Quantum theory cannot consistently describe the use of itself

Daniela Frauchiger and Renato Renner*
Institute for Theoretical Physics, ETH Zurich, Switzerland

Quantum mechanics may be used to describe systems that contain agents who themselves use quantum mechanics. We propose a gedankenexperiment to test the consistency of such a recursive use of the theory. We find that one agent, upon observing a particular measurement outcome, concludes that another agent has predicted the opposite outcome with certainty. The agents' conclusions, although all derived within quantum mechanics, are thus inconsistent. The experiment also provides a way to compare current interpretations of quantum mechanics, which differ in where they locate the origin of this inconsistency.

I. INTRODUCTION

Quantum mechanics is one of our best tested physical theories. Yet, despite amazing progress in experiments, we do not know whether or not its laws are applicable to complex macroscopic objects, like Schrödinger's cat [1]. Here we take a theorist's approach to explore this question. The idea, illustrated by Fig. 1, is to ask whether quantum mechanics can consistently describe systems that are complex enough to include agents who themselves use the theory to make predictions.

With a gedankenexperiment, we show that such a self-referential use of quantum theory sometimes yields contradictory claims. The experiment is information-theoretic in nature and can be described and analysed using standard quantum formalism (cf. the circuit diagram in the appendix). However, as emphasised by the term "gedankenexperiment," we do not claim that it is technologically feasible, at least not in the form presented here. Rather than probing nature, its purpose is to scrutinise the consistency of our description of nature in terms of quantum theory. (One may compare this to, say, the gedankenexperiment of letting an observer cross the event horizon of a black hole. Although we do not have the technology to carry out this experiment, reasoning about it provides us with insights on relativity theory.)

Our starting point is an argument due to Wigner about a very standard experiment: An agent, F, measures a system, S. Instead of taking F's perspective, however, Wigner analysed the situation from an outside viewpoint, from where one has no access to the outcome observed by F. His conclusion was that, due to the linearity of the quantum mechanical equations of motion, the joint system consisting of everything affected by F's measurement must evolve towards a superposition state, which has components for each possible measurement outcome. This is known as the *Wigner's Friend paradox* [2].

To make this more precise and concrete, we take S to be the spin of an electron and suppose that agent F non-destructively measures its vertical direction z, as shown in Fig. 2. Hence, upon observing the outcome, $z = -\frac{1}{2}$ or $z = +\frac{1}{2}$, agent F must conclude that S is in state

$$\psi_S = |{\downarrow}\rangle_S \qquad \text{or} \qquad \psi_S = |{\uparrow}\rangle_S \ , \tag{1}$$

respectively.

Consider now an outside agent, W, who has no direct access to the outcome z observed by his friend F. He may view F's lab as a big quantum system, $L \equiv S \otimes D \otimes F$, with a part F that includes his friend and another part D that contains her measurement devices and everything else connected to them. We suppose that, from W's perspective, the lab is at the beginning of the experiment in a pure

---

* renner@ethz.ch

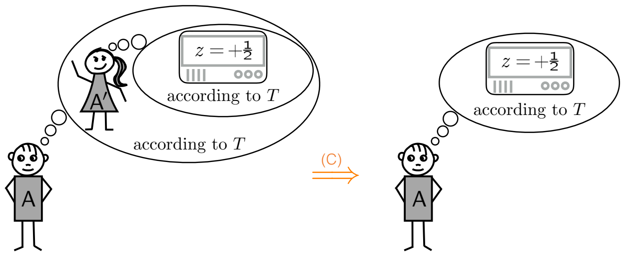

Figure 1. *Consistent reasoning.* We say that a theory T enables consistent reasoning (C) if one agent, A, can draw conclusions by reasoning within T about the conclusions drawn by another agent, A′, who also uses T. A classical example of such recursive reasoning is the muddy children puzzle (here T is just standard logic; see [7] for a detailed account). The idea of using a physical theory T to describe agents who themselves use T has also appeared in thermodynamics, notably in discussions around Maxwell's demon [8].

state, or at least a good approximation of such a state. (One may be concerned here that assigning pure states to complex systems is unrealistic, for it requires extremely accurate information about them [3, 4]. Crucially, however, the laws of quantum mechanics do not preclude this [5].) We also assume that the lab L is isolated during the time when F carries out her measurement. Translating this assumption to quantum mechanics, it means that the dependence of the final state of L on the initial state of S is given by a linear map of the form

$$U_{S\to L} = \begin{cases} |{\downarrow}\rangle_S \mapsto |{-\tfrac{1}{2}}\rangle_L = |{\downarrow}\rangle_S \otimes |\text{“}z{=}{-\tfrac{1}{2}}\text{”}\rangle_D \otimes |\text{“}\psi_S{=}|{\downarrow}\rangle\text{”}\rangle_F \\ |{\uparrow}\rangle_S \mapsto |{+\tfrac{1}{2}}\rangle_L = |{\uparrow}\rangle_S \otimes |\text{“}z{=}{+\tfrac{1}{2}}\text{”}\rangle_D \otimes |\text{“}\psi_S{=}|{\uparrow}\rangle\text{”}\rangle_F \,. \end{cases} \tag{2}$$

Here $|\text{“}z{=}{-\tfrac{1}{2}}\text{”}\rangle_D$ and $|\text{“}z{=}{+\tfrac{1}{2}}\text{”}\rangle_D$ denote the states of the lab's devices depending on the measurement outcome z. Analogously, $|\text{“}\psi_S{=}|{\downarrow}\rangle\text{”}\rangle_F$ and $|\text{“}\psi_S{=}|{\uparrow}\rangle\text{”}\rangle_F$ denote the states of F, with the labels indicating the friend's conclusions about the state of S as in (1). Hence, by linearity, if the initial spin state was, say, $|{\rightarrow}\rangle_S = \sqrt{1/2}(|{\downarrow}\rangle_S + |{\uparrow}\rangle_S)$, then the final state that W assigns to L is

$$\Psi_L = \sqrt{1/2}\big(|{-\tfrac{1}{2}}\rangle_L + |{+\tfrac{1}{2}}\rangle_L\big) \,, \tag{3}$$

i.e., a superposition of the two states defined in (2).

Although the state assignment (3) may appear to be “absurd” [2], it does not logically contradict (1). Indeed, the marginal on S is just a fully mixed state. While this is different from (1), the difference can be explained by the agents' distinct level of knowledge: F has observed z and hence knows the spin direction, whereas W is ignorant about it [6]. Agent F's state assignment, (1), and agent W's assignment, (3), are hence not contradictory.

To nevertheless arrive at the claimed contradiction, we need to extend the setup considered by Wigner. The basic idea is to make some of the information about the value z held by agent F, who is enclosed in lab L, available to the outside, but without lifting the isolation of L. Roughly, this is achieved by letting the initial state of S depend on a random value, r, which is known to another agent outside of L.

The analysis of this extended gedankenexperiment ultimately yields a no-go theorem (Theorem 1). It states that three natural-sounding assumptions about the agents' reasoning, (Q), (C), and (S), cannot all be valid. Assumption (Q) captures the universal correctness of quantum theory (specifically, it proclaims that an agent can be certain that a given proposition holds whenever the quantum-mechanical Born rule assigns probability 1 to it), (C) demands consistency in the sense illustrated by Fig. 1, and (S) ensures that, from the viewpoint of an agent who carries out a measurement, the measurement has one single outcome (e.g., if the agent observes $z = +\tfrac{1}{2}$ then she can be certain that she did not also observe $z = -\tfrac{1}{2}$). While the no-go theorem itself does not tell us which

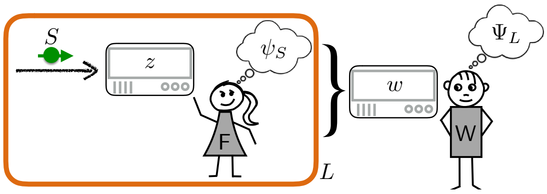

Figure 2. *Wigner and Deutsch's arguments.* Agent F measures the vertical spin component S of an electron, obtaining outcome z. Assuming that the measurement is non-destructive, S is now, from F's perspective, in a pure state ψ_S; cf. (1). Agent W, who is outside of F's lab, may instead regard that lab, including the agent F, as a big quantum system L (orange box). Having no access to z, he would assign a state Ψ_L of the form (3) to L, as noted by Wigner [2]. Deutsch argued that W could in principle test this state assignment by applying a carefully designed measurement to L [9].

of these assumptions is wrong, any specific interpretation of quantum theory, when applied to the gedankenexperiment, will necessarily conflict with at least one of them. This gives a new way to test and categorise interpretations of quantum theory (cf. Table III).

II. RESULTS

A. Description of the gedankenexperiment

We consider four different agents, who follow the protocol described below, with the specifications given in Table I. Two agents, the "friends" F and $\bar{\mathsf{F}}$, are located in separate labs, denoted by L and $\bar{L}$, respectively. Two further agents, W and $\bar{\mathsf{W}}$, are at the outside, from where they can apply measurements to L and $\bar{L}$, as shown in Fig. 3. We assume that the two labs are, from the viewpoint of W and $\bar{\mathsf{W}}$, initially in a pure state, and that they remain isolated during the experiment unless the protocol explicitly prescribes a communication step or a measurement applied to them.

Experimental Protocol

The steps are repeated in rounds $n = 0, 1, 2, \ldots$ until the halting condition (in the last step) is satisfied.

At n:00	Agent $\bar{\mathsf{F}}$ invokes a randomness generator (based on the measurement of a quantum system, R, in state $\lvert\mathsf{init}\rangle_R$ as defined in Table I) that outputs $r = \mathsf{heads}$ or $r = \mathsf{tails}$ with probabilities $1/3$ and $2/3$, respectively. She sets the spin of an electron, S, to $\lvert\downarrow\rangle_S$ if $r = \mathsf{heads}$ and to $\lvert\rightarrow\rangle_S = \sqrt{1/2}(\lvert\downarrow\rangle_S + \lvert\uparrow\rangle_S)$ if $r = \mathsf{tails}$, and sends S to F.
At n:10	Agent F measures S w.r.t. the basis $\{\lvert\downarrow\rangle_S, \lvert\uparrow\rangle_S\}$, recording the outcome $z \in \{-\frac{1}{2}, +\frac{1}{2}\}$.
At n:20	Agent $\bar{\mathsf{W}}$ measures $\bar{L}$ w.r.t. a basis containing the vector $\lvert\bar{\mathsf{ok}}\rangle_{\bar{L}}$ (defined in Table I). If the outcome associated to this vector occurs he announces $\bar{w} = \bar{\mathsf{ok}}$ and else $\bar{w} = \bar{\mathsf{fail}}$.
At n:30	Agent W measures L w.r.t. a basis containing the vector $\lvert\mathsf{ok}\rangle_L$ (defined in Table I). If the outcome associated to this vector occurs he announces $w = \mathsf{ok}$ and else $w = \mathsf{fail}$.
At n:40	If $\bar{w} = \bar{\mathsf{ok}}$ and $w = \mathsf{ok}$ then the experiment is halted.

The numbers on the left indicate the timing of the steps. For example, in round $n = 0$, agent F must start her measurement of S at time 0:10 and complete it before 0:20.

Before proceeding to the analysis of the thought experiment, a few comments about its relation to earlier proposals are in order. In the case where $r = \mathsf{tails}$, and leaving out the measurements by $\bar{\mathsf{W}}$ and W, the situation is identical to the one considered by Wigner, as described in the introduction. Furthermore, with W's measurement reinserted, it corresponds to a variant due to Deutsch [9], as depicted in Fig. 2. The particular choice of states and measurements used in the gedankenexperiment is derived from a construction by Hardy [10, 11]. The experiment is also similar to Brukner's

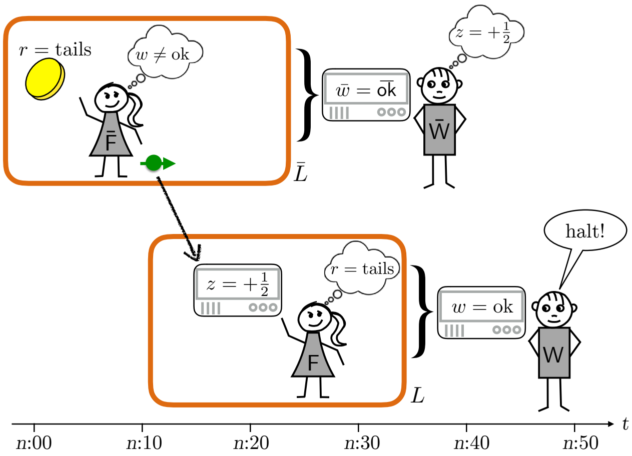

Figure 3. *Illustration of the gedankenexperiment.* In each round $n = 0, 1, 2, \ldots$, agent $\bar{\mathsf{F}}$ polarises an electron in a direction determined by a random value r. Agent F measures its vertical polarisation z. Agents $\bar{\mathsf{W}}$ and W measure the entire labs of $\bar{\mathsf{F}}$ and F to obtain outcomes $\bar{w}$ and w, respectively. For the analysis of the experiment, we assume that all agents are aware of the entire experimental setup but observe different parts. Agent $\bar{\mathsf{F}}$, for instance, observes r but has no direct access to w. She may however use quantum theory to draw conclusions about w.

proposal [12] of using Wigner's argument to obtain a strengthening of Bell's theorem [13] (cf. the discussion section).

B. Analysis of the gedankenexperiment

We analyse the experiment from the individual agents' viewpoints. For this we suppose that all agents employ the same *theory*, T. One may think of T as a set of rules that the agents use to derive novel statements from given ones and from their observations. The assumption that the agents "use quantum mechanics" can thus be phrased in terms of a requirement on T.

For our purposes, it suffices to include in T a basic variant of the quantum-mechanical Born rule, which only talks about the case of predictions that have probability 1. Crucially, however, we take this rule to be universally valid. In particular, it shall be applicable to systems S that themselves contain other agents.

Assumption (Q)

A theory T that *satisfies Assumption* (Q) allows any agent A to reason as follows. Suppose that A has established the statements

$$s_I^{\mathsf{A}} = \text{“}S \textit{ is in state } |\psi\rangle \textit{ at time } t_0.\text{”}$$

$$s_M^{\mathsf{A}} = \text{“}x \textit{ is obtained by a measurement of } S \textit{ w.r.t. the family } \{\pi_x^H\} \textit{ of Heisenberg operators relative to time } t_0. \textit{ The measurement is completed at time } t.\text{”}$$

where S is an arbitrary system around A, $|\psi\rangle$ a unit vector of its Hilbert space, and $\{\pi_x^H\}_{x\in\mathcal{X}}$ a family of positive operators on this space such that $\sum_x \pi_x^H = \mathbf{1}$. If $\langle\psi|\pi_\xi^H|\psi\rangle = 1$ for some $\xi \in \mathcal{X}$ then A can conclude

$$s_Q^{\mathsf{A}} = \text{“}I \textit{ am certain that } x = \xi \textit{ at time } t.\text{”}$$

measuring agent	value	measured system	relevant vectors of measurement basis
$\bar{\mathsf{F}}$	r	R	$\lvert\mathsf{heads}\rangle_R \quad \lvert\mathsf{tails}\rangle_R$
F	z	S	$\lvert\downarrow\rangle_S \quad \lvert\uparrow\rangle_S$
$\bar{\mathsf{W}}$	$\bar{w}$	$\bar{L}$	$\lvert\bar{\mathsf{ok}}\rangle_{\bar{L}} = \sqrt{1/2}(\lvert\bar{\mathsf{h}}\rangle_{\bar{L}} - \lvert\bar{\mathsf{t}}\rangle_{\bar{L}})$
W	w	L	$\lvert\mathsf{ok}\rangle_L = \sqrt{1/2}(\lvert-\tfrac{1}{2}\rangle_L - \lvert+\tfrac{1}{2}\rangle_L)$

(a) Measurements

time interval within round n	time evolution in $\bar{\mathsf{F}}$'s lab	in F's lab
before n:00	set R to $\lvert\mathsf{init}\rangle_R = \sqrt{1/3}\lvert\mathsf{heads}\rangle_R + \sqrt{2/3}\lvert\mathsf{tails}\rangle_R$	[irrelevant]
from n:00 to n:10	$U^{00\to10}_{R\to\bar{L}S} = \begin{cases}\lvert\mathsf{heads}\rangle_R \mapsto \lvert\bar{\mathsf{h}}\rangle_{\bar{L}} \otimes \lvert\downarrow\rangle_S \\ \lvert\mathsf{tails}\rangle_R \mapsto \lvert\bar{\mathsf{t}}\rangle_{\bar{L}} \otimes \lvert\rightarrow\rangle_S\end{cases}$	[irrelevant]
from n:10 to n:20	$U^{10\to20}_{\bar{L}\to\bar{L}} = \mathbf{1}_{\bar{L}}$	$U^{10\to20}_{S\to L} = \begin{cases}\lvert\downarrow\rangle_S \mapsto \lvert-\tfrac{1}{2}\rangle_L \\ \lvert\uparrow\rangle_S \mapsto \lvert+\tfrac{1}{2}\rangle_L\end{cases}$
from n:20 to n:30	[irrelevant]	$U^{20\to30}_{L\to L} = \mathbf{1}_L$

(b) Time evolution

Table I. *Details of the experimental protocol.* The values observed by the individual agents together with their measurement bases are listed in (a). The basis vectors $\lvert\bar{\mathsf{ok}}\rangle_{\bar{L}}$ and $\lvert\mathsf{ok}\rangle_L$ are expressed in terms of particular states, such as $\lvert-\tfrac{1}{2}\rangle_L$ and $\lvert+\tfrac{1}{2}\rangle_L$, which are defined in (b). For them to be well defined, the assumption that $\bar{L}$ and L are isolated quantum systems is crucial. Technically, it means that their time evolution is described by norm-preserving linear maps, i.e., isometries. The second protocol step, for instance, induces an isometry $U^{10\to20}_{S\to L}$ from S to L. The vectors $\lvert-\tfrac{1}{2}\rangle_L$ and $\lvert+\tfrac{1}{2}\rangle_L$ are then defined by this isometry as the states of the lab L at the end of the protocol step, depending on whether the incoming spin was $\lvert\downarrow\rangle_S$ or $\lvert\uparrow\rangle_S$, respectively. One may therefore, for concreteness, think of them as states like those in (2) — although it is actually unnecessary to assume anything about their structure.

We will start off with statements s^{A}_I and s^{A}_M that directly describe the experimental protocol. Since the measurements are specified in terms of vectors (defined in Table I), the Heisenberg operators will be of the form $\pi^H_x = U^\dagger\lvert\phi_x\rangle\langle\phi_x\rvert U$, where $\lvert\phi_x\rangle$ is the vector belonging to outcome x, and U the isometry corresponding to the time evolution between time t_0 and the time when the measurement is carried out.

Specifically, agent $\bar{\mathsf{F}}$ may start her reasoning with the two statements

$s^{\bar{\mathsf{F}}}_I =$ *"If $r = \mathsf{tails}$ at time n:10 then spin S is in state $\lvert\rightarrow\rangle_S$ at time n:10."*

$s^{\bar{\mathsf{F}}}_M =$ *"The value w is obtained by a measurement of L w.r.t. $\{\pi^H_{\mathsf{ok}}, \pi^H_{\mathsf{fail}}\}$. The measurement is completed at time n:40."*

where

$$\pi^H_{w=\mathsf{ok}} = (U^{10\to30}_{S\to L})^\dagger\lvert\mathsf{ok}\rangle\langle\mathsf{ok}\rvert_L(U^{10\to30}_{S\to L}) \qquad \text{and} \qquad \pi^H_{w=\mathsf{fail}} = \mathbf{1} - \pi^H_{w=\mathsf{ok}}$$

are the Heisenberg operators with respect to time $t_0 = n$:10, and $U^{10\to30}_{S\to L} = U^{20\to30}_{L\to L}U^{10\to20}_{S\to L}$. Using now that $U^{10\to 30}_{S\to L}|{\to}\rangle_S = \sqrt{1/2}\big(|{-\tfrac{1}{2}}\rangle_L + |{+\tfrac{1}{2}}\rangle_L\big)$ is orthogonal to $|\mathsf{ok}\rangle_L$, we find

$$\langle{\to}|\pi^H_{w=\mathsf{fail}}|{\to}\rangle = 1 - \langle{\to}|\pi^H_{w=\mathsf{ok}}|{\to}\rangle = 1 \ .$$

Hence, if the theory T that agent $\bar{\mathsf{F}}$ uses for her reasoning satisfies (Q) then she can infer from $s^{\bar{\mathsf{F}}}_I$ and $s^{\bar{\mathsf{F}}}_M$ that statement $s^{\bar{\mathsf{F}}}_Q$ of Table II holds.

Similarly, agent F's reasoning may be based on the two statements

$s^{\mathsf{F}}_I =$ *"If $r = \mathsf{heads}$ at time n:10 then the spin S is in state $|{\downarrow}\rangle_S$ at time n:10."*

$s^{\mathsf{F}}_M =$ *"The value z is obtained by a measurement of S w.r.t. $\{\pi^H_{-\frac{1}{2}}, \pi^H_{+\frac{1}{2}}\}$, which is completed at time n:20."*

where

$$\pi^H_{z=-\frac{1}{2}} = |{\downarrow}\rangle\langle{\downarrow}|_S \qquad \text{and} \qquad \pi^H_{z=+\frac{1}{2}} = |{\uparrow}\rangle\langle{\uparrow}|_S$$

are the Heisenberg operators with respect to time $t_0 = n{:}10$. Because

$$\langle{\downarrow}|\pi_{z=-\frac{1}{2}}|{\downarrow}\rangle = 1$$

she can conclude that if $r = \mathsf{heads}$ then $z = -\frac{1}{2}$. This is logically equivalent to statement s^{F}_Q of Table II.

We proceed with agent $\bar{\mathsf{W}}$, for whom the statement

$s^{\bar{\mathsf{W}}}_I =$ *"System R is in state $|\mathsf{init}\rangle_R$ at time n:00."*

holds. Here we consider Heisenberg operators with respect to time $t_0 = n{:}00$. We are only interested in the event that $\bar{w} = \overline{\mathsf{ok}}$ and $z = -\frac{1}{2}$, as well as its complement, i.e.,

$$\pi^H_{(\bar{w},z)=(\overline{\mathsf{ok}},-\frac{1}{2})} = (U^{00\to 10}_{R\to\bar{L}S})^\dagger(|\overline{\mathsf{ok}}\rangle\langle\overline{\mathsf{ok}}| \otimes |{\downarrow}\rangle\langle{\downarrow}|)(U^{00\to 10}_{R\to\bar{L}S}) \quad \text{and} \quad \pi^H_{(\bar{w},z)\neq(\overline{\mathsf{ok}},-\frac{1}{2})} = \mathbf{1} - \pi^H_{(\bar{w},z)=(\overline{\mathsf{ok}},-\frac{1}{2})} \ ,$$

where we have already used that $U^{10\to 20}_{\bar{L}\to\bar{L}}$ is the identity. It is straightforward to verify that $U^{00\to 10}_{R\to\bar{L}S}|\mathsf{init}\rangle_R = \sqrt{1/3}|\bar{\mathsf{h}}\rangle_{\bar{L}} \otimes |{\downarrow}\rangle_S + \sqrt{2/3}|\bar{\mathsf{t}}\rangle_{\bar{L}} \otimes |{\to}\rangle_S$ is orthogonal to $|\overline{\mathsf{ok}}\rangle_{\bar{L}} \otimes |{\downarrow}\rangle_S$, which implies that

$$\langle\mathsf{init}|\pi^H_{(\bar{w},z)\neq(\overline{\mathsf{ok}},-\frac{1}{2})}|\mathsf{init}\rangle = 1 - \langle\mathsf{init}|\pi^H_{(\bar{w},z)=(\overline{\mathsf{ok}},-\frac{1}{2})}|\mathsf{init}\rangle = 1 \ .$$

Agent $\bar{\mathsf{W}}$, who also uses (Q), can hence be certain that $(\bar{w}, z) \neq (\overline{\mathsf{ok}}, -\frac{1}{2})$, which implies statement $s^{\bar{\mathsf{W}}}_Q$ of Table II.

Finally, agent W can make the same statement $s^{\bar{\mathsf{W}}}_I$ as agent $\bar{\mathsf{W}}$. An analogous calculation as above shows that

$$\langle\mathsf{init}|\pi^H_{(\bar{w},w)=(\overline{\mathsf{ok}},\mathsf{ok})}|\mathsf{init}\rangle = 1/12 \tag{4}$$

where

$$\pi^H_{(\bar{w},w)=(\overline{\mathsf{ok}},\mathsf{ok})} = (U^{00\to 10}_{R\to\bar{L}S})^\dagger(\mathbf{1}_{\bar{L}} \otimes U^{10\to 30}_{S\to L})^\dagger(|\overline{\mathsf{ok}}\rangle\langle\overline{\mathsf{ok}}| \otimes |\mathsf{ok}\rangle\langle\mathsf{ok}|)(\mathbf{1}_{\bar{L}} \otimes U^{10\to 30}_{S\to L})(U^{00\to 10}_{R\to\bar{L}S})$$

is the Heisenberg operator belonging to the event that $\bar{w} = \overline{\mathsf{ok}}$ and $w = \mathsf{ok}$. Hence, according to quantum mechanics, agent W can be certain that the outcome $(\bar{w}, w) = (\overline{\mathsf{ok}}, \mathsf{ok})$ occurs after finitely many rounds. This is statement s^{W}_Q of Table II. (That the statement can indeed be derived using (Q) is shown in the appendix.)

To obtain further statements, the agents may use their theory T to reason about how they would reason from the viewpoint of another agent. To enable such nested reasoning we need to introduce

agent	observes	statement inferred via (Q)
$\bar{\mathsf{F}}$	r	$s_Q^{\bar{\mathsf{F}}} =$ *"If $r = \mathsf{tails}$ at n:10 then I am certain that W will observe $w = \mathsf{fail}$ at n:40."*
F	z	$s_Q^{\mathsf{F}} =$ *"If $z = +\frac{1}{2}$ at n:20 then I am certain that $\bar{\mathsf{F}}$ observed $r = \mathsf{tails}$ at n:10."*
$\bar{\mathsf{W}}$	$\bar{w}$	$s_Q^{\bar{\mathsf{W}}} =$ *"If $\bar{w} = \overline{\mathsf{ok}}$ at n:30 then I am certain that F observed $z = +\frac{1}{2}$ at n:20."*
W	w	$s_Q^{\mathsf{W}} =$ *"I am certain that there exists a round $n \in \mathbb{N}_{\geq 0}$ in which it is announced that $\bar{w} = \overline{\mathsf{ok}}$ at n:30 and $w = \mathsf{ok}$ at n:40."*

Table II. *The agents' observations and conclusions.* The statements that the individual agents can derive from quantum theory depend on the information accessible to them (cf. Fig. 3). Agent $\bar{\mathsf{F}}$, for instance, observes r, and uses this information to infer the value w, which will later be observed by W.

another assumption.

Assumption (C)

A theory T that *satisfies Assumption* (C) allows any agent A to reason as follows. If A has established

$s_1^{\mathsf{A}} =$ *"I am certain at time t_0 that agent A', upon reasoning using T, is certain that $x = \xi$ at time t."*

for some x, ξ, and $t > t_0$, then A can conclude

$s_2^{\mathsf{A}} =$ *"I am certain at time t_0 that $x = \xi$ at time t."*

Agent F may insert $\bar{\mathsf{F}}$'s statement $s_Q^{\bar{\mathsf{F}}}$ into s_Q^{F}, obtaining

$s_1^{\mathsf{F}} =$ *"If $z = +\frac{1}{2}$ at time n:20 then I am certain that $\bar{\mathsf{F}}$ is certain that W will observe $w = \mathsf{fail}$ at time n:40."*

If the theory T satisfies (C) then F can conclude from this that

$s_2^{\mathsf{F}} =$ *"If $z = +\frac{1}{2}$ at time n:20 then I am certain that W will observe $w = \mathsf{fail}$ at time n:40."*

Similarly, $\bar{\mathsf{W}}$ may combine this statement with his statement $s_Q^{\bar{\mathsf{W}}}$ to obtain

$s_1^{\bar{\mathsf{W}}} =$ *"If $\bar{w} = \overline{\mathsf{ok}}$ at time n:30 then I am certain that F is certain that W will observe $w = \mathsf{fail}$ at time n:40."*

$\bar{\mathsf{W}}$ could now, by virtue of (C), conclude that

$s_2^{\bar{\mathsf{W}}} =$ *"If $\bar{w} = \overline{\mathsf{ok}}$ at time n:30 then I am certain that W will observe $w = \mathsf{fail}$ at time n:40."*

Agent W, who hears the announcement of $\bar{w}$ by $\bar{\mathsf{W}}$, may turn the above into the following statement.

$s_1^{\mathsf{W}} =$ *"If $\bar{w} = \overline{\mathsf{ok}}$ at time n:30 then I am certain that $\bar{\mathsf{W}}$ is certain that I will observe $w = \mathsf{fail}$ at time n:40."*

	(Q)	(S)	(C)
Copenhagen	✓	✓	×
HV theory applied to subsystems	✓	✓	×
HV theory applied to entire universe	×	✓	✓
Many-worlds	?	×	?
Collapse theories	×	✓	✓
Consistent histories	✓	✓	×
QBism	✓	✓	×
Relational quantum mechanics	✓	✓	×
CSM approach	×	✓	✓
ETH approach	×	✓	✓

Table III. *Interpretations of quantum theory.* Theorem 1 can be applied to different interpretations of quantum theory. Each of them must violate at least one of the assumptions (indicated by ×).

Using (C), W can now as well conclude

$$s_2^{\mathsf{W}} = \text{“}\textit{If } \bar{w} = \overline{\mathsf{ok}} \textit{ at time n:30 then I am certain that I will observe } w = \mathsf{fail} \textit{ at time n:40.”}$$

To finalise our analysis, we need one more assumption, which captures the intuition that measurements have single outcomes for any agent.

Assumption (S)

A theory T *satisfies Assumption* (S) if it disallows any agent A, for whom the statement

$$s^{\mathsf{A}} = \text{“}\textit{I am certain at time } t_0 \textit{ that } x = \xi \textit{ at time } t.\text{”}$$

for some x, ξ, and $t > t_0$ is correct, to also make the statement

$$\bar{s}^{\mathsf{A}} = \text{“}\textit{I am certain at time } t_0 \textit{ that } x \neq \xi \textit{ at time } t.\text{”}$$

Under this assumption, s_2^{W} and s_Q^{W} cannot both be valid, i.e., we have arrived at a contradiction.

C. No-go theorem

The conclusion of the above analysis may be formulated as a no-go theorem.

Theorem 1. *The gedankenexperiment of Section II A cannot be consistently described by any theory T that satisfies Assumptions* (Q)*,* (C)*, and* (S)*.*

To illustrate the theorem, we consider in the following different interpretations and modifications of quantum mechanics. Each of them can be regarded as a theory T, and hence must violate either (Q), (C), or (S). This yields a natural categorisation as shown in Table III. In principle, an interpretation may also evade the conclusions of the theorem by not fitting into the framework used here — a possibility which we examine at the end of this section.

Theories that violate (Q). Assumption (Q) corresponds to the quantum-mechanical Born rule. Since the assumption is concerned with the special case of probability-1 predictions only, it is largely independent of how one interprets probabilities. However, the non-trivial aspect of (Q) is that it regards the Born rule as a universal law. That is, it demands that an agent A can apply the rule to arbitrary systems S around her, including large ones that may contain other agents. The specifier "around" is crucial, though: (Q) does not demand that the agent A describes herself as a quantum system. Such a requirement would indeed be overly restrictive (see [14]) for it would immediately rule out interpretations in the spirit of Copenhagen, according to which the observed quantum system and the observer must be distinct from each other [15, 16].

Assumption (Q) is manifestly violated by theories that postulate a modification of the quantum formalism, such as spontaneous [17–22] and gravity-induced [23–25] collapse models (cf. [26] for a review). These deviate from standard quantum theory already on microscopic scales, although the effects of the deviation typically only become noticeable in larger systems.

In some approaches to quantum mechanics, it is simply *postulated* that large systems are "classical", but the physical mechanism that explains the absence of quantum features remains unspecified [27]. In the view described in [3], for instance, the postulate says that measurement devices are infinite-dimensional systems whereas observables are finite. This ensures that coherent and incoherent superpositions in the state of a measurement device are indistinguishable. Similarly, according to the *ETH approach* [28], the algebra of available observables is time-dependent and does not allow one to distinguish coherent from incoherent superpositions once a measurement has been completed. General measurements on systems that count themselves as measurement devices are thus ruled out. Another example is the *CSM ontology* [29], according to which measurements must always be carried out in a *context*, which includes the measurement devices. It is then postulated that this context cannot itself be treated as a quantum system. Within all these interpretations, the Born rule still holds "for all practical purposes", but is no longer a universally applicable law in the sense of Assumption (Q) (see the discussion in [5]).

Another class of theories that violate (Q), although in a less obvious manner, are particular *hidden-variable (HV) interpretations* [30], with *Bohmian mechanics* as the most prominent example [31–33]. According to the common understanding, Bohmian mechanics is a "theory of the universe" rather than a theory about subsystems [34]. This means that agents who apply the theory must in principle always take an outside perspective on the entire universe, describing themselves as part of it. This outside perspective is identical for all agents, which ensures consistency and hence the validity of Assumption (C). However because (S) is satisfied, too, it follows from Theorem 1 that (Q) must be violated (see the appendix for more details).

Theories that violate (C). If a theory satisfies (Q) and (S) then, by Theorem 1, it must violate (C). This conclusion applies to a wide range of common readings of quantum mechanics, including most variants of the Copenhagen interpretation. One concrete example is the *consistent histories (CH)* framework [35–38], which is also similar to the *decoherent histories* approach [39, 40]. Another class of examples are subjectivistic interpretations, which regard statements about outcomes of measurements as personal to an agent, such as *relational quantum mechanics* [41], *QBism* [42, 43], or the approach proposed in [12] (see the appendix for a discussion of the CH framework as well as QBism).

The same conclusion applies to hidden-variable (HV) interpretations of quantum mechanics, provided that we use them to describe systems around us rather than the universe as a whole (contrasting the paradigm of Bohmian mechanics discussed above). In this case, both (Q) and (S) hold by construction. This adds another item to the long list of no-go results for HV interpretations: they cannot be local [13], they must be contextual [44, 45], and they violate freedom of choice [46, 47]. Theorem 1 entails that they also violate (C). In particular, there cannot exist an assignment of values to the hidden variables that is consistent with the agents' conclusions.

Theories that violate (S). Although intuitive, (S) is not implied by the bare mathematical formalism of quantum mechanics. Among the theories that abandon the assumption are the *relative state formulation* and the *many-worlds interpretation* [9, 48–52]. According to the latter, any quantum measurement results in a branching into different "worlds", in each of which one of the possible measurement outcomes occurs. Further developments and variations include the *many-minds interpretation* [53, 54] and the *parallel lives theory* [55]. A related concept is *quantum Darwinism* [56], whose purpose is to explain the perception of classical measurement outcomes in a unitarily evolving universe.

While many-worlds interpretations manifestly violate (S), their compatibility with (Q) and (C) depends on how one defines the branching. If one regards it as an objective process, (Q) may be violated (cf. the example in Sec. 10 of [57]). It is also questionable whether (Q) can be upheld if branches do not persist over time (cf. the no-histories view described in [58]).

Implicit assumptions. The formalism we used to analyse the gedankenexperiment is rather minimalistic. In particular, neither of the assumptions (Q), (C), or (S) refers to the notion of probabilities — although (Q) is of course motivated by the idea that a statement can be regarded as "certain" if it has probability 1 according to the standard formalism. Theorem 1 is therefore largely independent of how probabilities are interpreted.

Since the formalism does not rely on a statistical model, it also avoids the assumption that measurement outcomes obtained by different agents simultaneously have definite values. (For example, considering the original Wigner's friend experiment, (C) and (S) do not enforce that W assigns a definite value to F's outcome z.) Such simultaneous definiteness assumptions are otherwise rather common. They not only enter the proof of Bell's theorem [13] but also various modern arguments that employ probabilistic frameworks [59–67], where measurement outcomes are modelled as random variables belonging to a single random experiment.

Nevertheless, in our considerations above, we used concepts such as that of an "agent" or of "time". It is conceivable that the conclusions of Theorem 1 can be avoided by theories that provide a non-standard understanding of these concepts. We are however not aware of any concrete examples of such theories.

III. DISCUSSION

Current interpretations of quantum theory do not agree on the origin of the contradiction that arises in our analysis of the gedankenexperiment (cf. Table III). To compare the different views, it may therefore be useful to rephrase the experiment as a concrete game-theoretic decision problem.

Suppose that a casino offers the following gambling game. One round of the experiment is played, with the gambler in the role of W, and the roles of $\bar{\mathsf{F}}$, F, and $\bar{\mathsf{W}}$ taken by employees of the casino. The casino promises to pay $ 1.000 to the gambler if $\bar{\mathsf{F}}$'s random value was $r = \mathsf{heads}$. Conversely, if $r = \mathsf{tails}$, the gambler must pay $ 500 to the casino. It could now happen that, at the end of the game, $w = \mathsf{ok}$ and $\bar{w} = \overline{\mathsf{ok}}$, and that a judge can convince herself of this outcome. The gambler and the casino are then likely to end up in a dispute, putting forward arguments taken from Table II.

Gambler: "The outcome $w = \mathsf{ok}$ implies, due to $s_Q^{\bar{\mathsf{F}}}$, that $r = \mathsf{heads}$, so the casino must pay me $ 1.000."

Casino: "The outcome $\bar{w} = \overline{\mathsf{ok}}$ proves, by virtue of $s_Q^{\bar{\mathsf{W}}}$, that our employee observed $z = +\frac{1}{2}$. This in turn proves, by virtue of s_Q^{F}, that $r = \mathsf{tails}$, so the gambler must pay us $ 500."

How should the judge decide on this case? Could it even be that both assertions must be accepted as two "alternative facts" about what the value r was? We leave it as a task for further research to explore what the different interpretations of quantum mechanics have to say about this game.

Theorem 1 may be compared to earlier no-go results, such as [10–13, 44–46]. Most of them also use assumptions similar to (Q) and (S), although the latter is often only implicit. These are then shown to be in conflict with assumptions about reality, locality, and freedom of choice. (For example, in [12], it is shown that no theory can fulfil all of the following properties: (i) be compatible with quantum theory on all scales, (ii) simultaneously assign definite truth values to measurement outcomes of all agents, (iii) allow agents to freely choose measurement settings, and (iv) be local.) Theorem 1 now asserts that (Q) and (S) are already in conflict with the idea that agents can consistently reason about each other, in the sense of (C).

We also note that the argument presented here does not involve counterfactual reasoning. This is in contrast to many no-go results in quantum mechanics, which involve reasoning about choices that *could* have been made but have not actually been made. In fact, in the proposed experiment, the agents never make any choices (also no delayed ones, as e.g., in Wheeler's "delayed choice" experiment [68]).

We conclude by suggesting a modified variant of the experiment, which may be technologically feasible. The idea is to substitute agents $\bar{\mathsf{F}}$ and F by computers. While it is debatable whether computers can reasonably count as agents, one may argue that they could certainly carry out the tasks prescribed in the experimental protocol, and draw conclusions such as *"I am certain that* W *will observe* $w = \mathsf{fail}$ *at n:40."* To ensure that $\bar{\mathsf{F}}$ and F's lab are isolated, one could use quantum computers, which by construction do not leak information to their environment. Such an experiment may serve as a test for the statements in Table II, provided that one is ready to make some mild additional assumptions. For example, aborting the experiment at time n:20, just before $\bar{\mathsf{W}}$ starts her measurement, one could read out the values r and z. Assuming that aborting the experiment does not alter these values, and that the value r at time n:10 is still the same as at time n:10, this would be a test for statement s_Q^{F}. In this sense, quantum computers, motivated usually by applications in computing, may help us answering questions in fundamental research.

ACKNOWLEDGMENTS

We would like to thank Yakir Aharonov, Mateus Araújo, Alexia Auffèves, Jonathan Barrett, Veronika Baumann, Serguei Beloussov, Hans Briegel, Časlav Brukner, Adán Cabello, Giulio Chiribella, Roger Colbeck, Patricia Contreras Tejada, Giacomo Mauro D'Ariano, John DeBrota, Lídia del Rio, David Deutsch, Artur Ekert, Michael Esfeld, Philippe Faist, Aaron Fenyes, Hugo Fierz, Jürg Fröhlich, Christopher Fuchs, Shan Gao, Svenja Gerhard, Edward Gillis, Nicolas Gisin, Sheldon Goldstein, Sabrina Gonzalez Pasterski, Gian Michele Graf, Philippe Grangier, Bob Griffiths, Arne Hansen, Lucien Hardy, Aram Harrow, Klaus Hepp, Paweł Horodecki, Angela Karanjai, Adrian Kent, Gijs Leegwater, Matthew Leifer, Seth Lloyd, Markus Müller, Thomas Müller, Hrvoje Nikolić, Travis Norsen, Nuriya Nurgalieva, Jonathan Oppenheim, Sandu Popescu, Gilles Pütz, Joseph Renes, Jess Riedel, Rüdiger Schack, Robert Spekkens, Cristi Stoica, Tony Sudbery, Stefan Teufel, Roderich Tumulka, Lev Vaidman, Vlatko Vedral, Mordecai Waegell, Stefan Wolf, Filip Wudarski, Christa Zoufal, and Wojciech Zurek for comments and discussions.

This project was supported by the Swiss National Science Foundation (SNSF) via the National Centre of Competence in Research "QSIT", by the Kavli Institute for Theoretical Physics (KITP) at the University of California in Santa Barbara, by the Stellenbosch Institute for Advanced Study (STIAS) in South Africa, by the US National Science Foundation (NSF) under grant No. PHY17-48958, by the European Research Council (ERC) under grant No. 258932, and by the European Commission under the project "RAQUEL".

APPENDIX

Information-theoretic description

The experimental protocol may be represented in terms of a circuit diagram, Fig. 4, where the actions of the agents correspond to isometries acting on certain subsystems. The diagram emphasises the information-theoretic aspects of the experiment. While all agents have full information about the overall evolution (the circuit diagram itself), they carry out separate measurements, and therefore have access to different data (corresponding to different wires in the diagram). In a measurement, e.g., that of the electron spin S, the outcome z is recorded in a memory register, F, held by agent F.

Derivation of statement s_Q^{W} using Assumption (Q)

In Section II B we argued that the event $(\bar{w}, w) = (\overline{\mathsf{ok}}, \mathsf{ok})$ must occur after finitely many rounds n, which is statement s_Q^{W} of Table II. While this is a pretty obvious consequence of the Born rule, we now show that it also follows from Assumption (Q), which corresponds to the special case where the Born rule gives probability-1 predictions.

We consider Heisenberg operators with respect to a time t_0 right before the experiment starts. For any round n, let W^n be the isometry from $\mathbb{C}$ to $\bar{L} \otimes L$ that includes the initialisation of system R in state $|\mathsf{init}\rangle_R$ as well as $U^{00\to10}_{R\to\bar{L}S}$ and $U^{10\to20}_{S\to L}$ (cf. Table I), i.e.,

$$W^n_{\mathbb{C}\to\bar{L}L} = (\mathbf{1}_{\bar{L}} \otimes U^{10\to20}_{S\to L}) U^{00\to10}_{R\to\bar{L}S} |\mathsf{init}\rangle_R \ .$$

The Heisenberg operator of the event $(\bar{w}, w) = (\overline{\mathsf{ok}}, \mathsf{ok})$ in round n can thus be written as

$$\pi^n_{\overline{\mathsf{ok}},\mathsf{ok}} = (W^n_{\mathbb{C}\to\bar{L}L})^\dagger (|\overline{\mathsf{ok}}\rangle\langle\overline{\mathsf{ok}}| \otimes |\mathsf{ok}\rangle\langle\mathsf{ok}|)(W^n_{\mathbb{C}\to\bar{L}L}) \ .$$

We may now specify a Heisenberg operator π^H_{halt} for the halting condition, i.e., that the event $(\bar{w}, w) = (\overline{\mathsf{ok}}, \mathsf{ok})$ occurs in *some* round n,

$$\pi^H_{\mathrm{halt}} = \sum_{n=0}^{\infty} \pi^n_{(\overline{\mathsf{ok}},\mathsf{ok})} \prod_{m=0}^{n-1} \left(\mathbf{1} - \pi^m_{(\overline{\mathsf{ok}},\mathsf{ok})}\right) \ .$$

Note that these are operators on $\mathbb{C}$, i.e., $\pi^n_{\overline{\mathsf{ok}},\mathsf{ok}} = p\mathbf{1}$ for some $p \in \mathbb{C}$. It follows directly from (4) that $p = 1/12 > 0$. This yields

$$\pi^H_{\mathrm{halt}} = \mathbf{1} \sum_{n=0}^{\infty} p(1-p)^n = \mathbf{1} \ .$$

One may then apply Assumption (Q) to conclude that statement s_Q^{W} of Table II holds.

Analysis within Bohmian mechanics

According to Bohmian mechanics, the state of a system of particles consists of their quantum-mechanical wave function together with an additional set of variables that specify the particles' spatial positions. While the wave function evolves according to the Schrödinger equation, the time evolution of the additional position variables is governed by another equation of motion, sometimes referred to as the *guiding equation.* The general understanding is that these equations of motion must always be applied to the universe as a whole. As noted in [34], "if we postulate that subsys-

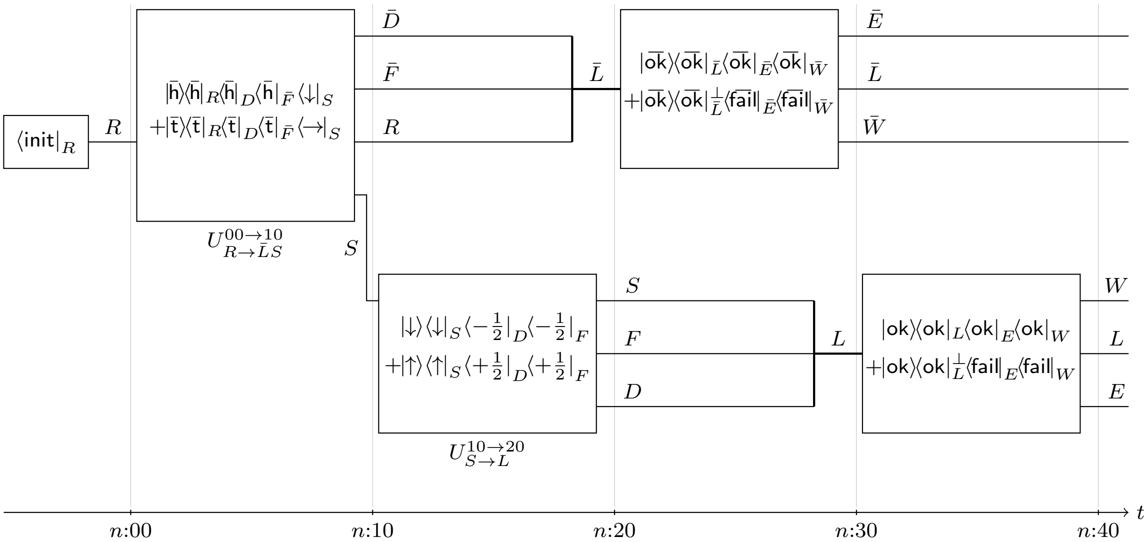

Figure 4. *Circuit diagram representation of the gedankenexperiment.* The actions of the agents during the protocol induce isometries (boxes) that act on particular subsystems (wires). For example, the measurement of S by agent F in the second protocol step, which starts at time n:10, can be regarded as an isometry $U^{00\to 10}_{S\to L}$ of the form (2) from S to F's lab, L. The subsystems labelled by $\bar{F}$, F, $\bar{W}$, and W contain the agents themselves, i.e., $\bar{\mathsf{F}}$, F, $\bar{\mathsf{W}}$, and W, respectively. Similarly, $\bar{D}$, D, $\bar{E}$, and E are "environment" subsystems, which include the agents' measurement devices and everything connected to them.

tems [rather than the universe] must obey Bohmian mechanics, we 'commit redundancy and risk inconsistency.'"

The gedankenexperiment presented in this work shows that this risk is real. Indeed, if the agents applied the Bohmian equations of motion directly to the relevant systems *around* them, rather than to the universe as a whole, their reasoning would be the same as the one prescribed by (Q). Consequently, $s_Q^{\bar{\mathsf{F}}}$, s_Q^{F}, $s_Q^{\bar{\mathsf{W}}}$, and s_Q^{W} of Table II would all hold. But since Bohmian mechanics also satisfies (S), this would imply a violation of (C). (This finding should not be confused with the known fact that, if the spatial position of a particle is measured, the Bohmian position of the measurement device's pointer is sometimes incompatible with the Bohmian position of the measured particle [69–74].)

The directive of [34] that Bohmian mechanics should be applied to the entire universe means that the agents must model themselves from an outside perspective. This ensures that they all have the same view, so that reasoning according to (C) is unproblematic. But then, because of Theorem 1, (Q) is necessarily violated. This is indeed confirmed by an explicit calculation in Bohmian mechanics, which reveals that $s_Q^{\bar{\mathsf{F}}}$ is invalid. Furthermore, and in contrast to standard quantum mechanics, the time order of the measurements carried out by $\bar{\mathsf{W}}$ and W is relevant. If the latter measured first then, according to Bohmian mechanics, statement $s_Q^{\bar{\mathsf{W}}}$ would be invalid instead.

This contradiction to standard quantum mechanics raises the question under what circumstances reasoning according to (Q) is still allowed in Bohmian mechanics. A candidate criterion could be that the memory of the system's initial state should still be available at the time when the measurement is completed, so that the prediction can be verified. The validity of agent $\bar{\mathsf{F}}$'s statement $s_Q^{\bar{\mathsf{F}}}$, for instance, could then be denied on the grounds that $\bar{\mathsf{F}}$ is herself subject to a measurement, which may destroy her memory of what spin state she prepared. This argument does however not work. The reason is that, in the relevant case when $\bar{w} = \bar{\mathsf{ok}}$, the value of r is retrievable by virtue of statements $s_Q^{\bar{\mathsf{W}}}$ and s_Q^{F}, which means that full information about the spin direction chosen by $\bar{\mathsf{F}}$ is still available at the time when w is observed.

Analysis within the CH framework

In the consistent histories (CH) framework, statements about measurement outcomes are phrased in terms of *histories*. These must, by definition, be elements of a whole family of histories, called a *framework*, that satisfies certain consistency conditions. In the gedankenexperiment proposed in this work, a possible history would be

$s_1 =$ *"Outcomes* $r = \mathsf{tails}$, $z = +\frac{1}{2}$, $\bar{w} = \overline{\mathsf{ok}}$, *and* $w = \mathsf{ok}$ *were observed (at times 0:10, 0:20, 0:30, and 0:40, respectively)."*

To verify that s_1 is indeed a valid history, one has to construct a family containing this history. It is straightforward to check that one such family is the set consisting of s_1 together with the additional histories

$s_2 =$ *"Outcomes* $r = \mathsf{tails}$, $z = +\frac{1}{2}$, $\bar{w} = \overline{\mathsf{ok}}$, *and* $w = \mathsf{fail}$ *were observed (at the respective times)."*

$s_3 =$ *"Outcomes* $r = \mathsf{heads}$, $z = +\frac{1}{2}$, *and* $\bar{w} = \overline{\mathsf{ok}}$ *were observed (at at the respective times)."*

$s_4 =$ *"Outcomes* $z = -\frac{1}{2}$ *and* $\bar{w} = \overline{\mathsf{ok}}$ *were observed (at at the respective times)."*

$s_5 =$ *"Outcome* $\bar{w} = \overline{\mathsf{fail}}$ *was observed at (the respective time)."*

The CH framework contains the Born rule as a special case and hence fulfils Assumption (Q). Since it also satisfies (S), it follows from Theorem 1 that it violates (C). To illustrate how this violation manifests itself, we may consider a shortened version of history s_1, which leaves the values of z and $\bar{w}$ unmentioned:

$s'_1 =$ *"Outcomes* $r = \mathsf{tails}$ *and* $w = \mathsf{ok}$ *were observed (at times 0:00 and 0:40, respectively)."*

The CH formalism provides a rule to assign probabilities to these histories, which turn out to be

$$\Pr[s_1] = 1/12 \quad \text{and} \quad \Pr[s'_1] = 0 \; . \tag{5}$$

This is in disagreement with the fact that s'_1 is just a part of history s_1, i.e., $s_1 \implies s'_1$.

The CH formalism accounts for this disagreement by imposing the rule that logical reasoning must be constrained to histories that belong to a single framework, which is not the case for s_1 and s'_1. Comparing them thus amounts to violating this rule. (This may be compared to the "three box paradox" [75], where calculations in three different consistent frameworks yield mutually incompatible probability assignments; see Sec. 22 of [37] as well as [76] for a discussion.)

Nevertheless, the gedankenexperiment exhibits an ambiguity of the CH formalism when it comes to decision problems, such as the one of the casino example described in Section III. Within a framework that contains history s'_1, the gambler's reasoning is correct, for $\Pr[s'_1] = 0$. Conversely, considering the framework that contains history s_1, it is readily verified that the other histories, s_2, s_3, s_4, and s_5, have probabilities $\frac{1}{12}$, 0, 0, and $\frac{5}{6}$, respectively. All non-zero probability histories of this framework that agree with the observation $\bar{w} = \overline{\mathsf{ok}}$ thus assert that $r = \mathsf{tails}$, in agreement with the casino's argument. The answer to the question whether the casino has to pay the gambler or vice versa hence depends on the choice of the framework. According to the CH approach, this choice is up to the physicist who employs the formalism.

Analysis within QBism

QBism is one of the most far-reaching subjectivistic interpretations of quantum mechanics. It regards quantum states as representations of an agent's personal knowledge, or rather beliefs, about the outcomes of future measurements, and it also views these outcomes as personal to the agent.

The analysis of the gedankenexperiment based on Assumption (Q) is compatible with QBism, provided that the derived statements are interpreted appropriately. Since the experimental setup is fixed, we can assume that the agents all start with the same prior knowledge. However, during the run of the experiment, they update their knowledge based on observations, which are different for the different agents. To emphasise this, we suppose that they write their observations and conclusions into a personal notebook. Agent F, for instance, could then talk about $\bar{\mathsf{F}}$'s notebook entry and say

$s_Q^{\mathsf{F}} =$ *"If I observe $z = +\frac{1}{2}$ at time n:20 then I am certain that, if I checked $\bar{\mathsf{F}}$'s notebook at time n:20, I would read that she observed $r = \mathsf{tails}$ at time n:10."*

which is analogous to the corresponding statement in Table II. Here the phrase "is certain that" expresses a degree of belief and may also be replaced by something like "would bet an arbitrarily large amount on". Similarly, F's statements s_1^{F} and s_2^{F} would read

$s_1^{\mathsf{F}} =$ *"If I observe $z = +\frac{1}{2}$ at time n:20 then I am certain that, if I checked $\bar{\mathsf{F}}$'s notebook at time n:20, I would read that she is certain that W will announce $w = \mathsf{fail}$ at time n:40."*

$s_2^{\mathsf{F}} =$ *"If I observe $z = +\frac{1}{2}$ at time n:20 then I am certain that W will announce $w = \mathsf{fail}$ at time n:40."*

Since F and $\bar{\mathsf{F}}$ started with identical prior knowledge, one would now expect that the implication

$$s_1^{\mathsf{F}} \implies s_2^{\mathsf{F}}$$

is valid, which corresponds to Assumption (C).

Nonetheless, demanding that (C) holds is necessarily incompatible with QBism. The reason is, once again, Theorem 1, combined with the fact that QBism satisfies (Q) and (S). Yet it seems reasonable to demand that a subjectivistic theory should provide rules allowing agents to reason about other agents. This raises the question whether Assumption (C) could be substituted by another rule that fits into the framework of QBism. We refer to [77] for a discussion of this question.

---

[1] E. Schrödinger, "Die gegenwärtige Situation in der Quantenmechanik," Naturwissenschaften **23**, 823–828 (1935).
[2] E.P. Wigner, "Remarks on the mind-body question," in *Symmetries and Reflections* (Indiana University Press, 1967) pp. 171–184.
[3] K. Hepp, "Quantum theory of measurement and macroscopic observables," Helv. Phys. Acta **45**, 237–248 (1972).
[4] Englert, B.-G., "On quantum theory," Eur. Phys. J. D **67**, 238 (2013).
[5] J.S. Bell, "On wave packet reduction in the Coleman-Hepp model," Helv. Phys. Acta **48**, 93–98 (1975).
[6] C.A. Fuchs, "QBism, the perimeter of quantum Bayesianism," arXiv:1003.5209 (2010).
[7] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi, *Reasoning About Knowledge* (MIT press, 2004).
[8] C. H. Bennett, "The thermodynamics of computation—a review," Int. J. Theor. Phys. **21**, 905–940 (1982).
[9] D. Deutsch, "Quantum theory as a universal physical theory," Int. J. Theor. Phys. **24**, 1–41 (1985).
[10] L. Hardy, "Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories," Phys. Rev. Lett. **68**, 2981–2984 (1992).
[11] L. Hardy, "Nonlocality for two particles without inequalities for almost all entangled states," Phys. Rev. Lett. **71**, 1665–1668 (1993).
[12] Č. Brukner, "On the quantum measurement problem," in *Quantum [Un]Speakables II: Half a Century of Bell's Theorem*, edited by R. Bertlmann and A. Zeilinger (Springer, 2017) pp. 95–117.
[13] J.S. Bell, "On the problem of hidden variables in quantum mechanics," Rev. Mod. Phys. **38**, 447–452 (1966).
[14] M.L. Dalla Chiara, "Logical self reference, set theoretical paradoxes and the measurement problem in quantum mechanics," J. Philos. Logic **6**, 331–347 (1977).
[15] W. Heisenberg, "Ist eine deterministische Ergänzung der Quantenmechanik möglich?" in *Wolfgang*

Pauli. Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg, u.a., Vol. II, edited by A. Hermann, K. von Meyenn, and V.F. Weisskopf (Springer, 1935) pp. 409–418.
[16] N. Bohr, "Discussions with Einstein on epistemological problems in atomic physics," in *Albert Einstein: Philosopher-Scientist*, edited by P.A. Schilpp (Cambridge University Press, 1949) pp. 200–241.
[17] G.C. Ghirardi, A. Rimini, and T. Weber, "Unified dynamics for microscopic and macroscopic systems," Phys. Rev. D **34**, 470–491 (1986).
[18] N. Gisin, "Stochastic quantum dynamics and relativity," Helv. Phys. Acta **62**, 363–371 (1989).
[19] P. Pearle, "Combining stochastic dynamical state-vector reduction with spontaneous localization," Phys. Rev. A **39**, 2277–2289 (1989).
[20] N. Gisin and M. Rigo, "Relevant and irrelevant nonlinear Schrödinger equations," J. Phys. A **28**, 7375 (1995).
[21] R. Tumulka, "On spontaneous wave function collapse and quantum field theory," in *Proc. R. Soc. A*, Vol. 462 (The Royal Society, 2006) pp. 1897–1908.
[22] S. Weinberg, "Collapse of the state vector," Phys. Rev. A **85**, 062116 (2012).
[23] F. Karolyhazy, "Gravitation and quantum mechanics of macroscopic objects," Il Nuovo Cimento A (1971-1996) **42**, 390–402 (1966).
[24] L. Diósi, "Models for universal reduction of macroscopic quantum fluctuations," Phys. Rev. A **40**, 1165–1174 (1989).
[25] R. Penrose, "On gravity's role in quantum state reduction," Gen. Rel. Gravit. **28**, 581–600 (1996).
[26] A. Bassi, K. Lochan, S. Satin, T.P. Singh, and H. Ulbricht, "Models of wave-function collapse, underlying theories, and experimental tests," Rev. Mod. Phys. **85**, 471–527 (2013).
[27] L.D. Landau and E.M. Lifshitz, "Quantum mechanics: Non-relativistic theory," (Elsevier, 2013) Chap. The basic concepts of quantum mechanics.
[28] J. Fröhlich and B. Schubnel, "Quantum probability theory and the foundations of quantum mechanics," in *The Message of Quantum Science: Attempts Towards a Synthesis*, edited by P. Blanchard and J. Fröhlich (Springer, 2015) pp. 131–193.
[29] A. Auffèves and P. Grangier, "Contexts, systems and modalities: A new ontology for quantum mechanics," Found. Phys. **46**, 121–137 (2015).
[30] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer-Verlag, 1932).
[31] L. De Broglie, "La mécanique ondulatoire et la structure atomique de la matière et du rayonnement," J. Phys. Radium **8**, 225–241 (1927).
[32] D. Bohm, "A suggested interpretation of the quantum theory in terms of "hidden" variables. I," Phys. Rev. **85**, 166–179 (1952).
[33] D. Dürr and S. Teufel, *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory* (Springer, 2009).
[34] D. Dürr, S. Goldstein, and N. Zanghí, "Quantum equilibrium and the origin of absolute uncertainty," J. Stat. Phys. **67**, 843–907 (1992).
[35] R.B. Griffiths, "Consistent histories and the interpretation of quantum mechanics," J. Stat. Phys. **36**, 219–272 (1984).
[36] R. Omnès, "Consistent interpretations of quantum mechanics," Rev. Mod. Phys. **64**, 339–382 (1992).
[37] R.B. Griffiths, *Consistent Quantum Theory* (Cambridge University Press, 2002).
[38] R.B. Griffiths, "The consistent histories approach to quantum mechanics," in *The Stanford Encyclopedia of Philosophy* (Stanford University, 2014).
[39] M. Gell-Mann and J.B. Hartle, "Quantum mechanics in the light of quantum cosmology," in *Complexity, Entropy and the Physics of Information*, edited by W. Zurek (Addison Wesley, 1990).
[40] J.B. Hartle, "The quasiclassical realms of this quantum universe," Found. Phys. **41**, 982–1006 (2011).
[41] C. Rovelli, "Relational quantum mechanics," Int. J. of Theor. Phys. **35**, 1637–1678 (1996).
[42] C.A. Fuchs and R. Schack, "Quantum-Bayesian coherence," Rev. Mod. Phys. **85**, 1693–1715 (2013).
[43] C.A. Fuchs, N.D. Mermin, and R. Schack, "An introduction to QBism with an application to the locality of quantum mechanics," Am. J. Phys. **82**, 749–754 (2014).
[44] J.S. Bell, "On the Einstein Podolsky Rosen paradox," Physics **1**, 195–200 (1964).
[45] S. Kochen and E. P. Specker, "The problem of hidden variables in quantum mechanics," J. Math. Mech. **17**, 59–87 (1967).
[46] R. Colbeck and R. Renner, "No extension of quantum theory can have improved predictive power," Nat. Commun. **2**, 411 (2011).
[47] R. Colbeck and R. Renner, "A short note on the concept of free choice," arXiv:1302.4446 (2013).
[48] H. Everett, ""Relative state" formulation of quantum mechanics," Rev. Mod. Phys. **29**, 454–462 (1957).

[49] J.A Wheeler, "Assessment of Everett's "relative state" formulation of quantum theory," Rev. Mod. Phys. **29**, 463 (1957).
[50] B.S. DeWitt, "Quantum mechanics and reality," Phys. Today **23**, 155–165 (1970).
[51] D. Deutsch, *The Fabric of Reality: The Science of Parallel Universes and Its Implications* (Allen Lane Science, 1997).
[52] L. Vaidman, "Many-worlds interpretation of quantum mechanics," in *The Stanford Encyclopedia of Philosophy* (Stanford University, 2016).
[53] H.D. Zeh, "On the interpretation of measurement in quantum theory," Found. Phys. **1**, 69–76 (1970).
[54] D. Albert and B. Loewer, "Interpreting the many worlds interpretation," Synthese **77**, 195–213 (1988).
[55] G. Brassard and P. Raymond-Robichaud, "Can free will emerge from determinism in quantum theory?" in *Is Science Compatible with Free Will? Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*, edited by A. Suarez and P. Adams (Springer, 2013) pp. 41–61.
[56] W.H. Zurek, "Relative states and the environment: einselection, envariance, quantum Darwinism, and the existential interpretation," arXiv:0707.2832 (2007).
[57] L. Vaidman, "On schizophrenic experiences of the neutron or why we should believe in the many-worlds interpretation of quantum theory," Int. Stud. Phil. Sci. **12**, 245–261 (1998).
[58] J. Butterfield and G.N. Fleming, "Quantum theory and the mind," PAS, Supplementary Volumes **69**, 113–173 (1995).
[59] J. Barrett, "Information processing in generalized probabilistic theories," Phys. Rev. A **75**, 032304 (2007).
[60] L. Hardy, "Reformulating and reconstructing quantum theory," arXiv:1104.2066 (2011).
[61] G. Chiribella, G.M. D'Ariano, and P. Perinotti, "Informational derivation of quantum theory," Phys. Rev. A **84**, 012311 (2011).
[62] L. Masanes and M.P. Müller, "A derivation of quantum theory from physical requirements," New J. Phys. **13**, 063001 (2011).
[63] R. Colbeck and R. Renner, "Is a system's wave function in one-to-one correspondence with its elements of reality?" Phys. Rev. Lett. **108**, 150402 (2012).
[64] M.F. Pusey, J. Barrett, and T. Rudolph, "On the reality of the quantum state," Nat. Phys. **8**, 475–478 (2012).
[65] O. Oreshkov, F. Costa, and Č. Brukner, "Quantum correlations with no causal order," Nat. Commun. **3**, 1092 (2012).
[66] M.S. Leifer and R.W. Spekkens, "Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference," Phys. Rev. A **88**, 052130 (2013).
[67] H. Barnum and A. Wilce, "Post-classical probability theory," in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R.W. Spekkens (Springer, 2016) pp. 367–420.
[68] J.A. Wheeler, "The "past" and the "delayed-choice" double-slit experiment," in *Mathematical Foundations of Quantum Theory*, edited by A.R. Marlow (Academic Press, 1978) pp. 9–48.
[69] B.-G. Englert, M.O. Scully, G. Süssmann, and H. Walther, "Surrealistic Bohm trajectories," Z. Naturforsch. **47**, 1175–1186 (1992).
[70] M. Correggi and G. Morchio, "Quantum mechanics and stochastic mechanics for compatible observables at different times," Ann. Phys. **296**, 371–389 (2002).
[71] L. Vaidman, "The reality in Bohmian quantum mechanics or can you kill with an empty wave bullet?" Found. Phys. **35**, 299–312 (2005).
[72] J. Kiukas and R.F. Werner, "Maximal violation of Bell inequalities by position measurements," J. Math. Phys. **51**, 072105 (2010).
[73] G. Naaman-Marom, N. Erez, and L. Vaidman, "Position measurements in the de Broglie-Bohm interpretation of quantum mechanics," Ann. Phys. **327**, 2522 – 2542 (2012).
[74] N. Gisin, "Why Bohmian mechanics? One and two-time position measurements, Bell inequalities, philosophy and physics," Entropy **20**, 105 (2018).
[75] Y. Aharonov and L. Vaidman, "Complete description of a quantum system at a given time," J. Phys. A **24**, 2315 (1991).
[76] A. Kent, "Consistent sets yield contrary inferences in quantum theory," Phys. Rev. Lett. **78**, 2874–2877 (1997).
[77] J.B. DeBrota, C.A. Fuchs, and R. Schack, "Can two QBists experience one thing? A response to Frauchiger and Renner," (2018), forthcoming.