physical sciences forum

MDPI

Proceeding Paper

Recoverable Self-Coding: A Single-Server Queueing Model of Commitment Recoverability †

Pieter van Rooyen

Department of Electrical and Electronic Engineering, Stellenbosch University, Bosman St., Stellenbosch Central, Stellenbosch 7600, South Africa; pgwvanrooyen@sun.ac.za

† Presented at Entropy 2026: Exploring Complexity and Information in Science, Barcelona, Spain, 1–3 July 2026.

Abstract

Shannon's channel-coding theorem permits arbitrarily reliable communication at rates below the channel capacity and rules it out above it. This paper asks the corresponding question of a system that acts irreversibly on what it infers: can its commitments still be undone? This matters because a decision that is not reviewed in time becomes permanent, right or wrong. The framework is called Recoverable Self-Coding (RSC). Candidate commitments arrive at rate R_{self}, and one certifier checks them at rate C_{self}; both rates can be estimated from event counts. The framework proposes that irreversible action is admissible only while the margin $\mathcal{M} = C_{\text{self}} - R_{\text{self}}$ is positive. The certifier is modelled as a single-server queue of capacity ratio $\text{CR} = R_{\text{self}}/C_{\text{self}}$, whose uncertified backlog $\mathcal{N}$ is the mean number of commitments awaiting certification. An exact identity for $\mathcal{N}$ holds at every CR below one; for fixed shapes with positive variability, $c_a^2 + c_s^2 > 0$, it grows as $(1 - \text{CR})^{-1}$ near $\text{CR} = 1$, and three exact solutions agree with the simulation. Under the outcome model, the recoverable fraction of commitments falls from 0.75 to 0.04 as CR rises from 0.5 to 0.94 while accuracy stays at or above its floor p_{lo}, a separation that is derived rather than measured. Two limits follow: counting uncertainty sets the observation time needed to estimate the margin—centuries for the review board of the opening example—and the collapse belongs to pipelines that keep expired work. One application is offered as an untested hypothesis: artificial intelligence (AI) may raise the rate of proposed decisions faster than review capacity.

Keywords: recoverable self-coding; recoverability; rate–capacity constraint; irreversibility; effective bandwidth; human–AI systems

1. Introduction

Artificial intelligence (AI) is changing how consequential decisions are made. Candidate decisions have become cheap to produce, and fast and easy to scale, so institutions route more and more judgments through their most efficient decision step. That shift of tasks toward the cheaper route is documented for automation and labour more generally [1]. If those decisions arrive faster than the people and procedures that review them can check them, review falls behind. Approvals are still issued, but reviewers can lose the ability to trace a decision to its outcome and to reverse it before harm is done, and oversight becomes a formality. Section 8 states this sequence as a hypothesis, not as a measured fact. Safety science has described this failure for decades: systems drift toward the boundary of safe operation under pressure for efficiency [2], tightly coupled systems produce "normal accidents" [3], and layers of defence erode [4]. What the AI setting still lacks is a measurable quantity that indicates how close a system is to the point where its commitments can no longer be undone.

Check for updates

Academic Editor: Miguel Rubi

Published: 29 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

A small example shows the quantity proposed here. Suppose a review board can certify ten decisions a week while nine and a half arrive, so that the margin between the two rates is $\mathcal{M} = 0.5$ per week. Let the board behave as a single server with Poisson arrivals and exponential service times (the M/M/1 queue of Section 6), and let each decision need certification within a horizon of $\Delta t = 2$ weeks. A decision is then certified in time with probability $P_c = 1 - e^{-\mathcal{M}\Delta t} \approx 0.63$. Now consider a second board, identical except that arrivals are five percent higher. Its margin is 0.025 per week and its P_c about 0.05. The two numbers describe two stationary regimes, and the backlog of a board whose load rises takes time to settle into the second (Section 6.2). The quality of individual judgments is the same in both, yet the second board certifies few of its commitments before they become permanent. Distinguishing the two regimes by measurement is itself demanding, and Section 6.2 gives the observation span it takes. P_c is only the probability of a first certification in time. Recoverability, defined in Section 5, also requires a corrective pass and is smaller still.

The quantity rests on an analogy with Shannon's channel-coding theorem. A memoryless channel of capacity C permits arbitrarily reliable recovery of a message when the information rate R is below C, and rules it out when R is above C [5]. **The claim of this paper is that a similar limit applies to systems that act irreversibly on what they infer. The claim is developed here for a system whose commitments pass through a single certifier.** What fails near the limit in such a system is *recoverability*: whether a commitment can still be corrected or cancelled before the options it removes are gone (Section 5 gives the definition). The failure is measurable. As the system is driven toward its capacity, the accuracy of its decisions falls no lower than a floor, while the fraction of its commitments that remain recoverable falls toward zero. The floor is the accuracy of the decisions that were not certified in time (Section 7). In an institution, this separation would mean that decisions remain mostly correct while reviewers lose the ability to correct them in time.

The framework is called Recoverable Self-Coding (RSC), and the system it describes is a *self-decoder*. Candidate commitments arrive at rate R_{self}, which the framework calls the induced flux, and the certifier can check them at rate C_{self}, the integrative capacity. Their ratio $\text{CR} = R_{\text{self}}/C_{\text{self}}$ is the capacity ratio, and the boundary $\text{CR} = 1$ plays the role that $R = C$ plays in Shannon's theorem. The application hypothesis of Section 8 is that AI can increase the arrival rate of proposed decisions faster than the capacity to review them, bringing many workflows closer to $\text{CR} = 1$. That change is not measured in this paper. The idea was introduced in the foundational preprint of this research programme [6], and an independent treatment has since reached the same thesis [7]. That treatment has no order parameter, that is, no measurable quantity that tracks the loss of recoverability, and no early-warning quantity. The present paper supplies the first and a candidate for the second.

The contributions are listed below so that what is new is distinguished from what is inherited.

1. *Inherited.* The mathematics is classical single-server queueing theory: the stability condition, the $\text{CR}/(1-\text{CR})$ backlog law, the heavy-traffic asymptotics and the effective-bandwidth sojourn exponent are all standard results [8–10]. Proposition 1 collects the exact identity for the backlog, the Pollaczek–Khinchine form, the GI/M/1 sojourn law and the heavy-traffic limit, and proves them for the model in this paper.
2. *New.* The backlog of uncertified commitments is identified as the order parameter of recoverability, and operational definitions make R_{self}, C_{self} and the margin $\mathcal{M}$ estimable from event counts alone (Section 3). Two limits on that proposal are then established. The counting precision of the margin degrades as the margin shrinks, so the observation span needed to resolve it grows as $\mathcal{M}^{-2}$ and, for the M/M/1 queue, is a fixed multiple

of the relaxation time (Section 6.2). And the collapse of recoverability at CR = 1 is a property of a pipeline that keeps expired work: where expired work leaves the queue, the backlog stays bounded at every load (Section 6.2).

3. *Distinguishing prediction.* An ordinary congestion model bounds delay. This model gives separate measures to the opportunity for correction and to the quality of the decision: the fraction of commitments that remain correctable, and the accuracy of the decisions themselves. Within the model of Section 7 the first falls toward zero under load while the second keeps its floor, and that separation is derived rather than measured. What a logged workflow can test is whether the model describes it, and the tests are stated at the end of Section 7.

The claim is narrower than a theorem. Treating one certification as one channel use is a modelling step, not a derived result (Section 2). Reading the two rates as information rates is an analogy, and the results are those of a single-server model within the scope stated at each formula. No law is asserted for arbitrary systems.

The paper is organized as follows. Section 2 states the constraint and shows the model in a diagram, and Section 3 shows that its two parameters can be estimated from event data. Section 4 places the work in the literature. Section 5 defines recoverability and states the full constraint, Section 6 derives the law of the backlog, and Section 7 compares simulated backlogs with the exact results under a controlled statistical protocol and illustrates the analytical outcome model. Section 8 states one application, the loss of human agency under AI acceleration, as a hypothesis. Signatures in operational field data are left to future work [6].

2. Decoders and Their Constraints

From channel reliability to action recoverability. A self-decoder is any physical or informational system that turns uncertain observations into updates of its internal state and then into commitments, some of which are irreversible. Nothing in the definition depends on the substrate, so biological cognition and physiology, artificial computation, institutions and combined human–AI systems all qualify. Figure 1 shows the model: candidate commitments arrive one by one, and each must be *certified*, that is, checked against evidence, before it becomes irreversible. Those that have arrived and are not yet certified are counted by $N_{\mathrm{uncert}}(t)$, whose mean is the backlog $\mathcal{N}$ of Section 6. A commitment certified within the horizon Δt is certified in time; one that misses Δt is committed uncertified and is still served, because nothing leaves the queue early (Section 6.2 treats the alternative). It is recoverable if it was certified in time and a corrective pass, drawn independently from the law of the sojourn time and adding no load, would also finish within an option-loss window of length H. The queue is open: arrivals are exogenous and no output returns as load, so the closed loop of the companion paper [11] lies outside the figure, as does information retention (Section 5).

Over a finite decision horizon Δt, the system is described by two rates. The induced informational flux $R_{\mathrm{self}}(t)$ is the arrival rate of candidate commitments that need certification, expressed in bits through the uncertainty quantum introduced below. It is measured through proxies such as queued decisions and uncertified candidate commitments, never through the raw data rate or a count of tokens per second. The effective integrative capacity $C_{\mathrm{self}}(t)$ is the rate at which candidate commitments can be certified. It is limited by throughput, buffering, the work of keeping the internal state consistent, and the latency of validation. The entity that certifies is the *certifier*; it is the single server of the queueing model of Section 6, and the review board above is one instance of it. A commitment awaiting certification is a *candidate commitment*, and one that has been certified or has missed its

https://doi.org/10.3390/psf2026016002

horizon is a *commitment*. Where every candidate commitment stays in the queue until it is certified, the framework proposes that recoverable operation requires

$$R_{\text{self}}(t) < C_{\text{self}}(t) \quad \Longleftrightarrow \quad \text{CR}(t) \equiv \frac{R_{\text{self}}(t)}{C_{\text{self}}(t)} < 1, \qquad \mathcal{M}(t) \equiv C_{\text{self}}(t) - R_{\text{self}}(t) > 0. \tag{1}$$

Equation (1) has the form of Shannon's condition $R < C$ with a different decoded quantity. Shannon's condition bounds how reliably a message is recovered, and Equation (1) is proposed as the condition under which an irreversible action can still be recovered in such a pipeline. In this sense, RSC states a rate–capacity condition for commitment. Both conditions concern feasibility, and neither limits speed. An encoder may emit symbols at an arbitrary rate, and a decision system may generate candidate commitments at an arbitrary rate. What the condition decides is whether the resulting stream can be decoded, or certified, as fast as it arrives.

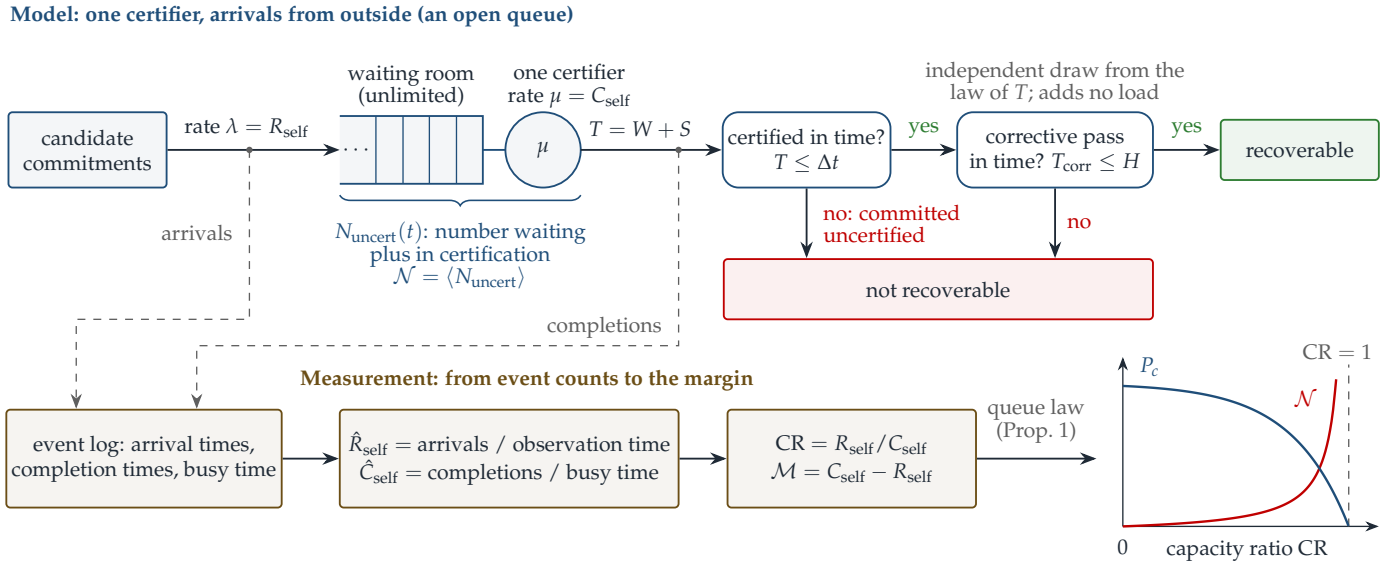


Figure 1. The certification pipeline modelled in this paper: candidate commitments arrive at rate $\lambda = R_{\text{self}}$ and wait for a single certifier of rate $\mu = C_{\text{self}}$ (top), and the two rates are estimated from an event log and combined with the law of the queue to give the backlog and the probability P_c of certification in time (bottom), with an inset for the M/M/1 case at $\mu\Delta t = 4$.

What the constraint is built on. Let λ be the arrival rate of candidate commitments and μ the certification rate. To draw the analogy with information rates, assign to each commitment the same amount of uncertainty, an *uncertainty quantum* of $\bar{w}$ bits that must be resolved to certify it. Multiplying by $\bar{w}$ expresses both rates in bits per unit of time, $R_{\text{self}} = \lambda\bar{w}$ and $C_{\text{self}} = \mu\bar{w}$. The quantum cancels in CR $= \lambda/\mu$, which is why counts are enough. Measured in channel uses instead of bits, with one channel use per certification, the two rates are $R_{\text{self}} = \lambda$ and $C_{\text{self}} = \mu$. Commitments with unequal quanta are handled by the load-weighted effective bandwidth [10,12,13].

The modelled system is allowed memory, since a person, an institution or a coupled human–AI system is not memoryless. Shannon's memoryless channel supplies only the form of the limit, and what this paper actually uses is the stability boundary of a service process. That boundary, CR < 1, exists for general stationary arrival and service processes. For the queue families of Proposition 1(e), which have fixed distribution shapes, finite second moments and positive variability, $c_a^2 + c_s^2 > 0$, the mean backlog grows as $(1 - \text{CR})^{-1}$ as CR approaches one (Section 6). Memorylessness is needed only for some of the explicit closed forms. When options are lost after a deadline Δt, the allowed arrival rate must also meet a specified probability of deadline success. Effective-capacity methods estimate this

relation under their traffic and service assumptions [13]. Section 6.2 gives exact deadline probabilities for the exponential-service queues used here.

The whole construction rests on one modelling step, which is to treat one certification as one channel use. This is a choice, not a theorem. Under it, the channel alphabet is the set of candidate commitments, and the coding operation is certification. The source uncertainty is the quantum $\bar{w}$ to be resolved for each commitment. The capacity is the certification service rate, which is not a supremum of mutual information. The analogy would carry more weight if a converse held, that is, a proof that no certification discipline can sustain recoverable operation at CR > 1. Loynes' theorem [8] supplies that converse for stability: above the boundary, no stationary backlog distribution exists under any work-conserving discipline that keeps all incoming work. Loynes' theorem gives no converse for the reading of the two rates as information rates. That reading therefore remains an analogy and is marked as one wherever it is used.

3. The Flux and the Capacity Are Measured Quantities

Because R_{self} and C_{self} are defined operationally as rates, both can be estimated from a stream of events without access to the internal model of the system. This section illustrates the estimation on physiological data, an example chosen because the data are openly available. The estimation has not been clinically validated. Physiological homeostasis is read as a biological self-decoder. A monitored variable that leaves its reference band is treated as a candidate commitment, so the rate of such crossings estimates the induced flux R_{self}. The rate at which excursions return to the band estimates the integrative capacity C_{self}.

The scope of the section is limited in two ways. The first limit is the meaning given to the events, which is a modelling choice. An out-of-range result can reflect the timing of measurements, a treatment or natural fluctuation, and need not be a commitment event at all. Validating the mapping against outcomes that represent restoration or failed recovery would take clinical expertise beyond the scope of this paper. The section therefore demonstrates that the rates can be measured and makes no claim of clinical validity. The second limit is that the restoration process is parallel, not serial, because several analytes can be out of range and return to range at the same time. The physiological data are therefore used only for the rates R_{self} and C_{self} and for their margin, quantities that are well defined for a parallel restoration process, which is an infinite-server process. In such a process, the stationary number of open excursions has mean $\lambda\,\mathbb{E}[D_{\text{exc}}]$, where D_{exc} is the duration of an excursion, and it stays finite at every load. The single-server law of Section 6 applies where certification passes through a finite shared bottleneck, as in the decision systems of Sections 7 and 8, and is left aside for these data. The distinction matters, because the divergence needs such a bottleneck, and an infinite-server process has none.

The rates were first estimated on synthetic longitudinal health records [14], where both are well defined and non-degenerate across the population (Figure 2a). The margin $\mathcal{M} = C_{\text{self}} - R_{\text{self}}$ follows directly for every unit (Figure 2b). The same computational procedure was then applied, with cohort-specific reference ranges, to real physiology. The data are the openly licensed demonstration subset of the Medical Information Mart for Intensive Care, version IV (MIMIC-IV) [15,16], read against its own per-assay reference bands. The procedure returns non-degenerate rates for 78 of its 100 patients, and Supplementary Table S1 gives the selection rules, the counts before and after exclusion, and the identity of each release. What the two cohorts establish is that the estimator runs on real event streams of two different kinds and returns finite rates for both. No comparison is drawn between them, for the reasons stated in the next paragraph.

Why no contrast between the cohorts is drawn. The two shares of units below the reference line, 90.3% of the synthetic units and 44.9% of the critically ill patients, measure

the observation scheme rather than a capacity property, and five analyses on the same pipeline establish it. As the estimator is coded, CR is the product of two factors: onsets per restoration, and the fraction of the observation span during which at least one analyte is out of range. The first factor exceeds one only when excursions are left open at the end of a record, and 97.4% of the MIMIC-IV units and 76.3% of the synthetic units do end on one, so the comparison is dominated by right-censoring. The median capacity ratio of the critically ill cohort, 1.02, is not significantly above one: a bootstrap over units gives $[0.90, 1.09]$. After the terminal censored excursions are excluded, onsets equal restorations, so the first factor is exactly one, CR equals the busy fraction of the observation span and cannot exceed one: the median falls to 0.50 in the synthetic records and 0.66 in MIMIC-IV, and both cohorts then read 100% below the line. That is a property of the exclusion rule rather than a finding about either cohort. The two arms are also not measured on the same panel, since nine analytes resolve a usable reference band in the synthetic records against thirteen in MIMIC-IV, and the per-analyte capacity ratio varies by more than an order of magnitude within a cohort, so a single pooled value per unit averages over very different series. A censoring-aware duration estimator, such as Kaplan–Meier on excursion durations, is the method a future cohort study should use; at these censoring rates it would rest almost entirely on extrapolation, and it is left to that study. Figure 2 is therefore an illustration that the two rates are computable from an event log, which is what this section claims.

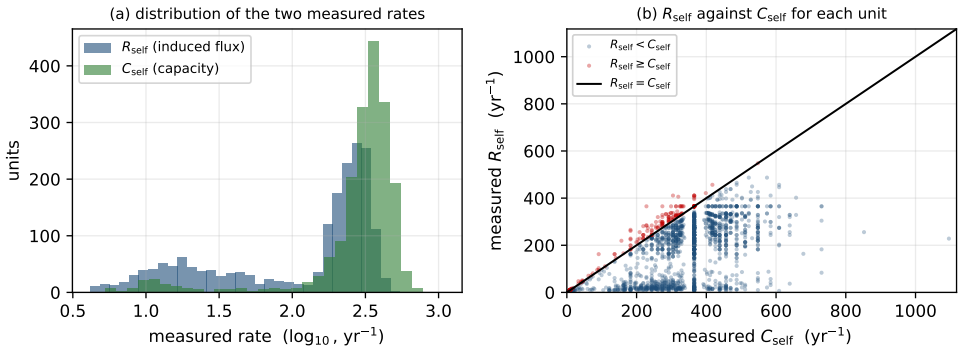


Figure 2. R_{self} and C_{self} measured from a longitudinal event stream of synthetic electronic health records [14], with homeostasis as the self-decoder and about 1900 units: (**a**) the distribution of the two measured rates, and (**b**) the two rates for each unit, with the line $R_{self} = C_{self}$ marked.

Because the restoration process is parallel, CR is here descriptive, and the line $R_{self} = C_{self}$ of Figure 2b is a reference marker, not the single-server stability boundary of Section 6; no region of that plane is a feasible set. With stationary arrivals and a finite mean excursion duration, an infinite-server model has no threshold in the capacity ratio at all. R_{self} counts onsets per unit of observation span, and C_{self} counts restorations per unit of time during which at least one excursion is open. CR therefore exceeds one only when counted excursions remain unresolved at the end of the series. Its value depends on the observation and censoring scheme, and no clinical or causal reading is drawn from it here. Supplementary Table S1 records the analysis details: the analyte panel, the eligibility and exclusion rules, the event definition, the censoring rules, and the dependence between events. The estimator also leaves the differences between the two observation processes in place. Sampling in intensive care is more frequent and is triggered by illness, the synthetic records are sampled on a different schedule, and CR may change with either. Both pipelines report rates per year, but the cohorts are observed over very different spans (days in intensive care, years in the synthetic records) and are sampled differently. Their absolute rates are therefore not compared. Matched-cadence analysis, sensitivity to downsampling, a pre-specified analysis plan and a receiver operating characteristic (ROC) analysis of detection remain

future work; the censoring sensitivity, the per-analyte results and the bootstrap interval for the median capacity ratio are reported above.

4. Relation to Existing Work

Equation (1) draws on several established fields. The mechanism is familiar, and the contribution of this paper is to make it operational. Classical decision frameworks include expected utility and optimal control, risk management [17], real options [18], robust and safe-set methods [19,20] and AI alignment [21]. They address decision quality, risk and safety, and real options also address the cost of irreversible commitment. The present construction adds a different constraint: the certifier is congested, and the time available for correction is finite. The safety-science tradition [2–4], recently restated for AI [7], describes this failure qualitatively and gives no measure of it. The systems literature documents the same family of failures at engineering scale. Metastable failures in distributed systems are degraded states that persist after the trigger has been removed [22].

Within this research programme, the foundational preprint [6] introduces the idea and the estimator programme at length. The present paper adds a compact information-theoretic statement, a demonstration of measurability with its scope limits (Section 3), the derivation of the backlog law (Section 6), and the controlled ordering of accuracy and recoverability (Section 7). The companion physics paper [11] treats the closed-loop problem, in which uncertified output is consumed again as load and the transition becomes first-order, with hysteresis. Here, the queue is open, and the approach to the boundary is continuous (Figure 1). The independent treatment of [7] reaches the agency thesis without an order parameter.

Five neighbouring frameworks measure quantities close to the two described in this paper. *Queues with deadlines and impatient customers* study the mechanism of Section 6 directly: work that must be served before a deadline, and work that leaves when it waits too long. Barrer's constant-deadline queue [23] and the general single-server treatment [24] supply the stationary laws used in Section 6.2, the many-server case with general patience is [25], and the earliest-deadline-first discipline has a heavy-traffic theory [26]. That literature computes the fraction of work that misses its deadline, and stops there. *Age of information* measures the staleness of what a monitor knows as the time since its last update, and optimizes update policies against it [27,28]; it is a freshness measure for information rather than an admissibility condition for action. *Queueing models of human supervisory control* treat an operator as a server and predict capacity from utilization and the wait times that interaction imposes [29]; the quantity at risk there is the supervisor's performance rather than the reversibility of what was approved. *Automation bias* concerns the quality of the human judgment itself under automated advice [30,31], and the *speed–accuracy trade-off* concerns the accuracy of one decision as a function of the time taken over it [32]. Both of those are about judgment, whereas the mechanism of this paper leaves individual judgment untouched and removes the opportunity to correct it. Table 1 states what each framework measures, whether it can be estimated from event counts alone, and what RSC adds.

RSC contributes two elements to these approaches: recoverability as a single operationally defined quantity, and, for the lossless pipeline of Section 6, a feasibility boundary read from event counts. It is meant to complement them.

Table 1. Neighbouring frameworks, the quantity each one measures, and what Recoverable Self-Coding adds to it.

Framework	What It Measures	From Event Counts?	What RSC Adds
Queues with deadlines and impatient customers [23–26]	the fraction of work served before its deadline, and the loss when it is not	yes	an opportunity to correct after certification, and an admissibility rule for irreversible action
Age of information [27,28]	the staleness of a monitor's information as time since its last update	yes	what the stale information was already used to do, and whether that can be undone
Human supervisory control [29]	operator utilization and the wait times that bound how many tasks one person can supervise	yes	the reversibility of the approved action, rather than the capacity of the supervisor
Automation bias [30,31]	errors of commission and omission when a human follows automated advice	no: labelled judgments are needed	a failure that holds with the quality of individual judgment unchanged
Speed–accuracy trade-off [32]	the accuracy of one decision as a function of the time taken over it	no: outcomes per decision are needed	a separation at the level of the system: accuracy at its floor while recoverability falls

Overlap with the author's other work. This paper is the proceedings contribution for the peer-reviewed conference abstract [33] and poster [34] for Entropy 2026. The poster states the same simulated values for accuracy and recoverability as Section 7 and the same measured rates as Section 3, with no analysis beyond them; the derivation, the closed forms, the precision analysis and the scope limits of this paper are new here. The foundational preprint [6] introduces the framework and the estimator programme and is not peer-reviewed. The companion physics paper [11] treats the closed loop; it shares the queueing background and none of the results of this paper. An extended version is in preparation for a journal Special Issue and will cite this paper as its proceedings antecedent.

5. The Recoverability Constraint

RSC separates two questions that conventional tools treat as one. Accuracy concerns whether beliefs are correct, whereas feasibility concerns whether an irreversible action is admissible under constraints of rate, latency, and option loss. A wrong belief can be revised at no permanent cost, but an irreversible commitment cannot. What matters is therefore whether the commitment can still be corrected before it becomes permanent, and whether the belief was right (message error) is a separate matter. Even a decoder that makes no errors can fail this test, because the world can change after a commitment that can no longer be corrected.

The terms used from here on are fixed as follows. *Correctness* is a property of a belief: the content of the commitment matches the world. *Certification* is the act of checking a candidate commitment against evidence before it becomes irreversible. A commitment is *traceable* when the record that links the decision to its outcome survives certification. *Recoverability* is a property of the path of the system, not of a single belief. A commitment is recoverable if an admissible sequence of future actions can still correct or cancel it before the options it removes are gone. *Information retention* is the condition defined next. Each measured event in this paper is tied to exactly one of these terms.

The framework's second condition concerns how much information about the environment the certified record still carries. Information is *retained* over Δt when, before commitment, perturbations δE of the environment remain informative about the certified macrostate Z. Operationally, the mutual information before commitment stays bounded away from zero:

$$I(\delta E; Z \mid \text{pre-commit}) \geq \iota > 0. \tag{2}$$

Positive mutual information means that Z retains information about δE. It does not establish local invertibility of the map from measurement to action. A many-to-one mapping can retain information and still be non-invertible. The stronger property needs a local inverse condition, either a full-rank Jacobian of the map or an observability and identifiability criterion over the horizon, and that condition is not verified here. This paper therefore calls Equation (2) information retention, which is what it measures, and leaves the stronger term, invertibility, to the foundational treatment [6]. The framework proposes the following admissibility rule. It says when irreversible action is allowed, and optimization lies outside it.

> **Recoverability Constraint.** Irreversible action is admissible only while $\mathcal{M}(t) > 0$ on the relevant horizon. If that condition fails, the framework withholds authorization for irreversible commitment, because such a commitment can remove future options that no later revision of belief restores. The rule is a modelling proposal, and the results below establish its queueing consequences within their stated scope.

6. The Uncertified Backlog and Its Law

The margin condition is the capacity axis of the framework, and information retention is its retention axis. The capacity axis is measured by the *uncertified backlog*:

$$\mathcal{N} \equiv \langle N_{\text{uncert}} \rangle. \tag{3}$$

$\mathcal{N}$ is the stationary mean number of uncertified commitments, counting those waiting for certification together with the one being certified (Figure 1). It has units of commitments, so it is a mean count, not a dimensionless ratio. The neutral symbol is used here because the quantity is the standard mean number in the system. Equation (3) is read from observable counts and depends on no horizon or threshold.

The index also has a thermodynamic reading, offered as an analogy only. If certifying a commitment were implemented as a logically irreversible resolution of at least one bit, Landauer's principle would associate the backlog of Equation (3) with dissipation that has not yet been incurred [35]. Since neither a computational substrate nor an erasure protocol is modelled here, this remains an interpretive analogy, and no quantitative bound is claimed.

6.1. Analytical Derivation for the Single-Server Bottleneck

The model is a lossless single-server queue with unlimited waiting room. Commitments are certified in order of arrival (first come, first served, or FCFS). The certifier works whenever work is present and completes each certification without interruption. There are no vacations, setup times, rejected commitments or departures before service. Arrivals are exogenous, and completed or timed-out commitments generate no additional load. A missed deadline changes the outcome label only. The commitment remains in the queue until it is served; Section 6.2 gives the setting in which that holds and the consequences of relaxing it.

The interarrival times $A_1, A_2, \ldots$ are independent and identically distributed (i.i.d.), with mean $1/\lambda$, variance $\mathrm{Var}(A)$ and squared coefficient of variation $c_a^2 = \lambda^2 \mathrm{Var}(A)$. The certification times $S_1, S_2, \ldots$ are i.i.d., with mean $1/\mu$, variance $\mathrm{Var}(S)$, and $c_s^2 = \mu^2 \mathrm{Var}(S)$.

The two entire sequences are independent. Both times are nonnegative, and both means are finite and strictly positive. Simultaneous arrivals are ordered by their indices, with service completions processed before the next arrival in that order. These conventions also allow zero service times. Deterministic and lattice distributions are included.

The capacity of the bottleneck is its service rate μ, and the buffer has no size limit. In Kendall's notation, this is the GI/GI/1 queue. The first symbol describes the interarrival times, the second describes the service times, and the number counts the servers. M stands for an exponential distribution, D for a deterministic one, H_2 for a two-phase hyperexponential one, and GI (or G) for a general distribution with independent draws. Let W_n be the time that commitment n waits before its certification starts. The Lindley recursion [36] is

$$W_{n+1} = \max(0,\ W_n + U_n), \qquad U_n = S_n - A_{n+1}. \tag{4}$$

Let $\eta_{n+1} = \max(0,\ -(W_n + U_n))$ be the idle time within the interval ending at arrival $n+1$. This variable includes its mass at zero. It can be zero even when the next arrival starts service immediately, if that arrival coincides with the preceding completion. For $0 < \mathrm{CR} < 1$, negative mean drift gives a unique proper stationary waiting-time law [8]. Supplementary S3.1 constructs this law from past increments and establishes the moments used below.

The symbols W, U, and η denote stationary versions indexed by arrivals, and $T = W + S$ is a typical arrival's sojourn time. In contrast, $\mathcal{N}$ is the mean number in the system at an equilibrium time, counting waiting commitments and the one in service. Little's law relates these two ways of sampling [37]. Arrivals need not see the time-stationary queue distribution. The equilibrium construction also applies to lattice arrivals, and Supplementary S3.1 explains the time-phase convention.

Proposition 1. *Under the model above, let* $0 < \mathrm{CR} = \lambda/\mu < 1$, $\mathbb{E}[A^2] < \infty$ *and* $\mathbb{E}[S^2] < \infty$. *Then, the following statements hold.*

(a) Exact identity. At every subcritical load,

$$\mathcal{N} = \mathrm{CR} + \frac{c_a^2 + \mathrm{CR}^2 c_s^2 - \lambda^2 \,\mathrm{Var}(\eta)}{2\,(1 - \mathrm{CR})}. \tag{5}$$

(b) Upper bound at every load [38]. $\mathcal{N} \le \mathrm{CR} + \dfrac{c_a^2 + \mathrm{CR}^2 c_s^2}{2\,(1 - \mathrm{CR})}$.

(c) Poisson arrivals (M/G/1). $\lambda^2 \mathrm{Var}(\eta) = 1 - \mathrm{CR}^2$, *so that*

$$\mathcal{N} = \mathrm{CR} + \frac{\mathrm{CR}^2\,(1 + c_s^2)}{2\,(1 - \mathrm{CR})}, \tag{6}$$

which is the Pollaczek–Khinchine formula [39]. For M/M/1 it reduces to $\mathcal{N} = \mathrm{CR}/(1 - \mathrm{CR})$, *and for M/D/1 to* $\mathcal{N} = \mathrm{CR} + \mathrm{CR}^2/(2(1 - \mathrm{CR}))$.

(d) Exponential service (GI/M/1). $\mathcal{N} = \mathrm{CR}/(1 - \sigma)$, *where* σ *is the unique root strictly inside* $(0, 1)$ *of* $\sigma = \mathbb{E}[e^{-\mu(1-\sigma)A}]$. *The sojourn time is exponential with rate* $\mu(1 - \sigma)$, *so that* $P_c \equiv \Pr(T \le \Delta t) = 1 - e^{-\mu(1-\sigma)\Delta t}$ *for* $\Delta t \ge 0$. *For M/M/1,* $\sigma = \mathrm{CR}$.

(e) Limit at the boundary. Fix the service law and write $A = B/\lambda$, *where the law of* B *is fixed,* $\mathbb{E}[B] = 1$ *and* $\mathbb{E}[B^2] < \infty$. *Assume* $c_a^2 + c_s^2 > 0$. *As* $\lambda \uparrow \mu$, $(1 - \mathrm{CR})\,\mathcal{N} \to \frac{1}{2}\,(c_a^2 + c_s^2)$, *so* $\mathcal{N}$ *diverges as* $(1 - \mathrm{CR})^{-1}$.

Proof. Put $d = -\mathbb{E}[U] = (1 - \mathrm{CR})/\lambda > 0$. The stationary construction makes W independent of the current S and the next A; hence, it is independent of U. Before using any stationary moment balance, Supplementary S3.1 proves $\mathbb{E}[W] \le \mathbb{E}[U^2]/(2d) < \infty$ from the empty-start recursion. This establishes the integrability needed for the first balance.

Write $X = W + U$, $W' = \max(0, X)$ and $\eta = \max(0, -X)$. Then, $W' - \eta = X$, $W'\eta = 0$ and $0 \le \eta \le A$. Stationarity and finite $\mathbb{E}[W]$ give $\mathbb{E}[\eta] = d$. A second-order balance needs

care because $\mathbb{E}[W^2]$ can be infinite. For $K > 0$, let $f_K(x) = \min(x^2, K)$. Stationarity gives $\mathbb{E}[f_K(W') - f_K(W)] = 0$, and

$$|f_K(W') - f_K(W)| \le |W'^2 - W^2| \le 2W|U| + U^2.$$

The last expression is integrable by independence and the established finite mean of W. Dominated convergence therefore gives $\mathbb{E}[W'^2 - W^2] = 0$ as an expectation of an integrable difference. Since $W'^2 - W^2 = 2WU + U^2 - \eta^2$, it follows that

$$2d\,\mathbb{E}[W] = \mathbb{E}[U^2] - \mathbb{E}[\eta^2] = \mathrm{Var}(U) - \mathrm{Var}(\eta).$$

The last equality uses $\mathbb{E}[U] = -d$ and $\mathbb{E}[\eta] = d$. Independence also gives $\mathrm{Var}(U) = \mathrm{Var}(A) + \mathrm{Var}(S)$. Little's law gives $\mathcal{N} = \lambda\mathbb{E}[T] = \mathrm{CR} + \lambda\mathbb{E}[W]$, which is finite. Substitution proves (a), and $\mathrm{Var}(\eta) \ge 0$ proves (b).

For (c), an independent exponential interarrival time gives, for $x \ge 0$,

$$\Pr(\eta > x) = \Pr(A > W + S + x) = e^{-\lambda x}\,\mathbb{E}[e^{-\lambda(W+S)}].$$

Thus, the positive part of the idle-time law is exponential with rate λ. Its probability mass is $\lambda\mathbb{E}[\eta] = 1 - \mathrm{CR}$. Consequently, $\mathbb{E}[\eta^2] = 2(1 - \mathrm{CR})/\lambda^2$ and $\lambda^2\mathrm{Var}(\eta) = 1 - \mathrm{CR}^2$. With $c_a^2 = 1$, substitution in (a) gives the Pollaczek–Khinchine formula [39] and its two stated special cases.

For (d), let Q be the number of other commitments found by a typical arrival. Supplementary S3.4 proves that the root σ exists and is unique, and that $\Pr(Q = n) = (1 - \sigma)\sigma^n$. Conditional on $Q = n$, exponential memorylessness makes T a sum of $n + 1$ independent exponentials with rate μ. For $s \ge 0$, its Laplace transform is

$$\mathbb{E}[e^{-sT}] = (1 - \sigma)\sum_{n=0}^{\infty}\sigma^n\left(\frac{\mu}{\mu + s}\right)^{n+1} = \frac{\mu(1 - \sigma)}{s + \mu(1 - \sigma)}.$$

This is the transform of an exponential variable with rate $\mu(1 - \sigma)$, which gives the stated P_c. Little's law gives $\mathcal{N} = \mathrm{CR}/(1 - \sigma)$. These classical GI/M/1 formulas are also stated in [40] (Section 5, pp. 7–8). The waiting time itself has an atom $\Pr(W = 0) = 1 - \sigma$, and the sojourn time has none.

For (e), use fixed independent sequences of B and S, and set $A = B/\lambda$ for each load. Put $J_\lambda = \lambda\eta$. Then, $0 \le J_\lambda \le B$ and $\mathbb{E}[J_\lambda] = 1 - \mathrm{CR}$. For every fixed $K > 0$,

$$\mathbb{E}[J_\lambda^2] \le K(1 - \mathrm{CR}) + \mathbb{E}[B^2; B > K].$$

Take the upper limit as $\lambda \uparrow \mu$, then let $K \to \infty$. The finite second moment of B gives $\mathbb{E}[J_\lambda^2] \to 0$. Hence, $\lambda^2\mathrm{Var}(\eta) \le \mathbb{E}[J_\lambda^2] \to 0$. Multiplying Equation (5) by $1 - \mathrm{CR}$ proves the limit. □

The scope of Proposition 1 is stated here. Its results are inherited queueing theory, collected here because the rest of the paper rests on them. The idle-time representation is due to Marshall [41], and Chen and Whitt give it explicitly in [42] (Theorem 8.1, Equations (8.1)–(8.3)). Supplementary S3.2 converts their positive-idle-period convention to the variable η used here. Equation (5) contains $\mathrm{Var}(\eta)$, which depends on the full arrival and service laws. Thus, CR, c_a^2, and c_s^2 determine the bound (b) and the limit (e), while finite-load $\mathcal{N}$ also depends on the idle-time term. All parts of the proposition hold under the stated second moments. Supplementary S3.1 also proves that $\mathbb{E}[S^3] < \infty$ suffices for $\mathbb{E}[W^2] < \infty$, without a third moment of A.

Part (e) is a limit of means. It yields Kingman's heavy-traffic mean approximation $\mathcal{N} \approx \frac{1}{2}(c_a^2 + c_s^2)\,(1 - \mathrm{CR})^{-1}$ [9] under the stated fixed-law scaling and positive variability. The exponent -1 is shared within this class, and the amplitude depends on the distributions. If both distributions are deterministic, $W = 0$ and $\mathcal{N} = \mathrm{CR}$ at every subcritical load.

The proof of (e) establishes no distributional limit or general deadline-probability formula. An exponential tail bound at a fixed load needs the additional hypothesis of Supplementary S3.5 [43]. Section 7 tests parts (c) and (d) against simulation for three queues.

6.2. The Boundary, the Deadline and the Margin

The boundary CR = 1 is the stability boundary of the certification queue. Under the independent arrival and service assumptions of Section 6.1, a stationary backlog exists when CR < 1 and fails to exist when CR > 1 [8]. At CR $= 1$, the queues studied here, which have positive variability, $c_a^2 + c_s^2 > 0$, have no stationary waiting-time distribution. Balanced deterministic arrivals and service are the exception. In the lossless pipeline, stability is necessary for recoverable operation but not sufficient, because a stable queue can still miss its deadlines; Section 6.2 shows that a pipeline which drops expired work instead keeps a positive recoverable fraction above the boundary. The certification horizon therefore enters recoverability and leaves the uncertified backlog untouched. For the M/M/1 case, Proposition 1(d) gives an exponential sojourn time with rate $\theta^\star = \mu(1 - \mathrm{CR})$, so a commitment is certified within Δt with probability exactly $P_c = 1 - e^{-\theta^\star \Delta t}$. For exponential service and general arrivals, the same form holds exactly with $\theta^\star = \mu(1 - \sigma)$. For the fixed-shape GI/GI/1 families of Proposition 1(e), the standard heavy-traffic approximation replaces the sojourn-time distribution by an exponential one with rate $\theta^\star \approx 2\mu(1-\mathrm{CR})/(c_a^2 + c_s^2)$, which has the form of the effective-bandwidth quality-of-service exponent [10,12,13,44]. In that case, the expression for P_c is an approximation, not an identity, and an exponential bound on the tail at a fixed load needs the additional condition stated in Supplementary S3.5. The approximation can fail at a short deadline: for M/D/1 with $\Delta t < 1/\mu$, the true probability of certification in time is zero.

The uncertified backlog of Equation (3) is the observable of the capacity axis alone. The second axis is information retention, measured by the mutual information of Equation (2), and it can fail independently, since the latency of certification can exceed the time over which the environmental cause drifts. Everything that follows bears on the capacity axis alone. Information retention is defined in Equation (2) because the paper uses the term, and it is left to future work rather than carried into the admissibility rule: this paper gives no estimator of $I(\delta E; Z \mid \text{pre-commit})$, no way to identify δE and Z in an event log, and no value or selection rule for ι, and the conditioning on pre-commitment is an event label rather than a random variable. The formulas above hold for a stationary queue, and they are applied here while the load changes slowly. This is valid only if the queue settles to its stationary state much faster than the load changes, $\tau_{\text{queue}} \ll \tau_{\text{load}}$. This is a real restriction. For the M/M/1 queue, the relaxation time is $\tau_{\text{queue}} \sim \mu^{-1}(1 - \sqrt{\mathrm{CR}})^{-2}$, which diverges at the boundary. This divergence is known as critical slowing down, and the end of this section returns to it.

From the margin to a candidate predictor. For the M/M/1 queue in units of counts ($\bar{w} = 1$), the decay rate is the feasibility margin itself, $\theta^\star = \mu(1 - \mathrm{CR}) = C_{\text{self}} - R_{\text{self}} = \mathcal{M}$, and under the heavy-traffic approximation, $\theta^\star$ is proportional to $\mathcal{M}$. The margin is thus more than the sign of condition (i): for the M/M/1 queue it is the rate at which the probability of a long sojourn time decays. It follows that the probability of first-pass certification, $P_c = 1 - e^{-\mathcal{M}\Delta t}$, can be read directly from the current margin (the two-pass recoverability of Section 7 also requires a corrective pass within H). Its complement $e^{-\mathcal{M}\Delta t}$ is the probability of missing the certification deadline. By analogy, it corresponds to the error probability in Shannon's theorem, since its exponent $\mathcal{M}$ vanishes at CR $= 1$ just as the channel's exponent vanishes at $R = C$. A system approaches the boundary $\mathcal{M} \to 0$ gradually, so a future predictor could use the present margin and its rate of change, $(\mathcal{M}(t), \dot{\mathcal{M}}(t))$, to estimate the risk that the margin reaches zero soon. In the language of stochastic processes, the two

quantities are the proposed inputs to a first-passage hazard for the loss of recoverability. The hazard is the instantaneous rate at which the margin first reaches zero, conditional on it not having reached zero already. This is a proposal, not a validated predictor. No hazard function is derived or tested empirically here, and calibration, lead time, false-alarm rates and ROC behaviour on labelled event streams are future work.

How precisely the margin can be measured. Near the boundary, the margin is a small difference between two large rates, so its precision decides whether it can be used at all. Let the two rates be estimated over an observation span τ_{obs} by the estimators of Figure 1, $\hat{R}_{\text{self}}$ from arrivals per unit of elapsed time and $\hat{C}_{\text{self}}$ from completions per unit of server-busy time. For independent renewal arrivals and service, over a window long enough for the effect of its two edges to be negligible and short enough for both rates to be treated as constant, counting variability gives

$$\begin{aligned} \operatorname{Var}(\hat{R}_{\text{self}}) \simeq \frac{c_a^2 \lambda}{\tau_{\text{obs}}}, \qquad \operatorname{Var}(\hat{C}_{\text{self}}) \simeq \frac{c_s^2 \mu^2}{\lambda \tau_{\text{obs}}}, \\ \operatorname{Var}(\hat{\mathcal{M}}) = \operatorname{Var}(\hat{C}_{\text{self}}) + \operatorname{Var}(\hat{R}_{\text{self}}) - 2\operatorname{Cov}(\hat{R}_{\text{self}}, \hat{C}_{\text{self}}), \end{aligned} \tag{7}$$

and, by the same first-order propagation, $\operatorname{Var}(\widehat{\text{CR}}) = [\operatorname{Var}(\hat{R}_{\text{self}}) + \text{CR}^2 \operatorname{Var}(\hat{C}_{\text{self}}) - 2\,\text{CR}\operatorname{Cov}]/\mu^2$, while the variance of the implied P_c is $(\Delta t\, e^{-\mathcal{M}\Delta t})^2 \operatorname{Var}(\hat{\mathcal{M}})$ for the M/M/1 queue, in which the sojourn decay rate is the margin itself. The first two expressions of Equation (7) are large-window approximations rather than finite-window identities, while the expression for $\operatorname{Var}(\hat{\mathcal{M}})$ is an identity; the residual variability of the window edges is why the deterministic-service estimate has any spread at all, where the leading term is zero. For the GI/M/1 queues of Proposition 1(d) the mapping from the rates to P_c runs through $\theta^\star = \mu(1-\sigma)$ and its gradient is different: at CR = 0.94, the same margin gives $P_c = 0.213$ for M/M/1 and 0.095 for the H_2/M/1 queue of Section 7. The covariance is carried through all three expressions. For this pair of estimators, it is measured to be negligible (Section 7), because at a given busy exposure, the completion count varies with the service process alone; a capacity estimated instead as throughput per unit of elapsed time would be strongly correlated with $\hat{R}_{\text{self}}$, since in a stable queue, throughput equals the arrival rate.

Precision therefore depends on the exposure, on the variability of both processes, and on whether the capacity is estimated or known independently. For the review boards of the Introduction ($\mu = 10$ per week), one year of counts gives a standard error of 0.62 per week, which is larger than the margin of 0.5 per week that it estimates, and 0.43 per week when μ is known. A statement about the span needed to resolve a margin requires its criterion. For a standard error of $\varepsilon\mathcal{M}$,

$$\tau_{\text{obs}} \gtrsim \frac{c_a^2 \lambda + c_s^2 \mu^2/\lambda}{(\varepsilon \mathcal{M})^2}, \tag{8}$$

and a 95% half-width of $\varepsilon\mathcal{M}$ multiplies that span by 1.96^2. Its numerator keeps the two counting terms alone, which the measured covariance of this pair of estimators justifies (Section 7). At $\mathcal{M} = 0.5$ per week, the three criteria of one standard error equal to the margin, a 95% half-width equal to the margin, and a 95% half-width of a tenth of the margin need 1.5, 5.9 and 592 years. At $\mathcal{M} = 0.025$ per week, they need 615, 2364 and 236,000 years. The first of these is the arithmetic behind the statement that several centuries of observation would be needed, and it is the weakest of the three criteria.

For the M/M/1 queue, as $\lambda \uparrow \mu$, the observation span and the relaxation time both grow as the inverse square of the margin: the span for one standard error tends to $2\mu/\mathcal{M}^2$ and the relaxation time to $4\mu/\mathcal{M}^2$, so the span for a 95% half-width of $\varepsilon\mathcal{M}$ is $1.92\,\tau_{\text{queue}}/\varepsilon^2$. For other renewal arrival and service laws, the counting-span numerator of Equation (8) tends to $\mu(c_a^2 + c_s^2)$, which preserves the inverse-square scaling of the span; the ratios to the relaxation

time stated here are specific to M/M/1. The two rates must also stay approximately constant across the window, $\tau_{\text{obs}} \leq \tau_{\text{load}}$, which for M/M/1, bounds the attainable relative precision by $\varepsilon \gtrsim 1.4\,(\tau_{\text{queue}}/\tau_{\text{load}})^{1/2}$: a relative precision of ten percent requires a load that drifts about two hundred times more slowly than the queue relaxes. These are necessary-condition estimates read off the asymptotic expressions above under the stated window condition, not a general queueing theorem. For the second board of the Introduction, the span for one standard error, 615 years, is half the relaxation time of 1229 years.

Two consequences follow for small margins. First, an interval for $\hat{\mathcal{M}}$ near zero contains negative values, where the queue has no stationary sojourn law and P_c is undefined. The interval is therefore reported together with a normal-approximation weight for $\lambda \geq \mu$, formed from the estimated margin and its standard error as a diagnostic rather than as a probability that the margin is negative, and P_c is quoted only on the part of the interval with $\lambda < \mu$. Forcing the margin positive before forming P_c reports $P_c = 0.28$ for the second board of the Introduction, against the stationary value of 0.05, an artefact of the estimator's own noise. Second, over 2000 independent years of that board, the estimated margin has the wrong sign in 48.6% of years and its 95% interval contains zero in 93.8%. What survives is the margin as a load diagnostic reported with its uncertainty, and the early-warning use above carries the condition that a window exists which is long enough for the counts and short enough for the rates. Supplementary S2.6 gives the protocol and the tables, and Section 7 checks the expressions above against the simulation.

What the collapse assumes: expired work stays in the queue. The model of Section 6.1 keeps a commitment in the queue after it misses Δt and certifies it late. A commitment that has missed its horizon is already irreversible, so certifying it afterwards restores no option; what the late certification produces is the record. The assumption therefore fits a pipeline in which review is an obligation that the deadline does not discharge, and the deadline decides only whether the commitment can still be reversed. Case assessment in pharmacovigilance, occurrence reporting in aviation safety and post-trade compliance review are of that kind. Where a pipeline instead drops expired items or approves them without review, the model changes, and the change is large enough to state here.

Give every candidate commitment a deterministic patience equal to Δt, so that one which cannot begin certification within its horizon leaves the queue and is committed without review. Under first come, first served, its waiting time is the workload it finds on arrival, so the departure can be settled on arrival; this is the constant-deadline queue of Barrer [23], whose general single-server treatment is [24]. Under Poisson arrivals, the workload that an arrival finds is its waiting time W, and a level-crossing balance gives the stationary density $\lambda \Pr(W = 0)\, e^{-\mathcal{M}x}$ on $(0, \Delta t]$ and $\lambda \Pr(W = 0)\, e^{-\mathcal{M}\Delta t} e^{-\mu(x-\Delta t)}$ above Δt. No arrival joins above the patience level, so the density decays at rate μ there whatever the load, and a stationary workload exists for $\lambda \geq \mu$ as well. An arrival joins if $W \leq \Delta t$ and is certified in time if $W + S \leq \Delta t$, so that

$$P_c = \Pr(W = 0)\left[1 + \frac{\lambda\,(1 - e^{-\mathcal{M}\Delta t})}{\mathcal{M}} - e^{-\mathcal{M}\Delta t}\right],$$

$$\Pr(W = 0)^{-1} = 1 + \frac{\lambda\,(1 - e^{-\mathcal{M}\Delta t})}{\mathcal{M}} + \mathrm{CR}\, e^{-\mathcal{M}\Delta t},$$

both continuous at $\mathcal{M} = 0$ and reducing to $P_c = 1 - e^{-\mathcal{M}\Delta t}$ as the patience grows. Recoverability is P_c^2 at $\Delta t = H$, since the corrective pass meets the same queue. The consequences are plain. At CR = 0.94, the first-pass probability is 0.70 against 0.21 in the lossless model, and recoverability is 0.49 against 0.046. At CR = 1, they are 2/3 and 0.44, and the mean backlog is 2.17 rather than divergent. At CR = 1.2, they are 0.56 and 0.31, with a mean backlog of 2.72. **In this variant, the boundary** CR = 1 **has no special status.** What degrades instead is the share of candidate commitments that are committed with no review at all,

which rises smoothly from 1.3% at CR = 0.3 to 17% at the boundary and 24% at CR = 1.2. The dashed curve in panel (a) of Figure 3 is this recoverability, and Supplementary S2.7 gives the derivation and the simulation check. The divergence of the backlog and the collapse of recoverability are therefore properties of the lossless pipeline, not of congestion as such.

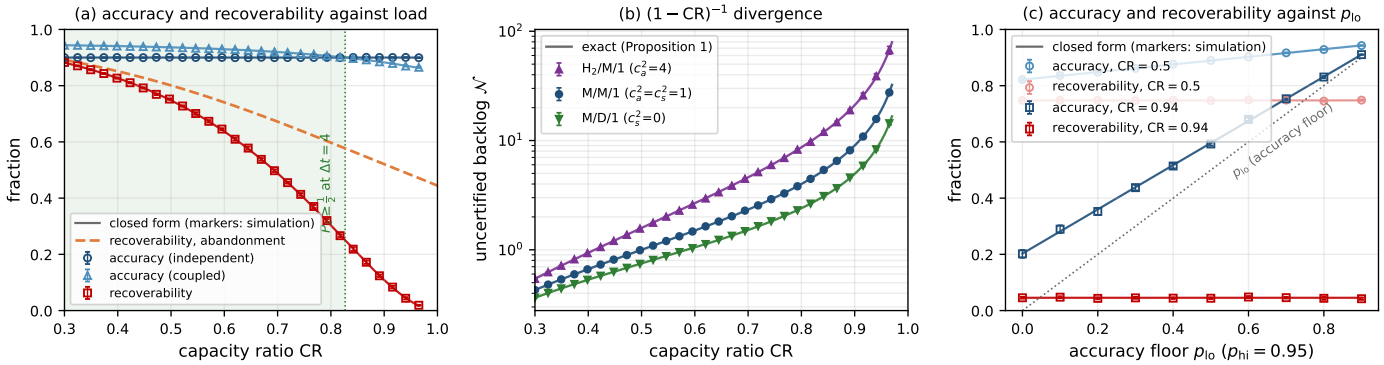


Figure 3. A single certifier of finite service rate driven toward the feasibility boundary, at $\mu = 1$ and $\Delta t = H = 4$ with the protocol of Table 2: (**a**) accuracy and recoverability against load, markers from simulation and lines from Equations (9) and (10), with the exact deadline-abandonment curve dashed; (**b**) the uncertified backlog for three arrival and service laws, markers from simulation and lines from Proposition 1; and (**c**) the same two quantities against the accuracy floor p_{lo} at two loads, under the same convention.

The order of service. Everything above is first come, first served. For non-preemptive work-conserving disciplines that ignore service times, the mean waiting time is the same [39], so $\mathcal{N}$ is invariant to these discipline changes with the arrival and service laws fixed; the distribution of sojourn times varies with the discipline, so the deadline probability depends on the order of service as well. Running the same arrival and service streams under last come, first served leaves the mean backlog unchanged at CR = 0.94 (15.14 against 15.15 over twelve replications) while in-time certification rises from 0.215 to 0.692, because it serves recent arrivals first and delays earlier ones, including those still within their horizons, which are the commitments a late certification can no longer make recoverable. A pipeline can therefore raise its in-time certification without lowering R_{self}, by changing the order of service or by dropping expired work. Reordering service preserves the stationary mean backlog under the assumptions above; dropping expired work reduces it, as the variant above shows.

A further qualification belongs to the closed-loop companion paper and lies outside the open queue treated here. When uncertified output is consumed again as further load, the crossing becomes first-order and can occur before CR = 1 [11], whereas in the open queue, the approach to the boundary is continuous. The margin also differs from model-free early-warning statistics, which read rising variance and autocorrelation near a bifurcation [45,46]. In the M/M/1 model, the two measured rates determine the deadline probability through a closed-form law. The same model predicts that the backlog relaxes more slowly as the arrival rate approaches the service capacity, so in this account, the slowing down follows from the two measured rates rather than being read off the backlog series itself. The two accounts can be compared on data. Hold C_{self} constant and let R_{self} rise. The M/M/1 model then predicts rising autocorrelation in the backlog. A model-free indicator would read that rise as an approaching bifurcation, while the margin attributes it to the measured rise in load.

https://doi.org/10.3390/psf2026016002

7. Numerical Demonstration

This section states the outcome model in closed form, shows its consequences, and compares the simulated queue with Proposition 1. The two kinds of comparison are different in status. Panels (a) and (c) of Figure 3 do not test the ordering of accuracy and recoverability, because that ordering follows analytically from the outcome model defined below; they display the resulting curves and check that the implementation reproduces them. Panel (b) is a numerical benchmark, in which simulated backlogs are compared with the exact results of Proposition 1 for three arrival and service distributions. The simulated system is the pipeline of Figure 1. Candidate commitments arrive at rate λ and are certified by a single server of rate μ, so the capacity ratio is CR $= \lambda/\mu$, and waiting times follow the Lindley recursion of Equation (4). A commitment is certified if its sojourn time (waiting plus service) is at most the certification horizon Δt. Otherwise, it is committed uncertified. It is recoverable if it is certified and a corrective pass could also finish within an option-loss window H. Recoverability is thus the availability of an opportunity for correction. It is scored for every commitment, whether or not the commitment turns out to be wrong. It therefore measures a capacity for correction and ignores correctness. It is not the fraction of wrong commitments that were recovered. The corrective pass is drawn from the stationary sojourn distribution, independently of the first pass, and adds no load, which is an idealization of the construction.

The outcome model in closed form. Let $F_T(x) = \Pr(T \le x)$ be the stationary sojourn distribution of the queue, so that $P_c = F_T(\Delta t)$. The corrective pass is an independent draw from the same distribution and adds no load, so recoverability is the product

$$P_{\text{rec}} = F_T(\Delta t)\, F_T(H), \tag{9}$$

and accuracy in the coupled model is the mixture

$$\text{acc} = p_{\text{lo}} + (p_{\text{hi}} - p_{\text{lo}})\, F_T(\Delta t), \tag{10}$$

while accuracy in the independent model is p_{acc} at every load. For the M/M/1 queue, Proposition 1(d) gives $F_T(x) = 1 - e^{-\mathcal{M}x}$ with $\mathcal{M} = \mu - \lambda$, so that $P_{\text{rec}} = (1 - e^{-\mathcal{M}\Delta t})(1 - e^{-\mathcal{M}H})$, which at $\Delta t = H$ is $(1 - e^{-\mathcal{M}\Delta t})^2$. Equations (9) and (10) determine every curve in panels (a) and (c), where they are drawn as lines through the simulated points.

Two consequences follow from this algebra alone. Accuracy is at least p_{lo} whenever $p_{\text{hi}} \ge p_{\text{lo}}$, because Equation (10) is a mixture of the two correctness probabilities; this is a statement about expectations, not a guarantee for a finite sample. Recoverability is formed from the sojourn times alone, so it is independent of p_{hi} and p_{lo}. The separation of the two quantities under load is therefore a property of the outcome model, and the runs in panels (a) and (c) illustrate it rather than establish it. The separation holds for a stationary subcritical queue, the stated corrective-pass law and fixed horizons. It does not extend to workflows with abandonment, feedback or horizon rules that change with load.

Correctness of belief is drawn under two models. In the independent model, a commitment is correct with a fixed probability $p_{\text{acc}} = 0.9$, which leaves congestion as the only cause of failure. In the coupled model, a commitment that is certified within Δt is correct with probability p_{hi}, and one that times out is correct with probability p_{lo}. Accuracy then depends on the same queue, and p_{lo} is its floor. Panel (c) of Figure 3 varies p_{lo} across its range at fixed p_{hi}.

The statistical protocol is controlled. Panels (a) and (c) use the M/M/1 queue. Each estimate discards a warm-up of 5000 commitments, because the queue starts empty and relaxes slowly near CR $= 1$, and each load point uses 12 independent replications of 4×10^4 commitments. Panel (b) compares three queues with their exact curves close to the boundary, where a run of that length is too short to give reliable estimates, because the

backlog relaxes slowly there. It therefore uses 12 replications of 4×10^5 commitments with a warm-up of 5×10^4. Every panel reports 95% Student-t confidence intervals over the replications (11 degrees of freedom). A run-length check at 48 replications compares the two protocols against the exact values (Supplementary S2.8). For the deadline fraction of panels (a) and (c), the exact M/M/1 value lies inside the replication interval of both protocols at the four loads tested, CR = 0.5, 0.793, 0.94 and 0.965 at $\Delta t = 4$, and the two protocol means differ by less than the shorter protocol's own half-width at all four; what the longer protocol buys there is an interval about three times narrower. The check covers those four loads at $\Delta t = 4$. For the uncertified backlog the shorter runs are both noisier and more erratic near the boundary: at CR = 0.965, the spread over replications for H_2/M/1 is about 34 in $\mathcal{N}$ units against 8 at the longer protocol, which is how a twelve-replication mean of 47 ± 8 arises against an exact 67.5. That is why panel (b) alone uses the longer runs. Table 2 collects the parameters.

Table 2. Simulation parameters for Figure 3. The quantities reported are *accuracy*, the fraction of commitments whose content is correct; *recoverability*, the fraction of all commitments that are certified within Δt and whose corrective pass would also finish within H; $\mathcal{N}$, the mean number of uncertified commitments; and P_c, the fraction certified within Δt.

Parameter	Value	Meaning
μ	1	certification rate; it fixes the time unit
Δt	4	certification horizon
H	4	option-loss window
CR grid	28 points in $[0.30, 0.965]$	load
replications	12	independent runs at every point
run length	4×10^4; 4×10^5 in (b)	commitments per replication
warm-up	5×10^3; 5×10^4 in (b)	discarded before any estimate
p_{acc}	0.90	correctness probability, independent model
p_{hi}	0.95	correctness when certified in time
p_{lo}	0.85; $[0, 0.9]$ in (c)	correctness when the deadline is missed
c_a^2	4 for H_2/M/1, 1 otherwise	squared coefficient of variation, interarrival times
c_s^2	0 for M/D/1, 1 otherwise	squared coefficient of variation, certification times

Claude (Anthropic) and Codex (OpenAI) assisted the author in drafting and editing text and in checking derivations and analysis code. The author reviewed the assisted outputs.

Panel (a) of Figure 3 shows Equations (9) and (10) against the simulated points, and the two agree to within replication noise at every load, which is the implementation check this panel supplies. Its shaded region marks the loads at which the first-pass probability $P_c = 1 - e^{-\mathcal{M}\Delta t}$ is at least $\frac{1}{2}$; since $P_c \geq \frac{1}{2}$ is $\mathcal{M} \geq \ln 2/\Delta t$, the boundary is CR $\leq 1 - \ln 2/(\mu\Delta t) = 0.827$ at $\mu\Delta t = 4$. The plotted recoverability also requires a corrective pass and is the smaller, two-pass quantity. Under the independent model, accuracy stays flat at 0.900 ± 0.001 while recoverability falls from 0.75 at CR = 0.5 to 0.044 ± 0.004 at CR = 0.94. In this model, congestion alone, with the quality of belief untouched, removes the opportunity to undo. At the reference coupling ($p_{\mathrm{hi}} = 0.95$, $p_{\mathrm{lo}} = 0.85$), accuracy drifts from 0.944 to 0.870 as CR rises from 0.3 to 0.94. Over the same range, recoverability falls from 0.88 to 0.044. This is a twenty-fold collapse against a drift of seven percentage points.

Panel (c) varies the coupling at CR = 0.5 and CR = 0.94, with coupled accuracy bounded below by the dotted diagonal p_{lo}. The flatness of recoverability in p_{lo} that it displays is imposed by the scoring rule rather than discovered, since the recoverability statistic never reads the correctness draws, and the lower bound on accuracy is Equation (10). What the panel adds is the sampling variability around those two expressions. The sweep covers $p_{\mathrm{lo}} \in [0, 0.9]$ at fixed $p_{\mathrm{hi}} = 0.95$. Coupled accuracy follows its floor, acc $\to P_c\, p_{\mathrm{hi}} + (1 - P_c)\, p_{\mathrm{lo}} \geq p_{\mathrm{lo}}$, and ranges over $[0.20, 0.91]$ at CR = 0.94, while recoverability is unchanged by the coupling

to within replication noise (values in Supplementary S2). Accuracy can be made to vanish at the boundary only by sending $p_{\mathrm{lo}} \to 0$, that is, by defining every timed-out commitment as wrong. Even then, the two quantities vanish at different orders in the stationary limit. Accuracy vanishes linearly in the margin, $\mathrm{acc} \to P_c\, p_{\mathrm{hi}} = O(1 - \mathrm{CR})$, whereas recoverability needs two passes to finish within their deadlines and vanishes quadratically, as $O((1-\mathrm{CR})^2)$, so recoverability collapses faster. For a fixed $p_{\mathrm{lo}} > 0$, accuracy is bounded below by p_{lo} while recoverability is set by load alone and falls toward zero. These are relative orders in the stationary limit. They do not mark a separate, earlier threshold at which recoverability fails, and the finite runs illustrate them without proving them.

Panel (b) is the one comparison in this section that goes beyond the closed forms above, because it checks the simulated queue against Proposition 1 rather than against the outcome model. Its markers are simulated backlogs and its lines are the exact results, Equation (6) for M/M/1 and M/D/1 and Proposition 1(d) for H_2/M/1. For all three queues, namely, M/M/1, M/D/1 with deterministic service and H_2/M/1 with hyperexponential arrivals of $c_a^2 = 4$, the simulated backlog lies on the exact curves of Proposition 1. The exact value lies inside the 95% confidence interval at 76 of the 84 load points, and the largest relative difference is 2.2%. The eight misses are at low and middle loads: CR = 0.448, 0.497, 0.620 and 0.768 for M/M/1, 0.645, 0.669, 0.719 and 0.891 for M/D/1, and none for H_2/M/1. Their relative gaps are between 0.2% and 1.0%, against replication intervals a few parts in a thousand wide, and none is at CR $\geq$ 0.90, where the intervals are wide enough to contain the exact value at every load (Supplementary S2.3). Separating a finite-run bias from the variability of the intervals themselves would take a coverage study of its own, so the misses are reported here as locations and gaps. All three curves diverge with the same $(1 - \mathrm{CR})^{-1}$ exponent and differ in amplitude. Averaged over CR $\geq$ 0.90, the amplitude $(1 - \mathrm{CR})\ \mathcal{N}$ is 0.94 for M/M/1, 0.50 for M/D/1 and 2.28 for H_2/M/1 in simulation, against 0.94, 0.50 and 2.26 in the same order from Proposition 1. The boundary limits of part (e) are 1, 0.5 and 2.5 (Supplementary S2). The collapse of recoverability is also not an artefact of the horizon. At $\Delta t = H = 2$, 4 and 8, recoverability falls thirty-fold, seventeen-fold and seven-fold, respectively, between CR = 0.5 and 0.94, so the collapse softens as the horizon grows and persists at every horizon tried (Supplementary S2).

Recovering the parameters from an event log. Sections 3 and 6 are connected by one further check. The estimator of Figure 1 was applied to event logs drawn from the model itself, so that the generating values are known. A separate event-driven run records the arrival, the start of service and the completion of every commitment, and the four quantities are then estimated from those records alone: the arrival rate from arrivals per unit of elapsed time, the service rate from completions per unit of server-busy time, the uncertified backlog by integrating the number in system over the window rather than through Little's law, and the deadline-success probability from the observed sojourn times. Work that straddles either end of the window is counted only for the part inside it. At loads 0.5, 0.7, 0.85 and 0.94, with twelve replications of 4×10^5 commitments and a warm-up of 5×10^4, all sixteen generating values of the M/M/1 queue lie inside the 95% interval over replications. The largest relative difference is 0.5%, and for the two rates, it is below 0.1% at every load. For the M/D/1 queue, eleven of the twelve available comparisons contain the generating value; the exception is the service rate at load 0.85, where the mean estimate differs from 1 by 7.65×10^{-7} against a replication half-width of 7.37×10^{-7}, because deterministic service makes that estimate almost exact in every replication and the interval becomes narrower than the residual effect of the window edge. No M/D/1 deadline-success benchmark is supplied in this calibration, so that queue contributes comparisons for the two rates and the backlog only.

https://doi.org/10.3390/psf2026016002

The same logs check the precision expressions of Section 6.2. Over 48 replications at each load, the standard errors predicted by Equation (7) for the margin, the capacity ratio and the implied P_c stand to the observed spread across replications at a ratio between 0.91 and 1.22 over the thirty-two available comparisons on the three queues; the largest is M/D/1 at CR = 0.85. A spread estimated from 48 replications carries about a tenth of itself in uncertainty, which is one standard error rather than an agreement band. Over the eight queue-and-load cases with stochastic service, the correlation between the two rate estimates lies between −0.16 and +0.11, and all eight 95% intervals contain zero, so the covariance term of Equation (7) is negligible for these two estimators. Deterministic service leaves the capacity estimate almost constant, so no correlation interval is available in those four cases. The standard error of the margin changes little with load, from 2.9×10^{-3} to 3.5×10^{-3} across the M/M/1 grid, while the margin itself falls eight-fold, which is why the relative precision degrades from 0.6% to 5.9% as the boundary is approached. Deterministic service removes the second term of Equation (7) and the spread of $\hat{C}_{\text{self}}$ falls by nearly three orders of magnitude, from about 2.5×10^{-3} to about 2.9×10^{-6}, which accounts for the one missed interval above. A deadline-success fraction read directly from the sojourn times is less precise than its count suggests, because sojourn times are autocorrelated: at CR = 0.94, its variance is 123 times the binomial value, so 150,000 commitments carry the information of about 1200. A moving-block bootstrap on a single log reproduces both results (Supplementary S2.6).

This establishes that the estimator recovers what it should, with the stated precision, when the data satisfy the model. It does not establish that any real workflow satisfies the model, since these logs were generated by it.

Because Equations (9) and (10) are consequences of the outcome model, no run of that model can refute them. What can be tested on a logged workflow is whether the model applies to it. Estimate the parameters on one window and test the predictions on another, and the account fails if any of the following holds. The event stream is not described by a single lossless first-come, first-served bottleneck, because work is abandoned at its deadline, served out of order, or fed back as further load. The deadline-success probability or the finite-load backlog measured on held-out windows disagrees with the value predicted from the estimated rates, beyond the stated uncertainty and the settling-time condition of Section 6.2. Commitments labelled on time and late do not keep their calibrated correctness probabilities when the load changes. Correction opportunities observed independently of the score, through logged cancellations or audits rather than through a simulated second pass, do not have the assumed duration law or are not independent of first-pass success. A mismatch of this kind rejects the model or its operational reading, and not the queueing results of Proposition 1.

8. One Application, as a Hypothesis: Decision Systems Under AI Acceleration

This section uses the framework to interpret human–AI decision systems. It reports no direct human–AI measurements, and its claims form a hypothesis that the estimator of Section 3 could test. For a person or an organization, C_{self} would be the bounded bandwidth of attention and institutional review [47]. For an automated pipeline or agent, it would be the rate at which proposed actions can be verified before commitment. The hypothesis is that AI raises R_{self}, because options and feedback arrive faster and decisions are made more often. The evidence for that step is indirect. The automation of decision tasks is documented for labour markets [1], where it describes the displacement of tasks toward the cheaper route rather than a measured rise in the arrival rate of decisions awaiting review. Direct measurement of review load under AI assistance is scarce: an industrial study of

an automated code-review assistant over 4335 pull requests reports that the average time to close a request rose from under six hours to over eight after adoption [48], which is a congestion signal rather than an estimate of the two rates. No study known to the author estimates R_{self} and C_{self} for a review pipeline. If the arrival rate does rise faster than the capacity to review, the result would be to drive CR $\to 1$ in the human, in the AI and in the coupled system.

Equation (1) then predicts a specific failure. As arrivals approach the certification capacity from below, the stationary mean backlog grows without bound under the conditions of Proposition 1(e). The opportunity to correct a commitment before it becomes permanent would then be lost, whatever the motivation, the competence or the accuracy of individual judgments. Under the lossless model, a rising load removes the opportunity for timely correction even where the quality of individual judgment is unchanged, so the erosion would be a property of the structure of the system rather than a personal failing. In this reading, an inability to keep up corresponds to CR $\to 1$, commitments made before their review is complete correspond to a rising backlog, and losing the record that links a decision to its outcome corresponds to a loss of traceability as defined in Section 5. The two boards of the Introduction supply the numbers: five percent more arrivals separates a stationary P_c of 0.63 from one of 0.05, with the quality of individual judgments unchanged. Whether a given workflow sits in either regime is a measurement question, and Section 6.2 gives the span that answering it takes. Testing the hypothesis needs a workflow in which proposal arrivals, review capacity, reversals and option-loss events are logged, which is the kind of data the estimator of Section 3 is designed for. Until it has been applied there, the mechanism is hypothesized, not shown.

9. Discussion

The results separate what has been established from what has not. The rates R_{self} and C_{self} can be estimated from event counts alone, within the scope limits of Section 3. The uncertified backlog obeys the law of Proposition 1. Its exponent is shared by the fixed-shape queue families of Proposition 1(e) that have positive variability, $c_a^2 + c_s^2 > 0$, and its amplitude depends on the distributions. For the M/M/1 queue, the ordering of accuracy and recoverability follows from the construction: accuracy is bounded below by its floor p_{lo}, while recoverability is set by load alone. These results are established here, within the lossless model of Section 6.1; Section 6.2 states what changes when expired work leaves the queue, and what precision the margin can be measured to.

Other parts are hypothesized or deferred. The clinical validity of the physiological event semantics is one. The margin-and-drift quantity as a validated early-warning predictor is another, since it is defined here, has not been tested anywhere, and carries the measurement limit of Section 6.2. The information-retention axis is defined but not exercised, with full invertibility unverified, and the human–AI application of Section 8 remains a hypothesis.

If the deferred parts hold, the framework has three uses. As an instrument, it tracks the margin M and its drift from observable counts, within the precision of Equations (7) and (8), which restrict its use to systems whose load drifts slowly against the relaxation time of their backlog. As a controller, it would shape rates. A candidate commitment can be deferred, checked again before any action is taken (repeat-first), escalated to a higher level of review, or suppressed. Used as a schedule, these four responses lower R_{self} step by step until the margin is restored, in the way an annealing schedule lowers a temperature. This is the remedy that the framework prescribes, and it is not modelled here. Other responses are open to a congested pipeline: dropping commitments that have passed their horizon, and serving the most recent arrival first, both raise in-time certification without changing either rate (Section 6.2). As a complement to uncertainty quantification, which states how

uncertain a belief is, RSC indicates whether irreversible action is admissible given the surrounding rate and capacity.

10. Conclusions

This paper models how the workload of a single certifier limits the opportunity to correct commitments before their deadlines. The comparison with Shannon's theorem concerns the form of this rate–capacity limit. The paper gives an exact identity for the certification backlog at every subcritical load, with closed forms for three arrival and service distributions, and the simulated queue agrees with them under a controlled statistical protocol. It also shows that the two parameters of the model, R_{self} and C_{self}, can be estimated from event data within stated scope limits, and that the estimator recovers the arrival rate, the service rate, the backlog and the deadline-success probability from event logs generated by the model itself. Within that model, recoverability falls toward zero under load while accuracy keeps its floor p_{lo}; this separation is derived from Equations (9) and (10) rather than established by measurement. Two results limit the proposal, and they are the paper's main findings. The first is a limit on measurement: the counting standard error of the margin hardly moves with load while the margin itself shrinks, so the observation span needed to resolve it grows as M^{-2}, runs to centuries for the review board of the Introduction, and is usable only if the rates stay constant across it and the stationary model applies. The second is a limit on scope: the collapse of recoverability belongs to a pipeline that keeps expired work; where a commitment that misses its horizon leaves the queue instead, the backlog stays bounded at every load and CR = 1 has no special status. What remains open is empirical. No logged workflow has yet been shown to satisfy the model, no real correction opportunity has been measured, and the early-warning use of the margin has no demonstrated lead time or false-alarm behaviour; the tests that would settle each are stated at the end of Section 7. The companion physics paper [11] develops the nonequilibrium thermodynamic basis and the first-order collapse of the closed loop. Optimal control of the arrival rate and the application of the estimator to logged decision workflows remain open [6].

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/psf2026016002/s1: Table S1: Analysis details for the two physiological cohorts; Section S2: Simulation protocol and robustness details (S2.1 Protocol; S2.2 Panel (a): values at the loads quoted in the text; S2.3 Panel (b): simulated backlog against the exact results of Proposition 1; S2.4 Panel (c): coupling sweep; S2.5 Horizon robustness; S2.6 Precision of the margin estimate; S2.7 Deadline abandonment and the order of service; S2.8 Run-length check); Section S3: Supplements to the proof of Proposition 1 (S3.1 Stationarity, finite moments and the identity; S3.2 Idle-time conventions and the Poisson case; S3.3 The heavy-traffic family and the limit of means; S3.4 The GI/M/1 law and the simulated H_2/M/1 root; S3.5 A deadline bound under an exponential-moment condition). The eight references cited in the Supplementary Materials [8,36,38,40–43] are all cited in the main text as well.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable: the study analysed only openly released, de-identified data and involved no human subjects.

Informed Consent Statement: Not applicable.

Data Availability Statement: Source code for the analyses and figures, with the derived Synthea estimates, is in the project repository [49], at the tagged revision `proceedings-v3.2` (commit `1a6048460800d8e5bbc0d96fb1aae7fbe53b21e8`; MIT license), archived at Zenodo under DOI `10.5281/zenodo.23333341`. The exact execution environment is recorded at the tag in `requirements-lock.txt` (Python 3.14.2, NumPy 2.4.1, Matplotlib 3.10.8, pandas 3.0.0); `requirements.txt` gives minimum-version floors; run commands are in the per-directory `README` files. The simulation

(Figure 3) uses no external data and fixed seeds. Physiological inputs are not redistributed: the Synthea synthetic cohort [14] and the MIMIC-IV Clinical Database Demo v2.2 [15,16] are openly licensed. The `README` gives the Synthea download command. That bundle is not versioned by its publisher, so it is identified here by content: `10k_synthea_covid19_csv.zip`, 56,851,927 bytes, SHA-256 `559757dc849f4361a328f456d2c0a20c6df72419068321c753c6be787161e937`, file dates 2020-05-26 inside the archive. The MIMIC-IV demo script downloads its three input tables (analysis runs of 14 July 2026); the derived per-unit estimates shipped with the repository carry SHA-256 `14ded000f45fed42f4647193bea8d3f519ad6597bff76f17c0027ecf4401e336`.

Acknowledgments: A fuller development of the framework is given in the foundational preprint [6]. During the preparation of this manuscript, the author used Claude (Anthropic; Claude Code with the Claude Opus 4.8, Opus 5, Opus 5.5, Fable 5 and Fable 5.1 models, June–October 2026) and Codex (OpenAI; Codex CLI with the GPT-6 Astra and GPT-6.1 Sol models, September–October 2026) for the purposes of drafting and editing text under the author's direction and checking the derivations and the code. The author has reviewed and edited the output and takes full responsibility for the content of this publication.

Conflicts of Interest: The author is an entrepreneur and inventor involved in the development of artificial intelligence and computational systems; these activities did not influence the analysis or conclusions presented in this work. The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	artificial intelligence
CR	capacity ratio, $R_{\text{self}}/C_{\text{self}}$
FCFS	first come, first served
i.i.d.	independent and identically distributed
MIMIC-IV	Medical Information Mart for Intensive Care, version IV
ROC	receiver operating characteristic
RSC	Recoverable Self-Coding
M/M/1 and similar	single-server queues in Kendall's notation, written as interarrival law / service law / number of servers: M/M/1, M/D/1, M/G/1, H_2/M/1, GI/M/1 and GI/GI/1, where M is exponential, D deterministic, H_2 hyperexponential, and G or GI general (Section 6.1)

The following symbols are used in this manuscript and the supplementary proof:

R, C	information rate and capacity of a channel in Shannon's theorem
R_{self}	induced flux: the rate at which candidate commitments arrive for certification
C_{self}	integrative capacity: the rate at which candidate commitments can be certified
$\hat{R}_{\text{self}}$, $\hat{C}_{\text{self}}$	estimates of R_{self} and C_{self} from an event log (Figure 1)
CR	capacity ratio, $R_{\text{self}}/C_{\text{self}}$; equal to λ/μ in the queue model
$\mathcal{M}$, $\dot{\mathcal{M}}$	margin, $C_{\text{self}} - R_{\text{self}}$, and its rate of change in time
$\mathcal{N}$	uncertified backlog: the mean number of uncertified commitments, $\langle N_{\text{uncert}} \rangle$
$N_{\text{uncert}}(t)$	number of uncertified commitments at time t: those waiting plus the one in certification
$\langle \cdot \rangle$	equilibrium time mean in the definition of $\mathcal{N}$
t	time
Δt	certification horizon: the deadline for certification
H	option-loss window: the deadline for the corrective pass
λ	arrival rate of candidate commitments
μ	certification (service) rate
$\bar{w}$	uncertainty quantum: the number of bits to be resolved for each commitment

n, j, k, m, r	integer indices or finite summation limits; n indexes a commitment and r an idle-time moment
A_n, A	time between two successive arrivals, and its stationary version
S_n, S	certification (service) time of commitment n, and its stationary version
W_n, W, W'	waiting time of commitment n before its certification starts, its stationary version, and the waiting time of the next commitment
U_n, U	$S_n - A_{n+1}$, and its stationary version
η_n, η	idle time within the interval ending at arrival n, and its stationary version, including zeros
η_+, q	idle time conditional on being positive, and $q = \Pr(\eta > 0)$ (Supplementary S3.2)
X	$W + U$
d, m_2	$-\mathbb{E}[U] = (1 - \mathrm{CR})/\lambda > 0$, and $\mathbb{E}[U^2]$
V_n	waiting-time recursion started empty, used to establish integrability (Supplementary S3.1)
Y_k	sum of the first k increments U_j, with $Y_0 = 0$ (Supplementary S3)
$K, f_K(x)$	positive truncation level, and $f_K(x) = \min(x^2, K)$
B_n, B	scaled interarrival time λA, with a fixed law of mean one in Proposition 1(e)
J_λ	scaled idle increment $\lambda\eta$ in the heavy-traffic family
Q_n, Q, Q'	count found by arrival n, its stationary version, and the count found by the next arrival (GI/M/1)
L_n, L	potential service-clock ticks in an interarrival interval, and their common law (Supplementary S3.4)
N	$N_{\mathrm{uncert}}(0)$ sampled at an equilibrium time (Supplementary S3.4)
$F(z)$	$\mathbb{E}[e^{-\mu(1-z)A}] - z$, whose interior zero is σ
z, s	root-equation argument and nonnegative Laplace-transform argument
p_1, p_2, r_1, r_2	mixture probabilities and exponential rates of the H_2 interarrival law (Supplementary S3.4)
T	sojourn time, $W + S$: waiting plus certification
$F_T(x)$	stationary distribution of the sojourn time, $\Pr(T \le x)$; $F(z)$ denotes a different function
T_{corr}	sojourn time of the corrective pass (Figure 1)
D_{exc}	duration of a physiological excursion (Section 3)
c_a^2, c_s^2	squared coefficients of variation of the interarrival and certification times
σ	root in $(0, 1)$ of the GI/M/1 equation of Proposition 1(d)
$\theta^\star$	decay rate of the sojourn-time tail
θ, Λ_n	positive parameter with $\mathbb{E}[e^{\theta U}] \le 1$, and the supermartingale $e^{\theta Y_n}$ (Supplementary S3.5)
P_c	probability that a commitment is certified within Δt (first pass)
P_{rec}	recoverability: the fraction of commitments certified within Δt whose corrective pass also finishes within H
$\tau_{\mathrm{queue}}, \tau_{\mathrm{load}}$	relaxation time of the backlog, and the time scale on which the load drifts
τ_{obs}	observation span: the length of the window from which the two rates are estimated
τ_{pat}	deterministic patience: the wait after which a candidate commitment leaves the queue (Supplementary S2.7)
ε	relative precision of the margin estimate: a half-width of $\varepsilon\mathcal{M}$
δE	perturbation of the environment
Z	certified macrostate
$I(\delta E; Z \mid \text{pre-commit})$	mutual information between δE and Z before commitment
ι	positive lower bound on that mutual information
p_{acc}	probability that a commitment is correct in the independent model

https://doi.org/10.3390/psf2026016002

$p_{\mathrm{hi}}, p_{\mathrm{lo}}$	probability that a commitment is correct when it is certified in time, and when it times out (coupled model); p_{lo} is the accuracy floor
acc	accuracy: the fraction of commitments that are correct
x, u, v, a, c	real arguments or constants in the proof; $v, a \geq 0$ and $c > 0$ where used
x^+	$\max(0, x)$, the positive part of a real number
$\mathbb{E}[\,\cdot\,]$, $\mathrm{Var}(\cdot)$, $\mathrm{Cov}(\cdot,\cdot)$, $\Pr(\cdot)$	expectation, variance, covariance and probability
$\mathbb{E}[Y;\, \mathcal{E}]$	expectation of a random variable Y restricted to an event $\mathcal{E}$
$O(\cdot)$	order of magnitude as CR $\to 1$

References

1. Acemoglu, D.; Restrepo, P. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *J. Econ. Perspect.* **2019**, *33*, 3–30. [CrossRef]
2. Rasmussen, J. Risk management in a dynamic society: A modelling problem. *Saf. Sci.* **1997**, *27*, 183–213. [CrossRef]
3. Perrow, C. *Normal Accidents: Living with High-Risk Technologies*; Basic Books: New York, NY, USA, 1984.
4. Reason, J. *Human Error*; Cambridge University Press: Cambridge, MA, USA, 1990. [CrossRef]
5. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
6. van Rooyen, P. Entropy, Annealing, and the Continuity of Agency in Human–AI Systems. *Preprint* **2026**. [CrossRef]
7. Shu, W.; Wei, P. AI Safety as Control of Irreversibility: A Systems Framework for Decision-Energy and Sovereignty Boundaries. *arXiv* **2026**. [CrossRef]
8. Loynes, R.M. The stability of a queue with non-independent inter-arrival and service times. *Math. Proc. Camb. Philos. Soc.* **1962**, *58*, 497–520. [CrossRef]
9. Kingman, J.F.C. The single server queue in heavy traffic. *Math. Proc. Camb. Philos. Soc.* **1961**, *57*, 902–904. [CrossRef]
10. Kelly, F.P. Notes on effective bandwidths. In *Stochastic Networks: Theory and Applications*; Kelly, F.P., Zachary, S., Ziedins, I., Eds.; Royal Statistical Society Lecture Notes Series; Oxford University Press: Oxford, UK, 1996; Volume 4, pp. 141–168. [CrossRef]
11. van Rooyen, P. First-Order Recoverability Collapse in Self-Referential Information Decoders: The Operating Loop of an AI System as a Driven Nonequilibrium Steady State. *arXiv* **2026**. [CrossRef]
12. Chang, C.S. Stability, queue length, and delay of deterministic and stochastic queueing networks. *IEEE Trans. Autom. Control* **1994**, *39*, 913–931. [CrossRef]
13. Wu, D.; Negi, R. Effective capacity: A wireless link model for support of quality of service. *IEEE Trans. Wirel. Commun.* **2003**, *2*, 630–643. [CrossRef]
14. Walonoski, J.; Kramer, M.; Nichols, J.; Quina, A.; Moesel, C.; Hall, D.; Duffett, C.; Dube, K.; Gallagher, T.; McLachlan, S. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *J. Am. Med Inform. Assoc.* **2018**, *25*, 230–238. [CrossRef] [PubMed]
15. Johnson, A.; Bulgarelli, L.; Pollard, T.; Horng, S.; Celi, L.A.; Mark, R. MIMIC-IV Clinical Database Demo (version 2.2). *PhysioNet* **2023**. [CrossRef]
16. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [CrossRef] [PubMed]
17. Rockafellar, R.T.; Uryasev, S. Optimization of Conditional Value-at-Risk. *J. Risk* **2000**, *2*, 21–41. [CrossRef]
18. Dixit, A.K.; Pindyck, R.S. *Investment Under Uncertainty*; Princeton University Press: Princeton, NJ, USA, 1994.
19. Ben-Haim, Y. *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*, 2nd ed.; Academic Press: Oxford, UK, 2006. [CrossRef]
20. Ames, A.D.; Coogan, S.; Egerstedt, M.; Notomista, G.; Sreenath, K.; Tabuada, P. Control Barrier Functions: Theory and Applications. In Proceedings of the 2019 18th European Control Conference (ECC), Naples, Italy, 25–28 June 2019; pp. 3420–3431. [CrossRef]
21. Russell, S. *Human Compatible: Artificial Intelligence and the Problem of Control*; Viking: New York, NY, USA, 2019.
22. Bronson, N.; Aghayev, A.; Charapko, A.; Zhu, T. Metastable Failures in Distributed Systems. In Proceedings of the Workshop on Hot Topics in Operating Systems (HotOS '21), Ann Arbor, MI, USA, 1–3 June 2021; pp. 221–227. [CrossRef]
23. Barrer, D.Y. Queuing with Impatient Customers and Ordered Service. *Oper. Res.* **1957**, *5*, 650–656. [CrossRef]
24. Baccelli, F.; Boyer, P.; Hébuterne, G. Single-server queues with impatient customers. *Adv. Appl. Probab.* **1984**, *16*, 887–905. [CrossRef]
25. Zeltyn, S.; Mandelbaum, A. Call Centers with Impatient Customers: Many-Server Asymptotics of the M/M/n+G Queue. *Queueing Syst.* **2005**, *51*, 361–402. [CrossRef]
26. Doytchinov, B.; Lehoczky, J.; Shreve, S. Real-time queues in heavy traffic with earliest-deadline-first queue discipline. *Ann. Appl. Probab.* **2001**, *11*, 332–378. [CrossRef]

27. Kaul, S.; Yates, R.; Gruteser, M. Real-time status: How often should one update? In Proceedings of the 2012 Proceedings IEEE INFOCOM, Orlando, FL, USA, 25–30 March 2012, pp. 2731–2735. [CrossRef]
28. Yates, R.D.; Sun, Y.; Brown, D.R.; Kaul, S.K.; Modiano, E.; Ulukus, S. Age of Information: An Introduction and Survey. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 1183–1210. [CrossRef]
29. Cummings, M.L.; Mitchell, P.J. Predicting Controller Capacity in Supervisory Control of Multiple UAVs. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2008**, *38*, 451–460. [CrossRef]
30. Parasuraman, R.; Riley, V. Humans and Automation: Use, Misuse, Disuse, Abuse. *Hum. Factors* **1997**, *39*, 230–253. [CrossRef]
31. Skitka, L.J.; Mosier, K.L.; Burdick, M. Does automation bias decision-making? *Int. J. Hum.-Comput. Stud.* **1999**, *51*, 991–1006. [CrossRef]
32. Heitz, R.P. The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Front. Neurosci.* **2014**, *8*, 150. [CrossRef] [PubMed]
33. van Rooyen, P. Entropy, Annealing, and the Continuity of Agency in Human–AI Systems. In Proceedings of the Entropy 2026: Exploring Complexity and Information in Science, Barcelona, Spain, 1–3 July 2026; Abstract.
34. van Rooyen, P. Entropy, Capacity, and the Continuity of Agency in Human–AI Systems. Poster presented at Entropy 2026: Exploring Complexity and Information in Science, Barcelona, Spain, 1–3 July 2026. Available online: https://github.com/Pietervr/recoverable-self-coding/tree/proceedings-v3.2/poster (accessed on 27 October 2026).
35. Landauer, R. Irreversibility and Heat Generation in the Computing Process. *IBM J. Res. Dev.* **1961**, *5*, 183–191. [CrossRef]
36. Lindley, D.V. The theory of queues with a single server. *Math. Proc. Camb. Philos. Soc.* **1952**, *48*, 277–289. [CrossRef]
37. Little, J.D.C. A Proof for the Queuing Formula: L = λW. *Oper. Res.* **1961**, *9*, 383–387. [CrossRef]
38. Kingman, J.F.C. Some Inequalities for the Queue GI/G/1. *Biometrika* **1962**, *49*, 315–324. [CrossRef]
39. Asmussen, S. *Applied Probability and Queues*, 2nd ed.; Stochastic Modelling and Applied Probability; Springer: New York, NY, USA, 2003. [CrossRef]
40. Núñez Queija, R. *Note on the GI/GI/1 Queue with LCFS-PR Observed at Arbitrary Times*; Technical Report PNA-R0009; Centrum voor Wiskunde en Informatica: Amsterdam, The Netherlands, 2000.
41. Marshall, K.T. Some Inequalities in Queuing. *Oper. Res.* **1968**, *16*, 651–668. [CrossRef]
42. Chen, Y.; Whitt, W. Algorithms for the Upper Bound Mean Waiting Time in the GI/GI/1 Queue. *Queueing Syst.* **2020**, *94*, 327–356. [CrossRef]
43. Kingman, J.F.C. A martingale inequality in the theory of queues. *Math. Proc. Camb. Philos. Soc.* **1964**, *60*, 359–361. [CrossRef]
44. Glynn, P.W.; Whitt, W. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *J. Appl. Probab.* **1994**, *31A*, 131–156. [CrossRef]
45. Scheffer, M.; Bascompte, J.; Brock, W.A.; Brovkin, V.; Carpenter, S.R.; Dakos, V.; Held, H.; van Nes, E.H.; Rietkerk, M.; Sugihara, G. Early-warning signals for critical transitions. *Nature* **2009**, *461*, 53–59. [CrossRef] [PubMed]
46. Dakos, V.; Carpenter, S.R.; van Nes, E.H.; Scheffer, M. Resilience indicators: Prospects and limitations for early warnings of regime shifts. *Philos. Trans. R. Soc. B* **2015**, *370*, 20130263. [CrossRef]
47. Simon, H.A. Designing Organizations for an Information-Rich World. In *Computers, Communications, and the Public Interest*; Greenberger, M., Ed.; The Johns Hopkins Press: Baltimore, MD, USA, 1971; pp. 37–72.
48. Cihan, U.; Haratian, V.; İçöz, A.; Gül, M.K.; Devran, Ö.; Bayendur, E.F.; Uçar, B.M.; Tüzün, E. Automated Code Review In Practice. *arXiv* **2024**, arXiv:2412.18531. [CrossRef]
49. van Rooyen, P. *Recoverable Self-Coding: Figure-Generation and Analysis Code*; Zenodo: Geneva, Switzerland, 2026. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.