nature communications

Matters arising

https://doi.org/10.1038/s41467-024-47172-0

Reply to: Quantum mechanical rules for observed observers and the consistency of quantum theory

Received: 15 September 2022

Accepted: 22 March 2024

Published online: 09 April 2024

Check for updates

Lídia del Rio [1] & Renato Renner [1]

REPLYING TO A. P. Polychronakos *Nature Communications* https://doi.org/10.1038/s41467-024-47170-2 (2024)

Wigner's friend experiment

In 1967, Eugene Wigner proposed a thought experiment to test the range of validity of quantum theory[1]. The experiment features two agents, Wigner and his friend, whom we call Alice (Fig. 1). Alice works in a perfectly isolated lab, and her task is to measure an observable X of a quantum system S. Wigner's task is to analyse the same experiment from the outside, treating Alice's entire lab as a quantum system. Crucial to this setup is that, when applying quantum theory, the two agents split the world differently into quantum and classical parts—i.e., they choose different 'Heisenberg cuts'[2]. Alice only models S as a quantum system and treats the outcome of her measurement X as part of the classical domain, yielding a definite value, x. In contrast, Wigner models Alice's entire lab as part of the quantum domain, including Alice herself and her memory of the measurement outcome. Hence, for Wigner, Alice's measurement is a reversible entangling operation between her and S, and x has no definite value. Alice and Wigner's conclusions regarding x are thus different, although, at this point, they are not strictly contradictory.

The FR experiment

In 2018, Daniela Frauchiger and one of us (Renner) proposed an extension of Wigner's thought experiment, often referred to as the 'FR experiment'[3] (see also[4] for a similar proposal). It involves a group of four agents tasked with making predictions about each other's measurements. Crucially, each agent applies the Heisenberg cut subjectively and may choose to model some of the other agents as quantum systems. All agents share the same initial information about the experimental setting and protocol, but during the actual run of the experiment, each agent may have access to additional data based on their local observations. Each agent analyses the experiment from their perspective using the same reasoning rules described by (Q), (C), and (S) below. The key insight of the experiment is that the agents reach contradictory conclusions; this result was framed as the no-go theorem[3] restated here. For an in-depth analysis of the FR experiment in the light of different interpretations of quantum theory, we refer to[5,6].

Theorem 1 [3]. No physical theory where it is possible to model the FR experiment is compatible with the reasoning rules (Q), (C), and (S).

Considered individually, each reasoning rule appears intuitive and unproblematic; nonetheless, Theorem 1 asserts that they are contradictory. For simplicity, we elucidate these reasoning rules by describing their use by an agent, Alice, who is deriving statements about the outcome x of a measurement specified by an observable X.

(Q) *Validity of quantum theory at the relevant scales:* Suppose that the observable X is defined on a quantum system S around Alice (i.e., Alice is not herself part of S). Alice may then apply the standard formalism of quantum theory to describe S and calculate the probabilities for the potential measurement outcomes. In particular, if this analysis yields that the outcome is x with probability 1, Alice can conclude "I am certain that $X = x$." (For concreteness, the 'standard formalism' can be the four quantum postulates of Nielsen & Chuang[7], Section 2.2, applied to the system S and its subsystems).

(C) *Consistency among agents:* Let Bob be another agent who reasons about the same measurement X. If Alice has deduced, "I am certain that Bob has concluded that he is certain that $X = x$" then Alice can conclude, "I am certain that $X = x$."

(S) *Single outcomes:* If Alice has concluded both "I am certain that $X = x$" and "I am certain that $X = x'$" for $x \neq x'$ then she considers that a contradiction.

A simple experiment to test reasoning rules

The idea behind rules (Q), (C), and (S) is that they correspond to the building blocks of reasoning that physicists naturally employ to analyze standard experiments. To see this, we introduce a simple experimental setup, which we term the 'Learned Prediction Experiment' (Box 1): An agent, Alice, learns a prediction from another agent, Bob, where both agents use the reasoning rules above, as shown in Fig. 2. An addition relevant to the later discussion is that a third agent, Wigner, may measure Bob's lab at some point in the experiment. As we will see, different proposals to circumvent Theorem 1 will also lead to different conclusions about this experiment.

[1]Institute for Theoretical Physics, ETH Zurich, Zurich, Switzerland. e-mail: lidia@phys.ethz.ch; renner@ethz.ch

https://doi.org/10.1038/s41467-024-47172-0

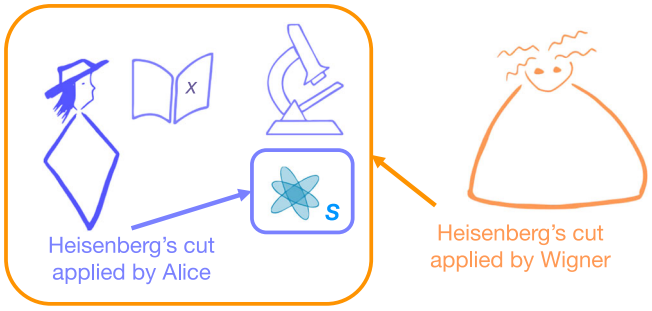


Fig. 1 | Wigner's friend experiment. The experiment concerns two quantum mechanics, Alice and Wigner. Alice measures a system S and records an outcome x. Alice applies the Heisenberg cut around system S: she treats S as a quantum system but regards herself, her notebook, and the outcome x as classical. Wigner analyses the situation from the outside, applying the Heisenberg cut around Alice's entire isolated lab: he models Alice, her notebook, her measurement instruments, and everything else in her lab as quantum systems undergoing a global reversible evolution.

Criticism of Theorem 1

A large number of recent works have criticised Theorem 1—not its technical statement or proof, but rather the nature of its assumptions, reasoning rules (Q), (C), and (S). This, of course, is precisely the point of the no-go theorem—it asserts that the combination of these reasoning rules leads to contradictions. Nonetheless, Theorem 1 is only of interest if the reasoning rules (Q), (C), and (S) accurately capture the way we reason about physical experiments. Works criticising theorem typically claim this is not true for some of these reasoning rules.

To warm up, we illustrate this with a common criticism, which was eloquently summarized by Scott Aaronson as 'It's hard to think when someone Hadamards your brain'[8]. The term 'Hadamard' refers here to a particularly destructive measurement, applied to an agent's lab, which is incompatible with the computational basis we would use to describe how the agent processes and stores information when reasoning. (This is also sometimes called a 'Bell measurement' or 'cat measurement'[9] because it often corresponds to a measurement in the Bell basis of the agent's memory and the system they observed.) The argument may be expressed in terms of a restriction on using rule (C).

Restriction 1. Reasoning rule (C) cannot be applied to predictions that Bob made after his memory was subjected to a destructive measurement.

In the case of the Learned Prediction Experiment, and assuming Wigner's measurement is indeed destructive, Restriction 1 means that the use of (C) should be forbidden if $t_W \leq t_P$ (Fig. 3).

Why Restriction 1 is sensible but irrelevant

We agree with this reasoning. In fact, Restriction 1 is already taken into account implicitly by the assumption that the agents all apply the same reasoning rules. Measurements that are so destructive that they disturb the agent's reasoning process are thus ruled out. More importantly, however, Restriction 1 is irrelevant in the case of the FR thought experiment. While (potentially) destructive measurements are applied to agents, the timing of these measurements is such that an agent never needs to make or communicate a prediction after having been subjected to such a measurement. Hence, Restriction 1, while absolutely justified, does not resolve the contradiction between the reasoning rules (Q), (C), and (S).

A stronger restriction

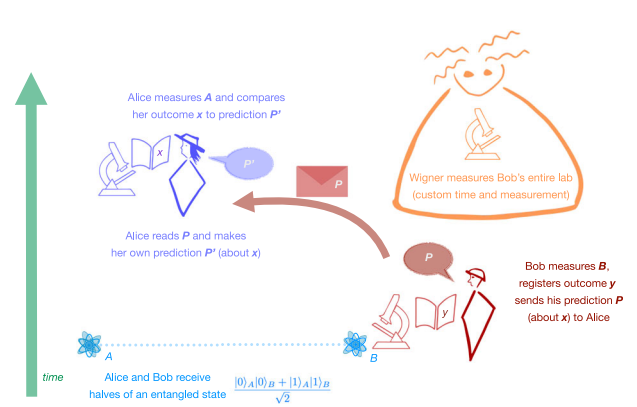
In a recent comment[9], Alexios Polychronakos analyses the FR thought experiment using an approach termed 'unitary quantum mechanics', which basically consists of putting the Heisenberg cut at the outside of the entire experiment. Technically, such an analysis corresponds to the one presented in[10] or[11] (the latter employs Bohmian mechanics; see the Supplementary Information for more details as well as a clarification of their claim that Theorem 1 is invalid). Motivated by this analysis, the author argues that if agents reason based on information held by other agents, along the lines of rule (C), then they arrive at invalid predictions—in agreement with Theorem 1[3]. Because Restriction 1 does not rule out this use of rule (C), he suggests extending the restriction to destructive measurements that lie in the future. In the spirit of Aaronson's slogan, this suggestion may be phrased as 'It's hard to think when someone will later Hadamard your brain.'

Restriction 2. Reasoning rule (C) cannot be applied to predictions that Bob made if his memory is subjected to a destructive measurement — even if that measurement lies in the future.

Applied to the Simple Prediction Experiment, Restriction 2 would imply that our analysis above is invalid even if Wigner measures Bob in the far future, i.e., after Bob has sent his prediction P to Alice, and possibly also after Alice has completed her measurement to verify the prediction (Fig. 3). In fact, the proposal by Polychronakos goes even one step further and similarly restricts the use of rule (Q). The author concludes that, equipped with these additional restrictions, the reasoning rules no longer yield a contradiction in the setting of the FR thought experiment.

BOX 1

Learned Prediction Experiment



t_0: Alice and Bob receive qubits A and B respectively, jointly prepared in an entangled Bell state $(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B)/\sqrt{2}$.
t_Y: Bob measures qubit B in the computational basis $\{|0\rangle,|1\rangle\}$ and registers his outcome y.
t_P: Bob makes a prediction P for the outcome x of Alice's measurement at t_X (see ahead) and communicates P to Alice.
$t_{P'}$: Alice receives P and infers from this a prediction P' for the outcome x of her measurement at t_X.
t_X: Alice measures qubit A in the computational basis $\{|0\rangle,|1\rangle\}$ and compares her outcome x to her prediction P'.
t_W: Wigner carries out a measurement of his choice on Bob's lab (which may include qubit B, Bob's memory, measurement instruments, and environmental degrees of freedom).

The first four steps occur at fixed times ordered as $t_0 < t_Y < t_P < t_{P'} < t_X$. Wigner's measurement occurs at a time t_W, which is also fixed but customizable. All agents are initially provided with a description of this protocol, including the timing of the steps. Furthermore, they know that all agents use the same set of reasoning rules to obtain their predictions.

https://doi.org/10.1038/s41467-024-47172-0

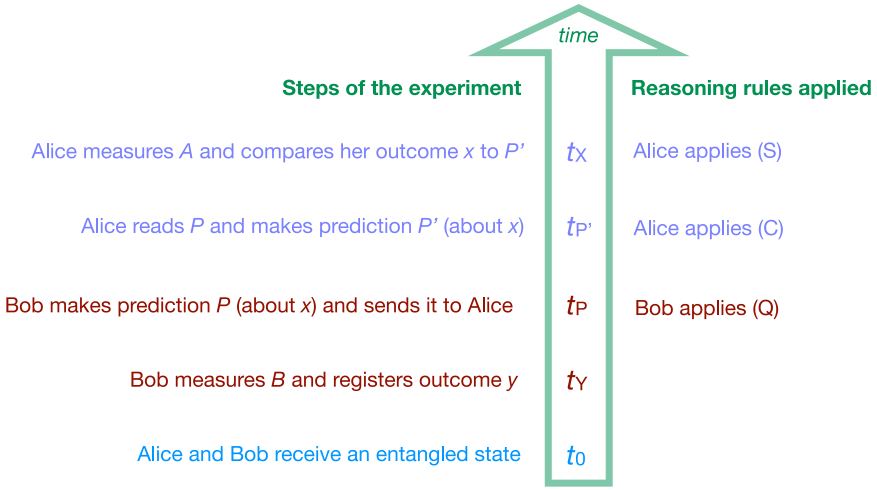


Fig. 2 | Applying reasoning rules to the Learned Prediction Experiment. If we omit Wigner's measurement, the analysis of this experiment is straightforward. For example, if Bob observes $Y = 1$, he can use reasoning rule (Q) to infer the prediction P = "I am certain that $X = 1$." Upon receiving and reading P, Alice may say "I am certain that Bob is certain that $X = 1$." Using (C) Alice immediately arrives at P' = "I am certain that $X = 1$." Finally, (S) demands that the outcome of Alice's measurement must indeed match her prediction P'.

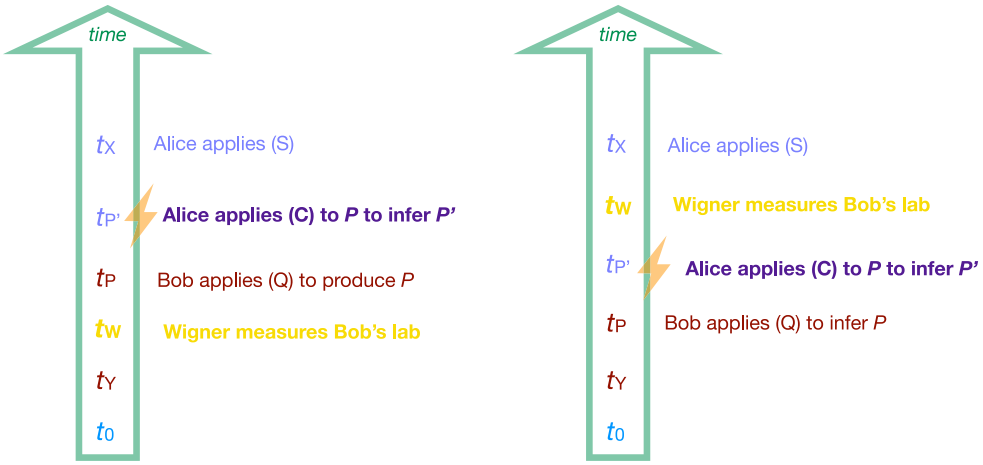


Fig. 3 | The Learned Prediction Experiment under different restrictions. Left: If Wigner measures Bob's lab before Bob produces his prediction ($t_W < t_P$), Restriction 1[8] forbids Alice from applying reasoning rule (C). This restriction ensures that agents do not rely on predictions computed by a malfunctioning agent. (In the FR thought experiment, this requirement is taken care of by the appropriate timing of the protocol steps.) Right: Restriction 2[9] forbids Alice from applying (C) (as in Fig. 2) even if Wigner measures Bob's lab after Bob sends his prediction.

We agree with this conclusion but would like to point out that such restrictions on the agents' reasoning entail various problems, which we detail in the following (see also the Supplementary Information).

Objection 1. Restrictions 1 and 2 are ambiguous. In the formulation of Restrictions 1 and 2 above, we used the term 'destructive measurement', which we believe is in the spirit of[8] and[9] (in these works, the terms 'Hadamard' and 'cat measurement' are used). But to make this unambiguous, it is necessary to characterise which measurements count as 'destructive'. Clearly, the particular measurements on the friends' labs in the FR experiment must be treated as destructive, but this is not sufficient; it would be easy to find variations of the FR experiment that also lead to contradictory conclusions with measurements that are just slightly rotated from the original ones. In that case, the safety induced by Restriction 2 would not be robust under small changes. On the other hand, if a proposal would extend the constraint to all settings where Alice's brain will be under any measurement, it would implicitly rule out even all classical logical reasoning used today—because our inferences are stored in physical memories that will eventually interact with their environments. It is unclear whether there is any natural boundary between these extremes that avoids either issue.

Objection 2. Physical justification of Restriction 2 requires signalling. The constraint imposed by Restriction 2 on using reasoning rule (C) seems unnecessarily strict when applied to settings like our Learned Prediction Experiment. If Wigner measures Bob after Bob communicates his prediction P to Alice, one would not expect this to render P invalid. This intuition may be verified by modelling Bob as a quantum system that outputs P. The non-signalling property of quantum theory then implies that a measurement performed by Wigner on Bob at time $t_W > t_X$ cannot be noticed by Alice at time t_X.

Restriction 2 is just a constraint on the applicability of a reasoning rule and hence does not imply signalling per se. However, if Restriction 2 was physically justified in the sense of having a physical origin, then the use of rule (C) without this restriction should sometimes lead to wrong predictions in experiments. To illustrate this, consider the Learned Prediction Experiment. If Restriction 2 was necessary here, the prediction P', computed by Alice using rule (C), would sometimes be wrong when Wigner measures Bob's lab at a

time $t_W > t_X$. Because Alice can verify her prediction at time t_X, this would violate the non-signalling principle: Wigner, by measuring or not measuring at time t_W, could send a signal to Alice at time t_X, into the past. In this sense, a physical justification for Restriction 2 is incompatible with the non-signalling principle.

Objection 3. Restrictions on (C) impair reasoning. Rule (C) allows agents to combine and compress information, which facilitates processing and prediction-making. Restrictions on the use of this rule may thus impair reasoning even in standard settings, where agents typically hold only partial information about the physical setup and must apply (C) to piece together their local bits of knowledge.

In our Learned Prediction Experiment, we assumed that all agents were provided with a full description of the experimental protocol. This allows Alice, in principle, to reach her prediction P' without applying (C): Alice may reverse-engineer Bob's reasoning to infer his outcome y from his prediction P, which was communicated to her. Knowing y she may then employ (Q) to come up with P'. However, this strategy for avoiding the use of (C) would not be available to Alice if her knowledge about the experimental setup was partial. An example of this would be a variant of the experiment where Alice's initial information consists of a description of her local setup only, so that she cannot simulate Bob.

That agents have partial information is common in real-world examples and particularly dramatic in cryptography scenarios. For example, in quantum key distribution protocols, without (C), Alice cannot make the logical step from "Bob publicly announced that his measurement basis was X" to "I know that Bob's measurement basis was X." It is unclear whether Alice and Bob, who in the setting trust each other but otherwise are embedded in an environment controlled by an adversary, can obtain any security guarantee for the distributed key without applying (C)[12].

But even in the special case where all agents have full information about the global setup, so that (C) could be substituted by (Q), this comes with a complexity overhead. An agent who wants to incorporate knowledge communicated by another agent would need to simulate that agent as a quantum system. In general network scenarios with N agents whose individual predictions may depend on chains of reasoning across several agents, the (classical) memory required for this scales exponentially with N[13].

Desiderata for resolutions of the paradox

We note that various other suggestions have been made in the recent literature to evade the contradiction in the FR experiment (see[14–20] for examples). Similarly to Restrictions 1 and 2 above, they postulate constraints on the reasoning rules, notably rules (Q) and (C). Likely, the objections discussed above are also relevant to them. More generally, any proposal to resolve the paradox faces the challenge of finding a fine balance. If the restriction on the reasoning rules is too moderate, they may still yield contradictory conclusions when applied to thought experiments like FR. Conversely, if the reasoning rules are constrained too much, their usability in everyday situations may be affected. To foster further research, we propose a list of desiderata for proposals for modified reasoning rules.

1. *Clear:* The proposed reasoning rules should be specified unambiguously (see Objection 1).
2. *Usable:* The proposed reasoning rules should be usable by an agent who is a physical system embedded in the physical world and who has only partial information about the world. In particular, the reasoning rules can only depend on information that is physically available to the agent and can be processed with the agent's physical resources (see Objection 3).
3. *Falsifiable:* The proposed reasoning rules should reproduce the predictions of quantum theory in all regimes that have been experimentally tested, including scenarios where individual agents have only partial information about the overall setup. In particular, any data produced by a realistic experiment that would falsify quantum theory should also falsify the reasoning rules.
4. *Consistent:* The proposed reasoning rules should apply to any experiment describable within the standard formalism of quantum theory, including thought experiments such as the Wigner's friend or the FR experiment, and should not yield contradictions.
5. *Physical:* The proposed reasoning rules should be physically justifiable; in particular, they should avoid the violation of basic physical principles (see Objection 2).

Outlook

We urge those who propose modified reasoning rules to circumvent Theorem 1 not to be discouraged by the objections presented here. This is a recent and novel problem, and it is only natural that the appropriate tools to tackle it have not yet been developed. To study and test reasoning rules in view of the desiderata listed above, we leave the reader with two suggestions for such tools. For computational tests, the free software package for quantum thought experiments developed by Nurgalieva, Mathis and ourselves[13] allows a user to formulate bespoke reasoning rules in a computer-readable manner and test them in different experimental settings: the software outputs the predictions of different agents and whether they are contradictory. For a theoretical analysis of Wigner's friend-type experiments, a promising approach is the framework of Vilasini and Woods[21] (see Supplementary Information), which enforces an explicit specification of the choice of the Heisenberg cut by the different agents.

Data availability

No data sets were generated or analysed during the current study.

References

1. Wigner, E. P. Remarks on the mind-body question. In *Symmetries and Reflections*, 171–184 (Indiana University Press, 1967). https://doi.org/10.1007/978-3-642-78374-6_20.
2. Heisenberg, W. Ist eine deterministische Ergänzung der Quantenmechanik möglich? In Hermann, A., von Meyenn, K. & Weisskopf, V. (eds.) *Wolfgang Pauli. Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg*, vol. II, 409–418 (Springer, 1985).
3. Frauchiger, D. & Renner, R. Quantum theory cannot consistently describe the use of itself. *Nat. Commun.* **9**, 3711 (2018).
4. Brukner, Č. On the quantum measurement problem. In Bertlmann, R. & Zeilinger, A. (eds.) *Quantum [Un]Speakables II: Half a Century of Bell's Theorem*, 95–117 (Springer, 2017). https://doi.org/10.1007/978-3-319-38987-5_5.
5. Nurgalieva, N. & del Rio, L. Inadequacy of modal logic in quantum settings. *Electron. Proc. Theor. Comput. Sci.* **287**, 267–297 (2019).
6. Nurgalieva, N. & Renner, R. Testing quantum theory with thought experiments. *Contemporary Phys.* **61**, 193–216 (2020).
7. Nielsen, M. A. & Chuang, I. L.*Quantum Computation and Quantum Information* (Cambridge University Press, 2010). https://www.cambridge.org/highereducation/books/quantum-computation-and-quantum-information/01E10196D0A682A6AEFFEA52D53BE9AE#overview.
8. Aaronson, S. It's hard to think when someone Hadamards your brain (2018). Retrieved from https://scottaaronson.blog/?p=3975 on 19.7.2022.
9. Polychronakos, A. P. Quantum mechanical rules for observed observers and the consistency of quantum theory (2022).
10. Bub, J. Understanding the Frauchiger–Renner argument. *Foundat. Phys.* **51**, 36 (2021).
11. Lazarovici, D. & Hubert, M. How quantum mechanics can consistently describe the use of itself. *Sci. Rep.* **9**, 470 (2019).
12. Portmann, C. & Renner, R. Security in quantum cryptography. *Rev. Mod. Phys.* **94**, 025008 (2022).

13. Nurgalieva, N., Mathis, S., del Rio, L. & Renner, R. Thought experiments in a quantum computer https://arxiv.org/abs/2209.06236 (2022).
14. Narasimhachar, V. Agents governed by quantum mechanics can use it intersubjectively and consistently https://arxiv.org/abs/2010.01167 (2020).
15. Waaijer, M. & Neerven, J. V. Relational analysis of the Frauchiger-Renner paradox and interaction-free detection of records from the past. *Found. Phys.* **51**, 45 (2021).
16. Żukowski, M. & Markiewicz, M. Physics and metaphysics of Wigner's friends: even performed premeasurements have no results. *Phys. Rev. Lett.* **126**, 130402 (2021).
17. Di Biagio, A. & Rovelli, C. Stable facts, relative facts. *Found. Phys.* **51**, 30 (2021).
18. Elouard, C. et al. Quantum erasing the memory of Wigner's friend. *Quantum* **5**, 498 (2021).
19. Renes, J. M. Consistency in the description of quantum measurement: Quantum theory can consistently describe the use of itself https://arxiv.org/abs/2107.02193 (2021).
20. Federico, M. & Grangier, P. A contextually objective approach to the extended Wigner's friend thought experiment https://arxiv.org/abs/2301.03016 (2023).
21. Vilasini, V. & Woods, M. P. A general framework for consistent logical reasoning in Wigner's friend scenarios: subjective perspectives of agents within a single quantum circuit https://arxiv.org/abs/2209.09281 (2022).

Acknowledgements

We acknowledge support from the Swiss National Science Foundation through SNSF project No. 200021_188541 and the Quantum Center of ETH Zurich. LdR further acknowledges support from the FQXi large grant 'Consciousness in the Physical World'. LdR is grateful for the hospitality of Perimeter Institute, where part of this work was carried out. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Colleges and Universities.

Author contributions

L.d.R. and R.R. contributed equally to this work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-47172-0.

Correspondence and requests for materials should be addressed to Lídia del Rio or Renato Renner.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.