



**HAL**  
open science

# **A contextually objective approach to the extended Wigner's friend thought experiment**

Maxime Federico, Philippe Grangier

## ► **To cite this version:**

Maxime Federico, Philippe Grangier. A contextually objective approach to the extended Wigner's friend thought experiment. 2024. <hal-03952265>

**HAL Id: hal-03952265**

**<https://hal.science/hal-03952265v1>**

Preprint submitted on 12 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# A contextually objective approach to the extended Wigner's friend thought experiment

Maxime Federico

Laboratoire Interdisciplinaire Carnot de Bourgogne,  
CNRS - Université Bourgogne Franche-Comté, UMR 6303, BP 47870, 21078 Dijon, France

Philippe Grangier

Laboratoire Charles Fabry, IOGS, CNRS, Université Paris Saclay, F91127 Palaiseau, France.

(Dated: January 10, 2023)

We present a discussion of the extended Wigner's friend thought experiment proposed by Frauchiger and Renner in [1]. We show by using various arguments, including textbook quantum mechanics and the ontological approach of Contexts, Systems, Modalities (CSM), that no contradiction arises if one admits that agents must agree on what is considered as a system and what is not. In such a contextually objective approach of quantum mechanics, the apparent contradiction is automatically removed. We also discuss why this mutual agreement between agents is already implicit in the standard formulations of quantum mechanics, and why removing it leads to inconsistencies.

## I. INTRODUCTION

Wigner's friend thought experiment has been proposed by Wigner in 1967 [2]. Recently, an extended version was proposed [1] in order to test how quantum mechanics describes agents who themselves use that theory to predict results of experiments, that may include other agents. This article triggered a large number of reactions [3] and it is not our purpose here to analyse all of them. We will rather summarize the argument, and show that the claimed contradiction between three arguments is removed not by giving up one of these three arguments, but rather by adding a fourth one. We will justify this additional argument by different approaches, including e.g. textbook quantum mechanics (QM), and the ontological approach of Contexts, Systems, Modalities (CSM). It may be concluded that the theory presented as QM by Frauchiger and Renner is not the actual QM, but a different theory that is not consistent indeed.

In Sections II and III we present the argument and the contradiction as introduced in [1], and in Sec. IV how to remove it by adding a simple fourth requirement. Then we discuss how to relate this argument to Hardy's paradox [4, 5] (Sec. V), to textbook QM as discussed by Laloë [6] (Sec. VI), and finally to the CSM point of view (Sec. VII). We conclude in Sec. VIII, and present some additional issues in Appendices 1 and 2.

## II. THE FRAUCHIGER-RENNER THOUGHT EXPERIMENT

The experimental protocol is the following (see Fig. 1): an agent  $\bar{F}$  uses a quantum random generator (quantum coin) with output  $r \in \{heads, tails\}$ . The coin quantum state is

$$|r\rangle = \frac{1}{\sqrt{3}} |h\rangle + \sqrt{\frac{2}{3}} |t\rangle, \quad (1)$$

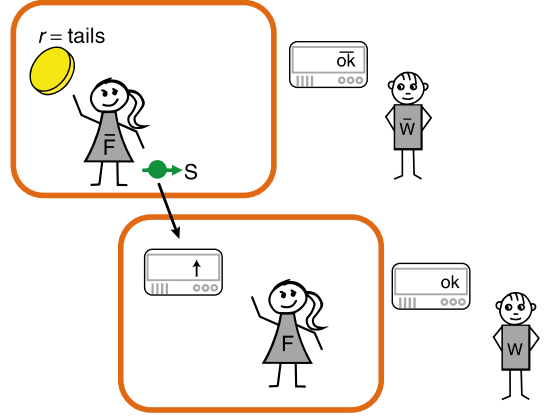


FIG. 1. Sketch of the thought experiment. Image adapted from [1].

where  $|h\rangle$  and  $|t\rangle$  are the states of *heads* and *tails* respectively. Probabilities are therefore  $1/3$  for *heads* and  $2/3$  for *tails*. During her measurement  $\bar{F}$  gets entangled with the coin, so the state of the total system including  $\bar{F}$  (and her lab) becomes

$$|\Psi\rangle = \frac{1}{\sqrt{3}} |h\rangle |\bar{F} : h\rangle + \sqrt{\frac{2}{3}} |t\rangle |\bar{F} : t\rangle, \quad (2)$$

using the notations [6] where  $|\bar{F} : h\rangle$  is the state of  $\bar{F}$  (and her lab) if she observes *heads* and  $|\bar{F} : t\rangle$  if she observes *tails*. Depending on the result, she prepares the spin of an electron in state  $|\downarrow\rangle$  if  $r = heads$  or in state  $|\rightarrow\rangle = \frac{1}{\sqrt{2}}(|\downarrow\rangle + |\uparrow\rangle)$  if  $r = tails$ . Therefore the total state becomes

$$|\Psi\rangle = \frac{1}{\sqrt{3}} |h\rangle |\bar{F} : h\rangle |\downarrow\rangle + \sqrt{\frac{2}{3}} |t\rangle |\bar{F} : t\rangle \frac{1}{\sqrt{2}} (|\downarrow\rangle + |\uparrow\rangle). \quad (3)$$

Then, she sends this electron to a second agent  $F$ , located in an other lab isolated from  $\bar{F}$  one (except during

the exchange of the electron assumed very short) who can measure the spin projection of that electron in the basis  $\{|\downarrow\rangle, |\uparrow\rangle\}$ . At this point, the total system including agents themselves is in the overall entangled state

$$|\Psi\rangle = \frac{1}{\sqrt{3}} \left[ |h\rangle |\bar{F}: h\rangle |\downarrow\rangle |F: \downarrow\rangle + |t\rangle |\bar{F}: t\rangle |\downarrow\rangle |F: \downarrow\rangle + |t\rangle |\bar{F}: t\rangle |\uparrow\rangle |F: \uparrow\rangle \right], \quad (4)$$

where  $|F: \uparrow\rangle$  and  $|F: \downarrow\rangle$  are the states of  $F$  (and her lab) observing spin *up* or *down*.

Two other separate agents  $\bar{W}$  and  $W$  can moreover perform measurements on the systems  $\bar{F}$  and  $F$  respectively with respect to the basis states

$$|\overline{OK}\rangle = \frac{1}{\sqrt{2}} \left[ |h\rangle |\bar{F}: h\rangle - |t\rangle |\bar{F}: t\rangle \right], \quad (5)$$

$$|\overline{fail}\rangle = \frac{1}{\sqrt{2}} \left[ |h\rangle |\bar{F}: h\rangle + |t\rangle |\bar{F}: t\rangle \right], \quad (6)$$

$$|OK\rangle = \frac{1}{\sqrt{2}} \left[ |\downarrow\rangle |F: \downarrow\rangle - |\uparrow\rangle |F: \uparrow\rangle \right], \quad (7)$$

$$|fail\rangle = \frac{1}{\sqrt{2}} \left[ |\downarrow\rangle |F: \downarrow\rangle + |\uparrow\rangle |F: \uparrow\rangle \right]. \quad (8)$$

Here we name the “sub-experiments” that composed the thought experiment by the name of agents corresponding to the different measurements  $\bar{F}$ ,  $F$ ,  $\bar{W}$  and  $W$ . However, it should be clear that  $\bar{F}$  and  $F$ , who were agents in the initial step, are then considered as systems from the point of view of  $\bar{W}$  and  $W$ . Therefore the states (5)-(8) written above contain agents, labs and measured systems on which  $\bar{W}$  and  $W$  are supposed to make quantum measurements, so that the whole ensemble of  $\bar{F}$  and  $F$  agents, labs and measured systems are projected into superposition states. This is clearly not feasible in practice, but the whole idea is to consider this operation feasible as a thought experiment, and to examine which consequences can be drawn from it.

Formally, considering the four situations where the active agents are either  $(\bar{F}, F)$ ,  $(\bar{W}, F)$ ,  $(\bar{F}, W)$ , or  $(\bar{W}, W)$ , by using and combining appropriate bases,  $|\Psi\rangle$  can be written equivalently as

$$|\Psi\rangle = \frac{1}{\sqrt{3}} \left[ |h\rangle |\bar{F}: h\rangle |\downarrow\rangle |F: \downarrow\rangle + |t\rangle |\bar{F}: t\rangle |\downarrow\rangle |F: \downarrow\rangle + |t\rangle |\bar{F}: t\rangle |\uparrow\rangle |F: \uparrow\rangle \right] \quad (9a)$$

$$= \sqrt{\frac{2}{3}} |\overline{fail}\rangle |\downarrow\rangle |F: \downarrow\rangle + \frac{1}{\sqrt{6}} \left[ |\overline{fail}\rangle - |\overline{OK}\rangle \right] |\uparrow\rangle |F: \uparrow\rangle \quad (9b)$$

$$= \frac{1}{\sqrt{6}} |h\rangle |\bar{F}: h\rangle \left[ |OK\rangle + |fail\rangle \right] + \sqrt{\frac{2}{3}} |t\rangle |\bar{F}: t\rangle |fail\rangle \quad (9c)$$

$$= \frac{1}{\sqrt{12}} |\overline{OK}\rangle |OK\rangle - \frac{1}{\sqrt{12}} |\overline{OK}\rangle |fail\rangle + \frac{1}{\sqrt{12}} |\overline{fail}\rangle |OK\rangle + \frac{\sqrt{3}}{2} |\overline{fail}\rangle |fail\rangle. \quad (9d)$$

Using these four expressions for the state vector  $|\Psi\rangle$ , one can deduce the following statements, which are close to the ones listed in [1]:

- $\bar{F}$  and  $F$  cannot get results *heads* and spin *up* [1.A]

- If  $\bar{W}$  gets  $\overline{OK}$ , then  $F$  gets spin *up* [1.B]

- If  $W$  gets  $OK$ , then  $\bar{F}$  gets *heads* [1.C]

- $\bar{W}$  and  $W$  can get  $\overline{OK}$  and  $OK$  with probability 1/12 [1.D]

If considered all true together, it is clear that these four statements lead to a contradiction, indeed from [1.D]  $\bar{W}$  and  $W$  can get  $\overline{OK}$  and  $OK$ , and in that case  $F$  and  $\bar{F}$  should get spin *up* and *heads* from [1.B] and [1.C], contradicting [1.A].

However this contradiction may likely be attributed to the undefined status of  $\bar{F}$  and  $F$ , who switch between being agents (able to make quantum measurements) and systems (being acted upon by quantum measurements). It is therefore required to clarify the definition and the role of agents. From here, several points of view exist and we are going to explore some of them.

### III. FRAUCHIGER & RENNER: GETTING A CONTRADICTION.

This point of view is based on three assumptions which allow authors of Ref [1] to show a “no-go theorem”.

**First assumption (Q) :** This defines what an agent must do to interpret and predict measurements. Suppose that a system  $S$  (external to the agent) is in state  $|\psi\rangle$  of a Hilbert space at time  $t_0$ ; an outcome  $x$  can be measured on  $S$  at time  $t$  with respect to a family of projectors  $\{\pi_x\}$  in Heisenberg representation. If  $\langle\psi|\pi_\xi|\psi\rangle = 1$ , then the agent can conclude that  $x = \xi$  at time  $t$ . From the point of view on this individual agent, this theory looks like quantum mechanics, and it will be used successively from the “subjective” point of view of different agents, in order to get a serie of statements.

First, if  $r = \textit{heads}$ , the spin is in state  $|\downarrow\rangle$ , and a measurement made by agent  $F$  with respect to the basis  $\{\pi_\downarrow = |\downarrow\rangle\langle\downarrow|, \pi_\uparrow = |\uparrow\rangle\langle\uparrow|\}$  gives  $\langle\downarrow|\pi_\downarrow|\downarrow\rangle = 1$ .

Therefore agents  $\bar{F}$  and  $F$  can conclude that if  $\bar{F}$  gets *heads*,  $F$  will get spin *down*; which is logically equivalent to: if  $F$  gets spin *up*,  $\bar{F}$  got *tails*, in agreement with eq. (9a). By considering successively the different pairs of agents, other statements can be obtained and lead to

- If  $F$  gets spin *up*, then  $\bar{F}$  gets *tails* [2.A]

- If  $\bar{W}$  gets  $\overline{OK}$ , then  $F$  gets spin *up* [2.B]

- If  $\bar{F}$  gets *tails*, then  $W$  gets *fail* [2.C]

- $\bar{W}$  and  $W$  can have  $\overline{OK}$  and  $OK$  [2.D]

which are either the same or logically equivalent statements compared to Section I (in the same order).

**Additional assumptions (C) and (S):** These two hypotheses stipulate that agents can trust predictions made by other agents, irrespective of their status, and that all predictions apply in the same universe, avoiding Everett’s multi world interpretation. It is under those three assumptions that a contradiction arises.

Indeed, if statement [2.D] is valid, [2.B] implies that  $F$  gets spin *up*, [2.A] that  $\bar{F}$  gets *tails*, and [2.C] that  $W$  gets *fail*. The last statement clearly contradicts the starting point that  $W$  gets *OK*. Using their assumptions as just shown, the authors of [1] can thus formulate the following no-go theorem:

$$(Q) + (C) + (S) \Rightarrow \text{Contradiction} \quad (10)$$

which has been formulated as the statement “quantum theory cannot consistently describe the use of itself” [1].

#### IV. REMOVING THE CONTRADICTION BY AN AGREEMENT BETWEEN AGENTS.

A major remark is that, though it is presented as equivalent to quantum mechanics (QM), the assumption (Q) of Frauchiger and Renner implies a very unusual use of the quantum formalism, where “self-proclaimed” agents may become systems for other agents. It means that the usual measurement postulate can hardly be applied in a consistent way [6]. A simple way to remove this problem, and also the contradiction, is to consider the additional assumption

(A) *All agents must agree on the definition of the quantum system to which they apply assumption (Q); as a consequence, no agent should be included in what another agent considers as the measured system.*

The second part of (A), i.e. that no agent should be included in what another agent considers as the measured system, is a joint consequence of (Q), stating that the agents apply QM to a system external to themselves, and of the first part of (A), i.e. the agents agreement on the definition of the system. Then the quantum system is the same for all agents, and contains no agent.

Given (A), it is clear that the previous contradiction vanishes since the four statements cannot be true together:

(i) if  $\bar{F}$  and  $F$  are agents, the first statement is true, but then the other three make no sense since  $\bar{F}$  and  $F$  cannot behave as (superposable) systems from the point of view of  $\bar{W}$  and  $W$ ;

(ii) if  $\bar{F}$  and  $F$  are systems, then the four lines correspond to the agents  $\bar{W}$  and  $W$  making four incompatible measurements on the global system composed of  $\bar{F}$ ,  $F$ , the coin and the spin. Incompatible measurements cannot be true together as it is usual in QM, such that there is no contradiction.

Therefore, with all of the four assumptions, we obtain

$$(Q) + (C) + (S) + (A) \Rightarrow \text{No contradiction} \quad (11)$$

which means that assumption (A) restricts (Q) to some safe range where no contradiction arises. This could be summarized as : “Objective quantum mechanics can describe the use of itself”, where “objective” is meant as requiring a mutual agreement between all the agents.

In the following we will develop these arguments, and in particular give more justifications and illustrations of the points (i) and (ii) above, as well as some intermediary situations. We will also see that assumption (A) can take different forms depending on which point of view we adopt, but always removes the contradiction.

#### V. LOOKING AT $\bar{W}$ AND $W$ AS AGENTS: LINK TO HARDY’S PARADOX

The Frauchiger and Renner (F&R) paradox is, as it was pointed out by Bub [4], not only a new formulation of Wigner’s friend experiment but also a twisted version of Hardy’s paradox originally presented to show a similar result as Bell’s theorem [7]. To do that, Hardy [5] used a Mach-Zehnder interferometer for electrons and positrons in order to get four entangled states which are similar to those of the F&R Gedankenexperiment. The difference arises only with the way one imposes that these four statements must be true together. In the following, we will assume that a hidden variable exists to describe the various states of eq. (9), in agreement with the formulation proposed in [5].

Let  $\lambda$  be the hidden variable which describes the states before measurements and assume that QM is a local realist theory. It means that results of measurements do not depend on the choice of measurement done by other agents. We introduce notations  $\bar{F}(h, \lambda) = 1$  if  $\bar{F}$  measures the state  $|h\rangle$ , and  $\bar{F}(h, \lambda) = 0$  otherwise. The same notations will be applied to other agents and their respective measurable states. The statements obtained in the preceding sections can be recovered using the expressions of  $|\Psi\rangle$  in the different bases. Indeed, from (9a) and for every experiment described by a given value of  $\lambda$ ,

$$\bar{F}(h, \lambda)F(\uparrow, \lambda) = 0. \quad (12)$$

From (9b) we additionally have

$$\text{if } \bar{W}(\overline{OK}, \lambda) = 1 \text{ then } F(\uparrow, \lambda) = 1, \quad (13)$$

from (9c)

$$\text{if } W(OK, \lambda) = 1 \text{ then } \bar{F}(h, \lambda) = 1, \quad (14)$$

and from (9d)

$$\bar{W}(\overline{OK}, \lambda)W(OK, \lambda) = 1 \text{ for } 1/12\text{th of experiments,} \quad (15)$$

which can be summed up in the four following statements completely equivalent to [1.A]-[1.D]:

- $\overline{F}(h, \lambda)$  and  $F(\uparrow, \lambda)$  cannot occur at the same time
- $\overline{W}(\overline{OK}, \lambda)$  is true  $\Rightarrow F(\uparrow, \lambda)$  is true
- $W(OK, \lambda)$  is true  $\Rightarrow \overline{F}(h, \lambda)$  is true
- $\overline{W}(\overline{OK}, \lambda)$  and  $W(OK, \lambda)$  occur one time out of 12

One can see that even if the assumptions and consequently the formalism are not exactly the same, the contradiction arises in the same way as when the four states of eq. (9) are considered simultaneously true. In that case inspired by Hardy's paradox, this is due to the hidden variable hypothesis. Therefore, it appears that in both situation, the contradiction is not due to QM itself but to the assumption that all conclusions drawn from eq. (9) are simultaneously true. This explains why the use of assumption (A) removes the contradiction, by making clear that they are incompatible in a quantum sense.

## VI. TEXTBOOK QM: PROJECTIVE MEASUREMENTS DEFINE AGENTS

Another approach to the F&R contradiction can be found in [6]. This point of view stipulates that defining agents is equivalent to define where the projective measurements are done, i.e. where the state is projected. For instance, without any projection, the state of the total system after "measurements" by  $\overline{F}$ ,  $F$ ,  $\overline{W}$  and  $W$  should be written as

$$\begin{aligned} |\Psi\rangle = & \frac{1}{\sqrt{12}} |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle |OK\rangle |W : OK\rangle \\ & - \frac{1}{\sqrt{12}} |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle |fail\rangle |W : fail\rangle \\ & + \frac{1}{\sqrt{12}} |fail\rangle |\overline{W} : fail\rangle |OK\rangle |W : OK\rangle \\ & + \frac{\sqrt{3}}{2} |fail\rangle |\overline{W} : fail\rangle |fail\rangle |W : fail\rangle. \end{aligned} \quad (16)$$

In this state, all observers are entangled as if we had considered Everett's multi-world interpretation. Then all possibilities can be investigated by choosing where the projections are made. The calculations are detailed in [6] and Appendix 2, here we simply describe qualitatively the results.

First, we consider  $\overline{F}$  and  $F$  as agent (point (i) of Section IV), and project states in agreement with their measurements. The result (see Appendix 2) is that whatever  $\overline{W}$  and  $W$  measure, they cannot obtain informations on spin state or coin state. Therefore, the statements [2.B] and [2.C] of Section III cannot be formulated and the contradiction does not arise.

In the second option which considers  $\overline{F}$  and  $F$  as systems, and only  $\overline{W}$  and  $W$  as agents,  $\overline{W}$  cannot say anything about the spin state, therefore using this way of defining agents and writing states after measurements, also prevent from any contradiction.

It is also instructive to consider the slightly modified situation where  $\overline{F}$  can "protect" her result by using an other qubit stored in her lab, in such a way that it escapes to  $\overline{W}$ 's measurement (referred as "hidden qubit" [6]). If  $\overline{F}$  observes "heads", the hidden qubit is in state  $|\overline{G} : h\rangle$  and in state  $|\overline{G} : t\rangle$  if she observes "tails". The overall state can thus be written as

$$\begin{aligned} |\Psi\rangle = & \frac{1}{\sqrt{3}} \left[ |h\rangle |\overline{F} : h\rangle |\overline{G} : h\rangle |\downarrow\rangle |F : \downarrow\rangle \right. \\ & + |t\rangle |\overline{F} : t\rangle |\overline{G} : t\rangle |\downarrow\rangle |F : \downarrow\rangle \\ & \left. + |t\rangle |\overline{F} : t\rangle |\overline{G} : t\rangle |\uparrow\rangle |F : \uparrow\rangle \right]. \end{aligned} \quad (17)$$

In the following, to simplify notations, we will replace  $|h\rangle |\overline{F} : h\rangle$  by  $|h\rangle$ ,  $|t\rangle |\overline{F} : t\rangle$  by  $|t\rangle$ ,  $|\downarrow\rangle |F : \downarrow\rangle$  by  $|\downarrow\rangle$  and  $|\uparrow\rangle |F : \uparrow\rangle$  by  $|\uparrow\rangle$  since agents (and their labs) are always in the state corresponding to the outcome they have measured. We also condense the hidden qubit notation in  $|h_G\rangle$  and  $|t_G\rangle$ . The state becomes

$$|\Psi\rangle = \frac{1}{\sqrt{3}} \left[ |h\rangle |h_G\rangle + |t\rangle |t_G\rangle \right] |\downarrow\rangle + \frac{1}{\sqrt{3}} |t\rangle |t_G\rangle |\uparrow\rangle, \quad (18)$$

and the bases used by  $\overline{W}$  and  $W$  to perform their respective measurements are

$$|\overline{OK}\rangle = \frac{1}{\sqrt{2}} (|h\rangle - |t\rangle), \quad |fail\rangle = \frac{1}{\sqrt{2}} (|h\rangle + |t\rangle), \quad (19)$$

$$|OK\rangle = \frac{1}{\sqrt{2}} (|\downarrow\rangle - |\uparrow\rangle), \quad |fail\rangle = \frac{1}{\sqrt{2}} (|\downarrow\rangle + |\uparrow\rangle). \quad (20)$$

The global state can then be put in the form

$$\begin{aligned} |\Psi\rangle = & |\overline{OK}\rangle |OK\rangle \frac{1}{\sqrt{12}} |h_G\rangle \\ & + |\overline{OK}\rangle |fail\rangle \left[ \frac{1}{\sqrt{12}} |h_G\rangle - \frac{1}{\sqrt{3}} |t_G\rangle \right] \\ & + |fail\rangle |OK\rangle \frac{1}{\sqrt{12}} |h_G\rangle \\ & + |fail\rangle |fail\rangle \left[ \frac{1}{\sqrt{12}} |h_G\rangle + \frac{1}{\sqrt{3}} |t_G\rangle \right]. \end{aligned} \quad (21)$$

As it was pointed out in [6], the state  $|t_G\rangle$  is only correlated with the result  $|fail\rangle$  which is not surprising since when  $\overline{F}$  sends a spin in the state  $|\rightarrow\rangle$ , it is orthogonal to  $|\overline{OK}\rangle$ . It means that the hidden qubit keeps a memory of  $\overline{F}$ 's measurement and therefore all statements derived before do not hold anymore. We emphasize that there is no need to read out the value of the hidden qubit: as in an interference experiment, it is enough that the "which path information" is stored somewhere to forbid the quantum superposition.

It is also interesting to investigate different scenarii for the two accessible states of the hidden qubit e.g. if  $\langle h_G | t_G \rangle = 1$  i.e.  $|h_G\rangle$  and  $|t_G\rangle$  are the same state, denoted for instance  $|G\rangle$ , then the overall system is in the

following state

$$\begin{aligned}
|\Psi\rangle &= \frac{1}{\sqrt{12}} |\overline{OK}\rangle |OK\rangle |G\rangle - \frac{1}{\sqrt{12}} |\overline{OK}\rangle |fail\rangle |G\rangle \\
&+ \frac{1}{\sqrt{12}} |fail\rangle |OK\rangle |G\rangle + \frac{\sqrt{3}}{2} |fail\rangle |fail\rangle |G\rangle,
\end{aligned} \tag{22}$$

which is the same as state (9d) without the hidden qubit.  $|G\rangle$  can be factored out and the hidden qubit plays no role since it is no more entangled with  $\overline{F}$ 's results. Otherwise, if  $\langle h_G | t_G \rangle = 0$  i.e. the qubit states are orthogonal and the overall state is identical to (21). If we project this global state onto each state of the hidden qubit

$$\begin{aligned}
\langle h_G | \Psi \rangle &= \frac{1}{\sqrt{12}} \left[ |\overline{OK}\rangle |OK\rangle + |\overline{OK}\rangle |fail\rangle \right. \\
&\quad \left. + |fail\rangle |OK\rangle + |fail\rangle |fail\rangle \right] \\
&= \frac{1}{\sqrt{3}} |h\rangle | \overline{F} : h \rangle \frac{|OK\rangle + |fail\rangle}{\sqrt{2}},
\end{aligned} \tag{23}$$

$$\begin{aligned}
\langle t_G | \Psi \rangle &= \frac{1}{\sqrt{3}} \left[ |fail\rangle |fail\rangle - |\overline{OK}\rangle |fail\rangle \right] \\
&= \sqrt{\frac{2}{3}} |t\rangle | \overline{F} : t \rangle |fail\rangle,
\end{aligned} \tag{24}$$

we recover the situation where  $\overline{F}$  is an agent, and  $\overline{W}$  cannot project her in a superposition.

This calculation has two consequences: first, it gives a way to make a transition between the different options, by allowing  $\langle h_G | t_G \rangle$  to take any value between 0 and 1. We emphasize that this calculation makes sense in a framework where the only agents are  $\overline{W}$  and  $W$ . If  $\langle h_G | t_G \rangle = 0$  then  $\overline{F}$  and  $F$  should be considered as decohered systems, as they appear in the usual theory of decoherence after tracing out the ‘‘ancilla’’ qubits. Second, it provides a ‘‘toy model’’ for a situation where a system actually does not behave as such: if  $\overline{W}$  tries to make a projective measurement on  $\overline{F}$ , including her lab and all the surrounding environment, it is enough that a single qubit escapes from the action of  $\overline{W}$  to make the projective measurement failed. When  $\overline{F}$  is a macroscopic system, it can be considered impossible for  $\overline{W}$  to get full control on every single qubit entangled with  $\overline{F}$ ; then considering  $\overline{F}$  as a system makes no sense.

Following the idea of Section IV, this can be seen as a reformulation of assumption (A):

(A) *All agents must agree on the definition of the quantum system to which they apply assumption (Q), that is to say where the state projection is done; as a consequence, no agent should be included in what another agent considers as the measured system.*

## VII. CONTEXTS, SYSTEMS AND MODALITIES: THE CSM APPROACH.

Here, we want to show that in the framework of the CSM (Contexts, Systems and Modalities) approach, developed by A. Auffèves and P. Grangier [8–11], the assumption (A) introduced and discussed above is no more an assumption but a theorem. We recall some basics of CSM which starts with a new physical ontology (i.e. a definition of all objects that are described by the theory), by considering three entities:

- System = subpart of the world, isolated well enough to have physical properties that can be studied.
- Context = all other physical systems external and possibly in contact with the studied one (ex: measurement apparatus, labs, etc); in CSM, the context corresponds to a classically defined measurement setup, and it acts like an interface between the system and the observer, see Fig. 2.
- Modalities = the set of answers (or values) that can be predicted with certainty and obtained repeatedly for a given system in a given context.

These definitions are bound together by the following rule (or axiom): in quantum mechanics, modalities are attributed jointly to the system and the context. It means that the quantum properties of a system do not ‘‘belong’’ to the system alone, but to the system within a context.

In order to give an operational content to CSM, we need a second axiom called Quantization Principle:

- (i) For each well-defined system and context, there is a discrete number  $N$  of mutually exclusive modalities. This number  $N$  depends on the system, but does not depend on any particular context.
- (ii) Modalities, when defined in different contexts, are generally not mutually exclusive, and they are said to be ‘‘incompatible’’.

CSM is based on a physically realistic point of view, i.e. physics is applied to objects which exist independently from any observer. However, the ‘‘object’’ for CSM is a system within a context, which makes it highly non-classical. As it can be shown on Figure 2(c), in CSM the observer is outside of the context, illustrating again that the (objective) state of the context must be taken into account in defining the modality. Such a ‘‘contextual objectivity’’ [8] is quite different from the classical ‘‘absolute objectivity’’ of Figure 2(a), but it fully agrees with quantum reality as deduced from empirical evidence.

Now, coming back to the original problem, CSM was hidden so far but in fact already there. In order to define systems and contexts according to CSM, we have to avoid mixing notations between agents, labs and experimental set ups. First, as it was pointed out before,

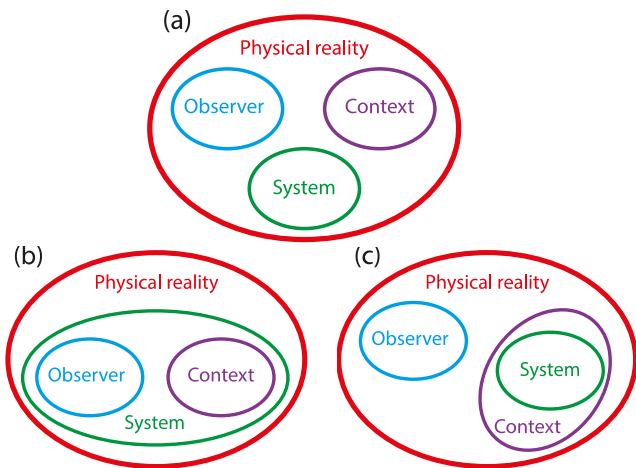


FIG. 2. Graphic representations of different ontologies: (a) Classical ontology: the observer can know the “real” physical properties of the system, and the context is only used as an auxiliary tool for measurements. (b) Usual quantum ontology: through successive “entangling” interactions and unitary evolution, the system will include the context, and also (ultimately) the observer or “agent”, whose status may be problematic. (c) CSM ontology: the context appears always between the system and the observer, and definite values of the relevant physical properties (modalities) are attributed jointly to the system and the context.

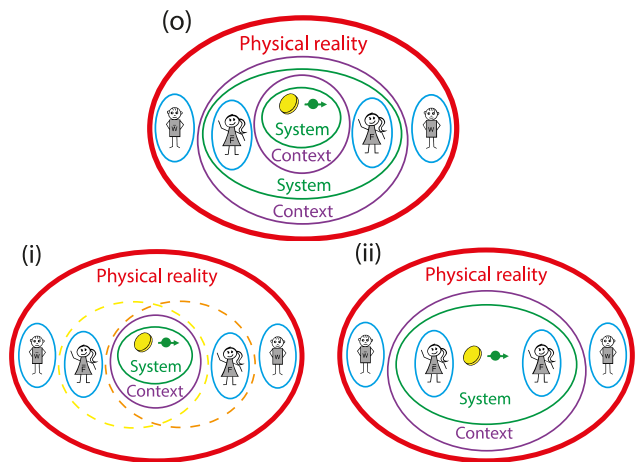


FIG. 3. (o) Naive way to see Wigner’s friends thought experiment. Such a picture is impossible for CSM, because a (quantum) system cannot include a (classical) context. (i) Case where  $\bar{F}$  and  $F$  are agents making quantum measurements on systems with their “own” classical contexts. (ii) Second option where  $\bar{F}$  and  $F$  are considered as systems.

there is only one ultimate reality which imposes that a context cannot be included in a system and vice versa (see Figure 3 (o)). For consistency, different agents must agree on the definition of systems and contexts which is exactly the purpose of assumption (A), formulated according to CSM as:

(A) All agents must agree on the definition of the system and context to which they apply assumption (Q); as a consequence, no agent should be included in what another agent considers as the measured system.

Again, several cases are possible, corresponding e.g. to point (i) and (ii) of Section IV, and are drawn in Figure 3. In case (i),  $\bar{F}$  and  $F$  perform quantum measurements on their systems within their respective contexts. Then  $\bar{W}$  and  $W$  cannot make quantum measurements on them (composite systems in yellow and orange dash lines of Figure 3), and the F&R reasoning does not hold. In case (ii),  $\bar{F}$  and  $F$  are considered as systems so  $\bar{W}$  and  $W$  can perform quantum measurements on those systems in two different contexts on each side (i.e. projecting  $\bar{F}$  or  $F$  in superposed or non-superposed states). The modalities defined in these different contexts are incompatible and thus not simultaneously true, no contradiction can arise.

To conclude, we have shown that within the CSM interpretation, the F&R contradiction is removed without additional assumption because CSM is already an “contextually objective” theory [8], where agents cannot become systems for other agents.

## VIII. CONCLUSION

To conclude, we have presented several ways to escape the “no-go theorem” introduced by F&R. Our arguments can be summed up by “**Contextually objective quantum mechanics can describe the use of itself**”.

This statement does not allow every single world interpretation of quantum mechanics to be self-consistent, but only those which rely on some form of objectivity. A consequence is that the (single world) quantum rules must introduce a clear distinction between agents and systems, embedded in the additional assumption (A). Then we have shown that in the framework of CSM approach, no F&R contradiction can arise since CSM already contains (A) as a theorem, deduced from the central concept of contextual objectivity [8] which governs quantum measurements in the CSM framework.

## APPENDIX 1: ABOUT THE TIME EVOLUTION OF THE EXPERIMENT

In their article [1], F&R insist on the temporal order of measurements made by successive agents, associated to changing the status of  $\bar{F}$  and  $F$  between system and context. In the CSM interpretation, changing the agents status in a same experiment cannot occur, therefore the chronology is not relevant. It could be if we had adopted Everett’s point of view also called “multi-world”; but this interpretation is in conflict with assumption (S), so the F&R no-go theorem also does not apply.

More precisely, through the usual quantum or CSM point of view, an ideal measurement takes place as follows: (a) entanglement of the system with the pointer (also quantum) and then (b) reading of the pointer, which brings the result at the single classical context level. From the Everett's point of view, there is no single classical context level but many possible results which are associated to different universe branches. Then it is, in principle, possible to reverse the evolution of these branches, rewriting the history as if no measurement were done, and to make another different measurement, creating again new branches. So in Everett's point of view, temporal order makes sense to describe first a measurement by  $\overline{F}$ , and then its "erasure" by  $\overline{W}$ , followed by a new measurement in the  $(\overline{OK}; \overline{fail})$  basis.

In CSM, the measurement can be reversed only if it has not reached the context level, i.e. between steps (a) and (b). When the measurement is over (step b), the result is macroscopic and unique and the modalities corresponding to other results in other contexts cannot co-exist. Explicitly, if  $\overline{F}$  and  $F$  are superposable systems, we can measure them in superposed bases

$$\overline{\mathcal{S}} = \{\overline{OK}; \overline{fail}\}, \quad \mathcal{S} = \{OK; fail\}$$

or in non superposed ones

$$\overline{\mathcal{N}} = \{heads; tails\}, \quad \mathcal{N} = \{up; down\}.$$

The four situations we have discussed correspond to four combined measurements  $(\overline{\mathcal{N}}, \mathcal{N})$ ,  $(\overline{\mathcal{N}}, \mathcal{S})$ ,  $(\overline{\mathcal{S}}, \mathcal{N})$ ,  $(\overline{\mathcal{S}}, \mathcal{S})$ . In that case, temporal order is not relevant since only one of these four measurements can be made, as in Hardy's paradox, such that there is a unique objective macroscopic world in which QM is consistent.

We note also that the CSM point of view makes a clear distinction between the usual "pre-measurement" stage, that is entangling the system with a probe in a reversible way, and the actual irreversible measurement that brings the result to the context level. It can be told that the F&R paradox fails because it mixes these two stages in an inappropriate way, as explained in details in [12].

## APPENDIX 2: ADDING A HIDDEN QUBIT

Another approach to the F&R contradiction can be found in [6]. This point of view stipulates that defining agents is equivalent to define where projective measurements are done, that is to say, where mathematically the state is projected. For instance, without any projection, the state of the total system after "measurements" by  $\overline{F}$ ,

$F$ ,  $\overline{W}$  and  $W$  is

$$\begin{aligned} |\Psi\rangle &= \frac{1}{\sqrt{12}} |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle |OK\rangle |W : OK\rangle \\ &- \frac{1}{\sqrt{12}} |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle |fail\rangle |W : fail\rangle \\ &+ \frac{1}{\sqrt{12}} |\overline{fail}\rangle |\overline{W} : \overline{fail}\rangle |OK\rangle |W : OK\rangle \\ &+ \frac{\sqrt{3}}{2} |\overline{fail}\rangle |\overline{W} : \overline{fail}\rangle |fail\rangle |W : fail\rangle. \end{aligned} \quad (25)$$

In this state, all observers are entangled as if we had considered Everett's multi-world interpretation.

Then all possibilities can be investigated for the projections. First, we consider  $\overline{F}$  and  $F$  as agent (point (i) of Section IV). Therefore, they have to project states in agreement with their measurements. According to  $\overline{F}$ , two states are available

$$|\Psi\rangle_{tails} = |t\rangle |\overline{F} : t\rangle \frac{1}{\sqrt{2}} (|\downarrow\rangle + |\uparrow\rangle), \quad (26)$$

$$|\Psi\rangle_{heads} = |h\rangle |\overline{F} : h\rangle |\downarrow\rangle. \quad (27)$$

If then we consider  $F$ 's projections which lead to many cases depending of the result she gets but also of  $\overline{F}$ 's results, we have the states

$$|\Psi_{\downarrow}\rangle_{tails} = |t\rangle |\overline{F} : t\rangle |\downarrow\rangle |F : \downarrow\rangle, \quad (28)$$

$$|\Psi_{\uparrow}\rangle_{tails} = |t\rangle |\overline{F} : t\rangle |\uparrow\rangle |F : \uparrow\rangle, \quad (29)$$

$$|\Psi_{\downarrow}\rangle_{heads} = |h\rangle |\overline{F} : h\rangle |\downarrow\rangle |F : \downarrow\rangle, \quad (30)$$

or equivalently written in the other bases

$$|\Psi_{\downarrow}\rangle_{tails} = \frac{1}{2} \left[ |\overline{fail}\rangle - |\overline{OK}\rangle \right] \left[ |fail\rangle + |OK\rangle \right], \quad (31)$$

$$|\Psi_{\uparrow}\rangle_{tails} = \frac{1}{2} \left[ |\overline{fail}\rangle - |\overline{OK}\rangle \right] \left[ |fail\rangle - |OK\rangle \right], \quad (32)$$

$$|\Psi_{\downarrow}\rangle_{heads} = \frac{1}{2} \left[ |\overline{fail}\rangle + |\overline{OK}\rangle \right] \left[ |fail\rangle - |OK\rangle \right]. \quad (33)$$

Here we see that all these states are products; whatever  $\overline{W}$  and  $W$  measure, they cannot obtain informations on spin state or coin state. Therefore, statements [2.B] and [2.C] of Section III are wrong and no contradiction can appear.

The second option is to consider  $\overline{F}$  and  $F$  as systems, and  $\overline{W}$  and  $W$  as agents (point (ii) of Section IV). Before they perform measurements, the system is in state (4). Then, two outcomes are possible for  $\overline{W}$  and give the states

$$\begin{aligned} |\Psi\rangle_{\overline{OK}} &= -|\uparrow\rangle |F : \uparrow\rangle |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle \\ &= \frac{1}{\sqrt{2}} \left[ |OK\rangle - |fail\rangle \right] |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle, \end{aligned} \quad (34)$$

$$\begin{aligned} |\Psi\rangle_{\overline{fail}} &= \left[ 2|\downarrow\rangle |F : \downarrow\rangle + |\uparrow\rangle |F : \uparrow\rangle \right] |\overline{fail}\rangle |\overline{W} : \overline{fail}\rangle \\ &= \left[ \frac{3}{\sqrt{2}} |fail\rangle |W : fail\rangle + \frac{1}{\sqrt{2}} |OK\rangle |W : OK\rangle \right] \\ &\otimes |\overline{fail}\rangle |\overline{W} : \overline{fail}\rangle. \end{aligned} \quad (35)$$

We see again, that the result  $\overline{OK}$  is only correlated to spin *up*, and that  $\overline{W}$  cannot predict anything on  $W$ 's measurement since both results *OK* and *fail* are still available. In the second option where  $\overline{W}$  gets  $\overline{fail}$ , no predictions can be obtained about the spin state. If we furthermore add the projection made by  $W$  we obtain

$$|\Psi_{OK}\rangle_{\overline{OK}} = |OK\rangle |W : OK\rangle |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle, \quad (36)$$

$$|\Psi_{fail}\rangle_{\overline{OK}} = -|fail\rangle |W : fail\rangle |\overline{OK}\rangle |\overline{W} : \overline{OK}\rangle, \quad (37)$$

$$|\Psi_{OK}\rangle_{\overline{fail}} = |OK\rangle |W : OK\rangle |\overline{fail}\rangle |\overline{W} : \overline{fail}\rangle, \quad (38)$$

$$|\Psi_{fail}\rangle_{\overline{fail}} = |fail\rangle |W : fail\rangle |\overline{fail}\rangle |\overline{W} : \overline{fail}\rangle, \quad (39)$$

or equivalently expressed in  $\overline{F}$  and  $F$  basis

$$|\Psi_{OK}\rangle_{\overline{OK}} = \frac{1}{2} [ |h\rangle - |t\rangle ] [ |\downarrow\rangle - |\uparrow\rangle ], \quad (40)$$

$$|\Psi_{fail}\rangle_{\overline{OK}} = \frac{1}{2} [ |h\rangle + |t\rangle ] [ |\uparrow\rangle - |\downarrow\rangle ], \quad (41)$$

$$|\Psi_{OK}\rangle_{\overline{fail}} = \frac{1}{2} [ |h\rangle - |t\rangle ] [ |\downarrow\rangle + |\uparrow\rangle ], \quad (42)$$

$$|\Psi_{fail}\rangle_{\overline{fail}} = \frac{1}{2} [ |h\rangle + |t\rangle ] [ |\downarrow\rangle + |\uparrow\rangle ]. \quad (43)$$

Again, since every state is a product state, statements [2.B] and [2.C] cannot be true and the contradiction disappears. Therefore this way to define agents and to write states after measurements, also erases the contradiction.

- 
- [1] D. Frauchiger and R. Renner, “Quantum theory cannot consistently describe the use of itself” *Nature Comm.* 9, 3711 (2018); a previous version is “Single-world interpretations of quantum theory cannot be self-consistent”, arXiv:1604.07422v1 (2016).
- [2] E.P. Wigner, “Symmetries and Reflections”, chapter Remarks on the Mind-Body Question, pages 171-184. Indiana University Press, 1967.
- [3] A search on arxiv in December 2022 gives more than 50 articles commenting or criticizing [1].
- [4] J. Bub, “In a defense of a “Single World” interpretation of quantum mechanics”, arXiv:1804.03267v1 (2018)
- [5] L. Hardy, “Quantum Mechanics, Local Realistic Theories, and Lorentz-Invariant Realistic Theories.”, *Physical Review Letters* 68, 2981 2984 (1992).
- [6] F. Laloë, “Can quantum mechanics be considered consistent? a discussion of Frauchiger and Renner’s argument.”, arXiv:1802.06396v2 (2018).
- [7] J.S. Bell, On the Einstein-Podolski-Rosen paradox, *Physics* 1, 195 (1964).
- [8] P. Grangier, “Contextual objectivity: a realistic interpretation of quantum mechanics”, *European Journal of Physics* 23:3, 331 (2002) [arXiv: quant-ph/0012122].
- [9] A. Auffèves and P. Grangier, “Contexts, Systems and Modalities: a new ontology for quantum mechanics”. *Found. Phys.* 46, 121 (2016), eprint arXiv:1409.2120 [quant-ph] (2014).
- [10] A. Auffèves and P. Grangier “Deriving Born’s rule from an Inference to the Best Explanation”, *Found Phys* 50, 1781 (2020) [arXiv:1910.13738 quant-ph]
- [11] P. Grangier, “Revisiting Quantum Mysteries”, <https://arxiv.org/abs/2105.14448>
- [12] M. Zukowski and M. Markiewicz, “Physics and Metaphysics of Wigner’s Friends: Even Performed Premeasurements Have No Results”, *Phys. Rev. Lett.* 126, 130402 (2021)