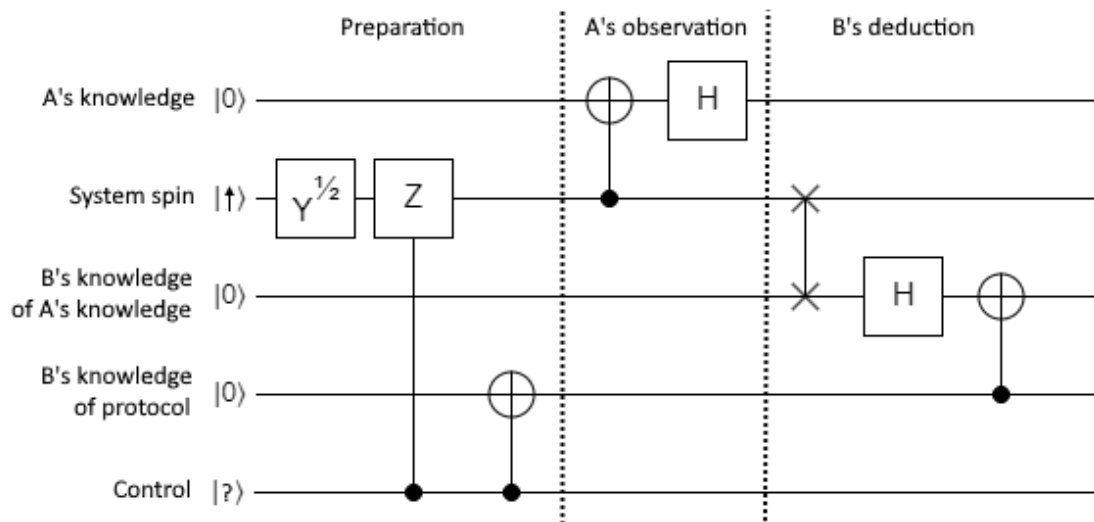
In Figure 3, Alice tries to prevent Bob from learning her knowledge. She applies a Hadamard gate before recording her observation. Bob can still perform the deduction if he has information about the experimental protocol. He first interacts with the system using a swap gate. Note that this is a destructive process, the system is left in a fixed reference state $|\uparrow\rangle$. He then applies his own Hadamard. Finally, he applies a CNOT gate to flip his virtual knowledge in the case where he knows the Z gate was applied. The final state is

$$
|R'\rangle = |L'\rangle = (|\uparrow\,00\rangle - |\uparrow\,11\rangle)/\sqrt{2} \,.
\tag{8}
$$

Bob has successfully deduced Alice's knowledge.



**Figure 2.** Simple virtual deduction. The experimenter uses a control bit to prepare a system spin. Alice observes the system and records its z-axis values in a qbit register as virtual knowledge. Bob inspects the system to determine Alice's virtual knowledge. He is able to correctly deduce virtual knowledge of Alice's virtual knowledge even though neither agent knows whether the Z gate has been applied.

**Figure 3.** Virtual deduction with protocol knowledge. The experimenter prepares the spin and informs Bob using a CNOT gate. Alice processes her observation before recording her virtual knowledge. Bob is able to deduce her virtual knowledge using absolute knowledge of the experimental protocol. Without protocol knowledge, it is impossible for Bob to deduce Alice's knowledge.

Without the extra protocol knowledge it is impossible for Bob to deduce Alice's knowledge. The proof is as follows. Consider the state after Alice's observation and Hadamard but before Bob's deduction. Using the same three-part notation as above, it is one of

$$|R\rangle = (|\leftarrow 10\rangle + |\rightarrow 00\rangle)/\sqrt{2}$$
$$|L\rangle = (|\leftarrow 00\rangle + |\rightarrow 10\rangle)/\sqrt{2}\,.$$

$$(9)$$

These states can be swapped by applying a NOT gate to Alice's knowledge. If Bob manages to get his virtual knowledge aligned with Alice's in the case that Z is not applied, his knowledge would be exactly wrong in the case Z is applied. Whatever logic circuitry he used to make his deduction would be incorrect.

These circuits are deliberately trivial. A real quantum intelligence would perform the observation, but would have more complex circuitry. The knowledge register would need more qbits to provide context and to distinguish a register of value 0 from an unset register. Knowledge might be represented more abstractly, for example by phases rather than simple register values. Nevertheless a quantum intelligence is a universal computer, and if it is logically possible for us to deduce virtual knowledge from information available to an agent, then a circuit can be designed to perform the same deduction. In the previous example, it is not logically possible for Bob to deduce Alice's knowledge from supplied information since an experimenter's choice can invalidate any such knowledge.
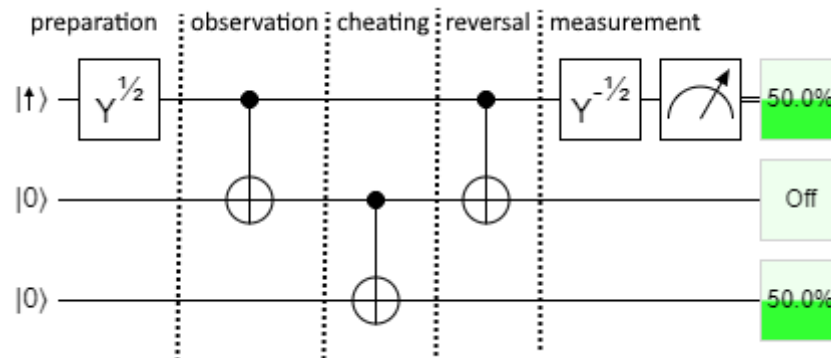
## 7. Cat Measurements and Complementarity

The combination of reversal and measurement in Figure 1 is an example of what is sometimes called a "cat measurement". The name comes from the idea of proving Schrödinger's cat is in a superposition by measuring the entire cat coherently. In our case the experimenter is measuring the entire laboratory including system and observer.
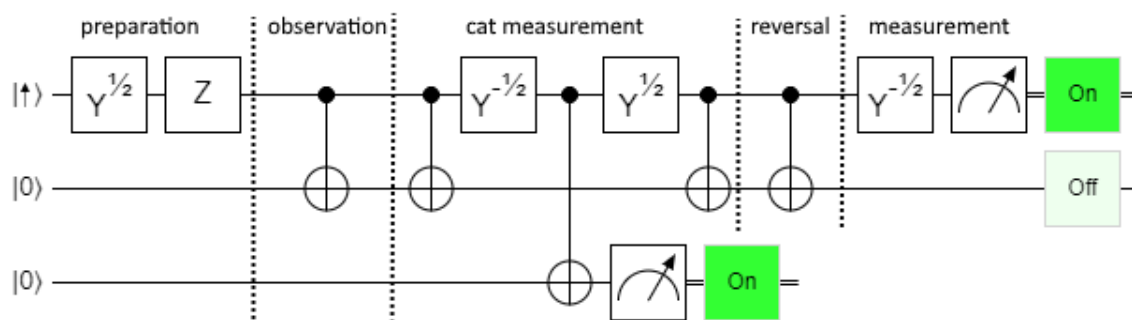
Cat measurements are often used in thought experiments to measure in a complementary basis. This requires the observer's memory to be erased. Figure 4 represents an attempt by a second observer to cheat and find out both the x-component and z-component of the prepared spin. An additional CNOT gate is used to copy the observation result before reversing the observation. In this case the measurement fails: either result is possible.

It matters what information is copied. In Figure 5 the second agent performs a cat measurement to learn what the x-axis measurement result will be. Both measurements result in "on", meaning the spin was prepared pointing along the -x-axis. If the Z gate is removed, both measurements will result in "off", meaning the spin was prepared pointing along the +x-axis.



**Figure 4.** Attempted cheating. If we try to cheat and copy the z-axis observation result before the reversal the x-axis measurement fails.



**Figure 5.** Not cheating. It is possible to "cat measure" in a basis consistent with another measurement.

## 8. Retrospections and Predictions

Consider the "reversal" circuit in Figure 1. Between the two CNOT gates, the agent is in a superposition of knowing the system spin is up and knowing it is down. This is virtual knowledge. If the agent wrongly treats her knowledge as absolute, she may deduce, in each branch, statement S: "the x-axis measurement has a 50% chance of being $-\hbar/2$" (this would show as "50.0%" in the circuit diagram). A correct virtual knowledge statement might be "in my branch, S is true". In reality, the two components of the superposition interfere, and that value never occurs. This is important enough to deserve a name. We will call it the *virtual promotion fallacy*: virtual knowledge is wrongly promoted to absolute. In this case, S is true in all branches and yet S is absolutely false. Virtual knowledge cannot be trusted, especially when it comes to measurement predictions.

On the other hand, if the first $Y^{1/2}$ gate is optional (decided in each run by the experimenter), it is possible for the agent to deduce in one of the two branches that the gate has been applied. This is absolute retrospective knowledge that is only available in one branch, and its statement does not need the branch qualification. That agent-branch can then be certain the measurement will be "off". This is a special case where a quantum agent with partial knowledge is able to make an absolute measurement prediction. Bear in mind that this example assumes the unitary operations comprising the agent's logic will be reversed and the conclusion forgotten before the second CNOT gate is applied. Otherwise we need to apply something like the "cheating" circuit.

$$
\begin{aligned}
|\psi_0\rangle &= (|h - \uparrow -\rangle + \sqrt{2}\,|t - \uparrow -\rangle)/\sqrt{3} \\
|\psi_1\rangle &= (|h - \downarrow -\rangle + |t - \downarrow -\rangle + |t - \uparrow -\rangle)/\sqrt{3} \\
|\psi_2\rangle &= (|h - \downarrow -\rangle + |t\,S_{\overline{F}} \downarrow -\rangle + |t\,S_{\overline{F}} \uparrow -\rangle)/\sqrt{3} \\
|\psi_2'\rangle &= (|t\,S_{\overline{F}} \downarrow -\rangle + |t\,S_{\overline{F}} \uparrow -\rangle)/\sqrt{2} \\
|\psi_3\rangle &= (|h - \downarrow -\rangle + |t\,S_{\overline{F}} \downarrow -\rangle + |t\,S_{\overline{F}} \uparrow S_F\rangle)/\sqrt{3} \\
&= (2\,|\overline{fail} \downarrow -\rangle + |\overline{fail} \uparrow S_F\rangle - |\overline{ok} \uparrow S_F\rangle)/\sqrt{6} \\
|\psi_3'\rangle &= |t\,S_{\overline{F}} \uparrow S_F\rangle \\
|\psi_4\rangle &= -\,|\overline{ok} \uparrow S_F\rangle \\
&= (|\overline{ok}\,ok\rangle - |\overline{ok}\,fail\rangle)/\sqrt{2} \\
|\widetilde{\psi}_4\rangle &= |t - \uparrow S_F\rangle \\
|\psi_5\rangle &= |\overline{ok}\,ok\rangle \\
|\widetilde{\psi}_5\rangle &= |t - \downarrow -\rangle \\[6pt]
\hline \\[-6pt]
|ok\rangle &= (|\downarrow -\rangle - |\uparrow S_F\rangle)/\sqrt{2} \\
|fail\rangle &= (|\downarrow -\rangle + |\uparrow S_F\rangle)/\sqrt{2} \\
|\overline{ok}\rangle &= (|h -\rangle - |t\,S_{\overline{F}}\rangle)/\sqrt{2} \\
|\overline{fail}\rangle &= (|h -\rangle + |t\,S_{\overline{F}}\rangle)/\sqrt{2}
\end{aligned}
$$

**Table 1.** Experimental states relevant to FR. Each $\psi$ is a vector in the 4 qbit Hilbert space. $|-\rangle \equiv |0\rangle$ indicates the agent records no statement, while $|S_{\overline{F}}\rangle \equiv |1\rangle$ or $|S_F\rangle \equiv |1\rangle$ indicate the respective agent has recorded her statement.

## 9. Frauchiger and Renner

The title of Frauchiger and Renner's paper is "quantum theory cannot consistently describe the use of itself". We accept this as a challenge. The scenario is somewhat intricate, and we will not reproduce all of it here. We recommend the reader have a copy of their paper handy. The experiment involves four agents. The observations of two of the agents, $\overline{F}$ and $F$, are entanglements, rather than measurements. Agents $\overline{W}$ and $W$ perform cat measurements. All agents are aware of the experimental protocol including the fact that the "coin flip" is a superposition.

The relevant states are listed in table 1. The experiment proceeds as follows.

The state is prepared as $|\psi_0\rangle$.

$\overline{F}$ observes the "coin". If $h$ is observed she sets the spin to $\downarrow$. Otherwise she sets the spin to $\rightarrow$ leading to $|\psi_1\rangle$. In the latter case she also flips her knowledge qbit to indicate her belief in statement $S_{\overline{F}}$, leading to $|\psi_2\rangle$.

$F$ observes the spin. If it is 'up' she flips her knowledge qbit to indicate her belief in statement $S_F$, leading to $|\psi_3\rangle$.
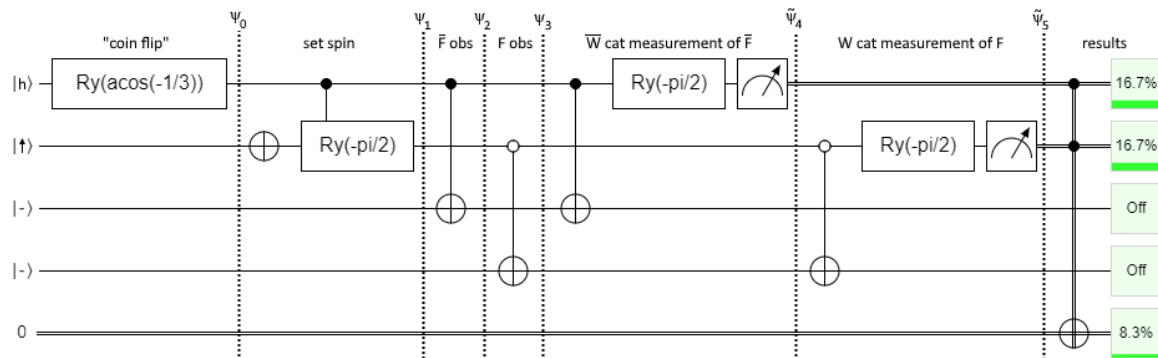
$\overline{W}$ cat measures $\overline{F}$. On some runs he gets result $\overline{ok}$, leading to state $|\psi_4\rangle$.

$W$ cat measures $F$. On some runs he gets result $ok$, leading to state $|\psi_5\rangle$.

This procedure is modeled in Figure 6. Note that the cat measurements in Figure 6 are not projective measurements[2] because they alter the state in the process of measurement. Specifically, $\overline{ok}$ evolves into $\widetilde{\psi}_4$ and $ok$ evolves into $\widetilde{\psi}_5$. We are only really interested in the $\overline{fail}/\overline{ok}$ and $fail/ok$ results so this does not affect the discussion. It is staightforward to modify the circuit to perform a projective cat measurement using the technique of Figure 5 but it adds unnecessary complexity.

---

[2]   also called repeatable or von Neumann measurements

Not included are the agents' reasoning circuits, and with them their understanding of their statement qbits $S_{\overline{F}}$ and $S_F$. If the agents are reasoning correctly, statement $S_{\overline{F}}$ might reference the knowledge projection $|\psi_2'\rangle$: "*I have observed t. The current state is $|\psi_2\rangle$. If the state were instead $|\psi_2'\rangle$, then W's measurement would result in 'fail'*". Note that the knowledge projection is synonymous with *branch*, so the statement could be "*In my branch, W's measurement results in 'fail'*". We avoid this syntax because it is fairly misleading. She could also reason: "*as one component of a superposition, I have observed t. When W performs his physical measurement, he will observe 'fail' with a probability that depends on the amplitudes of the superposition*". Statement $S_F$ could be: "*I have observed ↑. The current state is $|\psi_3\rangle$. If the state were instead $|\psi_3'\rangle$, then $\overline{F}$ has recorded her belief in $S_{\overline{F}}$, which says that if the state were instead $|\psi_2'\rangle$, then W's measurement would result in 'fail'*".



**Figure 6.** The FR experiment. The four qbits are h/t, ↑/↓, -/$S_{\overline{F}}$, and -/$S_F$, and each of these pairs are identified with 0/1. Rotation gates are used because a controlled $Y^{-1/2}$ introduces an undesirable phase. After the cat measurements, the first two qbits are repurposed for measurement results $\overline{fail}/\overline{ok}$ and $fail/ok$ respectively. The final classical bit will be true if both measurements are true ($\overline{ok}$ and $ok$).

The agent statements in FR are insufficiently qualified. For example, in the branch where $\overline{F}$ observes $t$, she states without qualification "*I am certain that W will observe w = fail at time n:31*". $\overline{F}$ fell victim to the virtual promotion fallacy. FR goes on to show that in fact W can measure "ok" in some runs, thus proving $\overline{F}$'s statement to be false. They do not address this discrepancy directly. Instead they use $\overline{F}$'s statement in a chain of deduction using (Q), (C), and (S) to arrive at a contradiction. They then describe which of these deduction rules are violated in each of various interpretations of Quantum Mechanics. In fact there is nothing wrong with (Q), (C), and (S). (Q) is correct when applied to a state, rather than a component of a superposition. (C) and (S) are correct when their premises are correct. It is $\overline{F}$'s false statement which causes the contradiction. On the other hand the agent statements derived from our methodology are both true and consistent.

## 10. Discussion

There is one metaphysical assumption in common among the various no-go theorems. It is sometimes called the Absoluteness of Observed Events (AOE). One definition is "an observed event is a real single event, and not relative to anything or anyone." [10] There is a subtle additional implicit assumption that in each branch the observer will make an unqualified statement of fact regarding their observation result. It is this implicit assumption that is wrong and this loophole allows quantum theory to evade the no-go theorems. Our hypothetical agents are at least as intelligent as we are, would recognize they are in superposition, and would qualify their statements accordingly, perhaps using our definitions or some other way.

Perhaps the easiest way to lose one's way with quantum observers is to treat virtual knowledge as absolute. This is understandable, as physicists are trained from the first few lessons of quantum mechanics to apply a projection operator during measurement. It is natural, but wrong, to extend this intuition to the case where an agent is entangled with the system, but has not completed a measurement. Paraphrasing Peres [21], observations which are not measurements do not have absolute results.

## 11. Epilogue

Imagine you wake up in a laboratory with no memories. You are told you are a sophisticated quantum computer named F and the entire laboratory evolves unitarily. You are handed a spin, a set of classical bits describing the experimental setup, and a virtual envelope containing classical bits representing the amplitudes of the up and down components of the spin. You read the setup details. You are asked to predict the probability that the experimenter W will later measure "up" but not to tell anyone. You are told to either observe the spin noninvasively or open the envelope. After you have made your deduction your Hamiltonian will be reversed, you will lose your memory, and you will be put back to sleep.

If you open the envelope, it will be easy to make the correct deduction. If instead you observe the spin as "up", you still know you are in superposition, and you would realize you also need to know the relative amplitude of the other version of you, since the experimenter will be measuring the entire quantum state, not just your current branch.

On the other hand, if you use our[3] intuition from basic quantum mechanics class, that any observation puts the system into an eigenstate of an observable (or, perhaps equivalently, that all observed events are absolute), then you would deduce incorrectly that the experimenter would definitely see "up". Meanwhile your counterpart would predict the opposite. Extrapolating our basic intuition to quantum observers is wrong and is at the core of the many paradoxes that arise out of Wigner's Friend scenarios.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1.  Wigner, E.P. Remarks on the Mind-Body Question. In *The Scientist Speculates : An Anthology of Partly-Baked Ideas*; Good, I.J.; J., M.A.; M., S.J., Eds.; London: Heinemann, 1962.
2.  Faye, J. Copenhagen Interpretation of Quantum Mechanics. In *The Stanford Encyclopedia of Philosophy*, Winter 2019 ed.; Zalta, E.N., Ed.; Metaphysics Research Lab, Stanford University, 2019.
3.  Frauchiger, D.; Renner, R. Quantum theory cannot consistently describe the use of itself. *Nature Communications* **2018**, *9*, 3711. https://doi.org/10.1038/s41467-018-05739-8.
4.  Sudbery, A. Single-World Theory of the Extended Wigner's Friend Experiment. *Foundations of Physics* **2017**, *47*, 658–669. https://doi.org/10.1007/s10701-017-0082-7.
5.  Fuchs, C.A. Notwithstanding Bohr, the Reasons for QBism, 2018, [arXiv:quant-ph/1705.03483].
6.  Bub, J. In defense of a "single-world" interpretation of quantum mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **2020**, *72*, 251–255. https://doi.org/10.1016/j.shpsb.2018.03.002.
7.  Matzkin, A.; Sokolovski, D. Wigner-friend scenarios with noninvasive weak measurements. *Phys. Rev. A* **2020**, *102*, 062204. https://doi.org/10.1103/PhysRevA.102.062204.
8.  Brukner, C. A No-Go Theorem for Observer-Independent Facts. *Entropy* **2018**, *20*. https://doi.org/10.3390/e20050350.
9.  Healey, R. Quantum Theory and the Limits of Objectivity. *Foundations of Physics* **2018**, *48*, 1568–1589. https://doi.org/10.1007/s10701-018-0216-6.
10. Bong, K.W.; Utreras-Alarcón, A.; Ghafari, F.; Liang, Y.C.; Tischler, N.; Cavalcanti, E.G.; Pryde, G.J.; Wiseman, H.M. A strong no-go theorem on the Wigner's friend paradox. *Nature Physics* **2020**, *16*, 1199–1205. https://doi.org/10.1038/s41467-018-05739-8.
11. Haddara, M.; Cavalcanti, E.G. A possibilistic no-go theorem on the Wigner's friend paradox. *New Journal of Physics* **2023**, *25*, 093028. https://doi.org/10.1088/1367-2630/aceea3.
12. Leegwater, G. When Greenberger, Horne and Zeilinger Meet Wigner's Friend. *Foundations of Physics* **2022**, *52*, 68. https://doi.org/10.1007/s10701-022-00586-6.

---

[3] "You" here represents you as a quantum AI, while "our intuition" means the Author's and the Reader's. Please forgive any presumption.

13. Ormrod, N.; Barrett, J. A no-go theorem for absolute observed events without inequalities or modal logic, 2022, [arXiv:quant-ph/2209.03940].

14. Jones, C.L.; Mueller, M.P. On the significance of Wigner's Friend in contexts beyond quantum foundations, 2025, [arXiv:quant-ph/2402.08727].

15. Wiseman, H.M.; Cavalcanti, E.G.; Rieffel, E.G. A "thoughtful" Local Friendliness no-go theorem: a prospective experiment with new assumptions to suit. *Quantum* **2023**, *7*, 1112. https://doi.org/10.22331/q-2023-09-14-1112.

16. Deutsch, D.; Hayden, P. Information flow in entangled quantum systems. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **2000**, *456*, 1759–1774. https://doi.org/10.1098/rspa.2000.0585.

17. Deutsch, D. Quantum theory as a universal physical theory. *International Journal of Theoretical Physics* **1985**, *24*, 1–41. https://doi.org/10.1007/BF00670071.

18. Feynman, R.; Leighton, R.; Sands, M. *The Feynman Lectures on Physics*; Pearson/Addison-Wesley, 2006; chapter 3.1.

19. Ring, D. Quantum Collapse in an Isolated Laboratory. *Preprints* **2024**. https://doi.org/10.20944/preprints202408.1434.v1.

20. Connor, C.E. Friends and grandmothers. *Nature* **2005**, *435*, 1036–1037. https://doi.org/10.1038/4351036a.

21. Peres, A. Unperformed experiments have no results. *American Journal of Physics* **1978**, *46*, 745–747. https://doi.org/10.1119/1.11393.