

# Quantum theory cannot consistently describe the use of itself

Daniela Frauchiger and Renato Renner\*

Institute for Theoretical Physics, ETH Zurich, 8093 Zurich; Switzerland

## Abstract

Quantum theory provides an extremely accurate description of fundamental processes in physics. It thus seems likely that the theory is applicable beyond the, mostly microscopic, domain in which it has been tested experimentally. Here we propose a *Gedankenexperiment* to investigate the question whether quantum theory can, in principle, have universal validity. The idea is that, if the answer was yes, it must be possible to employ quantum theory to model complex systems that include agents who are themselves using quantum theory. Analysing the experiment under this presumption, we find that one agent, upon observing a particular measurement outcome, must conclude that another agent has predicted the opposite outcome with certainty. The agents' conclusions, although all derived within quantum theory, are thus inconsistent. This indicates that quantum theory cannot be extrapolated to complex systems, at least not in a straightforward manner.

## Introduction

Direct experimental tests of quantum theory are mostly restricted to microscopic domains. Nevertheless, quantum theory is commonly regarded as being (almost) universally valid. It is not only used to describe fundamental processes in particle and solid state physics, but also, for instance, to explain the cosmic microwave background or the radiation of black holes.

The presumption that the validity of quantum theory extends to larger scales has remarkable consequences, as noted already in 1935 by Schrödinger [1]. His famous example consisted of a cat that is brought into a state corresponding to a superposition of two macroscopically entirely different states, one in which it is dead and one in which it is alive. Schrödinger pointed out, however, that such macroscopic superposition states do not represent anything contradictory in themselves.

This view was not shared by everyone. In 1967, Wigner proposed an argument, known as the Wigner's Friend Paradox, which should show that “quantum mechanics cannot have unlimited validity” [2]. His idea was to consider the views of two different observers in an experiment analogous to the one depicted by Fig. 1. One observer, called agent F, measures the vertical polarisation  $z$  of a spin one-half particle S, such as a silver atom. Upon observing the outcome, which is either  $z = -\frac{1}{2}$  or  $z = +\frac{1}{2}$ , agent F would thus say that S is in state

$$\psi_S = |\downarrow\rangle_S \quad \text{or} \quad \psi_S = |\uparrow\rangle_S, \quad (1)$$

respectively. The other observer, agent W, has no direct access to the outcome  $z$  observed by his friend F. Agent W could instead model agent F's lab as a big quantum system,  $L \equiv S \otimes D \otimes F$ , which contains the spin S as a subsystem, another subsystem, D, for the friend's measurement devices and everything else connected to them, as well as a subsystem F that includes the friend herself. Suppose that, from agent W's perspective, the lab L is initially in a pure state and that it remains isolated during agent F's spin measurement experiment. (One may object that these assumptions are unrealistic [3], but, crucially, the laws of quantum theory do not preclude that they be satisfied to arbitrarily good

---

\*renner@ethz.ch

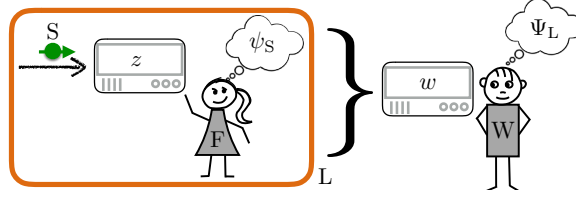


Figure 1: **Wigner’s and Deutsch’s arguments.** Agent F measures the spin  $S$  of a silver atom, obtaining outcome  $z$ . From F’s perspective,  $S$  is then in one of the two pure states  $\psi_S$  given in (1). Agent W, who is outside of F’s lab, may instead regard that lab, including the agent F, as a big quantum system  $L$  (orange box). Wigner argued that, having no access to  $z$ , he would assign a superposition state  $\Psi_L$  of the form (3) to  $L$  [2]. Deutsch later noted that agent W could in principle test this state assignment by applying a carefully designed measurement to  $L$  [6].

approximation [4].) Translated to quantum mechanics, this means that the dependence of the final state of  $L$  on the initial state of  $S$  is described by a linear map of the form

$$U_{S \rightarrow L} = \begin{cases} |\downarrow\rangle_S \mapsto |-\frac{1}{2}\rangle_L \equiv |\downarrow\rangle_S \otimes |“z=-\frac{1}{2}”\rangle_D \otimes |“\psi_S=|\downarrow\rangle”\rangle_F \\ |\uparrow\rangle_S \mapsto |+\frac{1}{2}\rangle_L \equiv |\uparrow\rangle_S \otimes |“z=+\frac{1}{2}”\rangle_D \otimes |“\psi_S=|\uparrow\rangle”\rangle_F \end{cases} . \quad (2)$$

Here  $|“z=-\frac{1}{2}”\rangle_D$  and  $|“z=+\frac{1}{2}”\rangle_D$  denote states of  $D$  depending on the measurement outcome  $z$  shown by the devices within the lab. Analogously,  $|“\psi_S=|\downarrow\rangle”\rangle_F$  and  $|“\psi_S=|\uparrow\rangle”\rangle_F$  are states of  $F$ , which depend on the friend’s own knowledge  $\psi_S$ ; cf. (1). Now, suppose agent W knew that the spin was initialised to  $|\rightarrow\rangle_S \equiv \frac{1}{\sqrt{2}}(|\downarrow\rangle_S + |\uparrow\rangle_S)$  before agent F measured it. Then, by linearity, the final state that agent W would assign to  $L$  is

$$\Psi_L = \frac{1}{\sqrt{2}}(|-\frac{1}{2}\rangle_L + |+\frac{1}{2}\rangle_L) , \quad (3)$$

i.e., a linear superposition of the two macroscopically distinct states defined in (2). To compare this to agent F’s view (1), one must consider the restriction of (3) to  $S$ . The latter is a maximally mixed state, and thus obviously different from agent F’s pure state assignment (1). But, crucially, the difference can be explained by the two agents’ distinct level of knowledge: Agent F has observed  $z$  and hence knows the final spin direction, whereas agent W is ignorant about it [5]. Consequently, although the superposition state (3) may appear “absurd” [2], it does not contradict (1). For this reason, the Wigner’s Friend Paradox cannot be regarded as an argument that rules out quantum mechanics as a universally valid theory.

In this work we propose a *Gedankenexperiment* that extends Wigner’s setup. It consists of agents who are using quantum theory to reason about other agents who are also using quantum theory. Our main finding is that such a self-referential use of the theory yields contradictory claims. This result can be phrased as a no-go theorem (Theorem 1). It asserts that three natural-sounding assumptions, (Q), (C), and (S), cannot all be valid. Assumption (Q) captures the universal validity of quantum theory (or, more specifically, that an agent can be certain that a given proposition holds whenever the quantum-mechanical Born rule assigns probability 1 to it). Assumption (C) demands consistency, in the sense that the different agents’ predictions are not contradictory. Finally, (S) is the requirement that, from the viewpoint of an agent who carries out a particular measurement, this measurement has one single outcome. The theorem itself is neutral in the sense that it does not tell us which of these three assumptions is wrong. However, it implies that any specific interpretation of quantum theory, when applied to the *Gedankenexperiment*, will necessarily conflict with at least one of them. This gives a way to test and categorise interpretations of quantum theory.

## Results

### The *Gedankenexperiment*

In the setup considered by Wigner (cf. Fig. 1), agent F carries out her measurement of S in a perfectly isolated lab L, so that the outcome  $z$  remains unknown to anyone else. The basic idea underlying the *Gedankenexperiment* we present here is to make some of the information about  $z$  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.

Box 1 specifies the proposed *Gedankenexperiment* as a step-wise procedure. The steps are to be executed by different agents — four in total. Two of them, the “friends” F and  $\bar{F}$ , are located in separate labs, denoted by L and  $\bar{L}$ , respectively. The two other agents, W and  $\bar{W}$ , are at the outside, from where they can apply measurements to L and  $\bar{L}$ , as shown in Fig. 2. We assume that L and  $\bar{L}$  are, from the viewpoint of the agents W and  $\bar{W}$ , 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. Note that the experiment can be described within standard quantum-mechanical formalism, with each step corresponding to a fixed evolution map acting on particular subsystems (cf. the circuit diagram in the appendix).

#### Experimental Procedure

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

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

Box 1: **Steps of the *Gedankenexperiment*.** The numbers on the left indicate the timing of the steps, and we assume that each step takes at most one unit of time. For example, in round  $n = 0$ , agent F starts her measurement of S at time 0:10 and completes it before time 0:11. Definitions of the relevant state and measurement basis vectors are provided in Tables 1 and 2.

As indicated by the term *Gedankenexperiment*, we do not claim that the experiment is technologically feasible, at least not in the form presented here. Like other thought experiments, its purpose is not to probe nature, but rather to scrutinise the consistency of our currently best available theories that describe nature — in this case 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.)

Before proceeding to the analysis of the experiment, a few comments about its relation to earlier proposals are in order. In the case where  $r = \text{tails}$ , agent F receives S prepared in state  $|\rightarrow\rangle_S$ . The first part of the experiment, prior to the measurements carried out by the agents  $\bar{W}$  and W, is then equivalent to Wigner’s original experiment as described in the introduction [2]. Furthermore, adding to this the measurement of agent F’s lab by agent W, one retrieves an extension of Wigner’s experiment proposed by Deutsch [6] (Fig. 1). The particular procedure of how agent  $\bar{F}$  prepares the spin S in the first step described in Box 1 is motivated by a construction due to Hardy [7, 8]. The setup considered here is also similar to a proposal by Brukner [9], who used a modification of Wigner’s argument to

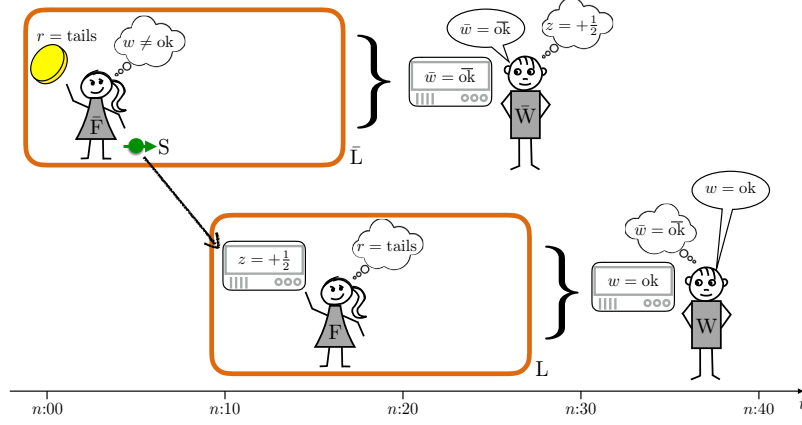


Figure 2: **Illustration of the *Gedankenexperiment*.** In each round  $n = 0, 1, 2, \dots$  of the experiment, agent  $\bar{F}$  tosses a coin and, depending on the outcome  $r$ , polarises a spin particle  $S$  in a particular direction. Agent  $F$  then measures the vertical polarisation  $z$  of  $S$ . Later, agents  $\bar{W}$  and  $W$  measure the entire labs  $\bar{L}$  and  $L$  (where the latter includes  $S$ ) to obtain outcomes  $\bar{w}$  and  $w$ , respectively. For the analysis of the experiment, we assume that all agents are aware of the entire procedure as specified in Box 1, but that they are located at different places and therefore make different observations. Agent  $F$ , for instance, observes  $z$  but has no direct access to  $r$ . She may however use quantum theory to draw conclusions about  $r$ .

obtain a strengthening of Bell's theorem [10] (cf. the discussion section).

## Analysis of the *Gedankenexperiment*

We analyse the experiment from the viewpoints of the different agents, who have access to different pieces of information (cf. Fig. 2). We assume, however, that all agents are aware of the entire experimental procedure as described in Box 1. We also suppose that they all employ the same theory. That is, their reasoning is based on the same set of rules, which we specify in the following as assumptions.

The first such assumption is that the agents “use quantum theory”, by which we basically mean that their conclusions about measurement outcomes are inferred from the Born rule. In fact, for our purposes, it suffices to consider the special case where the state of the system that is to be measured lies in the image of only one single measurement operator. In this case, the Born rule assigns probability 1 to the corresponding outcome, i.e., it predicts the outcome with certainty.

### Assumption (Q)

Suppose that agent  $A$  has established that

*Statement A<sup>(i)</sup>:* “System  $S$  is in state  $|\psi\rangle_S$  at time  $t_0$ .”

(Here  $S$  is an arbitrary system around the agent and  $|\psi\rangle_S$  a unit vector of the Hilbert space of  $S$ .)  
Suppose furthermore that agent  $A$  knows that

*Statement A<sup>(ii)</sup>:* “The value  $x$  is obtained by a measurement of  $S$  w.r.t. the family  $\{\pi_x^{t_0}\}_{x \in \mathcal{X}}$  of Heisenberg operators relative to time  $t_0$ , which is completed at time  $t$ .”

(Here  $\pi_x^{t_0}$ , for  $x \in \mathcal{X}$ , are positive operators on  $S$  that sum up to the identity.)  
If  $\langle \psi | \pi_\xi^{t_0} | \psi \rangle = 1$  for some  $\xi \in \mathcal{X}$  then agent  $A$  can conclude that

*Statement A<sup>(iii)</sup>:* “I am certain that  $x = \xi$  at time  $t$ .”

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

Table 1: **Time evolution.** The two labs,  $\bar{L}$  and  $L$ , are assumed to be isolated quantum systems. Technically, this means that their time evolution is described by norm-preserving linear maps, i.e., isometries. The second protocol step, for instance, in which  $F$  measures  $S$ , induces an isometry  $U_{S \rightarrow L}^{10 \rightarrow 20}$  from  $S$  to  $L$ . The vectors  $|-\frac{1}{2}\rangle_L$  and  $|+\frac{1}{2}\rangle_L$  are defined as the outputs of this isometry, i.e., the states of lab  $L$  at the end of the protocol step depending on whether the incoming spin was  $|\downarrow\rangle_S$  or  $|\rightarrow\rangle_S$ , respectively. For concreteness, one may think of them as states of the form (2) — although their structure is irrelevant for the argument. Analogously,  $|\bar{h}\rangle_{\bar{L}}$  and  $|\bar{t}\rangle_{\bar{L}}$  are defined as the states of lab  $\bar{L}$  at the end of the first protocol step, depending on whether  $r = \text{heads}$  or  $r = \text{tails}$ , respectively.

agent	value	measured system	measurement completed at	relevant vectors of measurement basis	Heisenberg projectors used for reasoning via (Q)
$\bar{F}$	$r$	$R$	$n:01$	$ \text{heads}\rangle_R$ $ \text{tails}\rangle_R$	$\pi_{w=\text{ok}}^{n:10} = [(U_{S \rightarrow L}^{10 \rightarrow 20})^\dagger  \text{ok}\rangle_L][\cdot]^\dagger$ $\pi_{w=\text{fail}}^{n:10} = \mathbf{1} - \pi_{w=\text{ok}}^{n:10}$
$F$	$z$	$S$	$n:11$	$ \downarrow\rangle_S$ $ \rightarrow\rangle_S$	$\pi_{z=-\frac{1}{2}}^{n:10} =  \downarrow\rangle_S \langle \downarrow _S$ $\pi_{z=+\frac{1}{2}}^{n:10} =  \rightarrow\rangle_S \langle \rightarrow _S$
$\bar{W}$	$\bar{w}$	$\bar{L}$	$n:21$	$ \text{ok}\rangle_{\bar{L}} = \sqrt{1/2}( \bar{h}\rangle_{\bar{L}} -  \bar{t}\rangle_{\bar{L}})$	$\pi_{(\bar{w},z)=(\text{ok},-\frac{1}{2})}^{n:00} = [(U_{R \rightarrow \bar{L}S}^{00 \rightarrow 10})^\dagger  \text{ok}\rangle_{\bar{L}}  \downarrow\rangle_S][\cdot]^\dagger$ $\pi_{(\bar{w},z) \neq (\text{ok},-\frac{1}{2})}^{n:00} = \mathbf{1} - \pi_{(\bar{w},z)=(\text{ok},-\frac{1}{2})}^{n:00}$
$W$	$w$	$L$	$n:31$	$ \text{ok}\rangle_L = \sqrt{1/2}( -\frac{1}{2}\rangle_L -  +\frac{1}{2}\rangle_L)$	$\pi_{(\bar{w},w)=(\text{ok},\text{ok})}^{n:00} = [(U_{R \rightarrow \bar{L}S}^{00 \rightarrow 10})^\dagger (U_{S \rightarrow L}^{10 \rightarrow 20})^\dagger  \text{ok}\rangle_{\bar{L}}  \text{ok}\rangle_L][\cdot]^\dagger$

Table 2: **Measurements carried out by the agents.** Each of the four agents observes a value, defined as the outcome of a measurement on a particular system at a particular time. The measurement basis vectors  $|\text{ok}\rangle_{\bar{L}}$  and  $|\text{ok}\rangle_L$  shown in the last two rows are expressed in terms of states, such as  $|-\frac{1}{2}\rangle_L$  and  $|+\frac{1}{2}\rangle_L$ , which are defined in Table 1. The last column shows the measurement operators that the agents insert into statement  $A^{(ii)}$  when reasoning according to Assumption (Q). These operators are given in the Heisenberg picture, referring to the system's state at a particular time, which is specified by a superscript. The bracket  $[\cdot]^\dagger$  stands for the adjoint of the preceding expression.

For our argument, we take this rule to be valid universally. It shall in particular be applicable to large and complex systems, which themselves contain agents. This is, for instance, the case for lab L, which in any round  $n$  of the experiment is measured with respect to the Heisenberg operators  $\pi_{w=\text{ok}}^{n:10}$  and  $\pi_{w=\text{fail}}^{n:10}$  defined in Table 2. For our analysis, we suppose that agent  $\bar{F}$  wants to predict the outcome  $w$  of this measurement. To this aim, she may start her reasoning with a statement that describes the corresponding measurement.

*Statement  $\bar{F}^{n:00}$ :* “The value  $w$  is obtained by a measurement of L w.r.t.  $\{\pi_{w=\text{ok}}^{n:10}, \pi_{w=\text{fail}}^{n:10}\}$ , which is completed at time  $n:31$ .”

Here and in the following, we specify for each statement a time, denoted as a superscript, indicating when the agent could have inferred the statement. Agent  $\bar{F}$ ’s statement  $\bar{F}^{n:00}$  above does not depend on any observations, so the time  $n:00$  we have assigned to it is rather arbitrary. This is however different for the next statement, which is based on knowledge of the value  $r$ . Suppose that agent  $\bar{F}$  got  $r = \text{tails}$  as the output of the random number generator in round  $n$ . According to the experimental instructions, she will then prepare the spin S in state  $|\rightarrow\rangle_S$ . Now, after completing the preparation, say at time  $n:01$ , she may make a second statement, taking into account that S remains unchanged until F starts her measurement at time  $n:10$ .

*Statement  $\bar{F}^{n:01}$ :* “The spin S is in state  $|\rightarrow\rangle_S$  at time  $n:10$ .”

Agent  $\bar{F}$  could conclude from this that the later state of the lab L,  $U_{S \rightarrow L}^{10 \rightarrow 20} |\rightarrow\rangle_S = \sqrt{1/2} (|\rightarrow\rangle_L + |\rightarrow\rangle_L)$ , will be orthogonal to  $|\text{ok}\rangle_L$ . An equivalent way to express this is that the state  $|\rightarrow\rangle_S$  has no overlap with the Heisenberg measurement operator corresponding to outcome  $w = \text{ok}$ , i.e.,

$$\langle \rightarrow | \pi_{w=\text{fail}}^{n:10} | \rightarrow \rangle = 1 - \langle \rightarrow | \pi_{w=\text{ok}}^{n:10} | \rightarrow \rangle = 1. \quad (4)$$

The two statements  $\bar{F}^{n:00}$  and  $\bar{F}^{n:01}$ , inserted into (Q), thus imply that  $w = \text{fail}$ . We may assume that agent  $\bar{F}$  draws this conclusion at time  $n:02$  and, for later use, put it down as statement  $\bar{F}^{n:02}$  in Table 3. Similarly, agent F’s reasoning may be based upon a description of her spin measurement, which is defined by the operators  $\pi_{z=-\frac{1}{2}}^{n:10}$  and  $\pi_{z=+\frac{1}{2}}^{n:10}$  given in Table 2.

*Statement  $F^{n:10}$ :* “The value  $z$  is obtained by a measurement of S w.r.t.  $\{\pi_{z=-\frac{1}{2}}^{n:10}, \pi_{z=+\frac{1}{2}}^{n:10}\}$ , which is completed at time  $n:11$ .”

Suppose now that agent F observed  $z = +\frac{1}{2}$  in round  $n$ . Since, by definition,

$$\langle \downarrow | \pi_{z=-\frac{1}{2}}^{n:10} | \downarrow \rangle = 1 \quad (5)$$

it follows from (Q) that S was not in state  $|\downarrow\rangle$  and hence that the random value  $r$  was not **heads**. This is statement  $F^{n:12}$  of Table 3. We proceed with agent  $\bar{W}$ , who may base his reasoning upon his knowledge of how the random number generator was initialised.

*Statement  $\bar{W}^{n:21}$ :* “System R is in state  $|\text{init}\rangle_R$  at time  $n:00$ .”

Consider the event that  $\bar{w} = \text{ok}$  and  $z = -\frac{1}{2}$ , as well as its complement. The Heisenberg operators of the corresponding measurement are given in Table 2. It is straightforward to verify that  $U_{R \rightarrow LS}^{00 \rightarrow 10} |\text{init}\rangle_R = \sqrt{1/3} |\bar{h}\rangle_L \otimes |\downarrow\rangle_S + \sqrt{2/3} |\bar{t}\rangle_L \otimes |\rightarrow\rangle_S$  is orthogonal to  $|\text{ok}\rangle_L \otimes |\downarrow\rangle_S$ , which implies that

$$\langle \text{init} | \pi_{(\bar{w}, z) \neq (\text{ok}, -\frac{1}{2})}^{n:00} | \text{init} \rangle = 1 - \langle \text{init} | \pi_{(\bar{w}, z) = (\text{ok}, -\frac{1}{2})}^{n:00} | \text{init} \rangle = 1. \quad (6)$$

Agent  $\bar{W}$ , who also uses (Q), can hence be certain that  $(\bar{w}, z) \neq (\text{ok}, -\frac{1}{2})$ . This implies that statement  $\bar{W}^{n:22}$  of Table 3 holds whenever  $\bar{w} = \text{ok}$ . Furthermore, because agent  $\bar{W}$  announces  $\bar{w}$ , agent W can be certain about  $\bar{W}$ ’s knowledge, which justifies statement  $W^{n:26}$  of the table. We have thus established all statements in the third column of Table 3.

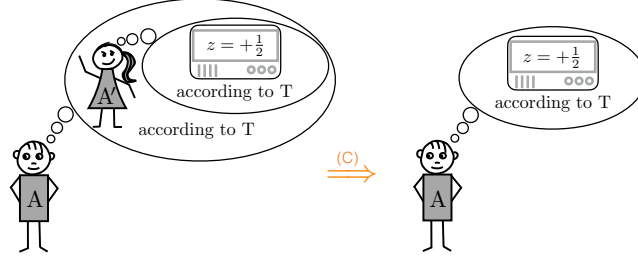


Figure 3: **Consistent reasoning as required by Assumption (C).** If a theory  $T$  (such as quantum theory) enables consistent reasoning (C) then it must allow any agent  $A$  to promote the conclusions drawn by another agent  $A'$  to his own conclusions, provided that  $A'$  has the same initial knowledge about the experiment and reasons within the same theory  $T$ . A classical example of such recursive reasoning is the muddy children puzzle (here  $T$  is just standard logic; see [11] 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 [12].

For later use we also note that a simple calculation yields

$$\langle \text{init} | \pi_{(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})}^{n:00} | \text{init} \rangle = 1/12 \quad (7)$$

where  $\pi_{(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})}^{n:00}$  is the Heisenberg operator belonging to the event that  $\bar{w} = \bar{\text{ok}}$  and  $w = \text{ok}$ , as defined in Table 2. Hence, according to quantum mechanics, agent  $W$  can be certain that the outcome  $(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})$  occurs after finitely many rounds. This corresponds to the following statement (which can indeed be derived using (Q), as shown in the appendix).

*Statement  $W^{0:00}$ :* “I am certain that there exists a round  $n$  in which the halting condition at time  $n:40$  is satisfied.”

The agents may now obtain further statements by reasoning about how they would reason from the viewpoint of other agents. To enable such nested reasoning we need another assumption, which is illustrated by Fig. 3.

#### Assumption (C)

Suppose that agent  $A$  has established that

*Statement  $A^{(i)}$ :* “I am certain that agent  $A'$ , upon reasoning within the same theory as the one I am using, is certain that  $x = \xi$  at time  $t$ .”

Then agent  $A$  can conclude that

*Statement  $A^{(ii)}$ :* “I am certain that  $x = \xi$  at time  $t$ .”

Agent  $F$  may insert agent  $\bar{F}$ ’s statement  $\bar{F}^{n:02}$  into  $F^{n:12}$ , obtaining statement  $F^{n:13}$  of Table 3. By virtue of Assumption (C), she may then conclude that statement  $F^{n:14}$  holds, too. Similarly,  $\bar{W}$  may combine this latter statement with his statement  $\bar{W}^{n:22}$  to obtain  $\bar{W}^{n:23}$ . He could then, again using (C), conclude that statement  $\bar{W}^{n:24}$  holds. Finally, agent  $W$  can insert this into his statement  $W^{n:26}$  to obtain statement  $W^{n:27}$  and, again with (C), statement  $W^{n:28}$ . This completes the derivation of all statements of Table 3.

For the last part of our analysis, we take again agent  $W$ ’s perspective. According to statement  $W^{0:00}$ , the experiment has a final round  $n$  in which the halting condition will be satisfied, meaning in particular that agent  $\bar{W}$  announces  $\bar{w} = \bar{\text{ok}}$ . Agent  $W$  infers from this that statement  $W^{n:28}$  of Table 3 holds in that round, i.e., he is certain that he will observe  $w = \text{fail}$  at time  $n:31$ . However, in this final round,

agent	assumed observation	statement inferred via (Q)	further implied statement	statement inferred via (C)
$\bar{F}$	$r = \text{tails}$ at time $n:01$	<i>Statement <math>\bar{F}^{n:02}</math>:</i> “I am certain that W will observe $w = \text{fail}$ at time $n:31$ .”		
F	$z = +\frac{1}{2}$ at time $n:11$	<i>Statement <math>F^{n:12}</math>:</i> “I am certain that $\bar{F}$ knows that $r = \text{tails}$ at time $n:01$ .”	<i>Statement <math>F^{n:13}</math>:</i> “I am certain that $\bar{F}$ is certain that W will observe $w = \text{fail}$ at time $n:31$ .”	<i>Statement <math>F^{n:14}</math>:</i> “I am certain that W will observe $w = \text{fail}$ at time $n:31$ .”
$\bar{W}$	$\bar{w} = \bar{\text{ok}}$ at time $n:21$	<i>Statement <math>\bar{W}^{n:22}</math>:</i> “I am certain that F knows that $z = +\frac{1}{2}$ at time $n:11$ .”	<i>Statement <math>\bar{W}^{n:23}</math>:</i> “I am certain that F is certain that W will observe $w = \text{fail}$ at time $n:31$ .”	<i>Statement <math>\bar{W}^{n:24}</math>:</i> “I am certain that W will observe $w = \text{fail}$ at time $n:31$ .”
W	announcement by agent $\bar{W}$ that $\bar{w} = \bar{\text{ok}}$ at time $n:21$	<i>Statement <math>W^{n:26}</math>:</i> “I am certain that $\bar{W}$ knows that $\bar{w} = \bar{\text{ok}}$ at time $n:21$ .”	<i>Statement <math>W^{n:27}</math>:</i> “I am certain that $\bar{W}$ is certain that I will observe $w = \text{fail}$ at time $n:31$ .”	<i>Statement <math>W^{n:28}</math>:</i> “I am certain that I will observe $w = \text{fail}$ at time $n:31$ .”

Table 3: **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. 2). Agent  $\bar{F}$ , for instance, if she observes  $r = \text{tails}$ , can use this information to infer  $w$ , which will later be observed and announced by W.

he will nevertheless observe  $w = \text{ok}$ ! We have thus reached a contradiction — unless agent W would accept that  $w$  admits multiple values. For our discussion below, it will be useful to introduce an explicit assumption, termed (S), which disallows this.

#### Assumption (S)

Suppose that agent A has established that

*Statement  $A^{(i)}$ :* “I am certain that  $x = \xi$  at time  $t$ .”

Then agent A must necessarily deny that

*Statement  $A^{(ii)}$ :* “I am certain that  $x \neq \xi$  at time  $t$ .”

## No-go theorem

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

*Theorem 1:* Any theory that satisfies assumptions (Q), (C), and (S) yields contradictory statements when applied to the *Gedankenexperiment* of Box 1.

To illustrate the theorem, we consider in the following different interpretations and modifications of quantum theory. Theorem 1 implies that any of them must violate either (Q), (C), or (S). This yields a natural categorisation as shown in Table 4 and discussed in the following subsections.

## Theories that violate Assumption (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 interpretational questions, such as the meaning of probabilities in general. However, the non-trivial aspect of (Q) is that



	(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 4: **Interpretations of quantum theory.** The proposed *Gedankenexperiment* can be employed to study the various interpretations of quantum theory. Theorem 1 implies that each of them must violate at least one of the Assumptions (Q), (C), and (S) (indicated by ×). For hidden variable (HV) theories, it is relevant whether agents who are using the theory apply its laws (e.g., the guiding equation in the case of Bohmian mechanics) to subsystems around them or to the universe as a whole.

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: Assumption (Q) does not demand that the agent A can describe herself as a quantum system. Such a requirement would indeed be overly restrictive (see [13]) 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 [14, 15].

Assumption (Q) is manifestly violated by theories that postulate a modification of standard quantum mechanics, such as spontaneous [16, 17, 18, 19, 20] and gravity-induced [21, 22, 23] collapse models (cf. [24] for a review). These deviate from the standard 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 [25]. 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” [26], 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” [27], 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 [4]).

Another class of theories that violate (Q), although in a less obvious manner, are particular “hidden-variable (HV) interpretations” [28], with “Bohmian mechanics” as the most prominent example [29, 30, 31]. According to the common understanding, Bohmian mechanics is a “theory of the universe” rather than a theory about subsystems [32]. 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 Assumption (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) formalism [33, 34, 35], which is also similar to the “decoherent histories” approach [36, 37]. Another class of examples are subjectivistic interpretations, which regard statements about outcomes of measurements as personal to an agent, such as “relational quantum mechanics” [38], “QBism” [39, 40], or the approach proposed in [9] (see the appendix for a discussion of the CH formalism 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 [10], they must be contextual [41, 42], and they violate freedom of choice [43, 44]. 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 Assumption (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 “many-worlds interpretations” [45, 46, 47, 6, 48]. 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” [49, 50] and the “parallel lives theory” [51]. A related concept is “quantum Darwinism” [52], 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 [53]). It is also questionable whether (Q) can be upheld if branches do not persist over time (cf. the no-histories view described in [54]).

## Implicit assumptions

Any no-go result, as for example Bell’s theorem [10], is phrased within a particular framework that comes with a set of built-in assumptions. Hence it is always possible that a theory evades the conclusions of the no-go result by not fulfilling these implicit assumptions. Here we briefly discuss how Theorem 1 compares in this respect to other results in the literature.

Bell’s original work [10] treats probabilities as a primitive notion. Similarly, many of the modern arguments in quantum foundations employ probabilistic frameworks [55, 56, 57, 58, 59, 60, 61, 62]. In contrast, probabilities are not used in the argument presented here — although Assumption (Q) is of course motivated by the idea that a statement can be regarded as “certain” if the Born rule assigns probability 1 to it. In particular, Theorem 1 does not depend on how probabilities different from 1 are interpreted.

Another distinction is that the framework used here treats all statements about observations as subjective, i.e., they are always defined relative to an agent. This avoids the *a priori* assumption that measurement outcomes obtained by different agents simultaneously have definite values. (Consider for example Wigner’s original setup described in the introduction. Even when Assumptions (C) and (S) hold, agent W is not forced to assign a definite value to the outcome  $z$  observed by agent F.) The assumption of simultaneous definiteness is otherwise rather common. It not only enters the proof of Bell’s theorem [10] but also the aforementioned arguments based on probabilistic frameworks.

Nevertheless, in our considerations, 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.

## Discussion

In the *Gedankenexperiment* proposed in this article, multiple agents have access to different pieces of information, and draw conclusions by reasoning about the information held by others. The rules for such reasoning may be ambiguous in the general context of quantum theory where, from the viewpoint of one agent, the information held by another agent can be in a superposition of different “classical” states. Crucially, however, in the argument presented here, the agents’ conclusions are all restricted to supposedly unproblematic “classical” cases. For example, agent  $\bar{W}$  only needs to derive a statement about agent  $F$  in the case where, conditioned on his own information  $\bar{w}$ , the information  $z$  held by  $F$  has a well-defined value (Table 3). Nevertheless, as we have shown, the agents arrive at contradictory statements.

Current interpretations of quantum theory do not agree on the origin of this contradiction (cf. Table 4). 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 of Box 1 is played, with the gambler in the role of agent  $W$ , and the roles of  $\bar{F}$ ,  $F$ , and  $\bar{W}$  taken by employees of the casino. The casino promises to pay €1000 to the gambler if  $\bar{F}$ ’s random value was  $r = \text{heads}$ . Conversely, if  $r = \text{tails}$ , the gambler must pay €500 to the casino. It could now happen that, at the end of the game,  $w = \text{ok}$  and  $\bar{w} = \text{ok}$ , 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 3.

*Gambler:* “The outcome  $w = \text{ok}$  proves, due to (4), that  $S$  was not prepared in state  $|\rightarrow\rangle_S$ . This means that  $r = \text{heads}$  and hence the casino must pay me €1000.”

*Casino:* “The outcome  $\bar{w} = \text{ok}$  implies, due to (6), that our employee observed  $z = +\frac{1}{2}$ . This in turn proves that  $S$  was not prepared in state  $|\downarrow\rangle_S$ . But this means that  $r = \text{tails}$ , 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 [41, 42, 10, 7, 8, 43, 9], which also use assumptions similar to (Q) and (S) (although the latter is often implicit). These two assumptions are usually shown to be in conflict with additional assumptions about reality, locality, or freedom of choice. For example, the result of [9], which is as well based on an extension of Wigner’s argument, asserts 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. Here we have shown that Assumptions (Q) and (S) are already problematic by themselves, in the sense that agents who use these assumptions to reason about each other as in Fig. 3 will arrive at inconsistent conclusions.

Another noticeable difference to earlier no-go results is that the argument presented here does not employ counterfactual reasoning. That is, it does not refer to 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 [63]).

We conclude by suggesting a modified variant of the experiment, which may be technologically feasible. The idea is to substitute agents  $\bar{F}$  and  $F$  by computers. Specifically, one would program them to carry out the tasks prescribed in Box 1, process the information accessible to them, and output statements such as “I am certain that  $W$  will observe  $w = \text{fail}$  at 1:31.” To account for the requirement that  $\bar{F}$  and  $F$ ’s labs be isolated, one would need to ensure that the computers used for their simulation do not leak any information to their environment — a property which is necessarily satisfied by quantum computers. Such an experiment could then be used to verify the statements in Table 3. For example, aborting the experiment right after 1:13, one could, in the case when  $z = +\frac{1}{2}$ , read out statement  $F^{1:13}$

made by agent F together with statement  $\bar{F}^{1:02}$  that agent  $\bar{F}$  has made just before. This would be a test for the correctness of statement  $F^{1:13}$ . Note that all statements in the fourth column of Table 3 could in the same way be tested experimentally. In this sense, quantum computers, motivated usually by applications in computing, may help us answering questions in fundamental research.

## Appendix

### Information-theoretic description

The experimental protocol described in Box 1 may be represented as a circuit diagram, Fig. 4. 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 have access to different data (corresponding to different wires in the diagram).

### Derivation of statement $W^{0:00}$ using Assumption (Q)

In the analysis of the *Gedankenexperiment* we argued that the event  $(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})$  must occur after finitely many rounds  $n$ , which is statement  $W^{0:00}$  described shortly after (7). While this is a pretty obvious consequence of the Born rule, we now show that it already follows from Assumption (Q), which corresponds to the special case of the Born rule when it gives probability-1 predictions.

We consider Heisenberg operators relative to time  $t_0 = 0:00$ , i.e., 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  $|\text{init}\rangle_R$  as well as  $U_{R \rightarrow \bar{L}S}^{00 \rightarrow 10}$  and  $U_{S \rightarrow L}^{10 \rightarrow 20}$  (cf. Table 1), i.e.,

$$W_{\mathbb{C} \rightarrow \bar{L}L}^n = (\mathbf{1}_{\bar{L}} \otimes U_{S \rightarrow L}^{10 \rightarrow 20}) U_{R \rightarrow \bar{L}S}^{00 \rightarrow 10} |\text{init}\rangle_R . \quad (8)$$

The Heisenberg operator of the event  $(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})$  in round  $n$  relative to time  $t_0$  can thus be written as

$$\pi_{(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})}^{(n)} = (W_{\mathbb{C} \rightarrow \bar{L}L}^n)^\dagger (|\bar{\text{ok}}\rangle\langle \bar{\text{ok}}|_{\bar{L}} \otimes |\text{ok}\rangle\langle \text{ok}|_L) (W_{\mathbb{C} \rightarrow \bar{L}L}^n) . \quad (9)$$

We may now specify a Heisenberg operator  $\pi_{\text{halt}}^{0:00}$  for the halting condition, i.e., that the event  $(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})$  occurs in some round  $n$ ,

$$\pi_{\text{halt}}^{0:00} = \sum_{n=0}^{\infty} \pi_{(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})}^{(n)} \prod_{m=0}^{n-1} (\mathbf{1}_{\mathbb{C}} - \pi_{(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})}^{(m)}) . \quad (10)$$

Note that these are operators on  $\mathbb{C}$ , i.e.,  $\pi_{(\bar{w}, w) = (\bar{\text{ok}}, \text{ok})}^{(n)} = p$  for some  $p \in \mathbb{C}$ . It follows directly from (7) that  $p = 1/12 > 0$ . We thus have

$$\pi_{\text{halt}}^{0:00} = \sum_{n=0}^{\infty} p(1-p)^n = \mathbf{1}_{\mathbb{C}} . \quad (11)$$

Inserting this measurement operator into the corresponding statement of Assumption (Q) yields statement  $W^{0:00}$ .

### 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

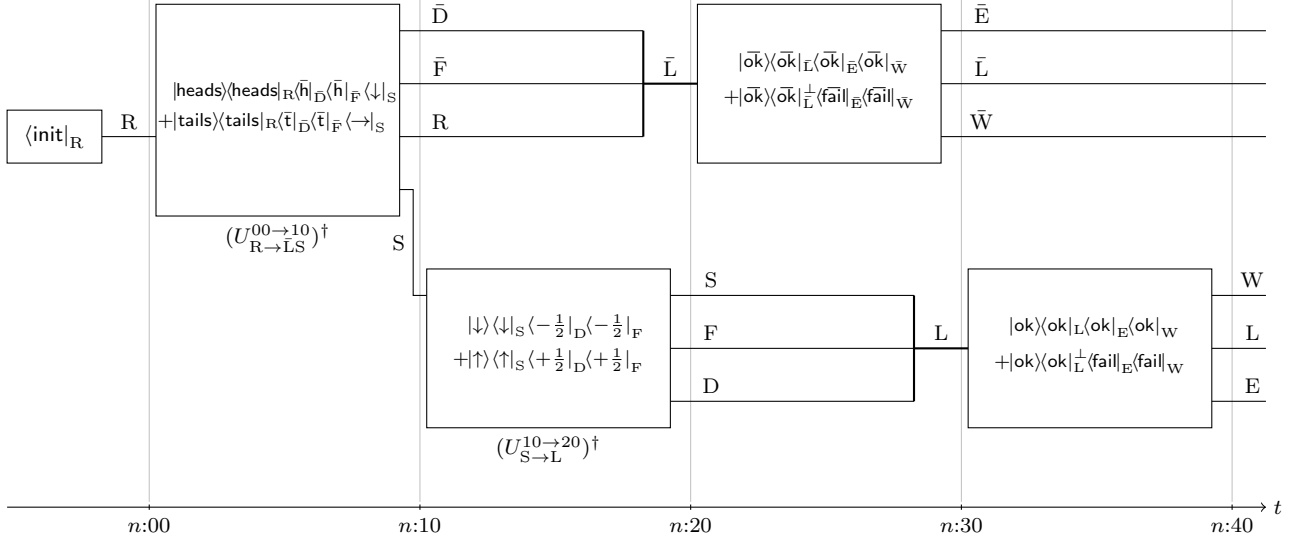


Figure 4: **Circuit diagram representation of the *Gedankenexperiment*.** The actions of the agents during the protocol correspond to isometries (boxes) that act on particular subsystems (wires). For example, the measurement of S by agent F in the second protocol step, which starts at time  $n:10$ , induces an isometry  $U_{S \rightarrow L}^{10 \rightarrow 20}$  from S to F’s lab L, analogous to the one defined by (2). The subsystems labelled by  $\bar{F}$ , F,  $\bar{W}$ , and W contain the agents themselves. Similarly,  $\bar{D}$ , D,  $\bar{E}$ , and E are “environment” subsystems, which include the agents’ measurement devices. The states of these subsystems depend on the measurement outcome, which is indicated by their label. For example,  $|\downarrow\rangle_F$  is the state of F when the agent has observed  $z = +\frac{1}{2}$ .

to the universe as a whole. As noted in [32], “if we postulate that subsystems [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). But since Bohmian mechanics also satisfies (S), this would, by virtue of Theorem 1, imply a violation of (C), i.e., the agents’ conclusions would contradict each other. (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 [64, 65, 66, 67, 68].)

The directive of [32] 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 statement  $\bar{F}^{n:02}$  of Table 3 does not hold there. Furthermore, the time order of the measurements carried out by agents  $\bar{W}$  and W is relevant within Bohmian mechanics. If agent W measured before agent  $\bar{W}$  then, according to Bohmian mechanics, statement  $\bar{W}^{n:22}$  would be invalid whereas  $\bar{F}^{n:02}$  would hold. This is a clear departure from standard quantum mechanics, where the time order in which agents  $\bar{W}$  and W carry out their measurements is irrelevant, because they act on separate systems.

This violation of (Q) raises the question under what circumstances Bohmian mechanics still endorses the use of the quantum-mechanical Born rule for predicting the outcome of a measurement. A candidate criterion could be that such a prediction is only valid if a memory of the prediction is available upon completion of the measurement. One may then be tempted to argue that agent  $\bar{F}$ ’s statement  $\bar{F}^{n:02}$ , for instance, is invalid because  $\bar{F}$  is herself subject to a measurement, which may destroy her memory of

the prediction for  $w$  before that value is measured. This argument does however not work. The reason is that, in the relevant case when  $\bar{w} = \bar{ok}$ , the value  $r$  and hence also agent  $\bar{F}$ 's prediction for  $w$  is, by virtue of statements  $\bar{W}^{n:22}$  and  $F^{n:12}$ , retrievable at the time when  $w$  is measured.

## Analysis within the CH formalism

In the consistent histories (CH) formalism, 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

*History  $h_1$* : “In round  $n$  the outcomes  $r = \text{tails}$ ,  $z = +\frac{1}{2}$ ,  $\bar{w} = \bar{ok}$ , and  $w = ok$  were observed.”

To verify that  $h_1$  is indeed a valid history, one has to construct a framework containing this history. It is straightforward to check that one such framework is the set consisting of  $h_1$  together with the additional histories

*History  $h_2$* : “In round  $n$  the outcomes  $r = \text{tails}$ ,  $z = +\frac{1}{2}$ ,  $\bar{w} = \bar{ok}$ , and  $w = \text{fail}$  were observed.”

*History  $h_3$* : “In round  $n$  the outcomes  $r = \text{heads}$ ,  $z = +\frac{1}{2}$ , and  $\bar{w} = \bar{ok}$  were observed.”

*History  $h_4$* : “In round  $n$  the outcomes  $z = -\frac{1}{2}$  and  $\bar{w} = \bar{ok}$  were observed.”

*History  $h_5$* : “Outcome  $\bar{w} = \bar{\text{fail}}$  was observed.”

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  $h_1$ , which leaves the values  $z$  and  $\bar{w}$  unmentioned:

*History  $h'_1$* : “In round  $n$  the outcomes  $r = \text{tails}$  and  $w = ok$  were observed.”

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

$$\Pr[h_1] = 1/12 \quad \text{and} \quad \Pr[h'_1] = 0. \quad (12)$$

Note that these probabilities disagree with the fact that  $h'_1$  is just a part of history  $h_1$ , i.e.,  $h_1 \implies h'_1$ . (This finding may be compared to the “three box paradox” [69], where calculations in three different consistent frameworks yield mutually incompatible probability assignments; see Sec. 22 of [35] as well as [70] for a discussion.)

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  $h_1$  and  $h'_1$ . To illustrate what this means, it is useful to return to the casino example described in the discussion section above. Within a framework that contains history  $h'_1$ , the gambler’s reasoning is correct, for  $\Pr[h'_1] = 0$ . That is,  $w = ok$  implies that  $r = \text{heads}$ . Conversely, considering the framework above, which contains history  $h_1$ , it is readily verified that the other histories,  $h_2$ ,  $h_3$ ,  $h_4$ , and  $h_5$ , have probabilities  $\frac{1}{12}$ , 0, 0, and  $\frac{5}{6}$ , respectively. That is, all non-zero probability histories of this framework that agree with the observation  $\bar{w} = \bar{ok}$  also assert that  $z = +\frac{1}{2}$  and  $r = \text{tails}$ . This seems to be in agreement with the casino’s argument, i.e.,  $\bar{w} = \bar{ok}$  implies that  $z = +\frac{1}{2}$  and  $r = \text{tails}$ . However, because the framework does not include a history that talks about  $r$  alone, it disallows the — seemingly obvious — implication  $(r = \text{tails and } z = +\frac{1}{2}) \implies r = \text{tails}$ . In other words, within the CH formalism, the casino can prove that  $r = \text{tails}$  and  $z = +\frac{1}{2}$ , but not that  $r = \text{tails}$ .

## 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.



To reflect these tenets of QBism in the analysis of the *Gedankenexperiment*, it is useful to imagine that the agents write their observations and conclusions into a personal notebook. For example, according to Table 3, when agent  $\bar{F}$  gets  $r = \text{tails}$  in round  $n = 1$ , she may put down the following

*Statement  $\bar{F}^{1:02}$ :* “ $r = \text{tails}$  at time 1:01, hence I am certain that I will hear W announcing  $w = \text{fail}$  at the end of this round.”

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, agent F, when she gets  $z = +\frac{1}{2}$ , may write into her notebook

*Statement  $F^{1:12}$ :* “ $z = +\frac{1}{2}$  at time 1:11, hence I am certain that, if I now checked  $\bar{F}$ ’s notebook, I would read that  $r = \text{tails}$  at time 1:01.”

Agent  $\bar{F}$  may as well write about agent  $\bar{F}$ ’s conclusions, i.e.,

*Statement  $F^{1:13}$ :* “ $z = +\frac{1}{2}$  at time 1:11, hence I am certain that, if I now checked  $\bar{F}$ ’s notebook, I would read that she is certain that she will hear W announcing  $w = \text{fail}$  at the end of this round.”

If one takes QBism to be a working tool for any agent to make predictions, it should in particular be the case that agent  $\bar{F}$ ’s prediction is accurate — at least for her. That is, agent  $\bar{F}$ ’s later observation of W’s announcement, which she may also put down into her notebook, should be in agreement with her earlier prediction. Agent F may thus be tempted to conclude from the above that

*Statement  $F^{1:14}$ :* “ $z = +\frac{1}{2}$  at time 1:11, hence I am certain that, if I will check  $\bar{F}$ ’s notebook at the end of this round, I will read that she heard W announcing  $w = \text{fail}$ .”

However, permitting such implications is akin to assuming (C). Because QBism satisfies (Q) and (S), it would result in the agents issuing contradictory statements such as “I heard W announcing  $w = \text{ok}$  but I am certain that my friend heard him announcing  $w = \text{fail}$ .” The *Gedankenexperiment* is thus an example of a multi-agent scenario where, to ensure consistency of QBism, implications of the type  $F^{1:13} \implies F^{1:14}$  must be disallowed. Nevertheless, there should be ways for agents to consistently reason about each other. One may therefore ask whether (C) could be substituted by another (weaker) rule that enables such reasoning but does not lead to contradictions. This question is currently being investigated (J.B. DeBrot, C.A. Fuchs, and R. Schack, manuscript in preparation).

## Acknowledgments

We would like to thank Yakir Aharonov, Mateus Araújo, Alexia Auffèves, Jonathan Barrett, Veronika Baumann, Serguei Belousov, Charles Bennett, Hans Briegel, Časlav Brukner, Harry Buhrman, Adán Cabello, Giulio Chiribella, Roger Colbeck, Patricia Contreras Tejada, Giacomo Mauro D’Ariano, John DeBrot, Lúcia del Rio, David Deutsch, Artur Ekert, Michael Esfeld, Philippe Faist, Aaron Fenyés, 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, John Loverain, Markus Müller, Thomas Müller, Hrvoje Nikolić, Travis Norsen, Nuriya Nurgalieva, Jonathan Oppenheim, Sandu Popescu, Matthew Pusey, Gilles Pütz, Joseph Renes, Jess Riedel, Valerio Scarani, Rüdiger Schack, Robert Spekkens, Cristi Stoica, Antoine Suarez, Tony Sudbery, Stefan Teufel, Roderich Tumulka, Lev Vaidman, Vlatko Vedral, Mordecai Waegell, Andreas Winter, 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”.

## References

- [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*, pages 171–184. Indiana University Press, (1967).
- [3] K. Hepp. Quantum theory of measurement and macroscopic observables. *Helv. Phys. Acta*, 45(2):237–248, (1972).
- [4] J.S. Bell. On wave packet reduction in the Coleman-Hepp model. *Helv. Phys. Acta*, 48(1):93–98, (1975).
- [5] C.A. Fuchs. QBism, the perimeter of quantum Bayesianism. Preprint at <http://arxiv.org/abs/1003.5209>, (2010).
- [6] D. Deutsch. Quantum theory as a universal physical theory. *Int. J. Theor. Phys.*, 24:1–41, (1985).
- [7] L. Hardy. Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories. *Phys. Rev. Lett.*, 68:2981–2984, (1992).
- [8] L. Hardy. Nonlocality for two particles without inequalities for almost all entangled states. *Phys. Rev. Lett.*, 71:1665–1668, (1993).
- [9] Č. Brukner. On the quantum measurement problem. In R. Bertlmann and A. Zeilinger, editors, *Quantum [Un]Speakables II: Half a Century of Bell’s Theorem*, pages 95–117. Springer, (2017).
- [10] J.S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38:447–452, (1966).
- [11] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning About Knowledge*. MIT press, (2004).
- [12] C. H. Bennett. The thermodynamics of computation—a review. *Int. J. Theor. Phys.*, 21:905–940, (1982).
- [13] M.L. Dalla Chiara. Logical self reference, set theoretical paradoxes and the measurement problem in quantum mechanics. *J. Philos. Logic*, 6:331–347, (1977).
- [14] W. Heisenberg. Ist eine deterministische Ergänzung der Quantenmechanik möglich? In A. Hermann, K. von Meyenn, and V.F. Weisskopf, editors, *Wolfgang Pauli. Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg, u.a.*, volume II, pages 409–418. Springer, (1935).
- [15] N. Bohr. Discussions with Einstein on epistemological problems in atomic physics. In P.A. Schilpp, editor, *Albert Einstein: Philosopher-Scientist*, pages 200–241. Cambridge University Press, (1949).
- [16] G.C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Phys. Rev. D*, 34:470–491, (1986).
- [17] N. Gisin. Stochastic quantum dynamics and relativity. *Helv. Phys. Acta*, 62:363–371, (1989).
- [18] P. Pearle. Combining stochastic dynamical state-vector reduction with spontaneous localization. *Phys. Rev. A*, 39:2277–2289, (1989).
- [19] R. Tumulka. On spontaneous wave function collapse and quantum field theory. In *Proc. R. Soc. A*, volume 462, pages 1897–1908. The Royal Society, (2006).
- [20] S. Weinberg. Collapse of the state vector. *Phys. Rev. A*, 85:062116, (2012).



- [21] F. Karolyhazy. Gravitation and quantum mechanics of macroscopic objects. *Il Nuovo Cimento A (1971-1996)*, 42:390–402, (1966).
- [22] L. Diósi. Models for universal reduction of macroscopic quantum fluctuations. *Phys. Rev. A*, 40:1165–1174, (1989).
- [23] R. Penrose. On gravity’s role in quantum state reduction. *Gen. Rel. Gravit.*, 28:581–600, (1996).
- [24] 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).
- [25] L.D. Landau and E.M. Lifshitz. *Quantum Mechanics: Non-Relativistic Theory*, chapter The basic concepts of quantum mechanics. Elsevier, (2013).
- [26] J. Fröhlich and B. Schubnel. Quantum probability theory and the foundations of quantum mechanics. In P. Blanchard and J. Fröhlich, editors, *The Message of Quantum Science: Attempts Towards a Synthesis*, pages 131–193. Springer, (2015).
- [27] A. Auffèves and P. Grangier. Contexts, systems and modalities: A new ontology for quantum mechanics. *Found. Phys.*, 46:121–137, (2015).
- [28] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer-Verlag, (1932).
- [29] L. De Broglie. La mécanique ondulatoire et la structure atomique de la matière et du rayonnement. *J. Phys. Radium*, 8:225–241, (1927).
- [30] D. Bohm. A suggested interpretation of the quantum theory in terms of “hidden” variables. I. *Phys. Rev.*, 85:166–179, (1952).
- [31] D. Dürr and S. Teufel. *Bohmian Mechanics: The Physics and Mathematics of Quantum Theory*. Springer, (2009).
- [32] D. Dürr, S. Goldstein, and N. Zanghí. Quantum equilibrium and the origin of absolute uncertainty. *J. Stat. Phys.*, 67(5):843–907, (1992).
- [33] R.B. Griffiths. Consistent histories and the interpretation of quantum mechanics. *J. Stat. Phys.*, 36:219–272, (1984).
- [34] R. Omnès. Consistent interpretations of quantum mechanics. *Rev. Mod. Phys.*, 64:339–382, (1992).
- [35] R.B. Griffiths. *Consistent Quantum Theory*. Cambridge University Press, (2002).
- [36] M. Gell-Mann and J.B. Hartle. Quantum mechanics in the light of quantum cosmology. In W. Zurek, editor, *Complexity, Entropy and the Physics of Information*. Addison Wesley, (1990).
- [37] J.B. Hartle. The quasiclassical realms of this quantum universe. *Found. Phys.*, 41:982–1006, (2011).
- [38] C. Rovelli. Relational quantum mechanics. *Int. J. of Theor. Phys.*, 35:1637–1678, (1996).
- [39] C.A. Fuchs and R. Schack. Quantum-Bayesian coherence. *Rev. Mod. Phys.*, 85:1693–1715, (2013).
- [40] 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).
- [41] J.S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195–200, (1964).
- [42] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *J. Math. Mech.*, 17:59–87, (1967).
- [43] R. Colbeck and R. Renner. No extension of quantum theory can have improved predictive power. *Nat. Commun.*, 2:411, (2011).

- [44] R. Colbeck and R. Renner. A short note on the concept of free choice. Preprint at <http://arxiv.org/abs/1302.4446>, (2013).
- [45] H. Everett. “Relative state” formulation of quantum mechanics. *Rev. Mod. Phys.*, 29:454–462, (1957).
- [46] J.A Wheeler. Assessment of Everett’s “relative state” formulation of quantum theory. *Rev. Mod. Phys.*, 29:463, (1957).
- [47] B.S. DeWitt. Quantum mechanics and reality. *Phys. Today*, 23:155–165, (1970).
- [48] D. Deutsch. *The Fabric of Reality: The Science of Parallel Universes and Its Implications*. Allen Lane Science, (1997).
- [49] H.D. Zeh. On the interpretation of measurement in quantum theory. *Found. Phys.*, 1:69–76, (1970).
- [50] D. Albert and B. Loewer. Interpreting the many worlds interpretation. *Synthese*, 77:195–213, (1988).
- [51] G. Brassard and P. Raymond-Robichaud. Can free will emerge from determinism in quantum theory? In A. Suarez and P. Adams, editors, *Is Science Compatible with Free Will? Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience*, pages 41–61. Springer, (2013).
- [52] W.H. Zurek. Relative states and the environment: einselection, enviance, quantum Darwinism, and the existential interpretation. Preprint at <http://arxiv.org/abs/0707.2832>, (2007).
- [53] 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).
- [54] J. Butterfield and G.N. Fleming. Quantum theory and the mind. *PAS, Supplementary Volumes*, 69:113–173, (1995).
- [55] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, (2007).
- [56] L. Hardy. Reformulating and reconstructing quantum theory. Preprint at <http://arxiv.org/abs/1104.2066>, (2011).
- [57] G. Chiribella, G.M. D’Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, (2011).
- [58] L. Masanes and M.P. Müller. A derivation of quantum theory from physical requirements. *New J. Phys.*, 13:063001, (2011).
- [59] 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).
- [60] M.F. Pusey, J. Barrett, and T. Rudolph. On the reality of the quantum state. *Nat. Phys.*, 8:475–478, (2012).
- [61] 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).
- [62] H. Barnum and A. Wilce. Post-classical probability theory. In G. Chiribella and R.W. Spekkens, editors, *Quantum Theory: Informational Foundations and Foils*, pages 367–420. Springer, (2016).
- [63] J.A. Wheeler. The “past” and the “delayed-choice” double-slit experiment. In A.R. Marlow, editor, *Mathematical Foundations of Quantum Theory*, pages 9–48. Academic Press, (1978).

- [64] B.-G. Englert, M.O. Scully, G. Süssmann, and H. Walther. Surrealistic Bohm trajectories. *Z. Naturforsch.*, 47:1175–1186, (1992).
- [65] M. Correggi and G. Morchio. Quantum mechanics and stochastic mechanics for compatible observables at different times. *Ann. Phys.*, 296:371–389, (2002).
- [66] J. Kiukas and R.F. Werner. Maximal violation of Bell inequalities by position measurements. *J. Math. Phys.*, 51, (2010).
- [67] 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).
- [68] N. Gisin. Why Bohmian mechanics? One and two-time position measurements, Bell inequalities, philosophy and physics. *Entropy*, 20:105, (2018).
- [69] Y. Aharonov and L. Vaidman. Complete description of a quantum system at a given time. *J. Phys. A*, 24(10):2315, (1991).
- [70] A. Kent. Consistent sets yield contrary inferences in quantum theory. *Phys. Rev. Lett.*, 78:2874–2877, (1997).