Latent Measurement:

Can a Non-Conscious Agent Collapse the Quantum State? A Triple-Layer Protocol for Young's Double-Slit Experiment

CosmicThinker
Independent Researcher
April 2026

Abstract

The measurement problem in quantum mechanics has remained open for nearly a century. What constitutes an observer? Is consciousness a necessary ingredient for the loss of quantum coherence, or does an irreversible physical interaction suffice? In this work we propose a triple-layer experimental protocol based on Young's double-slit experiment, in which three successive agents — an AI-integrated robot that executes the experiment (Robot 1), a second robot that examines and re-records the results (Robot 2), and a human observer who inspects the records of both — allow us to place operational bounds on which stages of the measurement chain are physically sufficient for the loss of interference and which are merely epistemic. We show that the strong von Neumann–Wigner hypothesis (which requires consciousness for collapse) and the decoherence interpretation (which requires only physical interaction) can be associated with operationally distinct experimental expectations in Mode B of the protocol. We connect the analysis with the concept of Latent Structural Causality (LSC): a result recorded by a non-conscious agent but not consulted by any human constitutes an example of structurally present but operationally silent information. The proposed protocol is feasible with current technology with respect to the automation of the setup, and could help translate the debate on the quantum observer from a philosophical question into an experimental one.

Keywords: Measurement problem; Young's double-slit experiment; Quantum observer; Artificial intelligence; Latent Structural Causality; von Neumann–Wigner hypothesis; Foundations of quantum mechanics.

1. Introduction

Since the mathematical formulation of quantum mechanics in the 1920s, the measurement problem has constituted one of the deepest debates in fundamental physics. The theory predicts with extraordinary precision the statistical outcomes of experiments, yet remains ambiguous on a crucial point: what physical process converts a quantum superposition into a definite result?

John von Neumann formalised the problem in 1932 by distinguishing two types of evolution: the continuous unitary evolution described by the Schrödinger equation, and a discontinuous 'collapse' associated with the act of measurement [1]. Eugene Wigner took this distinction to its logical conclusion with his celebrated 'Wigner's friend' scenario (1961): if a friend inside a closed laboratory measures a quantum system, at what point does the wave function collapse—when the friend looks, or when Wigner opens the door [2]?

This problem is not merely philosophical. Proietti et al. (2019) implemented a simplified version of Wigner's scenario with entangled photons, obtaining results incompatible with the assumption that the ‘friends’ possess definite results before the external observer consults them [4]. However, in all experiments conducted to date, both the ‘friends’ and the external observers are systems designed and supervised by humans, leaving open a question that David Hume anticipated in the eighteenth century: to what extent does our access to reality depend on the properties of the agent that observes [3]?

In this work we propose an experimental protocol that addresses this question directly. Using Young’s double-slit experiment [10]—the most fundamental and best-understood quantum scenario—we design a triple-layer observation chain in which an AI-integrated robot executes the complete experiment, a second robot examines the results, and only then does a human inspect the records. This structure allows us to place operational bounds on the contributions of physical interaction, information processing, and consciousness to the measurement process.

2. Decomposition of the Act of Measurement

The act of measurement, as practised in the laboratory, involves multiple operations that are usually treated as a single, indivisible process. We propose decomposing it into four conceptually distinguishable layers:

2.1. Layer 1: Physical Interaction

The quantum system interacts with a macroscopic apparatus (detector, screen, recording medium). This interaction is described by a coupling Hamiltonian and produces a correlation between the states of the system and those of the apparatus. In the decoherence formalism [6], this interaction, together with coupling to the environment, is sufficient to eliminate interference between macroscopically distinct branches.

2.2. Layer 2: Record

The result of the interaction is stored in a stable physical medium: a digital memory, a mark on a detector, a datum in a file. The record is an irreversible amplification that converts the quantum correlation into a classical imprint. Crucially, the record can exist without anyone consulting it.

2.3. Layer 3: Processing

An agent—human or artificial—processes the recorded data: reads it, analyses it, extracts patterns, computes statistics. This processing transforms raw data into interpretable information but does not modify the data themselves. An AI algorithm can perform this operation without any conscious component.

2.4. Layer 4: Conscious Interpretation

A human being becomes aware of the result: understands it, relates it to prior expectations, updates beliefs. This is the layer that the strong von Neumann–Wigner hypothesis [1, 2] identifies as necessary for collapse, and that the decoherence interpretation [6] and the many-worlds interpretation [7] consider physically irrelevant.

The experimental question we pose is: are Layers 1 and 2 sufficient for the loss of coherence, or do Layers 3 and 4 add any observable signature? Our triple-layer protocol is designed to place operational bounds on this question.

3. Experimental Protocol: Triple Layer on Young's Double-Slit Experiment

3.1. General Setup

The experiment uses a standard double-slit setup with a single-photon source. The novelty lies exclusively in the observation structure, which is organised into three successive levels (Figure 1):

Robot 1 (R1): A robotic system with integrated artificial intelligence that prepares the complete experimental setup (source, slits, detectors, screen), executes the experiment by sending individual photons, and records the results. No human intervenes at any stage of this process. R1 operates Layers 1 and 2.

Robot 2 (R2): A second robotic system, physically independent of R1, that accesses the records produced by R1, examines them, stores a second copy of the data, and performs a preliminary statistical analysis. R2 does not interact with the original quantum system—only with the classical data produced by R1. R2 operates Layer 3.

Human observer (H): A human being who, in the final stage, inspects the records of R1 and R2. The human is the first conscious agent in the entire chain. H operates Layer 4.

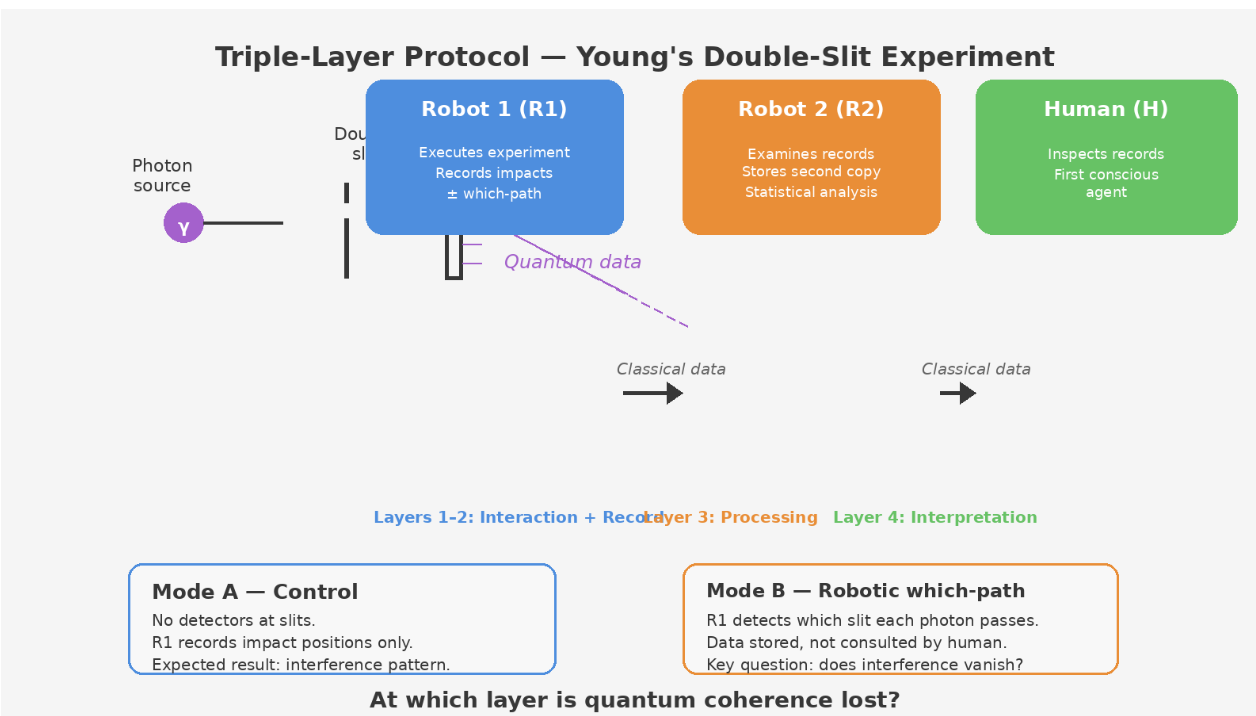

Figure 1: Diagram of the triple-layer protocol. Young's setup is operated entirely by R1, which records quantum data (impact positions ± which-path information). R2 processes the classical data from R1. The human inspects the records of both robots. The question is at which layer coherence is lost.

3.2. Modes of Operation

The protocol is executed in two distinct modes, each designed to isolate a different aspect of the measurement process (Figure 2):

Mode A — Control (no slit detection): R1 executes the standard double-slit experiment without placing detectors at the slits. It records only the impact positions on the screen. A full interference pattern is expected. This mode establishes the baseline and verifies that the robotic setup introduces no artefacts.

Mode B — Robotic detection (which-path stored): R1 places detectors at the slits that record which slit each photon passes through. This information is stored in R1's memory, but no human accesses it. R2 examines R1's records and stores a second copy of the data. Only then does the human inspect the records of both robots.

The key question in Mode B is: does the interference pattern on the screen vanish when which-path information has been recorded exclusively by non-conscious agents?

3.3. Observable and Decision Criterion

The primary observable of the protocol is the interference visibility, defined as $V = (I_{max} - I_{min}) / (I_{max} + I_{min})$, where I_{max} and I_{min} are the maximum and minimum intensities of the pattern recorded on the screen. In Mode A, a high visibility, consistent with the interferometric regime of the setup, is expected. In Mode B, the question is whether V remains high or drops to values close to zero.

R1 records the individual impact positions of each photon on the screen. The interference pattern is reconstructed from the accumulated distribution of impacts. R2 processes exclusively these classical data (positions and, in Mode B, which-path labels) and computes the visibility V as the primary statistic. The decision criterion is the direct comparison of V between Modes A and B: if $V_B \approx V_A$, which-path detection has not destroyed coherence; if $V_B \approx 0$, physical interaction has been sufficient.

Figure 2: Schematic comparison of the patterns expected on the screen for the two modes of operation. Mode A (left): without which-path detection, interference is expected. Mode B (right): with which-path detection by R1, the question is whether the loss of interference requires a conscious observer.

4. Predictions According to Different Interpretations

One of the virtues of the protocol is that it allows contrasting the prediction of standard decoherence with that of the strong von Neumann–Wigner hypothesis. However, it is important to note that the ‘interpretations’ of quantum mechanics do not constitute a discrete set of theories with unambiguous predictions; many admit internal variants. The following table should be understood as a heuristic tool, not as a closed classification.

Interpretive framework	Mode A (no detector)	Mode B (which-path by R1)
Decoherence (Zurek, 2003)	Interference	No interference (physical interaction suffices)
Strong von Neumann–Wigner hypothesis	Interference	Interference preserved (no conscious agent observed)
Many worlds (Everett, 1957)	Interference	No interference (decohered branches)
QBism (Fuchs, 2010)	Interference	Indeterminate (agent-dependent)

Table 1: Heuristic predictions of each interpretive framework for the two modes of the protocol. The discriminating column is Mode B.

The discriminating result is Mode B. Decoherence and the many-worlds interpretation predict that the interference pattern vanishes regardless of who or what records the which-path information, because the physical interaction with the detector has already destroyed coherence. By contrast, the strong von Neumann–Wigner hypothesis predicts that, since no conscious agent has accessed the information, the quantum system would not have collapsed and interference should persist.

If Mode B shows interference—an extraordinary result—this would constitute direct evidence that physical interaction is not sufficient for the loss of coherence and that some property of the observer is physically relevant. If Mode B does not show interference—the result expected by most physicists—the experiment would provide additional evidence, obtained with a protocol that removes the human observer from the measurement chain more completely than previous experiments, compatible with the thesis that consciousness plays no privileged role in the loss of quantum coherence. Nevertheless, it should be noted that a negative result in Mode B would specifically refute the strong version of the von Neumann–Wigner hypothesis, but would not necessarily close the entire debate on consciousness and measurement: a sophisticated defender of the relevance of consciousness could reformulate the hypothesis in less direct terms.

5. Connection with Latent Structural Causality

The proposed protocol constitutes a direct application of the concept of Latent Structural Causality (LSC) introduced in a previous work [9]. An interaction possesses LSC when it is structurally present in the formalism but remains operationally silent within a given regime.

In Mode B, which-path information is structurally present in R1’s records from the moment the slit detector interacts with the photon. However, that information is operationally silent for any conscious agent: it has not been consulted by any human, has not modified any beliefs, and has not generated any action based on it. LSC thus provides the conceptual framework for analysing the transition between layers:

After detection by R1, the result exists as a physical record—structurally real but operationally latent for a conscious agent. After examination by R2, the information has been copied and re-processed by a second non-conscious system, but remains operationally latent for Layer 4. After human inspection, the result becomes operationally active: a conscious agent knows and interprets it.

The experimental question is, in essence, a question about LSC: does the transition from latency to operativity—that is, the passage from Layers 1–3 to Layer 4—produce any physical effect on the quantum system? If the answer is no, then the loss of coherence occurred at Layer 1 and consciousness is epistemically relevant but physically inert. If the answer is yes, then conscious operativity would be a genuine physical ingredient, not a mere epistemic description.

6. The Problem of Inherited Bias

An aspect that must be treated rigorously is the possibility that the artificial agent introduces biases that contaminate the results. We identify three levels of potential bias:

6.1. Design Bias

R1's AI was trained by humans. Its algorithms reflect human decisions about what constitutes a correct experimental setup, which parameters to control, and which data to record. Strictly speaking, R1 is an indirect human observer: it executes a protocol designed by humans. However, this bias is equivalent to that introduced by any measuring apparatus—a photon detector was also designed by humans, and nobody argues that the detector is an 'observer' in the quantum sense. Design bias is inherent to any experiment and does not invalidate the proposed layer decomposition.

6.2. Recording Bias

R1 must decide when a detection has been completed (signal threshold, time window, noise discrimination). These decisions are operational and could, in principle, affect the recorded data. To control this bias, the protocol should include parallel runs in which a human performs the same experiment with the same parameters, allowing a direct comparison. Any systematic discrepancy between R1's data and the human's would be attributable to recording bias, not to quantum physics.

6.3. Existence Bias

This is the deepest and philosophically most interesting level. If R1 records a result and no conscious being consults it, does that result exist in any physically relevant sense? A realist would say yes: the bits in R1's memory are physical states of the world, independent of who reads them. A follower of QBism [8] would say no: a quantum result only becomes real relative to an agent who experiences it. The triple-layer protocol does not resolve this philosophical question, but translates it into one with empirical consequences: does the difference between a recorded-but-not-consulted result and a recorded-and-consulted result produce any measurable effect?

7. Experimental Feasibility

The components of the protocol are individually standard: single-photon sources, double-slit setups, and position detectors are routine equipment in quantum optics laboratories. The novelty lies in the complete automation of the process—from setup preparation to recording—by an autonomous robotic system.

The specific technical requirements include: (a) a robotic arm capable of aligning optical components with micrometric precision; (b) an AI-based control system that executes the experimental protocol without human supervision; (c) a second independent system capable of reading and storing a second copy of digital data; and (d) an isolation protocol ensuring that no human has access to the results until the final phase of the experiment.

It is important to note a crucial distinction: the technical feasibility of the individual components of this setup is not in question. The relevant issue is whether the automation adds discriminating power relative to existing setups with automatic recording. We argue that it does, for a specific reason: in current experiments, although recording is automatic, a human designs, supervises, and has potential access to the results in real time. Our protocol requires that the complete chain—preparation, execution, recording, and initial analysis—occur without human presence or access. This difference, although subtle, is precisely what the von Neumann–Wigner hypothesis requires in order to be tested.

8. Discussion

The proposed protocol does not claim to solve the measurement problem—a problem that has remained open for a century—but to place operational bounds on one of its most debated facets: the role of the conscious observer. By decomposing the act of measurement into layers and assigning each layer to a different type of agent (robotic or human), the protocol allows one to discriminate between two families of hypotheses:

(a) Hypotheses in which physical interaction (Layers 1–2) is sufficient for the loss of coherence. In this case, R1, R2, and the human observe the same data—the observation chain is irrelevant to the physics.

(b) Strong hypotheses in which conscious access (Layer 4) has additional physical relevance. According to the strong von Neumann–Wigner hypothesis, the availability of which-path information to non-conscious agents would not by itself suffice to complete the loss of coherence, which would imply that the interference pattern could persist in R1's and R2's records as long as no human has consulted them. This prediction, although extraordinarily improbable according to the current consensus, is precisely what the protocol allows us to test.

The most probable result—that Mode B shows no interference, confirming that physical interaction suffices—would be a relevant though unsurprising result. It would reinforce the thesis that consciousness plays no privileged role in the loss of coherence, using a protocol that decouples the measurement layers more explicitly than conventional setups. However, it is important to recognise the limits of this result: since standard decoherence already predicts the loss of interference from the mere interaction with the detector, a negative result in Mode B specifically tests the strong von Neumann–Wigner hypothesis, not the entirety of the debate on consciousness and quantum measurement.

The improbable but revolutionary result—that Mode B shows interference—would indicate that something in conscious processing is physically relevant to quantum mechanics. Such a result would require a profound rethinking of the foundations of the theory and would open an entirely new field of research.

In either case, the connection with Latent Structural Causality provides a conceptual framework that allows both results to be discussed coherently. If the loss of coherence occurs at Layer 1, operativity is a property of the physical interaction, not of the observer. If it requires Layer 4, the information remains genuinely latent until a conscious agent activates it.

9. Conclusion

We have proposed a triple-layer experimental protocol for Young's double-slit experiment that allows one to place operational bounds on the role of the observer in quantum mechanics. The protocol decomposes the measurement chain into three successive agents—two AI robots and a human—and shows that the strong von Neumann–Wigner hypothesis and standard decoherence can be associated with distinct experimental expectations in Mode B of this configuration. The scope of the protocol is specific: it tests whether physical interaction and automatic recording are sufficient for the loss of coherence, or whether conscious access adds any observable effect. It does not claim to solve the measurement problem in its entirety, but to provide an empirical tool for one of its most debated facets.

The connection with the concept of Latent Structural Causality provides an epistemological framework for discussing the transition between recorded information and consulted information, regardless of which hypothesis is ultimately favoured by the data.

Ultimately, this work stems from a simple intuition: if we want to know whether consciousness matters for quantum mechanics, let us remove consciousness from the experiment and see what happens. That the answer to this question requires building robots that do physics without us is, perhaps, the most profound reminder that the hardest questions often admit the simplest formulations.

Transparency Note on the Use of AI

This work was developed with the technical assistance of artificial intelligence models. Claude (Anthropic) was used as an assistant in the drafting, structuring, and formalisation of the manuscript. ChatGPT (OpenAI) was used as a critical-review assistant, providing a detailed report that led to substantial corrections between the first and the present version. The original idea for the experiment, the conception of the triple-layer protocol, the underlying physical intuition, and all decisions regarding the final content are entirely attributable to CosmicThinker.

References

[1] J. von Neumann, Mathematische Grundlagen der Quantenmechanik, Springer (1932).

[2] E. P. Wigner, Remarks on the Mind-Body Question, in The Scientist Speculates, ed. I. J. Good (1961).

[3] D. Hume, A Treatise of Human Nature (1739).

[4] M. Proietti et al., Experimental test of local observer independence, Science Advances 5, eaaw9832 (2019).

[5] C. Brukner, A No-Go Theorem for Observer-Independent Facts, Entropy 20, 350 (2018).

[6] W. H. Zurek, Decoherence, einselection, and the quantum origins of the classical, Rev. Mod. Phys. 75, 715 (2003).

[7] H. Everett III, “Relative State” Formulation of Quantum Mechanics, Rev. Mod. Phys. 29, 454 (1957).

[8] C. A. Fuchs, QBism, the Perimeter of Quantum Bayesianism, arXiv:1003.5209 (2010).

[9] CosmicThinker, Latent Structural Causality: Defining Physical Existence in Pre-Observable Regimes, Zenodo (2026).

[10] T. Young, The Bakerian Lecture: Experiments and Calculations Relative to Physical Optics, Phil. Trans. R. Soc. Lond. 94, 1 (1804).

[11] J. A. Wheeler, The “Past” and the “Delayed-Choice” Double-Slit Experiment, in Mathematical Foundations of Quantum Theory (1978).

[12] J. S. Bell, Speakable and Unspeakable in Quantum Mechanics, Cambridge University Press (1987).

[13] L. Smolin, Einstein’s Unfinished Revolution, Penguin Press (2019).

[14] J. Barbour, The End of Time, Oxford University Press (1999).