Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Forecasts that change the future: a closed timelike curve model on quantum hardware

**Peter Maurice Catt**

peter.catt@virtualblue.co.nz

Virtual Blue Limited

**Research Article**

**Keywords:**

**Posted Date:** November 20th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11790699/v1

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

**Additional Declarations:** No competing interests reported.

# Forecasts that change the future: a closed timelike curve model on quantum hardware

Peter M. Catt[1,*]

[1]Virtual Blue Limited, Auckland, New Zealand

[*]Corresponding author: peter.catt@virtualblue.co.nz

ORCID: 0009-0000-1324-0861

## Abstract

Published forecasts can bring about the outcomes they announce or provoke responses that avert them. We ask which forecasts remain consistent once people act on them, by conditioning forecasts on agreement with the outcomes they induce. Under a uniform forecast prior, each retained forecast is weighted by its probability of coming true. This rule can be implemented by quantum postselection in a closed timelike curve model built from quantum teleportation. Deutsch's stationary rule gives different weights. In a stylised recession model with a strongly self-defeating response, the recession forecast receives weight 0.0001 under agreement conditioning and 0.30 under the stationary rule. Under a uniform prior the two rules coincide when the outcome ignores the forecast. The rule is classically computable, so the hardware experiment tests how faithfully a processor implements it and gives no computational advantage. Its plan was published before two confirmatory sessions, which ran on IBM Kingston on consecutive days. After readout mitigation the plan's tests of leakage, acceptance and equivalence to a classical filter were met, and a separate witness certified entanglement in the primary circuits. Where the response always averts a forecast recession, the forecast that should be excluded retained weight 0.011 with one two-qubit gate, 0.027 (95% interval 0.015 to 0.039) with two to three and 0.062 with five to seven. Without readout mitigation the second figure is 0.070, above the limit of 0.05. A classically correlated pair in place of the entangled pair gave 0.021.

**Keywords:** Reflexive forecasting; Self-fulfilling prophecy; Forecast evaluation; Closed timelike curves; Postselection; Quantum computing

# 1 Introduction

Forecasts can change the outcomes they predict. An outbreak warning prompts the precautions that avert the outbreak. A recession forecast can lead firms to postpone investment and hiring, as they do when uncertainty rises [1], and a shock to one sector can spread to others through input–output links [2]. Merton [3] called a prediction that causes itself to come true a self-fulfilling prophecy. In economics, Lucas [4] argued that behaviour estimated under one policy changes when the policy changes.

Earlier work asks whether a single forecast can survive its own publication. Simon [5] showed that an election forecast can still be correct when publishing it sways voters. Grunberg and Modigliani [6] showed that a correct public prediction exists when the public's reaction is continuous, as a fixed point of that reaction. Performative prediction [7, 8] studies the same feedback for predictive models that are retrained on the data they influence. We take a set of published forecasts and a model of the response, and ask which forecasts remain when only those that came true are kept, and with what weights.

Quantum feedback has been proposed as a way to drive systems towards presumed states or trajectories [9]. We instead condition published forecasts on agreement with the outcomes they induce, which a postselected closed timelike curve model implements through quantum teleportation [10–12]. Huang et al. [13] simulated such curves on IBM and Quantinuum processors to decode scrambled quantum information, Morapakula and Ikeda (preprint, arXiv:2607.27473) ran related postselected decoders on IBM processors, and Ji et al. [14] derived how much information a noisy postselected curve can carry. The conditioning rule is Bayes' rule for the agreement event, and its postselected form is the classical case of known results [11, 12], including those in recent preprints by Carney (arXiv:2609.08716) and by Ferradini, Gitton and Vilasini (arXiv:2502.04171).

We read the rule as a consistency condition for forecasts that change their outcomes, and show that it and Deutsch's stationary rule [15, 16] weight such forecasts differently. In a stylised recession model, a forecast that is almost always averted receives almost no weight under agreement conditioning and 0.30 under the stationary rule. We then test how faithfully a quantum processor implements the model, in an experiment whose plan and analysis code were lodged publicly before the confirmatory sessions [17], following the practice of fixing an analysis plan before outcomes are seen [18]. One arm of that experiment replaces the entangled pair with a classically correlated pair.

## 2 Two consistency rules

### 2.1 Agreement conditioning

A forecast $f$, one of $d$ classes, is published with prior probability $\pi_f$, people respond, and an outcome $o$ follows. The response is summarised by a matrix $T$, whose element $T_{fo}$ is the probability that publishing $f$ leads to outcome $o$. Conditioning published forecasts on agreement between forecast and outcome, $O = F$, gives

$$P(f \mid O = F) = \frac{\pi_f T_{ff}}{\sum_h \pi_h T_{hh}}, \tag{1}$$

defined only when agreement has positive probability, $\sum_h \pi_h T_{hh} > 0$. With a uniform prior, $\pi_f = 1/d$, agreement occurs with probability $\mathrm{Tr}\, T/d$ and each forecast is retained in proportion to its chance of coming true:

$$w_f = \frac{T_{ff}}{\mathrm{Tr}\, T}. \tag{2}$$

The prior and the agreement condition fix this rule. Consistency alone does not determine a unique distribution of forecasts.

### 2.2 A worked example

Take two classes, no recession ($f = 0$) and recession ($f = 1$). Suppose a recession follows a forecast of no recession with probability 0.2. Suppose also that a recession forecast leads firms to take precautions, so that a recession follows it with probability 0.1. Then $T_{00} = 0.8$ and $T_{11} = 0.1$. If the two forecasts are published equally often, forecast and outcome agree in $(0.8 + 0.1)/2 = 0.45$ of cases. Of the forecasts that came true, the share forecasting no recession is $0.8/0.9 = 0.89$ and the share forecasting recession is $0.1/0.9 = 0.11$. Recessions follow 15% of all published forecasts, yet only 11% of the forecasts that came true were recession forecasts, because publishing a recession forecast makes a recession less likely.

### 2.3 The stationary rule

Deutsch's consistency condition [15] reduces, for the dephased classical response channel, to $\sigma = \sigma T$, a stationary distribution of the response [16]. We calculated this stationary distribution classically, applying Deutsch's maximum-entropy prescription [15] when it was non-unique. Deutsch's condition and the postselected model are known to be inequivalent in general [11, 12]. The stationary distribution is a mixture that reproduces itself through the response, and it does not require individual forecasts to come true. For $T_{11} = \eta$ and $T_{00} = 1 - q$,

$$\sigma_1 = \frac{q}{q + 1 - \eta}, \qquad w_1 = \frac{\eta}{1 - q + \eta}. \tag{3}$$

In the worked example $q = 0.2$ and $\eta = 0.1$, so $\sigma_1 = 0.18$ against $w_1 = 0.11$. As $\eta \to 0$ with $q < 1$, $w_1$ vanishes and $\sigma_1$ tends to $q/(1+q)$. At $q = 1$ and $\eta = 0$ no history is accepted and $w$ is undefined. For $q = 0$ and $\eta = 1$, maximum entropy selects $\sigma = (1/2, 1/2)$. Where both are defined, the difference between the two weights is

$$w_1 - \sigma_1 = \frac{(\eta - q)(1 - \eta - q)}{(1 - q + \eta)(1 + q - \eta)}. \tag{4}$$

The two rules therefore give the same weights if and only if $q = \eta$, when the outcome ignores the forecast, or $q = 1 - \eta$, when both forecasts are equally likely to come true. Agreement conditioning gives the recession forecast the larger weight when a recession is more likely after it is forecast ($\eta > q$) and the forecast of no recession is the more reliable of the two ($1 - q > \eta$), or when both inequalities are reversed.

### 2.4 Any number of classes

The condition for the two rules to coincide extends to $d$ classes. Let $t$ be the vector of diagonal elements of $T$, $t_f = T_{ff}$, so that $w = t/\operatorname{Tr} T$. Suppose $\operatorname{Tr} T > 0$ and $T$ has a unique stationary distribution, as it does when the Markov chain on the forecast classes is irreducible [19]. Then $w = \sigma$ if and only if

$$tT = t. \tag{5}$$

This holds because $w$ is a probability distribution and so equals $\sigma$ exactly when it is stationary, $wT = w$. Multiplying both sides by $\operatorname{Tr} T$ gives equation (5). For two classes the second component of equation (5) reads $(1 - q)q + \eta^2 = \eta$, which is $q(1 - q) = \eta(1 - \eta)$, the condition found above.

Two families of responses satisfy equation (5) for any $d$. If every row of $T$ is the same distribution $\rho$, the outcome ignores the forecast and both rules return $\rho$. If $T$ is doubly stochastic with a constant positive diagonal, both rules return the uniform distribution.

Bishop, Costa and Ralph [20] assign a state to a qubit on a postselected curve, the first iterate of Deutsch's map from the maximally mixed state, and give a condition for it to be a fixed point of that map under a unitary interaction. Equation (5) concerns a different object, the distribution of forecasts conditional on agreement, for a classical response with any number of classes.

### 2.5 What each rule describes

Agreement conditioning describes a record of forecasts that came true. Forecasts are published with frequencies $\pi$, the outcomes arrive, and those that came true are kept. Equation (1) gives the composition of what is kept. Every retained forecast is correct, and the fraction retained is the agreement probability $\sum_h \pi_h T_{hh}$, which is the proportion of the published fore-

casts that are correct, a measure of the correspondence between forecasts and outcomes [21]. The rule depends on how often each forecast was issued.

The stationary rule describes a process that repeats. Suppose the same response applies in every period and each period's forecast is the outcome of the period before, the naive forecast [22]. The forecasts then form a Markov chain with transition matrix $T$, and $\sigma$ is its long-run distribution, which involves no prior when it is unique. A forecast drawn from $\sigma$ comes true with probability $\sum_f \sigma_f T_{ff}$, which is less than one unless every forecast in the mixture is certain to come true.

The first rule gives the forecasts in a record of successes, and the second gives the forecasts issued by a process that repeats itself. Neither rule is correct in general, and which applies depends on how the forecasts are generated and selected.

### 2.6 Relation to fixed-point forecasts and the proportion correct

In this discrete setting, a forecast that is certain to survive publication is a class with $T_{ff} = 1$. The existence result of Grunberg and Modigliani [6] relies on a continuous variable, and with discrete classes no such forecast need exist. Equation (2) applies whether or not one exists and ranks every class by how often it survives.

The proportion correct scores the forecasts and the response together. Equation (1) divides it by forecast class, along the diagonal of the joint distribution of forecasts and outcomes [23]. Two sets of forecasts with the same proportion correct can retain different forecasts, and a class with a low $T_{ff}$ contributes little to the retained record unless it is issued much more often than the others.

Murphy [21] uses consistency for the correspondence between a forecast and the forecaster's own judgement. Here it means agreement between a published forecast and the outcome that follows it. Only categorical forecasts are treated.

### 2.7 Postselected closed timelike curve model

The agreement condition is imposed by teleporting the outcome register $O$ back into the forecast register $F$ (Fig. 1a). The register $F$ and a reference register $R$ are prepared in $|\Phi^+\rangle = d^{-1/2} \sum_f |f\rangle_F |f\rangle_R$, which supplies the uniform prior, with $d = 2^k$ for $k$ qubits. The unitary $U$ reads $F$, acts on a system register $S$ and writes the outcome to $O$. Projecting $O$ and $R$ onto $|\Phi^+\rangle$ leaves, for a fixed set of random draws $s$, the unnormalised state $d^{-1} \sum_{f:\, o(f,s)=f} |f\rangle_F |s'\rangle_S$, with $s'$ the final state of $S$. Because $F$ retains the published forecast, the terms are orthogonal and cannot interfere, so averaging over draws gives the weights of equation (2) and an acceptance probability $p = \mathrm{Tr}\, T/d^2$, the classical agreement probability multiplied by a factor $1/d$ from the Bell projection. Both are special cases of Carney's results for Bell postselection

of general classical loops, which give the success probability and the conditional map (arXiv:2609.08716). It also follows that a classically correlated pair gives identical statistics for these measurements, so they do not certify entanglement. The Bell measurement is a controlled-NOT gate from *O* to *R* and a rotation of *O*, and runs in which both qubits read 0 are kept (Fig. 1b). This postselected circuit is called the loop below.

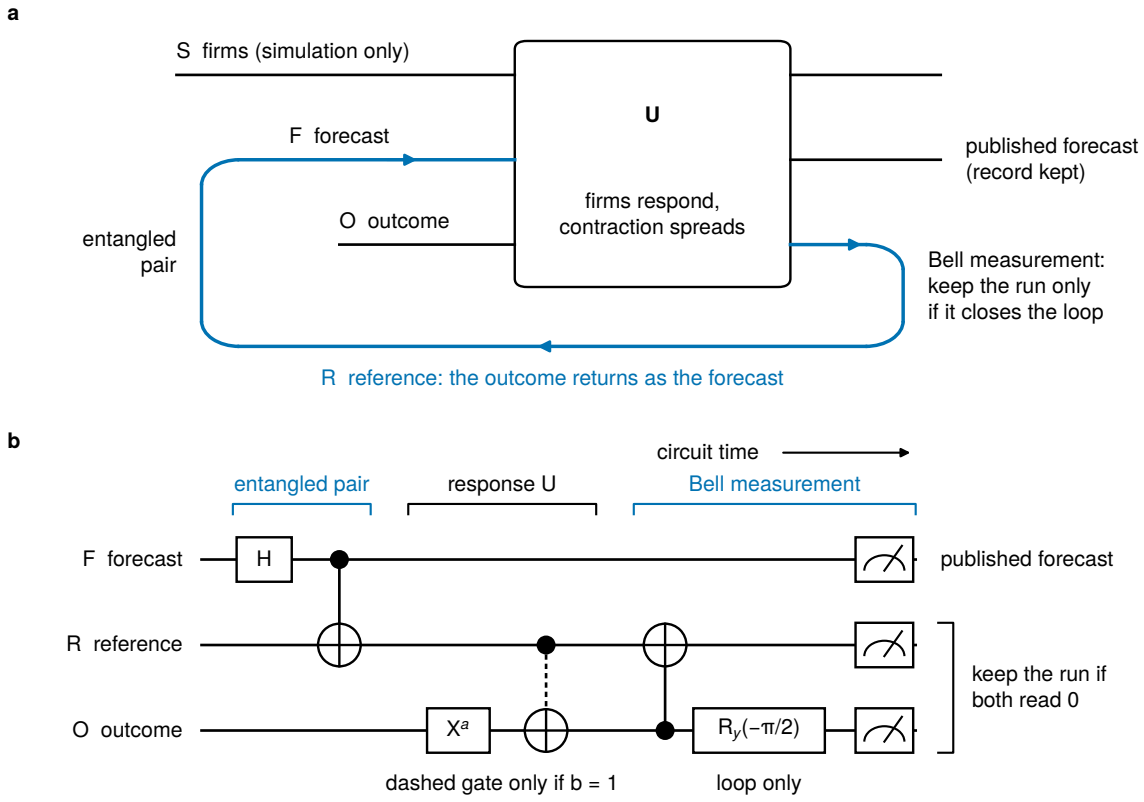



**Fig. 1** Postselected closed timelike curve model. **a** The entangled registers *F* and *R* and the outcome register *O* implement agreement conditioning through Bell postselection. On the processor the response *U* is the *X* gate and the controlled-NOT gate of **b**. **b** Loop circuit of the primary arm. *R* sits between *F* and *O*, so every two-qubit gate joins neighbouring qubits. The dashed gate applies when $b = 1$, and the final rotation is omitted in the control

# 3 Methods

## 3.1 Illustrative application: a recession forecast and business responses

The illustration is a generic model of contagion on a network, labelled as a recession. Ten firms on a Watts–Strogatz trading network [24] receive a forecast of recession or no recession. Contraction spreads with baseline probability $\beta$ per period. Responding firms use $\beta(1-\alpha c)$ after a recession forecast and $\beta(1+\max(\alpha, 0)\, r)$ otherwise, where *c* is the strength of precautions and *r* that of leaner running when no recession is forecast. Positive $\alpha$ represents precautions (cash reserves, credit lines) that slow transmission, and negative $\alpha$ represents

spending cuts that accelerate it. Transmission between two firms uses the geometric mean of their probabilities. Each firm ignores the forecast with probability $\varepsilon$. Recession, $o = 1$, means at least $m$ contracting firms after $H$ periods.

The network is generated once and used for every setting. A firm that starts contracting stays contracting. In each draw one firm, chosen uniformly at random, is contracting at the start, and each firm decides once, for all $H$ periods, whether to ignore the forecast. Updates are synchronous. In each period every contracting firm passes contraction to each neighbour that is not yet contracting, with the probability above evaluated on the state at the start of the period. A draw consists of the initially contracting firm, the set of firms that ignore the forecast and one uniform random number per directed link per period. The same draw is used for every forecast value, so for a fixed draw the outcome is a deterministic function of the forecast.

For $k = 3$, forecasts and outcomes are eight classes of the number of contracting firms, in steps of two, and the response scales with the forecast class $f$. A responding firm's transmission probability is multiplied by $(1 - \alpha c\, g)(1 + \max(\alpha, 0)\, r\, (1 - g))$ with $g = f/(d - 1)$, which reduces to the two-class rule at $d = 2$. With ten firms the count runs from 0 to 10, so only the first six classes can occur as outcomes. The two highest can be forecast but never come true, and both rules give them zero weight.

For two classes $T$ was computed exactly by enumerating the model, and the results below use the exact values. For eight classes, and for the instances run as circuits, $T$ was estimated by Monte Carlo with the parameters of Table 1. At the five reaction settings the Monte Carlo estimates of the recession-forecast weight (0.518, 0.443, 0.306, 0.114 and 0) differ from the exact values (0.514, 0.438, 0.299, 0.114 and 0.0001) by at most twice the Monte Carlo standard error.

Six settings are used on the processor: five reaction strengths $\alpha$ and a floor setting ($\alpha = 1$, $\varepsilon = 0$), in which every firm responds and a forecast recession is always averted, so the recession forecast should receive no weight. The comparison of the two rules uses 17 values of $\alpha$ from $-1$ to 1. Its sensitivity was examined by changing one parameter at a time: the lean-running strength, the probability of ignoring the forecast, the transmission probability, the recession threshold and the network seed.

### 3.2 Simulation

Each instance is one draw of the model's randomness, sampled classically and shared across forecast values. For $k = 1$, its outcome is $o(f) = a \oplus bf$, with $a, b \in \{0, 1\}$. Pooling the instances with equal shots gives the same measured statistics as a single circuit that holds the draws coherently in ancilla qubits and never measures them, because the draws label orthogonal states and cannot interfere. Circuits were built in Qiskit (Javadi-Abhari et

**Table 1** Model and simulation parameters

| Parameter | Symbol | Value |
|---|---|---|
| Firms | $N$ | 10 |
| Forecast register (processor; simulation) | $k$ | 1; 1 and 3 |
| Trading network mean degree; rewiring probability; seed | | 4; 0.1; 11 |
| Baseline transmission probability per period | $\beta$ | 0.20 |
| Precaution and lean-running strengths | $c$, $r$ | 1; 0.25 |
| Probability that a firm ignores the forecast | $\varepsilon$ | 0.10 |
| Periods; initially contracting firms | $H$ | 2; 1 |
| Recession threshold (firms contracting) | $m$ | 4 |
| Reaction settings on the processor | $\alpha$ | $-1$, $-0.5$, 0, 0.5, 1 |
| Floor setting | | $\alpha = 1$, $c = 1$, $\varepsilon = 0$ |
| Circuit instances per setting on the processor | | 48 |
| Simulation: circuit instances per setting ($k = 1$; $k = 3$) | | 480; 12 |
| Monte Carlo draws per setting for $T$ | | 20,000 |

al., arXiv:2405.08810). The simulated circuit writes each firm's final state onto ten qubits and counts the contracting firms with a Fourier-basis adder (Draper, arXiv:quant-ph/0008033). Exact outcome probabilities were compared with equation (2) for each instance and, pooled over instances, with the Monte Carlo estimate of $T$. Equation (5) was checked numerically on 2,000 random response matrices with two to six classes.

For $k = 1$ each instance is one of four one-bit functions of the forecast: constant 0, constant 1, identity ($o = f$) and NOT ($o = 1 - f$). Their error-free acceptance probabilities are 1/4, 1/4, 1/2 and 0. Each setting is a classical mixture of the four, and its error-free acceptance probability and weights follow from the mix (Table 2). The 48 instances are a finite draw, so these values differ from those of the model. At $\alpha = 0$ the mix gives a recession-forecast weight of 0.2292, against 0.299 in the model. The hardware tests compare each estimate with the value implied by the mix. At the floor setting only constant-0 and NOT instances occur. Leakage there comes from errors in constant-0 instances and from accepted runs of NOT instances, none of which should be accepted.

### 3.3 Arms of the hardware experiment

The hardware experiment had $k = 1$ and ran on three qubits of the IBM processor ibm_kingston, physical qubits 58, 51 and 52. The loop circuit encodes the classically calculated outcome function directly (Fig. 1b). The registers $F$ and $R$ are prepared in $|\Phi^+\rangle$, an $X$ gate acts on $O$ when $a = 1$, a controlled-NOT gate acts on $O$ when $b = 1$, and the Bell measurement of $O$ and $R$ follows. Each circuit ran with a matched control, the same circuit without the final rotation of $O$, so that loop and control share every two-qubit gate.

**Table 2** The 48 circuit instances of each setting by outcome function, with the error-free acceptance probability and recession-forecast weight they imply

| Setting | Constant 0 | Constant 1 | Identity | NOT | Acceptance | Weight |
|---|---|---|---|---|---|---|
| $\alpha = -1$ | 9 | 15 | 24 | 0 | 0.3750 | 0.5417 |
| $\alpha = -0.5$ | 22 | 18 | 8 | 0 | 0.2917 | 0.4643 |
| $\alpha = 0$ | 37 | 11 | 0 | 0 | 0.2500 | 0.2292 |
| $\alpha = 0.5$ | 32 | 4 | 0 | 12 | 0.1875 | 0.1111 |
| $\alpha = 1$ | 33 | 0 | 0 | 15 | 0.1719 | 0.0000 |
| Floor | 29 | 0 | 0 | 19 | 0.1510 | 0.0000 |

Four arms ran the same 48 instances per setting (Table 3). In the primary arm, *R* sits on the qubit between those of *F* and *O*, and for $b = 1$ the controlled gate takes its control from *R*, which holds the same value as *F* (Fig. 1b). The unlinked arm runs the compiled circuits of the primary arm with the rotation that prepares *F* set to give 0 for half the shots and 1 for the other half, in place of an equal superposition. This gives a classically correlated pair with the same gates and the same ideal statistics. The one-pair arm uses a single entangled pair and single-qubit gates. Because the response is classical, this arm has the same outcome distribution as the loop circuit. For $b = 0$ the pair is *F* and *R*, with *R* flipped when $a = 1$, and *O* holds *a*. For $b = 1$ the pair is *F* and *O*, with *O* flipped when $a = 1$, and *R* holds *a*. The original arm uses the layout of the pilot jobs described below, recompiled, in which *F* and *R* are not adjacent and the controlled gate takes its control from *F*.

**Table 3** Arms of the hardware experiment. Gate counts are two-qubit gates in the compiled circuits for $b = 0$ and $b = 1$

| Arm | Question | Gates |
|---|---|---|
| Primary | Does the loop of Fig. 1b, compiled with the pair on adjacent qubits, stay within the limits? | 2, 3 |
| Unlinked | Does any measured quantity depend on the pair being entangled? | 2, 3 |
| One pair | What do the loop's additional gates cost? | 1, 1 |
| Original | How much of the pilot result was the compilation? | 5, 7 |

Separate witness circuits checked for entanglement, which the benchmark statistics cannot certify. They are the primary circuits with $a = 0$ and no final rotation, which should leave $|\Phi^+\rangle$ on *F* and *R* ($b = 0$) or on *F* and *O* ($b = 1$). The fidelity is $(1 + \langle XX\rangle - \langle YY\rangle + \langle ZZ\rangle)/4$, and a value above $1/2$ certifies entanglement [25]. The same circuits were run unlinked, where the expected fidelity is $1/2$.

The 18 compiled circuits were fixed in a file lodged with the plan, and the jobs sent that file without recompiling. They are six benchmark circuits, one for each value of *b* in the primary,

one-pair and original compilations, six witness circuits, one for each value of $b$ and each of three measurement bases, and six readout-calibration circuits, one for each benchmark circuit. Parameters set $a$ and the choice of loop or control in each benchmark circuit, and the preparation of $F$ that distinguishes the unlinked arm from the primary arm. Two sessions ran, the first on 17 November 2026 (Coordinated Universal Time, UTC) and the second on 18 November 2026. The plan set the second at least 24 hours after the first. By the clock of the submitting script the second job was sent 24 h 0 min 27 s after the first. IBM's records put the creation of the two jobs 23 h 59 min 7 s apart, 53 s under 24 hours, and the start of their execution 24 h 0 min 7 s apart. Each session had 128 shots per instance, arm and circuit type (loop or control), and twice that at the floor setting of the primary, unlinked and one-pair arms, giving 405,504 shots per session with calibration and witness circuits. Each execution had 64 shots, as two Pauli-twirling randomisations of 32. Half of each readout calibration ran before the benchmark circuits and half after. Two-qubit gates were Pauli twirled [26] and dynamical decoupling [27] with the XpXm sequence was applied.

### 3.4 Classical filter and benchmark quantities

The control estimates the conditional response matrix $\hat{T}$, each row normalised by the number of runs with that forecast. With $\hat{w}^{\mathrm{L}}$ the weights measured through the loop,

$$\hat{w}_f^{\mathrm{F}} = \frac{\hat{T}_{ff}}{\mathrm{Tr}\,\hat{T}}, \qquad \Delta = \hat{w}_1^{\mathrm{L}} - \hat{w}_1^{\mathrm{F}}, \tag{6}$$

where $\hat{w}^{\mathrm{F}}$ is the classical filter under a uniform prior, estimated on the same hardware from the control, and $\Delta$ is the difference between loop and filter for the recession forecast. At the floor setting the error-free rule excludes the recession forecast, and the weight that the loop gives it is reported as leakage.

### 3.5 Tests, predictions and analysis fixed before the confirmatory sessions

The plan, the analysis script, the specification and the compiled circuits were published on Zenodo [17]. The repository's metadata gives the creation of the record as 03:19:43 UTC on 17 November 2026, and IBM's record gives the creation of the first job as 03:22:37 UTC, 174 seconds later. No circuit of the confirmatory experiment was run on the processor before the record was created. The four files of the record have not been changed since, and their checksums match the files used. Four tests apply to the primary arm, on the estimate pooled over both sessions, and each is decided on an interval limit. The first is that the upper limit of the 95% interval for the floor leakage is at most 0.05. The second is that, at every setting, the 90% interval for the acceptance probability minus its error-free value lies within $\pm 0.03$. The third is that, at every setting, the 90% interval for $\Delta$ lies within $\pm 0.05$. The second and third are the interval-inclusion form of two one-sided tests at the 5% level [28], hence the 90%

intervals. The fourth is that the lower limit of the 95% interval for the witness fidelity exceeds $1/2$ in both primary witness circuits, and that the upper limit is at most 0.55 in both unlinked ones. The other arms are reported with the same quantities and are not tested. The plan states that the author expected the primary arm to meet all four tests, that tests 1 to 3 do not depend on entanglement, and that test 4 is a manipulation check on the pair, entangled in the primary arm and not in the unlinked arm. A classically correlated pair gives a fidelity of $1/2$, and a value at or below 0.55 in the unlinked circuits is consistent with that but does not show that the state is separable.

Four predictions were also lodged, each to be reported as held or not held. Primary floor leakage lies between 0.010 and 0.045. Floor leakage is lower in the one-pair arm than in the primary arm, and lower in the primary arm than in the original arm. The 95% interval for the difference in floor leakage between the primary and unlinked arms includes zero. The ratio of original to primary floor leakage lies between 1.7 and 4.5. The predictions are judged on the pooled estimates.

Readout mitigation [29] applies the pseudoinverse of each compiled circuit's assignment matrix to pooled frequencies. Negative values are not set to zero, so estimates near zero can be negative. Each session uses its own calibration. The pooled estimate is formed from the equal-weight average of the two sessions' mitigated distributions, so it differs slightly from the mean of the two session estimates. Intervals are percentile intervals [30] from 2,000 bootstrap replicates that resample shots within each instance, arm, circuit type and session, and within each row of each calibration. The 48 instances are treated as fixed. The tests were applied to the two sessions pooled. The intervals reflect sampling of shots and calibrations within the two sessions and do not include variation between days. The plan also flags a tested quantity when the two sessions differ by more than a fixed multiple of the bootstrap standard error of their difference, 2 for the floor leakage and 2.64 for each setting of tests 2 and 3. A flag does not change the outcome of a test.

### 3.6 Noise model lodged with the plan

The plan document includes predictions from a noise model with a depolarising error [31] on each two-qubit gate, one tenth of that error on each single-qubit gate, and the readout errors measured in the pilot jobs. The two-qubit error was set to 1.0% and to 1.6%, chosen to match the floor leakage of the pilot circuits in the two pilot jobs that ran them, so the original arm is not an independent check of the model. These are effective errors per gate that absorb idling, crosstalk and state preparation. They are several times the two-qubit errors of 0.19% and 0.35% that IBM reported for the two couplers when the jobs were submitted. The model was not validated beyond the two pilot values, and the plan notes that it does not reproduce the control-arm values of the pilots closely. The model treats the primary and unlinked arms identically, and the values printed for them differ only by simulation noise.

### 3.7 Pilot jobs

Four jobs ran on the same processor on 12 November 2026, before the plan was lodged. They made up a first experiment, which did not meet its leakage criterion. Their criteria were fixed in stages during that day and recorded in a specification that was hashed privately, with no independent timestamp. Their loop circuits, here called the pilot circuits, had 5 and 7 two-qubit gates when compiled, although each has the same outcome distribution as one entangled pair with single-qubit gates. Their readout mitigation set negative values to zero, and the specification of the last job does not mention readout mitigation, which its analysis applied. They are reported as pilots, with their data [32], and carry no weight in the tests above.

The confirmatory experiment kept the response model, the 48 instances per setting, the three qubits, readout mitigation by inversion of the assignment matrix and the limits of the first three tests. It changed the compilation of the loop, added the unlinked and one-pair arms, ran two sessions on consecutive days, split each calibration around the benchmark circuits, stopped setting negative values to zero and fixed the circuits, tests and analysis in a public record beforehand. The original arm repeats the layout of the pilot circuits. Its circuits were recompiled for the lodged file and are not identical to those of the pilot jobs.

### 3.8 Use of generative artificial intelligence

The author used an artificial intelligence (AI) assistant (Claude, Anthropic) in drafting the text and in writing the code for the simulations, analysis and figures. The author reviewed all AI-assisted text and code and takes full responsibility for the content of this article.

## 4 Results

### 4.1 The circuit reproduces agreement conditioning

For every instance simulated with positive acceptance probability, at $k = 1$ and $k = 3$, the postselected circuit reproduced classical agreement conditioning to within $10^{-15}$. Instances in which no forecast agrees with its outcome, such as $o(f) = 1 - f$, have zero acceptance probability, and their conditional weights are undefined. For $k = 1$, pooled over 480 instances per setting, the weights lay within a total variation distance of 0.014 of equation (2) (Fig. 2a). The difference is sampling variation in the draws.

### 4.2 The two rules in the recession model

When the forecast has no effect ($\alpha = 0$) both rules give the recession forecast the base rate of recessions, 0.299 (Fig. 2a). As the response becomes self-defeating the rules separate.

The weight under agreement conditioning falls to 0.114 at $\alpha$ = 0.5 and to 0.0001 at $\alpha$ = 1. The stationary rule gives between 0.283 and 0.299 at every $\alpha \geq 0$. As $\eta \to 0$ it tends to $q/(1+q)$ and not to zero. In this model leaner running after a forecast of no recession also raises $q$ as precautions lower $\eta$, which holds the stationary weight near the base rate at the baseline parameters.

When the response is self-fulfilling both weights rise, to 0.514 under agreement conditioning and 0.536 under the stationary rule at $\alpha = -1$. The two curves cross between $\alpha = -1$ and $\alpha = -0.875$, where $q + \eta$ passes through one. Equation (4) gives the crossings for any two-class response. With $q$ held at its baseline value of 0.299 and $\eta$ varied from 0 to 1, the two weights cross at $\eta = q$ and at $\eta = 1 - q$ (Fig. 2b). Between the crossings agreement conditioning gives the recession forecast the larger weight, and outside them the stationary rule does. At $\eta$ = 0 the stationary rule still gives 0.23. At $\eta$ = 1, where a forecast recession always occurs, the stationary rule gives the recession forecast all the weight and agreement conditioning gives it 0.59, because forecasts of no recession still come true 70% of the time.

The rules also differ in how often their forecasts come true. Every forecast retained by agreement conditioning is correct, and the share of published forecasts retained falls from 0.72 at $\alpha = -1$ to 0.29 at $\alpha$ = 1. A forecast drawn from the stationary mixture comes true with probability 0.72 at $\alpha = -1$, 0.58 at $\alpha$ = 0 and 0.40 at $\alpha$ = 1.

### 4.3 Eight forecast classes and non-uniform priors

With eight classes, where $T$ is a Monte Carlo estimate, the rules differ in the same direction. At $\alpha$ = 1 agreement conditioning puts 0.61 of the weight on the class of two or three contracting firms, against 0.51 under the stationary rule, and 0.01 on classes of six or more, against 0.05 (Fig. 2c). The total variation distance between the two sets of weights is 0.10 at $\alpha$ = 1, 0.07 at $\alpha = -1$ and zero at $\alpha$ = 0. Forecasts of severe contraction are the ones most strongly averted, and agreement conditioning removes them. Equation (5) holds at $\alpha$ = 0, where the rows of $T$ are identical, and its two sides differ by 0.08 in their largest element at $\alpha$ = 1.

Agreement conditioning depends on the prior, and the stationary rule has none. At $\alpha = -0.5$ recession forecasts make up 0.44, 0.25 and 0.08 of the retained forecasts when they make up 0.5, 0.3 and 0.1 of those published, against 0.40 under the stationary rule (Fig. 2d). At $\alpha$ = 0, where the forecast has no effect, the stationary rule returns the base rate of recessions, 0.299, and agreement conditioning returns it only under the uniform prior.

### 4.4 Sensitivity to the model parameters

The comparison was repeated with one parameter changed at a time (Table 4). In all eleven cases, the baseline and ten changes, the recession forecast has weight at most 0.007 under agreement conditioning at $\alpha$ = 1, and between 0.178 and 0.402 under the stationary rule. The

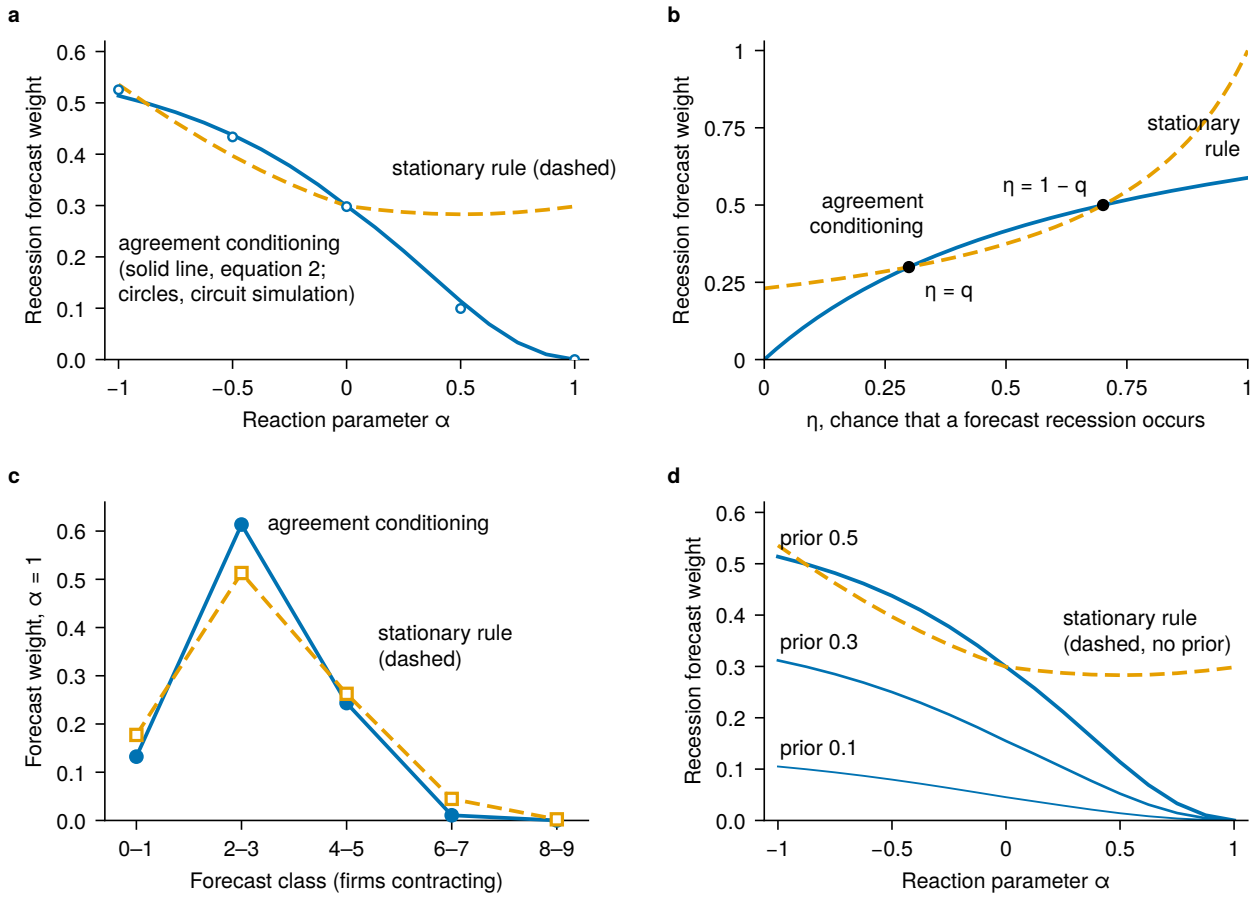



**Fig. 2** Agreement conditioning (solid blue lines) and the stationary rule (dashed orange lines) in the recession model. **a** Weight of the recession forecast from self-fulfilling ($\alpha = -1$) to self-defeating ($\alpha = 1$) responses. The line is equation (2) and the circles are the circuit simulation. **b** Both weights against the chance $\eta$ that a forecast recession occurs, with $q = 0.299$, from equation (3). The rules coincide at the two points marked by filled circles. **c** Weights of the first five of eight forecast classes at $\alpha = 1$, with filled circles for agreement conditioning and open squares for the stationary rule. **d** Weight of the recession forecast under agreement conditioning for three priors on the recession forecast

stationary weight is not tied to the base rate. In the baseline case it stays within 0.02 of the base rate at every $\alpha \geq 0$, and it moves away when the lean-running strength, the transmission probability or the recession threshold is changed. With no lean running ($r = 0$), $q$ does not change with $\alpha$ and the stationary weight falls from the base rate, 0.299, to $q/(1 + q) = 0.230$, and with $r = 0.5$ it rises to 0.353. The size of the difference between the rules changes with every parameter, and its direction at $\alpha = 1$ is the same in every case.

**Table 4** Weight of the recession forecast under agreement conditioning ($w_1$) and the stationary rule ($\sigma_1$) when one model parameter is changed, computed exactly. The base rate is the weight under both rules at $\alpha = 0$

| Case | Base rate | $\alpha = 0.5$ | | $\alpha = 1$ | |
|---|---|---|---|---|---|
| | | $w_1$ | $\sigma_1$ | $w_1$ | $\sigma_1$ |
| Baseline | 0.299 | 0.114 | 0.283 | 0.000 | 0.298 |
| Lean running $r = 0$ | 0.299 | 0.105 | 0.246 | 0.000 | 0.230 |
| Lean running $r = 0.5$ | 0.299 | 0.125 | 0.316 | 0.000 | 0.353 |
| Ignoring the forecast $\varepsilon = 0$ | 0.299 | 0.095 | 0.284 | 0.000 | 0.305 |
| Ignoring the forecast $\varepsilon = 0.3$ | 0.299 | 0.154 | 0.283 | 0.007 | 0.285 |
| Transmission $\beta = 0.15$ | 0.168 | 0.049 | 0.180 | 0.000 | 0.203 |
| Transmission $\beta = 0.25$ | 0.440 | 0.220 | 0.374 | 0.000 | 0.370 |
| Network seed 12 | 0.294 | 0.110 | 0.280 | 0.000 | 0.296 |
| Network seed 13 | 0.305 | 0.119 | 0.286 | 0.000 | 0.300 |
| Threshold $m = 3$ | 0.554 | 0.399 | 0.452 | 0.003 | 0.402 |
| Threshold $m = 5$ | 0.127 | 0.023 | 0.147 | 0.000 | 0.178 |

### 4.5 Implementation on the processor

The tests concern the primary arm, pooled over the two sessions (Table 5). All were met, with readout mitigation as the plan specifies. Floor leakage was 0.0270 (95% interval 0.0150 to 0.0390; Fig. 3a). The acceptance probability lay within 0.0042 of its error-free value at every setting, and the widest 90% interval reached 0.0114 in magnitude (Fig. 3b). The difference $\Delta$ between loop and filter was at most 0.0149 in magnitude, and the widest 90% interval reached 0.0287 (Fig. 3c). The witness fidelity was 0.9792 (0.9695 to 0.9882) for $b = 0$ and 0.9630 (0.9528 to 0.9734) for $b = 1$ in the primary circuits, and 0.4990 (0.4822 to 0.5153) and 0.4842 (0.4684 to 0.5009) in the unlinked ones. Both primary values exceeded $1/2$ in each session and without readout mitigation (ESM_1, Table S11). The $b = 1$ value was 0.9779 in the first session and 0.9480 in the second, a fall of 2.8 times the bootstrap standard error of the difference. The plan's flag does not cover the witness.

The 90% interval for $\Delta$ excluded zero at three of the six settings, $\alpha = -0.5$, $\alpha = 0$ and the floor, at the first two by 0.0002 and 0.0005. The difference between loop and filter had the

same sign in all four arms at five of the six settings. It was negative at $\alpha = -1$ and $-0.5$ and positive at $\alpha = 0.5$, $\alpha = 1$ and the floor (ESM_1, Tables S4 to S7). At the floor it was 0.0127 in the primary arm, 0.0088 in the unlinked arm, 0.0130 in the one-pair arm and 0.0314 in the original arm. It was present with one two-qubit gate and without entanglement, so it requires neither the longer circuits nor the entangled pair. Loop and control differ only in the final rotation of *O*, so the offset arises in that rotation and the measurement after it, or in a readout error that depends on the state and that the calibration does not capture.

The loop weight itself lay above the error-free value of the instance mix at four settings: 0.2459 against 0.2292 at $\alpha = 0$, 0.1338 against 0.1111 at $\alpha = 0.5$, 0.0155 against 0 at $\alpha = 1$ and 0.0270 against 0 at the floor. Its 95% interval excluded the error-free value at each of the four (Table 2; ESM_1, Table S4). The plan set no test on this comparison.

Floor leakage in the primary arm was 0.0193 (0.0034 to 0.0349) in the first session and 0.0352 (0.0177 to 0.0519) in the second, so the second session alone would not have met test 1. Both the constant-0 and the NOT instances contributed to the rise (ESM_1, section S5). The test is defined on the pooled estimate, and the pooled interval does not include variation between days. The lodged flag for disagreement between sessions covers tests 1 to 3 and was not raised.

Without readout mitigation, a calculation that was not part of the plan, the floor leakage of the primary arm is 0.0699 (0.0622 to 0.0783), above the limit. The correction comes mainly from two readout errors, a 0 read as 1 on the qubit carrying *F* and a 1 read as 0 on the qubit carrying *R* (Supplementary Information, ESM_1, section S5 and Table S10). Differences between arms are little affected. The original arm exceeds the primary arm by 0.0376 without mitigation and by 0.0346 with it.

The arms share the response model and the three qubits. The primary and unlinked arms share compiled circuits, and the others differ in gate sequence and qubit roles as well as gate count. Floor leakage was 0.0114 (0.0003 to 0.0225) in the one-pair arm, 0.0270 in the primary arm, 0.0206 (0.0092 to 0.0316) in the unlinked arm and 0.0616 (0.0444 to 0.0786) in the original arm (Fig. 3a; ESM_1, Table S9). All four predictions held. The plan gives no derivation for their ranges, and they are not severe tests. The original arm leaked 2.3 times as much as the primary arm (95% interval 1.4 to 4.3), a difference of 0.0346 (0.0138 to 0.0552). No difference between the primary and unlinked arms was detected. At the floor setting their leakage differed by 0.0064 ($-0.0056$ to 0.0191) and their acceptance by $-0.0008$ ($-0.0070$ to 0.0056), and the 95% interval for the difference in loop weight included zero at all six settings (ESM_1, Table S13). The primary arm exceeded the one-pair arm by 0.0156 ($-0.0005$ to 0.0316) at the floor setting, an interval that includes zero. At $\alpha = 1$, where the error-free weight is also zero, the two arms gave 0.0155 and 0.0153.

The floor leakage of the one-pair, primary and original arms lay within the range of the noise

**Table 5** Tests and predictions lodged before the confirmatory sessions, and their outcomes. Test 4 is the plan's manipulation check. Outcomes are after readout mitigation, as the plan specifies. Without it test 1 is not met

| Test or prediction | Result | Outcome |
|---|---|---|
| Test 1. Floor leakage, upper 95% limit at most 0.05 | 0.0270 (0.0150 to 0.0390) | Met |
| Test 2. Acceptance within $\pm 0.03$ at every setting (90%) | widest limit 0.0114 | Met |
| Test 3. $\Delta$ within $\pm 0.05$ at every setting (90%) | widest limit 0.0287 | Met |
| Test 4. Witness fidelity: lower 95% limit above $1/2$ primary; upper limit at most 0.55 unlinked | 0.9695, 0.9528; 0.5153, 0.5009 | Met |
| Prediction 1. Primary floor leakage between 0.010 and 0.045 | 0.0270 | Held |
| Prediction 2. One pair below primary below original (point estimates) | 0.0114, 0.0270, 0.0616 | Held |
| Prediction 3. Primary minus unlinked: 95% interval includes zero | 0.0064 ($-0.0056$ to 0.0191) | Held |
| Prediction 4. Original to primary ratio between 1.7 and 4.5 | 2.3 | Held |

model, and that of the unlinked arm lay just below it, at 0.0206 against a lower bound printed as 0.021 (Table 6). The model was fitted to the pilot circuits, and every measured interval is wider than the model's range, so the agreement is a weak check.

**Table 6** Floor leakage by arm: predicted in the lodged plan by the noise model, at two-qubit gate errors of 1.0% and 1.6%, and observed. The predicted values for the primary and unlinked arms differ only by simulation noise

| Arm | Gates | Predicted, 1.0% | Predicted, 1.6% | Observed |
|---|---|---|---|---|
| One pair | 1, 1 | 0.006 | 0.012 | 0.0114 |
| Primary | 2, 3 | 0.020 | 0.032 | 0.0270 |
| Unlinked | 2, 3 | 0.021 | 0.031 | 0.0206 |
| Original | 5, 7 | 0.055 | 0.084 | 0.0616 |

### 4.6 Pilot jobs

In the pilot job that held all its criteria in one specification (job 4), the pilot circuits, with 5 and 7 two-qubit gates, met the acceptance and equivalence criteria at all six settings. Floor leakage was 0.0535 (95% interval 0.0451 to 0.0729), and the upper limit exceeded 0.05. Without setting negative values to zero it was 0.0500 (0.0324 to 0.0663), and without readout

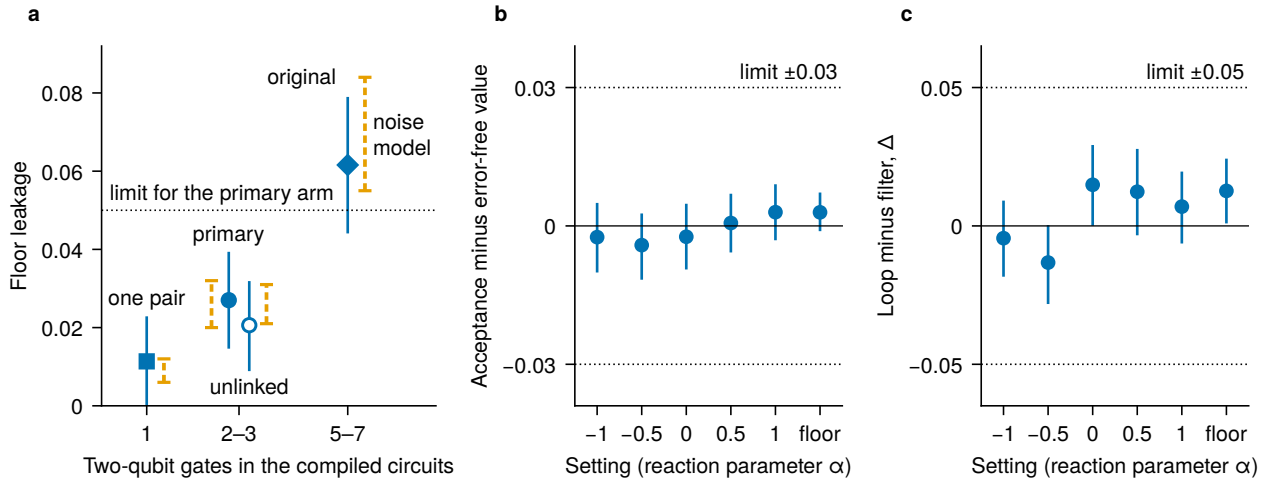



**Fig. 3** Hardware experiment, two sessions pooled. **a** Floor leakage by arm against the number of two-qubit gates, with 95% intervals. Filled symbols are the arms with an entangled pair and the open circle is the unlinked arm. Horizontal positions are nominal, with the primary and unlinked arms side by side. Each dashed orange bar gives the range that the noise model lodged with the plan predicted for the arm beside it, and the dotted line is the limit of test 1. **b** Acceptance probability of the primary arm minus its error-free value at each setting, with 90% intervals and the limits of test 2. **c** Difference Δ between loop and filter in the primary arm, with 90% intervals and the limits of test 3

mitigation 0.1022. An earlier job with the same compiled circuits gave 0.0812 (0.0681 to 0.0997) and missed equivalence at the floor setting. The two leakage estimates differ by about 2.6 times the standard error implied by their intervals, so intervals from one job understate variation between jobs. A circuit with 171 two-qubit gates, which computes the outcome on ten qubits, had floor leakage 0.323 (0.303 to 0.346) in job 4. ESM_1, section S9, gives the pilot results in full.

## 5 Discussion

Conditioning on agreement and requiring a stationary distribution are both consistency rules for a forecast that changes its outcome. Under a uniform prior they agree when the outcome ignores the forecast, and in this model they differ most when the response is self-defeating. In the recession model a forecast that is almost always averted receives a weight of 0.0001 under agreement conditioning and about 0.30 under the stationary rule, or between 0.178 and 0.402 across the parameter changes tested. Which rule applies depends on how forecasts are generated and selected, and a statement about which forecasts are consistent needs to say which rule it uses.

When a published forecast moves its outcome, the proportion correct mixes the forecaster’s

reading of events with the effect of publication. A warning that averts its event is scored as a false alarm, and a forecast that comes true because people acted on it is scored as a hit. Agreement conditioning does not separate the two. $T_{ff}$ is the fraction of published forecasts of class $f$ that come true, and equation (1) gives the share of the correct forecasts that each class contributes. The rule describes the composition of a record of correct forecasts. It is not a scoring rule or a method of forecasting.

These results depend on the specified forecast prior and response model. The recession example is illustrative and uncalibrated, and the response matrix is assumed known. The comparison of the two rules is exact for two classes and numerical for eight.

Postselection on the Bell outcome keeps a fraction $1/d$ of the runs in which forecast and outcome agree, the same for every forecast class, which gives the acceptance probability $\mathrm{Tr}\, T/d^2$ and leaves the weights of equation (2) unchanged. At the floor setting a recession forecast never comes true, so every accepted shot that carries one results from an error in preparation, gates, idling or readout, and the leakage is the share of accepted shots of that kind. Leakage was 0.011 with one two-qubit gate, 0.027 with two to three and 0.062 with five to seven, each consistent with the range of a noise model fitted to the pilot circuits. The original arm, which repeats the layout of the pilot circuits, exceeded the primary arm by 0.035 (0.014 to 0.055), so the larger leakage of the pilot circuits is consistent with their larger gate count. No difference between the primary and unlinked arms was detected at any setting. The witness certified entanglement in the primary circuits and certified none in the unlinked ones.

The filter is classically computable and the benchmark statistics need no entanglement, so the experiment tests the implementation of the circuit and shows no quantum advantage. The stationary rule was calculated classically and was not run on the processor. The hardware evidence comes from one processor and one set of three qubits on two consecutive days, with $d = 2$. The leakage test is met with the readout mitigation fixed in the plan and would not be met without it. The result also depends on readout that stays stable within a session. In the plan's noise model at the higher gate error, a readout error on one qubit that was 0.010 higher during the benchmark circuits than at calibration cut the share of synthetic experiments meeting the test from 80% to 8%.

Larger forecast registers need more shots and more gates. The acceptance probability is $\mathrm{Tr}\, T/d^2$, so with eight classes the recession model accepts between 1.3% and 1.8% of runs, against 14% to 36% with two, and the shots needed for a given precision on the retained weights rise in inverse proportion. The entangled register needs one two-qubit gate for each forecast qubit, and the outcome function needs more as classes are added. The original arm, with five to seven two-qubit gates, did not meet the limit used here even with readout mitigation (0.062, 95% interval 0.044 to 0.079), so larger registers on this processor would need shallower encodings of the outcome function or further error mitigation.

## Supplementary Information

ESM_1 (PDF) gives the records of the experiment, the compiled circuits, the results by arm, setting and session, the estimates without readout mitigation, the readout calibration, the witness circuits, the comparisons between arms, the noise model lodged with the plan and the pilot jobs.

## Declarations


### Acknowledgements

The author acknowledges the use of IBM Quantum services and thanks the IBM Quantum team for access to ibm_kingston and for Qiskit. The views expressed are those of the author and do not reflect the official policy or position of IBM or the IBM Quantum team.

### Funding

The author received no external funding. Processor time was paid for personally by the author.


### Competing interests

The author is a director and shareholder of Virtual Blue Limited, a consultancy that provides artificial intelligence and machine-learning services. The company did not fund this work and has no product or service based on it. The author declares no other competing interests.

### Ethics approval and consent to participate

Not applicable.

### Data availability

The plan of the hardware experiment, its specification and its compiled circuits were lodged before the first session in the Zenodo repository [17], https://doi.org/10.5281/zenodo.24297205 (version 1.0.0). The measurement counts, IBM's job records, the analysis outputs and the simulation outputs are in version 2.0.0 of that record [33], https://doi.org/10.5281/zenodo.24352180, which carries the four lodged files unchanged. The records of the four pilot jobs are in a separate Zenodo record [32], https://doi.org/10.5281/zenodo.24267517.

### Code availability

The script that compiled, submitted and analysed the hardware experiment was lodged with the plan [17]. The scripts for the analyses made after the run, the simulations and the figures are in version 2.0.0 of the same record [33].

### Author contributions

P.M.C. conceived and designed the study, ran the simulations and hardware experiments, analysed the results and prepared the manuscript.

## References


[1] Bloom N. The impact of uncertainty shocks. Econometrica. 2009;77(3):623–685. https://doi.org/10.3982/ECTA6248

[2] Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. The network origins of aggregate fluctuations. Econometrica. 2012;80(5):1977–2016. https://doi.org/10.3982/ECTA9623

[3] Merton RK. The self-fulfilling prophecy. Antioch Rev. 1948;8(2):193–210. https://doi.org/10.2307/4609267

[4] Lucas RE Jr. Econometric policy evaluation: a critique. Carnegie-Rochester Conf Ser Public Policy. 1976;1:19–46. https://doi.org/10.1016/S0167-2231(76)80003-6

[5] Simon HA. Bandwagon and underdog effects and the possibility of election predictions. Public Opin Q. 1954;18(3):245–253. https://doi.org/10.1086/266513

[6] Grunberg E, Modigliani F. The predictability of social events. J Polit Econ. 1954;62(6):465–478. https://doi.org/10.1086/257604

[7] Perdomo J, Zrnic T, Mendler-Dünner C, Hardt M. Performative prediction. In: Daumé H III, Singh A, editors. Proceedings of the 37th International Conference on Machine Learning. Proc Mach Learn Res. 2020;119:7599–7609. https://proceedings.mlr.press/v119/perdomo20a.html

[8] Hardt M, Mendler-Dünner C. Performative prediction: past and future. Stat Sci. 2025;40(3):417–436. https://doi.org/10.1214/25-STS986

[9] Uys H, Bassa H, du Toit P, Ghosh S, Konrad T. Quantum control through measurement feedback. Phys Rev A. 2018;97(6):060102. https://doi.org/10.1103/PhysRevA.97.060102

[10] Bennett CH, Brassard G, Crépeau C, Jozsa R, Peres A, Wootters WK. Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. Phys Rev Lett. 1993;70(13):1895–1899. https://doi.org/10.1103/PhysRevLett.70.1895

[11] Lloyd S, Maccone L, Garcia-Patron R, Giovannetti V, Shikano Y, Pirandola S, Rozema LA, Darabi A, Soudagar Y, Shalm LK, Steinberg AM. Closed timelike curves via postselection: theory and experimental test of consistency. Phys Rev Lett. 2011;106(4):040403. https://doi.org/10.1103/PhysRevLett.106.040403

[12] Lloyd S, Maccone L, Garcia-Patron R, Giovannetti V, Shikano Y. Quantum mechanics of time travel through post-selected teleportation. Phys Rev D. 2011;84(2):025007. https://doi.org/10.1103/PhysRevD.84.025007

[13] Huang YT, Huang HW, Lin JD, Miranowicz A, Lambert N, Chen GY, Nori F, Chen YN. Experimental simulation of postselected closed timelike curves for decoding scrambled quantum information. Phys Rev Res. 2026;8(2):023084. https://doi.org/10.1103/tm83-sxpm

[14] Ji K, Lloyd S, Wilde MM. Retrocausal capacity of a quantum channel: communicating through noisy closed timelike curves. Phys Rev Lett. 2026;136(23):230801. https://doi.org/10.1103/znyd-npk5

[15] Deutsch D. Quantum mechanics near closed timelike lines. Phys Rev D. 1991;44(10):3197–3217. https://doi.org/10.1103/PhysRevD.44.3197

[16] Aaronson S, Watrous J. Closed timelike curves make quantum and classical computing equivalent. Proc R Soc A. 2009;465(2102):631–647. https://doi.org/10.1098/rspa.2008.0350

[17] Catt PM. Forecasts that change the future: plan for the confirmatory experiment. Version 1.0.0. Zenodo. 2026. https://doi.org/10.5281/zenodo.24297205

[18] Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. Proc Natl Acad Sci USA. 2018;115(11):2600–2606. https://doi.org/10.1073/pnas.1708274114

[19] Norris JR. Markov chains. Cambridge: Cambridge University Press; 1997. https://doi.org/10.1017/CBO9780511810633

[20] Bishop LG, Costa F, Ralph TC. Quantum state tomography on closed timelike curves using weak measurements. Class Quantum Gravity. 2025;42(4):045018. https://doi.org/10.1088/1361-6382/ada90b

[21] Murphy AH. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather Forecast. 1993;8(2):281–293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2

[22] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. 3rd ed. Melbourne: OTexts; 2021. https://otexts.com/fpp3/

[23] Murphy AH, Winkler RL. A general framework for forecast verification. Mon Weather Rev. 1987;115(7):1330–1338. https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2

[24] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998;393(6684):440–442. https://doi.org/10.1038/30918

[25] Gühne O, Tóth G. Entanglement detection. Phys Rep. 2009;474(1–6):1–75. https://doi.org/10.1016/j.physrep.2009.02.004

[26] Wallman JJ, Emerson J. Noise tailoring for scalable quantum computation via randomized compiling. Phys Rev A. 2016;94(5):052325. https://doi.org/10.1103/PhysRevA.94.052325

[27] Viola L, Knill E, Lloyd S. Dynamical decoupling of open quantum systems. Phys Rev Lett. 1999;82(12):2417–2421. https://doi.org/10.1103/PhysRevLett.82.2417

[28] Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987;15(6):657–680. https://doi.org/10.1007/BF01068419

[29] Bravyi S, Sheldon S, Kandala A, Mckay DC, Gambetta JM. Mitigating measurement errors in multiqubit experiments. Phys Rev A. 2021;103(4):042605. https://doi.org/10.1103/PhysRevA.103.042605

[30] Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman and Hall/CRC; 1994. https://doi.org/10.1201/9780429246593

[31] Nielsen MA, Chuang IL. Quantum computation and quantum information. 10th anniversary ed. Cambridge: Cambridge University Press; 2010. https://doi.org/10.1017/CBO9780511976667

[32] Catt PM. Forecasts that change the future: a closed timelike curve model on quantum hardware (data and code). Version 1.0.0. Zenodo. 2026. https://doi.org/10.5281/zenodo.24267517

[33] Catt PM. Forecasts that change the future: plan and results of the confirmatory experiment. Version 2.0.0. Zenodo. 2026. https://doi.org/10.5281/zenodo.24352180

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf