Measuring time with stationary quantum clocks

Sergii Strelchuk (✉ ss870@cam.ac.uk)

University of Cambridge

Mischa Woods

Rämistrasse 101

Article

Keywords:

Posted Date: February 22nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1297227/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Measuring time with stationary quantum clocks

Sergii Strelchuk¹ and Mischa P. Woods²

¹*DAMTP, University of Cambridge, UK*

²*Institute for Theoretical Physics, ETH Zurich, Zurich, Switzerland*

Time plays a fundamental role in our ability to make sense of the physical laws in the world around us. The nature of time has puzzled people — from the ancient Greeks to the present day — resulting in a long running debate between philosophers and physicists alike to whether time needs change to exist (the so-called relational theory), or whether time flows regardless of change (the so-called substantival theory). One way to decide between the two is to attempt to measure the flow of time with a stationary clock, since if time were substantival, the flow of time would manifest itself in the experiment. Alas, conventional wisdom suggests that in order for a clock to function, it cannot be a static object, thus rendering this experiment seemingly impossible. Here we show, with the aid of counterfactual measurements, the surprising result that a quantum clock can measure the passage of time even while being switched off, thus lending constructive support for the substantival theory of time.

I. INTRODUCTION

Time is an essential ingredient in the world we inhabit. Not so surprisingly, it has played a central role in all of our physical theories. Yet this role was rather mundane until the advent of modern physics. In particular, out of the three pillars of modern physics — quantum mechanics, special and general relativity — it was only the latter two which forced us to change our preconceptions about the nature of time. Quantum mechanics, on the other hand, while very strange and mysterious in many ways, did not bring any novel insights as far as time is concerned: time is just a parameter which increases in line with any mundane classical clock — just like in Newton’s laws. Even more recently, it has been proven, in the context of quantum mechanics, that the time-analogue of Bell’s inequality always has a perfectly sound classical explanation [1, 2].

Conversely, in the philosophy of physics domain, the notion of time occupied a prominent role in debates dating back millennia which are still very active today. One of the ongoing debates is whether time necessitates change to exist. The substantival theory of time says that time exists and provides an invisible container in which matter lives, regardless of whether the matter is moving. On the other hand, the relational theory of time says that time is a set of relationships among the events of physical material in space — that is to say, time ceases to exist if matter ceases to change.

Views have continuously shifted through the ages: Many believe that Greek atomists such as Democritus thought time was substantival. The first written account dates back to Aristotle, who advocated for a relational theory: “*But neither does time exist without change;...*” [3]. Newton and Leibniz had opposed views, with Newton on the substantive camp [4], and Leibniz on the relational camp [5]. Ernst Mach attacked Newton’s arguments in favour of a more relative theory [6]. Einstein credited Mach’s views as being highly influential in his guiding principles when developing his theory of general relativity; although later

shifted his stance to a more substantival interpretation of his theory [7].

Arguments for and against either theories are still ongoing [8–11]. One of the biggest problems in this debate is the apparent inability to distinguish experimentally between the two scenarios — indeed, it is widely accepted that any clock (or physical matter used as a rudimentary clock), would need to change its state in order to measure the passing of time, hence rendering a clock useless for detecting the substantival’s hypothesised passage of time without change.

Here we show that nonrelativistic quantum mechanics renders this widely held belief wrong. This is to say, we demonstrate by designing special clocks and adhering to well established interpretations of quantum theory, that it is possible to measure the passage of time between two events even when said clock has been always off, in other words, never evolving. Our result thus allows one to precisely detect the presence of the flow of time without evolution, which we argue provides strong and experimentally testable evidence for the substantival theory of time.

II. RESULTS

Overview: We will start by considering the simplest systems naturally occurring in nature which can serve as elementary clocks when dynamically evolving. We then show how they can be used to tell the time while being switched off via a more advanced protocol and measurement scheme. This is followed by an analysis of the interpretations used and a demonstration of how this strange phenomena cannot be explained if one assumes an underlying “real” description of nature, using classical variables. Finally, in section III, we explore the implications of our results in relation to the long standing historical debates on the nature of time.

Elementary timekeeping systems: While we tend to think of timekeeping devices as engineered systems, we can use the dynamics of naturally occurring processes as elementary clocks. The simplest of which can

be thought of as an arbitrary state $|\psi\rangle$ at time $t = 0$ whose evolution takes it through a sequence of mutually orthogonal states $|\tau_0\rangle, |\tau_1\rangle, \dots, |\tau_{N_T}\rangle$ at times $\tau_0, \tau_1, \dots, \tau_{N_T}$ consecutively. If we know that the system is initially in the state $|\psi\rangle$, and we let it evolve for an unknown time t given the promise that t belongs to the set $\{\tau_0, \tau_1, \tau_2, \dots, \tau_{N_T}\}$, then we can determine precisely what the elapsed time is by measuring in the basis $|\tau_0\rangle, |\tau_1\rangle, |\tau_2\rangle, \dots, |\tau_{N_T}\rangle$, since at said times, the measurement will deterministically allow one to distinguish between said states. To use this timekeeping system in practice, consider two external events, which we will call *1st event* and *2nd event*, with some unknown elapsed time $t \in \{\tau_0, \tau_1, \tau_2, \dots, \tau_{N_T}\}$ between them. One initialises the system in the state $|\psi\rangle$ when the 1st event occurs and then waits until the 2nd event occurs, at which point the measurement would be performed; revealing the elapsed time. When a clock is revealed to have been dynamically evolving upon measurement as in this example, we say it is operating in *standard fashion*.

This is arguably one of the most elementary types of timekeeping devices one can conceive of. It has no inherently quantum features and should be quite ubiquitous in nature since the states of systems tend to become completely distinguishable if one waits long enough. As a simple illustration of how it could be used, imagine being on a boat at sea heading in to port. There is a lighthouse whose pulsing flashes of light reach you every 3 seconds, but thick fog rolls in obscuring the flashes for some time before clearing. By setting the times $\{\tau_n\}$ to coincide with the flashes and choosing the 1st and 2nd events to be flashes before and after the presence of the fog, the clock can deduce the duration of the foggy period.

Suppose that, in addition, there is a state denoted $|E\rangle$ which is invariant under time evolution, in other words, it remains in the state $|E\rangle$ at all times. Furthermore, suppose it is orthogonal to $|\psi\rangle$ and its time evolved state at all times. We refer to this state as an *off* state, in contrast to any state which evolves in time, which we refer to as an *on* state. The state $|E\rangle$ could be, for example, an energy eigenstate of the system. If one starts in the state $|E\rangle$ and then measures the state later, clearly there is no possible measurement one can perform which would reveal any information about the elapsed time — it would be analogous to taking the

batteries out of a wall clock, and then trying to use it to tell the time.

Telling the time when the clock is off: We will now show, with the aid of some additional stationary ancilla qubits, how quantum mechanics allows one to determine precisely which time τ_n it is, even though the timekeeping device was never on (in the runs of the experiment in which we determined τ_n). For the purpose of illustration, consider the most basic example of $N_T = 1$, in which the system in question merely has two times τ_0, τ_1 . We relegate the $N_T \geq 2$ version to supplementary material A. In this case we will only need one ancillary state, which we denote $|A\rangle$. It could be, for example, chosen to be another energy eigenstate which is orthogonal to the other states, or if such a state is not available, we can use an ancilla qubit with a trivial Hamiltonian so as to ensure its states do not evolve. In the latter case, denoting a basis for the ancilla by $|\uparrow\rangle, |\downarrow\rangle$, the states would then be associated with $|\psi(t)\rangle \equiv |\psi(t)\rangle |\uparrow\rangle$, $|E\rangle \equiv |E\rangle |\uparrow\rangle$, and the new ancilla state with $|A\rangle = |E\rangle |\downarrow\rangle$ (Here $|\psi(t)\rangle$ denotes $|\psi\rangle$ after evolving for a time t).

We will use the same retrodictive arguments as in other famous experiments, such as the Elitzur-Vaidman bomb tester or Hardy's paradox. However, let us first describe the protocol, before discussing the interpretation: Initially the system is set to the off state $|E\rangle$. Then, when the 1st event occurs, we apply a unitary U_0 such that the system is now in a superposition of on and off, namely, of $|E\rangle$ and $|\psi\rangle$. We use branching notation to indicate the orthogonal branches of the superposition associated with static and dynamical terms (Upper branch is static, lower branch is dynamic). The state reads:

$$\begin{array}{c} \bullet c|E\rangle \\ |E\rangle \swarrow \searrow \\ \bullet s|\psi\rangle \end{array}$$

where $c = \cos(\theta)$, $s = \sin(\theta)$. One then waits until the 2nd event occurs, at either time τ_0 or τ_1 , at which point one applies a judiciously chosen unitary U_m followed by immediately measuring in the $|E\rangle, |\tau_0\rangle, |\tau_1\rangle, |A\rangle$ basis, which we will call the measurement basis for short. This completes the protocol — modulo specification of the required constraints on U_0 and U_m . Diagrammatically, up to the point of measurement, this protocol and unitary U_m have the form:

$$\begin{array}{c} \bullet c|E\rangle \xrightarrow{\tau_0} c|E\rangle \xrightarrow{U_m} cA_1^0|E\rangle + cA_2^0|A\rangle + cA_3^0|\tau_0\rangle + cA_4^0|\tau_1\rangle \\ |E\rangle \swarrow \searrow \\ \bullet s|\psi\rangle \xrightarrow{\tau_0} s|\tau_0\rangle \xrightarrow{U_m} -cA_2^0|A\rangle + sA_3^1|\tau_0\rangle + sA_4^1|\tau_1\rangle \end{array} \quad (1)$$

at time τ_0 and

$$\begin{array}{l}
 |E\rangle \begin{cases} \xrightarrow{\tau_1} c|E\rangle \xrightarrow{U_m} cA_1^0|E\rangle + cA_2^0|A\rangle + cA_3^0|\tau_0\rangle + cA_4^0|\tau_1\rangle \\ \xrightarrow{\tau_1} s|\psi\rangle \xrightarrow{U_m} -cA_1^0|E\rangle + sA_3^{1'}|\tau_0\rangle + sA_4^{1'}|\tau_1\rangle \end{cases}
 \end{array} \quad (2)$$

at time τ_1 , where $|cA_1^0| > 0$, $|cA_2^0| > 0$ and other amplitudes are arbitrary. Squiggly arrows represent the passing of an amount of time τ_0 or τ_1 , while straight arrows the application of U_m .

Suppose the E outcome is obtained.¹ At time τ_0 , a non zero amplitude associated with $|E\rangle$ exists only for the off branch, while for time τ_1 , this amplitude from the off branch cancels with an on branch amplitude due to destructive interference. As such, when E is obtained, we can deduce that the clock has collapsed to a branch of the wave function that was always off *and* that the unknown time t must be τ_0 . Similarly, when outcome A is obtained, we deduce that the clock was collapsed to a branch of the wave function that was always off *and* that the time t must be τ_1 . However, if outcome τ_0 or τ_1 is obtained, we cannot conclude anything useful: the clock may have been on and we cannot deduce whether the time is τ_0 or τ_1 .

Hence whenever outcome E or A is obtained, we can deduce what the time is, even though the system was always off, that is to say, always in a stationary state (since $|E\rangle$ and $|A\rangle$ are orthogonal to all dynamical branches whenever E , A have a non zero probability of being obtained). At first sight, one may object to this interpretation, since the system has certainly been in a superposition of on and off, and in this sense, has actually been evolving in time. However, the outcomes E and A can only arise via an always off branch of the wave function so if it is seen then we have been confined to a part of the total quantum state in which the system is always stationary. Furthermore, this interpretation is independent of the basis used to represent the state during the different stages of the protocol — we represented it in the measurement basis purely for convenience.

Connection with previous ideas: Our explanation of the results has relied on a certain kind of retrodictive interpretation of quantum mechanics. This interpretation can be most readily captured by Schrödinger’s proverbial cat experiment: one starts with an alive cat, denoted $|\text{alive}\rangle$, which when put in the closed box, takes on the form $\frac{1}{\sqrt{2}}(|\text{alive}\rangle + |\text{dead}\rangle)$. Suppose that upon opening the box, we make a measurement of alive versus dead, and obtain the outcome “alive”. Then in the

conventional collapse of the wave function formalism, the state of the cat changes discontinuously into $|\text{alive}\rangle$ and the $|\text{dead}\rangle$ component ceases to have any further physical existence or further consequence — see fig. 1. This is a form of retrodiction, since it says that upon observing an alive cat we collapse to a state in which the cat was *always* alive, since the dead and alive branches were orthogonal at all times. In our setup, this retrodiction takes on the form of a so-called *counterfactual measurement*,² since when outcome E or A is obtained, it allows one to deduce properties of a dynamical clock by reasoning counterfactually, while collapsing the wave function to a state where those eventualities never actually took place. Counterfactual measurements were first discovered by Elitzur and Vaidman in their famous bomb tester experiment [12–14], and later applied to other scenarios including (most prominently) counterfactual computation [15–19], and experimentally verified [20–22]. Prior examples have played out in space, while ours is the first protocol to do so in time. Another example of retrodiction and counterfactual measurement in quantum mechanics is the celebrated Hardy paradox [23]. Weakly measuring the distinct branches can provide theoretical and experimental evidence for the validity of such retrodictive interpretations [24–26].

We will call our clock when operated in this retrodictive fashion a *counterfactual clock*. Likewise, the measurement outcomes of interest in a counterfactual measurement (E and A), are referred to as counterfactual outcomes. If the above protocol is implemented and a counterfactual outcome is obtained, we say that the clock was *always off*. In fig. 2 we compare one instance of the Elitzur and Vaidman bomb test with one instance of the counterfactual clock.

Our analysis thus far has been presented *forward in time*, namely starting from the initial state $|E\rangle$ and analysing the different steps of the protocol until post-selecting on a counterfactual outcome. We can also examine the *backwards in time* process where we start with one of the post-selected states, and examine how it evolves backwards in time to the initial state $|E\rangle$. Whenever analysing paradoxes involving post-selection, it is prudent to check that the paradox still exists when the problem is analysed backwards in time,

¹ We use the convention in which “outcome x ” means that the post-measurement state is $|x\rangle$.

² Also known as “interaction-free measurement” or “quantum interrogation”.

since this can reveal any “hidden” assumptions inadvertently made due to post-selecting. It is important to do so in the case of counterfactual experiments, as shown in [27]. We prove in methods section IV A that when the counterfactual outcomes are analysed backwards in time, the pre-selected wave function is collapsed onto an always off branch. This guarantees that our counterfactual measurements have been properly implemented.

Of course, one can apply alternative interpretations of quantum mechanics. In the many-worlds interpretation of Schrödinger’s cat, upon measurement, reality splits into two parallel worlds — one in which the component $|\text{alive}\rangle$ prevails and the cat is alive, the other, in which the component $|\text{dead}\rangle$ prevails and the cat is dead. Conversely, the explanation of the counterfactual clock also changes if one uses this latter interpretation. Specifically, in the language of many-worlds, whenever the outcome E or A is obtained, we will be living in a world in which the system never dynamically evolved (i.e. never in an on state) yet in this world we learn information about the dynamical states (i.e. the on branch), namely what the time is. The fact that the system was evolving in another world is of no consequence to *us*.

FIG. 1: **Schrödinger’s cat.** **A1)** Schrödinger’s cat starts out alive. At the 1st dotted line, the box is closed and the cat is in a superposition of dead and alive until just before the box is opened (2nd dotted line). **A2)** The box has now been opened, causing the cat to be measured. Here we show the case in which we see an alive cat. The history of the cat from before the box was closed to present is now set in stone, with the dead cat branch of the wave function ceasing to have ever existed.

The role of the ancilla is more subtle than may appear at first sight: indeed, one might envisage a situation with no ancilla and where we interpret outcomes E and τ_1 as counterfactual outcomes by choosing the unitary U_m differently to as presented. Notwithstanding, we prove in supplementary material A 5, that there does not exist a counterfactual clock for which there is no ancilla and the probabilities of the two counterfactual outcomes are both non zero. This implies a more fundamental role for the ancilla state: by including it, we can reach a larger class of unitary transformations which are necessary for the counterfactual clock

FIG. 2: **Comparison between the Elizur-Vaidman bomb test and the counterfactual clock**

A1) Mach-Zehnder interferometer with 50/50 beam splitters and a photosensitive bomb placed in the bottom arm. Photon trajectories are black lines. A photon enters at a) and splits into a superposition travelling down the upper and lower arms upon interacting with the 1st beam splitter. The photon in the bottom arm is coherently absorbed by the bomb, hence exploding. The photon in the upper arm encounters the 2nd beam splitter and coherently splits into trajectories b), c) without bomb exploding.

A2) Continuing on from A1), a measurement is now performed via the photodetectors D1 and D2. There are 3 possible outcomes: If photon took path c), D1 clicks. If photon took path b), D2 clicks. If bomb exploded, photon is lost and neither photodetector clicks. Fig A2 depicts the scenario that D2 clicks (the counterfactual outcome).

Upon clicking, the wave function collapses into an eventuality where the bomb never exploded. **B1)** The counterfactual clock starts out off. At the 1st dotted line, the 1st event occurs and the clock is put into a superposition of on and off until just before the 2nd event occurs (2nd dotted line). **B2)** The 2nd event now occurs, causing the clock to be measured. Here we show the case in which we see a counterfactual outcome. The history of the clock from before the 1st event occurred to present is not set in stone, with the on clock branch of the wave function ceasing to have ever existed. **Comparison**

between A and B: Both use a counterfactual measurement to avoid the eventualities of branch 1 of the wave function occurring when obtaining certain outcomes.

to function. This is important for understanding what is the minimal model for making the appropriate comparison with classical descriptions, which we will do shortly.

In this set-up, the optimal protocol is the one which maximises the probabilities of obtaining either measurement outcomes E or A over unitaries U_m and branching amplitudes c, s . In the case in which both outcomes are equally likely, $|cA_1^0|^2 = |cA_2^0|^2$; we prove in supplementary material A 5 that there exists a protocol which assigns probability $1/6$ to obtaining the outcome E at time τ_0 and probability $1/6$ to outcome A at time τ_1 .

One may wonder whether a way out of this conundrum is via the existence of an underlying description respecting realism, that is, a theory where the relevant physical properties have well defined values at all times prior to the measurement in our protocol. For example, such a theory could have hidden variables which evolve dynamically when the clock is supposedly off. We prove that there cannot exist a non-contextual ontic model for the relevant degrees of freedom for the above counterfactual clock with probability $1/6$ for each counterfactual outcome. It also rules out the possibility of assigning elements of reality (an ontic description) to both the on and off branches simultaneously. When a counterfactual outcome is obtained, we must assign an element of reality to the off branch of the wave function since it was observed, thus it is not possible to assign an element of reality (ontic description) to the dynamical branch under said circumstances. Thus our proof that a non-contextual ontic model cannot exist rules out all classical models under natural assumptions; see methods IV B for details.

III. DISCUSSION

With the aid of counterfactual measurements, we have shown that quantum mechanics allows one to predict the time passed between two events, even though no dynamics has taken place between the two events. The most elementary form of this clock presented in the main text should be obtainable from ubiquitous systems in nature (a four dimensional system with two stationary and two non-stationary states), while the more advanced engineered clocks providing a larger number of distinguishable times (large N_T) require more ancillas and should be realisable with state of the art current technologies such as those developed for quantum computation.

While it is readily clear that a classical clock cannot tell the time without dynamical evolution, it could have been that our quantum clock had hidden variables which were changing when the clock was supposedly off — we ruled out such a possibility under reasonable assumptions.

We will now discuss the implications for the rela-

tional and substantial theories of time which were presented in the introduction. As explained previously, one of the reasons why the debate has been ongoing for more than two thousand years, is due to the lack of the possibility to experimentally distinguish between the two theories due to an apparent inability to measure the flow of time without evolution. Our results can remedy this dilemma by the following experiment:

Alice has a random number generator which outputs heads or tails. If she finds heads (tails), she sends two light pulses to Bob separated by an elapsed time τ_0 (τ_1). Bob possesses a counterfactual clock in the energy eigenstate $|E\rangle$, which can distinguish between two elapsed times τ_0 and τ_1 , but does not know the outcome of the random number generator. Alice tunes her two light pulses so that when they come into contact with Bob's clock, the 1st pulse implements U_0 while the 2nd implements U_m . Bob measures his clock in the basis $|E\rangle, |\tau_0\rangle, |\tau_1\rangle, |A\rangle$ upon seeing the second light pulse. Alice repeats this experiment until Bob reports to her that he found a counterfactual outcome. There is a second system near Bob which remains in an energy eigenstate during the entire experiment and does not interact with the light pulses; Bob will use it for reference purposes only. Consider the last run of the experiment: the clock did not change between the two light pulses (relative to the second system) but was nevertheless responsible for the detection and quantification of an elapsed time which is identical to what a clock which was dynamically evolving would have predicted (had one been present). We argue that in a relative theory of time, from the perspective of the clock relative to the second system, any value for the elapsed time would have to be arbitrary, since time would not exist from its perspective. Contrarily, the experiment is perfectly consistent with a substantial theory of time, since in this case, from the clock's perspective, there is a well-defined elapsed time between the events, as time passes independently of the clock's state of motion.

We can formalise this observation via the following assumptions and theorem:

- (A) Counterfactual measurements exist.
- (B) If time can be measured via a clock which is always off, then time is substantial.

Theorem 1. Time is substantial if assumptions (A) and (B) hold.

We prove the theorem in methods IV C. If (A) is false, then all the other phenomena which uses counterfactual measurements, such as counterfactual computing or the bomb test, would require serious re-evaluation. Similarly, if (A) holds yet (B) is false, then arguably, the very meaning of a substantial theory of time requires substantial review.

Finally, we conclude with a discussion regarding how one might modify our conclusions. Our results relied on the validity of the counterfactual measurement interpretation in quantum mechanics, and we discussed

how in the alternative interpretation of many-worlds, when one of the counterfactual outcomes is obtained, while *we* would be in a world where the clock was always off, there would be another world where it would be on. Regarding the theory of time debate, this interpretation would still lead to difficulties for a relational theory of time, since in our thought experiment the second system remains in Bob's world, so when he obtains the counterfactual outcome, there would have been no relative change between the clock and the second sys-

tem in Bob's world, yet Bob would know how much time had passed in his world from the measurement outcome.

As alluded to in the introduction, our beliefs about the nature of time have been highly influential in the development of our physical theories. Going forward, perhaps one of the key ingredients to finding a correct theory of quantum gravity is contingent of asserting an appropriate belief about the nature of time. We hope that our work will bring much needed attention to this debate.

IV. METHODS

Here we provide more details concerning the claims made in the main text. There are three subsections. The first, section IV A, describes the two state formalism analysis in which the states evolve both forwards and backwards in time. The second, section IV B, describes the setup and how we performed our calculations which led us to conclude that our counterfactual clock has no non-contextual ontic description. The third, section IV C, contains a proof of theorem 1.

A. Backwards in time analysis

Here we analyse our protocol backwards in time for the post-selected outcomes of interest. The objective is to show that the clock was off when viewed from this perspective. For simplicity, we analyse the case $N_T = 1$ presented in the main text where the probability of obtaining either counterfactual outcome is $1/6$. The general case is relegated to supplementary material A 4. The physical interpretation is the same in all cases.

The amplitudes associated with obtaining the counterfactual outcomes are given by

$$\langle E|U_m U(\tau_0)U_0|E\rangle \quad \text{and} \quad \langle A|U_m U(\tau_1)U_0|E\rangle, \quad (3)$$

for times $t = \tau_0$ and $t = \tau_1$ respectively, where $U(\tau_j)$ is the unitary dynamics of the clock (when it is on) for a time τ_j . In the main text we interpreted these processes going forwards in time, namely starting in state $|E\rangle$ and analysing the state after applying U_0 , $U(\tau_j)$, U_m and finally post-selecting on measurement outcomes $|E\rangle$ and $|A\rangle$. Here we will analyse the reverse process: starting with the post-selected state $\langle E|$ or $\langle A|$, evolving backwards in time by analysing the affect of the sequence of unitaries U_m , $U(\tau_j)$, U_0 , followed by pre-selecting onto the initial state $|E\rangle$. This process is known as the two-state vector formalism. This formalism, and many other variants on this idea, permit time-symmetric interpretations of quantum mechanics (see review [28] for more details).

For the example under consideration, U_m is provided at the end of supplementary material A 3. Using it we find that the backwards evolved post-selected counterfactual outcome state just prior to dynamics take on the form:

$$\langle E|U_m = \frac{1}{\sqrt{3}}(\langle E| - \langle \tau_1| - \langle A|) \quad \text{and} \quad \langle A|U_m = \frac{1}{\sqrt{3}}(\langle E| - \langle \tau_0| - \langle A|). \quad (4)$$

We now have to evolve backwards in time for an amount τ_0 . Recall that $U(\tau_0)$ induces the following transitions: $U(\tau_0)|E\rangle = |E\rangle$, $U(\tau_0)|A\rangle = |A\rangle$, $U(\tau_0)|\psi\rangle = |\tau_0\rangle$. Consider the ansatz $\langle \tau_1|U(\tau_0) = c_0\langle \psi_\perp| + c_2\langle E| + c_3\langle \psi|$, where $|\psi_\perp\rangle$ is some state which is perpendicular to $|E\rangle$ and $|\psi\rangle$. We have $0 = \langle \tau_1|E\rangle = \langle \tau_1|U(\tau_0)U(\tau_0)^\dagger|E\rangle = c_2$. Similarly we find that $c_3 = 0$ and thus that $\langle \tau_1|U(\tau_0) = \langle \psi_\perp|$. Furthermore, since $U(\tau_1)|E\rangle = |E\rangle$, $U(\tau_1)|A\rangle = |A\rangle$, $U(\tau_1)|\psi\rangle = |\tau_1\rangle$, it follows in an analogous fashion that $\langle \tau_0|U(\tau_1) = \langle \psi'_\perp|$, where $\langle \psi'_\perp|$ is some state perpendicular to $|E\rangle$ and $|\psi\rangle$. Therefore, from eq. (4), the backwards in time dynamics yields the states

$$\langle E|U_m U(\tau_0) = \frac{1}{\sqrt{3}}(\langle E| - \langle \psi_\perp| - \langle A|) \quad \text{and} \quad \langle A|U_m U(\tau_1) = \frac{1}{\sqrt{3}}(\langle E| - \langle \psi'_\perp| - \langle A|). \quad (5)$$

The unitary U_0 only acts non trivially in the subspace spanned by $|E\rangle$ and $|\psi\rangle$: $U_0|E\rangle = c|E\rangle + s|\psi\rangle$, $U_0|\psi\rangle = -s|E\rangle + c|\psi\rangle$, and thus it acts trivially on the dynamical branches $\langle \psi_\perp|$ and $\langle \psi'_\perp|$. We thus have

$$\langle E|U_m U(\tau_0)U_0 = \frac{1}{\sqrt{3}}(c\langle E| - s\langle \psi| - \langle \psi_\perp| - \langle A|) \quad \text{and} \quad \langle A|U_m U(\tau_1)U_0 = \frac{1}{\sqrt{3}}(c\langle E| - s\langle \psi| - \langle \psi'_\perp| - \langle A|). \quad (6)$$

Finally, by post-selecting on $|E\rangle$, and recalling that $c = 1/\sqrt{2}$, we find that the amplitudes associated with the counterfactual outcomes is $1/\sqrt{6}$ in both cases, hence leading to a probability of $1/6$ for each counterfactual outcome as stated in the main text. Importantly, during the entire backwards evolution, we see that the dynamical branches of the wave function has not contributed to the pre-selected state $|E\rangle$. Another way to see this is to note that had we measured whether the clock was on or off during the evolution in either of the above pre and post selected cases, we would have found that the probability of finding the clock on was zero. Namely, we can define the binary outcome projective measurement $P_{\text{on}} := |\psi\rangle\langle\psi| + |\tau_0\rangle\langle\tau_0| + |\tau_1\rangle\langle\tau_1|$ and $P_{\text{on}} := |E\rangle\langle E| + |A\rangle\langle A|$, where P_{on} corresponds to a projection onto the on subspace. From [28], it follows that the probability of the clock being on at any time during the dynamical phase, that is any $t \in (0, \tau_0)$, for the run of the experiment corresponding to post-selecting on $|E\rangle$ is

$$\text{Prob}(\text{on}, E) := \frac{|\langle E|U_m U(\tau_0 - t)P_{\text{on}}U(t)U_0|E\rangle|^2}{|\langle E|U_m U(\tau_0 - t)P_{\text{on}}U(t)U_0|E\rangle|^2 + |\langle E|U_m U(\tau_0 - t)P_{\text{on}}U(t)U_0|E\rangle|^2}. \quad (7)$$

Note that since the ancilla and energy eigenstate are stationary, we have that $U(t) = U_{\perp E, A}(t) + \mathbb{1}_{E, A}$, where $\mathbb{1}_{E, A}$ is the identity operator on the subspace spanned by $|E\rangle$ and $|A\rangle$, and that $U_{\perp, A}(t)$ is orthogonal to said subspace. Therefore, if we assume that the dynamics is Markovian, namely $U(t_1)U(t_2) = U(t_1 + t_2)$ for all $t_1, t_2 \in \mathbb{R}$, it follows that $U(\tau_0 - t)P_{\text{on}}U(t) = U_{\perp E, A}(\tau_0)$ and hence that

$$\text{Prob}(\text{on}, E) = 0 \quad (8)$$

for all $t \in (0, \tau_0)$. Similarity, it follows that the probability of the clock being on at any time during the dynamical phase, that is any $t \in (0, \tau_1)$, for the run of the experiment corresponding to post-selecting on $|A\rangle$ is

$$\text{Prob}(\text{on}, A) := \frac{|\langle A|U_m U(\tau_1 - t)P_{\text{on}}U(t)U_0|E\rangle|^2}{|\langle A|U_m U(\tau_1 - t)P_{\text{on}}U(t)U_0|E\rangle|^2 + |\langle A|U_m U(\tau_1 - t)P_{\text{on}}U(t)U_0|E\rangle|^2} = 0. \quad (9)$$

B. No classical analogue

The interpretation of measuring time when the clock was off using counterfactual reasoning relied on a basic concept in quantum mechanics, namely, the superposition principle. From this principle one infers that if a system is in a superposition of two states, it cannot be regarded as being in either until measured. As we discussed, Schrödinger famously popularised this point with his thought experiment concerning a cat. However, what if underlying our counterfactual clock protocol, there was a “real” description; that is to say if one could associate the states in our clock protocols with probability distributions λ over some underlying states — a so-called *ontic state space* Λ — and the measurements with update rules for said distributions, in a meaningful way? While such a result would not invalidate the current interpretation provided, it would, at least in principle, provide for an alternative interpretation in which the clock might have been dynamically evolving even when a measurement collapsing the clock to the off branch is obtained. We will now consider such a possibility for the simplest elementary timekeeping systems.

To start with, let's consider the case where we only run the clock when it is on; that is, the clock is never in a superposition of on and off. Then the question is: can we find an ontic state space on which there exists appropriate probability distributions that represent our states $\{|E\rangle\langle E|, |\tau_0\rangle\langle\tau_0|, |\tau_1\rangle\langle\tau_1|\}$ and measurements (with PVM elements $\{|E\rangle\langle E|, |\tau_0\rangle\langle\tau_0|, |\tau_1\rangle\langle\tau_1|, |A\rangle\langle A|\}$)? Moreover, if an underlying ontic state space does exist, one should expect to be able to take probabilistic mixtures of the states and measurements in it — if your lab assistant walked in to your lab at some unknown stage of the protocol's implementation, the assistant would attribute such a description. Thus denoting the convex hull by conv , we required that when a state $\rho \in \mathbf{s}_{\text{on}} = \text{conv}(\{|E\rangle\langle E|, |\tau_0\rangle\langle\tau_0|, |\tau_1\rangle\langle\tau_1|\})$ is prepared in our protocol, the probability that the ontic state is in state λ , is $P[\lambda|\rho]$. Likewise, when we make the canonical measurement with corresponding POVM element $\mathbf{E} \in \mathbf{e}_{\text{on}} = \text{conv}(\{|E\rangle\langle E|, |\tau_0\rangle\langle\tau_0|, |\tau_1\rangle\langle\tau_1|, |A\rangle\langle A|\})$, then the probability that the outcome associated with $\mathbf{E}$ occurred, given that the ontic state is in state λ , is $P[\mathbf{E}|\lambda]$. Furthermore, we require the usual convexity relationships of mixtures of quantum states and measurements hold at the ontic level: for all $\lambda \in \Lambda$, for all $p \in [0, 1]$, for all $\rho_1, \rho_2 \in \mathbf{s}_{\text{on}}$: $P[\lambda|p\rho_1 + (1-p)\rho_2] = pP[\lambda|\rho_1] + (1-p)P[\lambda|\rho_2]$ and similarly, that for all $\lambda \in \Lambda$, for all $p \in [0, 1]$, for all $\mathbf{E}_1, \mathbf{E}_2 \in \mathbf{e}_{\text{on}}$: $P[p\mathbf{E}_1 + (1-p)\mathbf{E}_2|\lambda] = pP[\mathbf{E}_1|\lambda] + (1-p)P[\mathbf{E}_2|\lambda]$. Finally, we require that the ontic states can reproduce the measurement statistics of our protocol: for all $\rho \in \mathbf{s}_{\text{on}}$, for all $\mathbf{E} \in \mathbf{e}_{\text{on}}$,

$$\text{tr}[\rho \mathbf{E}] = \int_{\Lambda} d\lambda P[\lambda|\rho] P[\mathbf{E}|\lambda], \quad (10)$$

where $\text{tr}[\cdot]$ denotes the trace, and the r.h.s. is the probability of obtaining outcome associated with E for quantum state ρ , according to quantum mechanics.

Notice how we have assumed that the probability distributions $P[\lambda|\rho]$ do not depend on how the quantum state ρ was prepared nor how the distribution $P[E|\lambda]$ depends on how the measurements were implemented. Different preparation procedures which lead to the same quantum state and different measurement procedures which lead to the same POVM elements, are called different *contexts*. We are thus looking for a non-contextual ontic variable model. This notion of contextuality was pioneered by Spekkens [29]. Prior notions of classicality based on local realism [30] or its generalisation to non-contextual realism [31], cannot be tested in our case since we do not dispose of a local structure, nor relevant commuting measurements, in our clock protocols. However, in keeping with this philosophy of preparation and measurement non-contextuality, our current formulation of an ontic variable model requires some refinement: since the set $\mathbf{e}_{\text{on}}$ is not tomographically complete, nor the set $\mathbf{s}_{\text{on}}$ span all quantum states in the Hilbert space, neither all the d.o.f. of the density matrices nor those in the POVM elements, contribute to the measurement statistics. Indeed, $\text{tr}[\rho E] = \text{tr}[\mathcal{P}_{\mathcal{R}}(\rho)\mathcal{P}_{\mathcal{R}}(E)]$ for all $\rho \in \mathbf{s}_{\text{on}}$, $E \in \mathbf{e}_{\text{on}}$, where $\mathcal{P}_{\mathcal{R}}$ is a projection onto the vector space $\mathcal{R}$, generated by projecting the span of $\mathbf{s}_{\text{on}}$ onto the span of $\mathbf{e}_{\text{on}}$. We should thus make the replacements $P[\lambda|\cdot] \mapsto P[\lambda|\mathcal{P}_{\mathcal{R}}(\cdot)]$ and $P[\cdot|\lambda] \mapsto P[\mathcal{P}_{\mathcal{R}}(\cdot)|\lambda]$ in the above theory. This completes the description of our would-be non-contextual ontic variable model. Finally, if a classical description exists, one would also need a stochastic map describing the dynamics of the ontic variables throughout the protocols. This imposes an additional constraint on our would-be ontic description, which we will not need to consider.

In this case, it can readily be seen that such a model *does* exist; for example, one may simply choose the vectors $\{|E\rangle, |\tau_0\rangle, |\tau_1\rangle, |A\rangle\}$ as a basis for the ontic state space Λ with each step and measurement allocated to deterministic distributions λ on it.

We now turn to the case in which the clock can also be used to tell the time when off via counterfactual measurements. We need to supplement the sets $\mathbf{s}_{\text{on}}$, $\mathbf{e}_{\text{on}}$ with the other relevant elements which are now required, namely, for states,

$$\mathbf{s}_{\text{cf}} = \text{conv}(\mathbf{s}_{\text{on}} \cup \{|\text{cf}_0\rangle\langle\text{cf}_0|, |\text{cf}_1\rangle\langle\text{cf}_1|, U_m|\text{cf}_0\rangle\langle\text{cf}_0|U_m^\dagger, U_m|\text{cf}_1\rangle\langle\text{cf}_1|U_m^\dagger\}), \quad (11)$$

where $|\text{cf}_0\rangle := c|E\rangle + s|\tau_0\rangle$, $|\text{cf}_1\rangle := c|E\rangle + s|\tau_1\rangle$, since these are the states which appear in our protocol. When it comes to the measurements, in addition to those required in the final measurement, namely $\mathbf{e}_{\text{on}}$, we want to include measurements corresponding to ontic degrees of freedom which are able to describe the paradoxical aspects of our protocol. Before applying U_m , at times τ_0, τ_1 , when we measure in the measurement basis, the measurement determined which branch (off or on) we collapsed to, but we can only deduce the time when we happen to collapse onto an on branch. This situation is analogous to a statistical mixture over on and off. What is surprising, is that after U_m is applied, we can deduce not only which branch we were on, but also the time when we collapse to an off branch. So our would-be ontic model should have a variable which determines whether we are in the off branch, or the on branch at the times τ_0, τ_1 . In other words, a variable which, after the application of U_m , plays the same role that the measurement basis played before the application of U_m . Including this in the set of things which are measurable in our would-be non-contextual theory, gives us

$$\mathbf{e}_{\text{cf}} = \text{conv}(\mathbf{e}_{\text{on}} \cup \{U_m^\dagger|E\rangle\langle E|U_m, U_m^\dagger|\tau_0\rangle\langle\tau_0|U_m, U_m^\dagger|\tau_1\rangle\langle\tau_1|U_m, U_m^\dagger|A\rangle\langle A|U_m\}). \quad (12)$$

An alternative motivation for the inclusion of the additional ontic degrees of freedom, is that, quantum mechanically, the unitary channel generated by U_m could have been applied to rotate the measurement basis, rather than being applied to the state. This alternative protocol, is equivalent to the one we study here, from the perspective of quantum mechanics. We can thus think of these two alternative implementations of our protocol as different contexts which are indistinguishable according to the laws of quantum mechanics, in an analogous way to how our would-be ontic variable model is preparation and measurement non-contextual by design.

In [32] an algorithm was developed which, given a set $\mathbf{s}$ of states and set $\mathbf{e}$ of POVM elements, can determine whether an ontic variable model as described here exists. When said sets have finitely many extremal points, as in the case for $\mathbf{s}_{\text{cf}}, \mathbf{e}_{\text{cf}}$, it provably runs in finite time. For the simple case where the probability of each counterfactual outcome is $1/6$ as described in the main text, the unitary U_m is given by eq. (A7a) and c, s by eq. (A7b). For this case, we used the algorithm to provide a computer assisted proof that the sets $\mathbf{s}_{\text{cf}}, \mathbf{e}_{\text{cf}}$ — and hence our counterfactual clock — do *not* admit a non-contextual ontic variable model as per the above description; see supplementary material B for details.

A crucial aspect of our protocol for using the clock to tell the time when off, via counterfactual measurements, was the existence of negative amplitudes allowing for destructive interference between the on and off branches. Since probabilities are nonnegative, one might have thought that this aspect of our protocol automatically rules out any non-contextual realistic theory. However, this is definitely not the case since other experiments using

counterfactual measurements, such as the Elitzur-Vaidman bomb test [12], have been shown to admit a non-contextual ontic model description analogous to the one ruled out here for our setup; see [33].

C. Proof of Theorem 1

Here we prove the theorem from the main text:

Theorem 1. Time is substantial if assumptions (A) and (B) hold.

Note that while we will only need to consider the counterfactual clock explained in the main text (i.e. the $N_T = 1$ case), the proof is analogous in the case one uses the clocks generalised for $N_T \geq 2$ found in supplementary material A. In the latter case, the definitions of a *counterfactual measurement* and of an *always off* clock which are found in the main text and used in assumptions (A) and (B), would be substituted for the more general definitions found in the supplementary material; i.e. definitions 1 and 2.

Proof. Assumption (A) together with the example of a counterfactual clock for $N_T = 1$ provided in the main text (the values for U_m and c, s which fully determine the protocol are provided at the end of A3 in the supplementary material), prove that for any two elapsed times τ_0, τ_1 between two events, a counterfactual clock exists and can determine said elapsed time with the clock always off. Therefore, the theorem follows by invoking assumption (B). $\square$

ACKNOWLEDGMENTS

We thank Laura Burri for insightful feedback on an early draft of this manuscript. We are grateful to Thomas Galley, Victor Gitten and Richard Jozsa for discussions. M.W. acknowledges funding from the Swiss National Science Foundation (AMBIZIONE Fellowship, No. PZ00P2 179914) in addition to supported from the Swiss National Science Foundation via the National Centre for Competence in Research ‘‘QSIT’’. S.S. acknowledges support from the QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union’s Horizon 2020 Programme (QuantAlgo project), and administered through EPSRC Grant No. EP/R043957/1, and support from the Royal Society University Research Fellowship scheme. M.W. and S.S. both acknowledge funding from the Royal Society International Exchanges program (grant No. IES\R1\1801060).

AUTHOR CONTRIBUTIONS

M.W. derived the protocols and wrote the paper. S.S. implemented the numerical proof for the non existence of a non-contextual ontic model. Both authors contributed to ideas and discussions.

-
- [1] T. Purves and A. J. Short, ‘‘Quantum theory cannot violate a causal inequality,’’ (2021), arXiv:2101.09107.
 - [2] J. Wechs, H. Dourdent, A. A. Abbott, and C. Branciard, ‘‘Quantum circuits with classical versus quantum control of causal order,’’ (2021), arXiv:2101.08796.
 - [3] Aristotle, *Physics*, Book IV, Chapters 10 and 11. (Neeland Media, 2006) Translated by R.P. Hardie and R.K. Gaye.
 - [4] R. Rynasiewicz, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (Metaphysics Research Lab, Stanford University, 2014) summer 2014 ed.
 - [5] J. K. McDonough, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (Metaphysics Research Lab, Stanford University, 2019) fall 2019 ed., Section 5: Leibniz on the Space and Time.
 - [6] P. Pojman, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (Metaphysics Research Lab, Stanford University, 2020) winter 2020 ed.
 - [7] C. Hofer, *Studies in History and Philosophy of Science Part A* **25**, 287 (1994).
 - [8] T. Maudlin, in *The Metaphysics Within Physics* (Oxford University Press, 2007) pp. 104–142.
 - [9] C. Rovelli, *The Order of Time* (Penguin Books Limited, 2018).
 - [10] L. Smolin, *Time Reborn: From the Crisis in Physics to the Future of the Universe* (Houghton Mifflin Harcourt, 2013).
 - [11] J. Barbour, *The End of Time: The Next Revolution in Physics* (Oxford University Press, USA, 2001).
 - [12] A. C. Elitzur and L. Vaidman, *Foundations of Physics* **23**, 987 (1993).
 - [13] L. Vaidman, *Foundations of Physics* **33**, 491 (2003).
 - [14] R. Penrose, *Shadows of the Mind: A Search for the Missing Science of Consciousness* (Oxford University Press, 1994).
 - [15] G. Mitchison and R. Jozsa, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **457**, 1175 (2001).

- [16] R. Jozsa, in *1st NASA Conference on Quantum Computing and Quantum Communications* (1998) arXiv:quant-ph/9805086.
- [17] P. Kwiat, H. Weinfurter, T. Herzog, A. Zeilinger, and M. A. Kasevich, *Phys. Rev. Lett.* **74**, 4763 (1995).
- [18] C. Y.-Y. Lin and H.-H. Lin, in *Proceedings of the 30th Conference on Computational Complexity, CCC '15* (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, DEU, 2015) p. 537–566.
- [19] M. Hasenöhl and M. M. Wolf, “‘Interaction-Free’ Channel Discrimination,” (2020), arXiv:2010.00623.
- [20] H. Weinfurter, P. Kwiat, T. Herzog, A. Zeilinger, and M. Kasevich, in *Quantum Electronics and Laser Science Conference* (Optical Society of America, 1995) Session: Spatial Quantum Effects (QTuF).
- [21] P. G. Kwiat, A. G. White, J. R. Mitchell, O. Nairz, G. Weihs, H. Weinfurter, and A. Zeilinger, *Phys. Rev. Lett.* **83**, 4725 (1999).
- [22] A. G. White, P. G. Kwiat, O. Nairz, G. Weihs, and A. Zeilinger, in *International Quantum Electronics Conference* (Optical Society of America, 1998) p. QMF1.
- [23] L. Hardy, *Phys. Rev. Lett.* **68**, 2981 (1992).
- [24] Y. Aharonov, A. Botero, S. Popescu, B. Reznik, and J. Tollaksen, *Physics Letters A* **301**, 130 (2002).
- [25] J. S. Lundeen and A. M. Steinberg, *Phys. Rev. Lett.* **102**, 020404 (2009).
- [26] K. Yokota, T. Yamamoto, M. Koashi, and N. Imoto, *New Journal of Physics* **11**, 033011 (2009).
- [27] L. Vaidman, *Phys. Rev. Lett.* **98**, 160403 (2007).
- [28] Y. Aharonov and L. Vaidman, “The two-state vector formalism: An updated review,” in *Time in Quantum Mechanics*, edited by J. Muga, R. S. Mayato, and Í. Egusquiza (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) pp. 399–447.
- [29] R. W. Spekkens, *Phys. Rev. A* **71**, 052108 (2005).
- [30] J. S. Bell, *Physics Physique Fizika* **1**, 195 (1964).
- [31] S. Kochen and E. P. Specker, *Journal of Mathematics and Mechanics* **17**, 59 (1967).
- [32] V. Gitton and M. P. Woods, “Solvable criterion for the contextuality of any prepare-and-measure scenario,” (2020), arXiv:2003.06426.
- [33] R. Spekkens, M. Elliot, and M. Leifer, In preparation; see talk here <http://pirsa.org/16060102/>.
- [34] H. Havlicek and K. Svozil, *Entropy* **20** (2018), 10.3390/e20040284.
- [35] D. Avis, in *Polytopes-Combinatorics and Computation* (Springer, 2000) pp. 177–198.

Supplementary material

Supplementary material A: Elementary time keeping devices

In subsections A 1 and A 2, we detail the protocol one implements when using the clock to tell the time counterfactually. In subsection A 3, we characterise the measurement basis used at the end of the protocol from subsections A 1 and A 2. In subsection A 4 we prove some consequences of the protocols and definitions made in the previous subsections. Finally in subsection A 5, we provide two structure theorems which prove that ancillas are necessary and that counterfactual clocks exist for arbitrary $N_T \in \mathbb{N}_{>0}$.

1. Setup and protocol

In this subsection, we explain the full protocol which allows one to determine the time $t \in \{\tau_0, \tau_1, \tau_2, \dots, \tau_{N_T}\}$ using counterfactual measurements. For the purpose of performing general projective measurements, we append an m -dimensional ancilla space $\mathcal{H}_A$ to the Hilbert $\mathcal{H}_S$ of the system, leading to a total Hilbert space $\mathcal{H}_S \oplus \mathcal{H}_A$. We will require the number of ancillas to be greater or equal to the number of distinguishable time intervals, namely $m \geq N_T$. The ancilla states are always stationary, namely they do not change in time. When the clock is tuned on, it is initialised to another state denoted $|\psi\rangle$. Thus in total, we have the orthonormal basis states $\{|\psi\rangle, |E\rangle, |\tau_0\rangle, |\tau_1\rangle, \dots, |\tau_{N_T}\rangle, |A_1\rangle, |A_2\rangle, \dots, |A_m\rangle\}$, which we call the measurement basis (Since the clock never has support on the basis state $|\psi\rangle$ when measured, it will often be omitted for simplicity). Observe that the m ancillas could be other energy eigenstates of the system which are orthogonal to the states $|\tau_0\rangle, |\tau_1\rangle, \dots, |\tau_{N_T}\rangle, |E\rangle$, or they could be produced via a separate $m+1$ dimensional system via the identification $|\tau_0\rangle \equiv |\tau_0\rangle \otimes |d_0\rangle, |\tau_1\rangle \equiv |\tau_1\rangle \otimes |d_0\rangle, \dots, |\tau_{N_T}\rangle \equiv |\tau_{N_T}\rangle \otimes |d_0\rangle$, and $|E\rangle \equiv |E\rangle \otimes |d_0\rangle$, for original states of the system and $|A_1\rangle \equiv |E\rangle \otimes |d_1\rangle, |A_2\rangle \equiv |E\rangle \otimes |d_2\rangle, \dots, |A_m\rangle \equiv |E\rangle \otimes |d_m\rangle$, for the ancillary states, where $\{|d_0\rangle, |d_1\rangle, \dots, |d_m\rangle\}$, are orthonormal states of the separate $m+1$ dimensional system.

Initially we set the system to be in the energy eigenstate, $|E\rangle$. Then, when the 1st event occurs, a unitary is applied to take the state to a superposition of $|E\rangle$ and $|\psi\rangle$. Using the branching notation from the main text to distinguish between the two orthogonal states, this has the form:

$$\begin{array}{c}
 \bullet c|E\rangle \\
 \diagup \\
 |E\rangle \\
 \diagdown \\
 \bullet s|\psi\rangle
 \end{array} \tag{A1}$$

where $c := \cos \theta$ and $s := \sin \theta$. We then wait until the 2nd event occurs at unknown time τ_l . The system now is of the form

$$\begin{array}{c}
 \bullet c|E\rangle \rightsquigarrow^{\tau_l} c|E\rangle \\
 \diagup \\
 |E\rangle \\
 \diagdown \\
 \bullet s|\psi\rangle \rightsquigarrow^{\tau_l} s|\tau_l\rangle
 \end{array} \tag{A2}$$

where $\tau_l \in \{\tau_0, \tau_1, \dots, \tau_{N_T}\}$. We now perform a unitary U_m over the system. The quantum states have support on the ancillas after the application of the unitary. The system now takes on the form:

$$\begin{array}{c}
 \bullet c|E\rangle \rightsquigarrow^{\tau_l} c|E\rangle \xrightarrow{U_m} cA_0^0|E\rangle + \sum_{j=1}^{N_T} cA_j^0|A_j\rangle + \sum_{k=N_T+1}^m cB_{-1,k}|A_k\rangle + \sum_{k=1}^{N_T+1} cB_{-1,k}|\tau_{k-1}\rangle \\
 \diagup \\
 |E\rangle \\
 \diagdown \\
 \bullet s|\psi\rangle \rightsquigarrow^{\tau_l} s|\tau_l\rangle \xrightarrow{U_m} sA_0^1(\tau_l)|E\rangle + \sum_{j=1}^{N_T} sA_j^1(\tau_l)|A_j\rangle + \sum_{k=N_T+1}^m sB_{l,k}|A_k\rangle + \sum_{k=1}^{N_T+1} sB_{l,k}|\tau_{k-1}\rangle
 \end{array} \tag{A3}$$

where the amplitudes $A_q^0, A_q^1(\tau_l)$ are to be characterised in supplementary material A 3. To finalise the protocol, we immediately measure in the measurement basis after U_m is applied. Here we will use the convention that the counterfactual outcomes are associated with the post-measurement states $\{|E\rangle, |A_1\rangle, \dots, |A_{N_T}\rangle\}$.

2. Counterfactual measurement (Formal definition)

Given the presentation of the general protocol in supplementary material A 1, we can now formally define the counterfactual measurement which is appropriate for our setting. These form a generalisation of the main text to the $N_T > 1$ case. In the following, we label the energy E by $A_0 := E$ for simplicity of notation (since the symbol A_0 has not been used previously, this should not lead to confusion).

Definition 1 (Counterfactual measurement for clocks for $N_T \in \mathbb{N}_{>0}$). We call a projective measurement a *counterfactual measurement* if it has outcomes $A_0, A_1, A_2, \dots, A_{N_T}$ (e.g. post-measurement states $|A_0\rangle, |A_1\rangle, |A_2\rangle, \dots, |A_{N_T}\rangle$ respectively) of which at least one of these outcomes is a *counterfactual outcome*. Furthermore, we say that the outcome A_k is a counterfactual outcome if the following two equations are both satisfied.

$$A_k^1(\tau_k) = 0, \quad (\text{A4})$$

and

$$A_k^1(\tau_l) = -\frac{c}{s} A_k^0 \quad (\text{A5})$$

for all $l = 0, 1, 2, \dots, k-1, k+1, \dots, N_T$ and where $k \in \{0, 1, 2, \dots, N_T\}$.

Definition 2 (Clock always off for $N_T \in \mathbb{N}_{>0}$). When a counterfactual outcome A_k is obtained, we say that the elapsed time between the two events was τ_k , and that the clock was *always off*.

The reasoning behind these definitions is based on the usual arguments of counterfactual measurements: eq. (A4) allows us to conclude that if one obtains measurement outcome A_k and $t = \tau_k$, the clock must have been always off, since the amplitude in the on branch associate with the ket $|A_k\rangle$ is zero during said time interval (see eq. (A3)). Meanwhile, eq. (A5) guarantees that if τ_k is obtained, then we *must* have $t = \tau_k$, since the amplitude $A_k^1(\tau_l)$ cancels out with the amplitude A_k^0 for all times $t \neq \tau_k$, (see eq. (A3)). In supplementary material A 5 we show that these definitions are not vacuous, specifically, that clocks exist for any $N_T \in \mathbb{N}_{>0}$ with the probabilities of obtaining their counterfactual outcomes all non-zero. In supplementary material A 4 we will see that these definitions have other interpretational implications.

3. Characterisation of unitary U_m .

We consider the case in which we associate all of the measurement outcomes $E, \tau_1, \tau_2, \dots, \tau_{N_T}$ with the elapsed time being $\tau_0, \tau_1, \tau_2, \dots, \tau_{N_T}$ respectively, and the clock being off (this is to say, we associate $E, \tau_1, \tau_2, \dots, \tau_{N_T}$ with counterfactual outcomes). The constraints imposed by eqs. (A4) and (A5) constrain the matrix coefficients of the unitary matrix U_m . Specifically, by writing them out explicitly, we have that U_m is of the form $U_m = U_{\text{ex}} U'_m$ where U_{ex} implements a simple change of basis, defined via the relations: $U_{\text{ex}} |E\rangle = |E\rangle, U_{\text{ex}} |\tau_0\rangle = |A_1\rangle, U_{\text{ex}} |\tau_1\rangle = |A_2\rangle, \dots, U_{\text{ex}} |\tau_{N_T-1}\rangle = |A_{N_T}\rangle, U_{\text{ex}} |A_1\rangle = |\tau_0\rangle, U_{\text{ex}} |A_2\rangle = |\tau_1\rangle, \dots, U_{\text{ex}} |A_{N_T}\rangle = |\tau_{N_T-1}\rangle$, and $U_{\text{ex}} |\tau_{N_T}\rangle = |\tau_{N_T}\rangle, U_{\text{ex}} |A_{N_T+1}\rangle = |A_{N_T+1}\rangle, U_{\text{ex}} |A_{N_T+2}\rangle = |A_{N_T+2}\rangle, \dots, U_{\text{ex}} |A_m\rangle = |A_m\rangle$. The other matrix U'_m has the following matrix representation:

	$ E\rangle$	$ \tau_0\rangle$	$ \tau_1\rangle$	$ \tau_2\rangle$	$\dots$	$ \tau_{N_T-1}\rangle$	$ \tau_{N_T}\rangle$	$ A_1\rangle$	$\dots$	$ A_m\rangle$
$ E\rangle$	A_0^0	A_1^0	A_2^0	A_3^0	$\dots$	$A_{N_T}^0$	γ_0	$B_{-1,1}$	$\dots$	$B_{-1,m}$
$ \tau_0\rangle$	0	$-A_1^0 r$	$-A_2^0 r$	$-A_3^0 r$	$\dots$	$-A_{N_T}^0 r$	γ_1	$B_{1,1}$	$\dots$	$B_{1,m}$
$ \tau_1\rangle$	$-A_0^0 r$	0	$-A_2^0 r$	$-A_3^0 r$	$\dots$	$-A_{N_T}^0 r$	γ_2	$B_{2,1}$	$\dots$	$B_{2,m}$
$ \tau_2\rangle$	$-A_0^0 r$	$-A_1^0 r$	0	$-A_3^0 r$	$\dots$	$-A_{N_T}^0 r$	γ_3	$B_{3,1}$	$\dots$	$B_{3,m}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$ \tau_{N_T-1}\rangle$	$-A_0^0 r$	$-A_1^0 r$	$-A_2^0 r$	$-A_3^0 r$	$\dots$	$-A_{N_T}^0 r$	γ_{N_T}	$B_{N_T,1}$	$\dots$	$B_{N_T,m}$
$ \tau_{N_T}\rangle$	$-A_0^0 r$	$-A_1^0 r$	$-A_2^0 r$	$-A_3^0 r$	$\dots$	0	γ_{N_T+1}	$B_{N_T+1,1}$	$\dots$	$B_{N_T+1,m}$
$ A_1\rangle$	$B_{N_T+2,0}$	$B_{N_T+2,1}$	$B_{N_T+2,2}$	$B_{N_T+2,3}$	$\dots$	B_{N_T+2,N_T}	B_{N_T+2,N_T+1}	B_{N_T+2,N_T+2}	$\dots$	B_{N_T+2,N_T+m+1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$ A_m\rangle$	$B_{N_T+m+1,0}$	$B_{N_T+m+1,1}$	$B_{N_T+m+1,2}$	$B_{N_T+m+1,3}$	$\dots$	B_{N_T+m+1,N_T}	B_{N_T+m+1,N_T+1}	B_{N_T+m+1,N_T+2}	$\dots$	B_{N_T+m+1,N_T+m+1}

(A6)

where we have denoted $r := c/s = \cos(\theta)/\sin(\theta)$, and all matrix entries are arbitrary complex numbers such that $U'_m U_m{}^\dagger = U_m{}^\dagger U'_m = \mathbb{1}$. In writing matrix eq. (A6), we have used the convention that kets are row vectors and bras are column vectors. We will use this matrix representation convention throughout. The horizontal solid inner line and the vertical dotted inner line are visual aids only.

The probability of finding the register off and the time to be τ_l upon measurement is $P_{\text{cf}}^{(l)} = |cA_l^0|^2$, $l \in \mathbb{N}_{\geq 0}$.

The example in the main text where there are just two distinguishable times (i.e. $N_T = 1$) and requires one ancilla state corresponds to the unitary matrix $U_m = U_{\text{ex}}U'_m$, with U_{ex} determined by $U_{\text{ex}}|E\rangle = |E\rangle$, $U_{\text{ex}}|\tau_0\rangle = |A\rangle$, $U_{\text{ex}}|A\rangle = |\tau_0\rangle$, $U_{\text{ex}}|\tau_1\rangle = |\tau_1\rangle$. The unitary U'_m is given by

$$\begin{array}{c|cccc} & |E\rangle & |\tau_0\rangle & |\tau_1\rangle & |A\rangle \\ \hline |E\rangle & \sqrt{1/3} & \sqrt{1/3} & \sqrt{1/3} & 0 \\ |\tau_0\rangle & 0 & -\sqrt{1/3} & \sqrt{1/3} & -\sqrt{1/3} \\ |\tau_1\rangle & -\sqrt{1/3} & 0 & \sqrt{1/3} & \sqrt{1/3} \\ \hline |A\rangle & \sqrt{1/3} & -\sqrt{1/3} & 0 & \sqrt{1/3} \end{array} \quad (\text{A7a})$$

and coefficients c and s by

$$c = s = 1/\sqrt{2}. \quad (\text{A7b})$$

4. Backwards in time analysis in the $N_T \geq 2$ scenario

Here we extend the analysis presented in section IV A to the general case of multiple distinguishable times ($N_T \in \mathbb{N}_{>0}$). Since the results are quantitatively the same as in section IV A, we will not discuss at length the interpretation to avoid repetition. The amplitude corresponding to the k th counterfactual outcome is

$$\langle A_k | U_m U(\tau_k) U_0 | A_0 \rangle, \quad (\text{A8})$$

for $k = 0, 1, 2, \dots, N_T$ and where we are again using the convention $A_0 := E$. Using $U_m = U_{\text{ex}}U'_m$ and eq. (A6), we have

$$\begin{aligned} \langle A_k | U_m = & A_k^0 \langle A_0 | - A_k^0 r \langle \tau_0 | - A_k^0 r \langle \tau_1 | - \dots - A_k^0 r \langle \tau_{k-1} | - A_k^0 r \langle \tau_{k+1} | - A_k^0 r \langle \tau_{k+2} | - A_k^0 r \langle \tau_{k+3} | - \dots - A_k^0 r \langle \tau_{N_T} | \\ & + B_{N_T+2,k} \langle A_1 | + B_{N_T+2,k} \langle A_2 | + B_{N_T+3,k} \langle A_3 | + \dots + B_{N_T+m,k} \langle A_m | \end{aligned} \quad (\text{A9})$$

where $\langle \tau_{-1} | := 0$. Recall that $U(\tau_k)|A_0\rangle = |A_0\rangle$, $U(\tau_k)|A_1\rangle = |A_1\rangle$, $U(\tau_k)|A_2\rangle = |A_2\rangle$, $\dots$, $U(\tau_k)|A_m\rangle = |A_m\rangle$, and $U(\tau_k)|\psi\rangle = |\tau_k\rangle$. Therefore, analogously to section IV A, it follows that $\langle \tau_l | U(\tau_k) = \langle \psi_{\perp}^{(l,k)} |$ for all $k, l = 0, 1, \dots, N_T$ with $k \neq l$, where $\{\langle \psi_{\perp}^{(l,k)} | \}_{l,k}$ are orthogonal to $|A_0\rangle$ and $|\psi\rangle$. This follows from writing the ansatz $\langle \tau_l | U(\tau_k) = c_0^{(l,k)} \langle \psi_{\perp}^{(l,k)} | + c_2^{(l,k)} \langle A_0 | + c_3^{(l,k)} \langle \psi |$, followed by observing the two equalities $\delta_{k,l} = \langle \tau_l | \tau_k \rangle = \langle \tau_l | U(\tau_k) U(\tau_k)^\dagger | \tau_k \rangle = (c_0^{(l,k)} \langle \psi_{\perp}^{(l,k)} | + c_2^{(l,k)} \langle A_0 | + c_3^{(l,k)} \langle \psi |) |\psi\rangle = c_3^{(l,k)}$, where $\delta_{k,l}$ is the Kronecker delta, and $0 = \langle \tau_l | A_0 \rangle = \langle \tau_l | U(\tau_k) U(\tau_k)^\dagger | A_0 \rangle = (c_0^{(l,k)} \langle \psi_{\perp}^{(l,k)} | + c_2^{(l,k)} \langle A_0 | + c_3^{(l,k)} \langle \psi |) |A_0\rangle = c_2^{(l,k)}$. Thus from eq. (A9) we arrive at

$$\begin{aligned} \langle A_k | U_m U(\tau_k) = & A_k^0 \langle A_0 | - A_k^0 r \langle \psi_{\perp}^{(0,k)} | - A_k^0 r \langle \psi_{\perp}^{(1,k)} | - \dots - A_k^0 r \langle \psi_{\perp}^{(k-1,k)} | - A_k^0 r \langle \psi_{\perp}^{(k+1,k)} | - A_k^0 r \langle \psi_{\perp}^{(k+2,k)} | \\ & - \dots - A_k^0 r \langle \psi_{\perp}^{(N_T,k)} | + B_{N_T+2,k} \langle A_1 | + B_{N_T+2,k} \langle A_2 | + B_{N_T+3,k} \langle A_3 | + \dots + B_{N_T+m,k} \langle A_m | \\ = & A_k^0 \langle A_0 | + \langle \psi_{\perp}^{\prime\prime(k)} |, \end{aligned} \quad (\text{A10})$$

where $\langle \psi_{\perp}^{\prime\prime(k)} |$ is another (unnormalised) state orthogonal to $|A_0\rangle$ and $|\psi\rangle$. Thus recalling that U_0 only acts nontrivially on the subspace spanned by $|A_0\rangle$ and $|\psi\rangle$, we find that

$$\langle A_k | U_m U(\tau_k) = A_k^0 (c \langle A_0 | - s \langle \psi |) + \langle \psi_{\perp}^{\prime\prime(k)} |. \quad (\text{A11})$$

Thus upon pre-selecting onto the initial energy eigenstate $|A_0\rangle$, we observe that the only terms contributing to $\langle A_k | U_m U(\tau_k) | A_0 \rangle$ come from non dynamical branches of the wave function.

As with the case in which $N_T = 1$ presented in methods IV A, we can analyse the probability that a measurement of whether the clock was dynamically evolving during the dynamical stage of our pre and post selected setup. For the case of post-selecting on counterfactual outcome $|A_k\rangle$, following [28], we have the probability of the clock being on at time $t \in (0, \tau_k)$ is given by

$$\text{Prob}(\text{on}, A_k) := \frac{|\langle A_k | U_m U(\tau_k - t) P_{\text{on}} U(t) U_0 | A_0 \rangle|^2}{|\langle A_k | U_m U(\tau_k - t) P_{\text{on}} U(t) U_0 | A_0 \rangle|^2 + |\langle A_k | U_m U(\tau_k - t) P_{\text{on}} U(t) U_0 | A_0 \rangle|^2}, \quad (\text{A12})$$

where $P_{\text{on}} := |\psi\rangle\langle\psi| + |\tau_0\rangle\langle\tau_0| + |\tau_1\rangle\langle\tau_1| + \dots + |\tau_{N_T}\rangle\langle\tau_{N_T}|$ is the projection onto the on subspace and $P_{\text{on}} := |A_0\rangle\langle A_0| + |A_1\rangle\langle A_1| + \dots + |A_m\rangle\langle A_m|$ projects onto the complementary space (the subspace where no dynamics occurs). Note that since the ancillas and energy eigenstate are stationary, we have that $U(t) = U_{\perp A_0, \dots, A_m}(t) + \mathbb{1}_{A_0, \dots, A_m}$, where $\mathbb{1}_{A_0, \dots, A_m}$ is the identity operator on the subspace spanned by $|A_0\rangle, |A_1\rangle, \dots, |A_m\rangle$, and that $U_{\perp A_0, \dots, A_m}(t)$ is orthogonal to said subspace. Therefore, if we assume that the dynamics is Markovian, namely $U(t_1)U(t_2) = U(t_1 + t_2)$ for all $t_1, t_2 \in \mathbb{R}$, it follows that $U(\tau_k - t)P_{\text{on}}U(t) = U_{\perp A_0, \dots, A_m}(\tau_k)$ and hence that

$$\text{Prob}(\text{on}, A_k) = 0 \quad (\text{A13})$$

for all $t \in (0, \tau_k)$ and for all $k = 0, 1, 2, \dots, N_T$.

5. Structure and achievability proofs

In this subsection we present two propositions. The first one shows that ancillary states are necessary in order for the counterfactual clock to function, while the second shows that counterfactual clocks exist which can distinguish between an arbitrary number of times (i.e. for all $N_T \in \mathbb{N}_{>0}$).

To start with, we need to motivate why it could have been that the ancillary states were not strictly necessary. The most straightforward way to see this is by recalling eq. (A8) from the previous section, and noting that the unitary U_m has the form $U_m = U_{\text{ex}}U'_m$. It follows that

$$\langle A_k | U_m U(\tau_k) U_0 | E \rangle = \begin{cases} \langle E | U'_m U(\tau_0) U_0 | E \rangle & \text{if } k = 0 \\ \langle \tau_{k-1} | U'_m U(\tau_k) U_0 | E \rangle & \text{if } k = 1, 2, 3, \dots, N_T. \end{cases} \quad (\text{A14})$$

The physical interpretation of this is that rather than post-selecting on $|E\rangle = |A_0\rangle$ and ancillas $|A_1\rangle, \dots, |A_{N_T}\rangle$, we could have implemented the unitary U'_m , in stead of U_m , and post-selected on $|E\rangle$ and $|\tau_0\rangle, \dots, |\tau_{N_T-1}\rangle$. In this latter method, since the states $|\tau_k\rangle$ evolve in time, upon implementing the counterfactual measurement, the clock would start evolving unless it was immediately turned off (note that this does not happen when post-selecting onto the ancillas, since these are by definition stationary states). However, in the latter method, we see from the r.h.s. of eq. (A14) that one could have used unitaries in which the number of ancillas is zero, i.e. $m = 0$, since the ancillas only appear implicitly (no pre or post selection onto them is required). The following proposition proves that in the case of $N_T = 1$, the unitary U'_m needs at least one ancilla. This result implies that the ancillas are necessary.

Proposition 1 (Ancillas are necessary). Consider the counterfactual clock protocol described by the two state formalism on the r.h.s. of eq. (A14) for the case where the clock can distinguish between two times ($N_T = 1$) and the number of ancilla states is zero ($m = 0$), namely

$$\langle E | U'_m U(\tau_0) U_0 | E \rangle \quad \text{and} \quad \langle \tau_0 | U'_m U(\tau_1) U_0 | E \rangle. \quad (\text{A15})$$

There is no solution for which the probability of the elapsed time being τ_0 , namely $P_{\text{cf}}^{(0)} := |\langle E | U'_m U(\tau_0) U_0 | E \rangle|^2$, and the probability of the elapsed time being τ_1 , namely $P_{\text{cf}}^{(1)} = |\langle \tau_0 | U'_m U(\tau_1) U_0 | E \rangle|^2$, are both non-zero.

Proof. For the case $N_T = 1$, $m = 0$, the matrix representation of U'_m in eq. (A6) reduces to

$$\begin{array}{c|ccc} & |E\rangle & |\tau_0\rangle & |\tau_1\rangle \\ \hline |E\rangle & A_0^0 & A_1^0 & \gamma_0 \\ |\tau_0\rangle & 0 & -A_1^0 r & \gamma_1 \\ |\tau_1\rangle & -A_0^0 r & 0 & \gamma_2 \end{array} \quad (\text{A16})$$

Unitary matrices require their row and column vectors to be orthogonal. Applying this constraint to the 1st two columns, one finds $A_0^0 (A_1^0)^\dagger = 0$, which implies $|A_0^0| |A_1^0| = 0$ and hence $A_0^0 = 0$ and/or $A_1^0 = 0$. Therefore, since $P_{\text{cf}}^{(0)} = |cA_0^0|^2$, and $P_{\text{cf}}^{(1)} = |cA_1^0|^2$, there is no solution for which $P_{\text{cf}}^{(0)} > 0$ and $P_{\text{cf}}^{(1)} > 0$. $\square$

We now prove that for all $N_T \in \mathbb{N}^+$, there exists a finite dimensional ancilla system and unitary U_m such that there is a non zero probability of finding the clock off at all measurement times.

Proposition 2 (Counterfactual clocks with arbitrarily many distinguishable times exist). Let $N_T \in \mathbb{N}_{>0}$. Then, for any $\{\tilde{A}_0^0, \tilde{A}_1^0, \tilde{A}_2^0, \dots, \tilde{A}_{N_T}^0, \tilde{\gamma}_0, \tilde{\gamma}_1, \tilde{\gamma}_2, \dots, \tilde{\gamma}_{N_T+1}\} \in \mathbb{R}^{2N_T+3}$, and $r \neq 0$, there exists a $\gamma > 0$ for which a unitary matrix of the form eq. (A6) exists, with the amplitudes corresponding to counterfactual outcomes given by $A_0^0 = \gamma \tilde{A}_0^0$, $A_1^0 = \gamma \tilde{A}_1^0$, $A_2^0 = \gamma \tilde{A}_2^0$, $\dots$, $A_{N_T}^0 = \gamma \tilde{A}_{N_T}^0$, and gamma coefficients given by $\gamma_0 = \gamma \tilde{\gamma}_0$, $\gamma_1 = \gamma \tilde{\gamma}_1$, $\gamma_2 = \gamma \tilde{\gamma}_2$, $\dots$, $\gamma_{N_T+1} = \gamma \tilde{\gamma}_{N_T+1}$, and $m = 2(N_T + 2)$ ancillary states.

Proof. It is by construction, and hence can be used to work out particular values of γ for any instance of the problem. Denote by $\vec{F}_j = \gamma \vec{f}_j$ ($j = 1, 2, 3, \dots, m$) the j th column vector of eq. (A6). As is well known, since eq. (A6) is a square matrix, it follows that it is a unitary matrix if the vectors $\{\vec{F}_j\}_{j=1}^m$ form an orthonormal family. We therefore need to prove that the vectors $\vec{F}_j$ can be made orthonormal.

To start with, we will decompose the vector $\{\vec{f}_j\}_{j=1}^{N_T+2}$ into a direct sum of three other vectors, namely for $j = 1, 2, 3, \dots, N_T + 2$, let $\vec{f}_j = \vec{e}_j \oplus \vec{X}_j \oplus \vec{Y}_j$, where $\vec{e}_j, \vec{X}_j, \vec{Y}_j \in \mathbb{R}^{N_T+2}$. Since the matrix in eq. (A6) is a square matrix, this choice fixes the number of ancillas to be $m = 2(N_T + 2)$. Note that the vectors $\{\vec{e}_j\}_{j=1}^{N_T+2}$ are completely fixed, up to the constant γ , by the given parameters in the proposition statement. Meanwhile the vectors $\vec{X}_j, \vec{Y}_j$ are complete undetermined at this stage. We first fix the $\vec{Y}_j$ vectors: for $j, k = 1, 2, 3, \dots, N_T + 2$, let $[\vec{Y}_j]_k = \delta_{j,k} c_j$, where $\delta_{j,k}$ is the Kronecker delta, and the coefficients $\{c_j\}_{j=1}^{N_T+2} \in \mathbb{R}^{N_T+2}$ are to be determined. We will now fix the $\vec{X}_j$ vectors. We will use them for so-called *dimensional lifting* of the vectors $\vec{e}_j$ to an orthogonal set. The algorithm in [34] shows how to choose vectors $\{\vec{x}_j\}_{j=1}^{N_T+2}$ so that the vectors $\{\vec{e}_j \oplus \vec{x}_j\}_{j=1}^{N_T+2}$ form an orthogonal family for any given set $\{\vec{e}_j\}_{j=1}^{N_T+2}$. It thus follows due to the form of the vectors $\vec{Y}_j$, that the vectors $\{\vec{F}_j = \gamma \vec{e}_j \oplus \vec{X}_j \oplus \vec{Y}_j\}_{j=1}^{N_T+2}$ form an orthogonal family for all $\{c_j\}_{j=1}^{N_T+2} \in \mathbb{R}^{N_T+2}$ and for all $\gamma > 0$. Imposing normalisation on the vectors $\{\vec{F}_j\}_{j=1}^{N_T+2}$, we find for $j = 1, 2, 3, \dots, N_T + 2$:

$$\frac{1}{\gamma^2} - |c_j|^2 = (\vec{X}_j)^\dagger \vec{X}_j + (\vec{Y}_j)^\dagger \vec{Y}_j. \quad (\text{A17})$$

Now denote by $D := \max_{j \in \{1, 2, \dots, N_T+2\}} \{(\vec{X}_j)^\dagger \vec{X}_j + (\vec{Y}_j)^\dagger \vec{Y}_j\}$, and the value of j which solves the maximization by j^* . Furthermore, set $c_{j^*} = 0$ so that it follows from eq. (A17), that $\gamma = 1/\sqrt{D}$. We can solve eq. (A17) for all $j \neq j^*$, with a solution for the coefficients c_j satisfying $0 \leq |c_j|^2 \leq 1/\gamma^2 = D$.

We now have an orthonormal set of vectors $\{\vec{F}_j\}_{j=1}^{N_T+2}$, and all that remains is to find the remaining vectors $\{\vec{F}_j\}_{j=N_T+3}^{N_T+2+m}$, with $m = 2(N_T + 2)$. Since this latter set of vectors contains elements without any constraint on them, other than those which allow eq. (A17) to be a unitary matrix, we can simply apply the Gram-Schmidt orthonormalization procedure on the input sequence $(\vec{F}_j)_{j=1}^{N_T+2} \frown (\vec{z}_j)_{j=N_T+3}^{N_T+2+m}$, where $\{\vec{z}_j\}_{j=N_T+3}^{N_T+2+m}$ is an arbitrary set of vectors linearly independent of $\{\vec{F}_j\}_{j=1}^{N_T+2}$, and $\frown$ denotes sequence concatenation. The output of the Gram-Schmidt orthogonalisation procedure is then the orthonormal family $(\vec{F}_j)_{j=1}^{3(N_T+2)}$ where the 1st $N_T + 2$ elements are identical to the first $N_T + 2$ orthonormal vectors of the input sequence to the Gram-Schmidt orthogonalisation procedure. This completes the proof. $\square$

Supplementary material B: Implementation of the proof of no classical model

Here we describe the numerical implementation of the proof that there does not exist a non contextual ontic model for the counterfactual clock described in methods IV B. The entire problem is fully determined by the set $\mathbf{s}_{\text{cf}}$ of states (eq. (11)) and the set $\mathbf{e}_{\text{cf}}$ of effects (eq. (12)), with U_m and c, s given by eqs. (A7a) and (A7b).

We will follow the outline of the algorithm [32] and perform the following steps. The inner product is taken to be the Hilbert-Schmidt inner product:

1. Project the set of extreme points of the set $\mathbf{s}_{\text{cf}}$ onto $\text{span}(\mathbf{e}_{\text{cf}})$, where span denotes the linear span. Let $P_{\mathbf{e}}(\mathbf{s})$ denote the resulting set vectors. We call $\text{span}(P_{\mathbf{e}}(\mathbf{s}))$ the *Reduced space* and denote it by $\mathcal{R}$. It is the effective vector space of our clock.
2. Using the Gram-Schmidt orthogonalization procedure, construct an orthonormal basis for $\mathcal{R}$.
3. Project the extreme points of the sets $\mathbf{s}_{\text{cf}}$ and $\mathbf{e}_{\text{cf}}$ onto $\mathcal{R}$. Denote the corresponding new sets as $P_{\mathcal{R}}(\mathbf{e})$ and $P_{\mathcal{R}}(\mathbf{s})$. The elements of the latter sets represent rays (also known as half-lines) emanating from the origin (which is the point $\mathbf{0} := (0, \dots, 0)$ with dimension $\dim(\mathcal{R})$ entries in the orthonormal basis for $\mathcal{R}$).
4. Run Vertex Enumeration algorithm [35] twice. Once for rays $P_{\mathcal{R}}(\mathbf{e})$ and again for the rays $P_{\mathcal{R}}(\mathbf{s})$. The input to the algorithm is given using the V-representation which has the format (list of vertices, list of rays) and in our case it simplifies to $(\mathbf{0}, \text{list of rays})$. Note that it differs from the H-representation which is set of linear inequalities corresponding to the intersection of halfspaces. We used Matlab wrapper GeoCalcLib (<http://worc4021.github.io/>) and performed all the computations using Matlab R2020a. The two runs of the algorithm on inputs $(\mathbf{0}, P_{\mathcal{R}}(\mathbf{s}))$ and $(\mathbf{0}, P_{\mathcal{R}}(\mathbf{e}))$ produce the following sets of extreme rays: $\text{Ray}_S, \text{Ray}_E$ respectively.

5. Form a new set $\text{Ray}_{final} = \{a \otimes b \mid \forall a \in \text{Ray}_S \text{ and } \forall b \in \text{Ray}_E\}$.
6. Run Vertex Enumeration algorithm on $(\mathbf{0} \otimes \mathbf{0}, \text{Ray}_{final})$ using the V-representation to obtain a set of extreme rays W . The set W contains the list of potential non-classicality witnesses.
7. To verify that the pair of sets $\mathbf{s}_{cf}, \mathbf{e}_{cf}$ does not admit a non contextual ontic model, it suffices to check whether there exists $w \in W$ such that

$$\left\langle \sum_{j=1}^{\dim(\mathcal{R})} \mathcal{R}_j \otimes \mathcal{R}_j, w \right\rangle < 0, \quad (\text{B1})$$

where $(\mathcal{R}_j)_{j=1}^{\dim(\mathcal{R})}$ is the sequence of orthonormal basis elements for $\mathcal{R}$ calculated in step 2, and $\langle \cdot, \cdot \rangle$ represent the Hilbert-Schmidt inner product.

Note that computation times for the last instance of vertex enumeration (with input $(\mathbf{0} \otimes \mathbf{0}, \text{Ray}_{final})$) runs for an extremely long time, so we employed a slight optimization by preprocessing the sets $\text{Ray}_S, \text{Ray}_E$ by applying the GeoCalcLib vertex Reduction routine which removed non-extremal (redundant) rays of the polyhedrons in V-representation. This allowed us to run the Vertex Enumeration algorithm with input $(\mathbf{0}, T)$, where $T \subset \text{Ray}_{final}$.

We were able to successfully identify an element of $w \in W$ which provides a violation which is two orders of magnitude larger than any error due to rounding and approximations that stem from using floating-point arithmetic.