Governing Near Environmental Criticality: The Triple Bind and the Turn to Act-Regarding Restraint

Wanhong Huang
huangwanhong@serendip.ngo

Abstract

A governance arrangement approaching a critical threshold in the system it governs is in a situation this paper argues is structurally worse than the sum of its difficulties. Three quantities on which such an arrangement depends degrade as the threshold nears, and they are not three independent risks to be managed separately: they are monotone functions of a single approach parameter, so they degrade together and their unfavourable values coincide exactly where governance most needs them. Detection reliability falls, because the statistical signatures on which early warning depends become harder to estimate as the system's recovery slows. The margin within which an intervention is safe shrinks, because the distance to the boundary of the occupied basin is what shrinks as a threshold nears. And estimation of the system's position degrades, in a manner the paper argues is worse than mere imprecision: near-unit-root estimation of persistence is biased downward, and because the recovery rate is a logarithm of persistence, a small downward bias becomes a large overstatement of stability precisely where persistence approaches one. The paper reports a simulation confirming that the same bias which overstates stability by five per cent far from a threshold overstates it nearly threefold near one, so the error runs toward apparent safety and grows as safety diminishes. The governance consequence is the paper's central claim. The standard institutional response to uncertainty, which is to observe more closely, act more cautiously in light of what is observed, and learn from the response, has all three of its legs cut at once by the triple bind, since observing, acting, and learning degrade together and in the same direction. The paper argues that near criticality an arrangement must therefore transfer weight from inputs that describe the system to inputs that describe the proposed act, principally its reversibility, because a property of an act does not degrade as the system approaches a threshold. This is the sense in which the burden criterion of the series' first paper is not merely applicable near criticality but is the only governing input that survives it. The paper surveys the relevant statistical, dynamical, decision-theoretic and institutional literatures, states what each owns, and identifies the residue. It reports no field study.

Keywords: critical transitions; deep uncertainty; estimation bias; precaution; irreversibility

Notices

On the references. All 27 works cited in this paper have been checked against their sources at claim level: each source was consulted and confirmed to state what the citing sentence attributes to it. No entry in this paper's bibliography is unverified.
Status. This is a working draft circulated for discussion. It is a preliminary discussion paper and is not a finished statement of its author's position. Sections and statement numbers are subject to change.

Licence. This work is made available under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0).

Statement on the use of language models. Drafting, literature search and argumentative criticism for this paper were conducted in dialogue with large language models, specifically Claude (Anthropic) and ChatGPT (OpenAI). The claims, the structure, the selection of material and the position taken are the author's. References have been checked in stages, to two standards, and the process is incomplete. Part of the corpus has been verified at identity level, meaning author, title and publication coordinates were confirmed against the publisher of record. A smaller part has additionally been verified at claim level, meaning the cited source was consulted and confirmed to state what the citing sentence attributes to it. The remainder is unverified at either standard. The standard reached is marked on each bibliography entry: no mark for claim level, ‡ for identity level only, † for not yet verified. Corrections made so far are recorded in the project's citation-verification file; where a source proved not to bear the weight the text placed on it, the text was revised rather than the citation removed.

Companion papers. This paper is Paper 10 of a series on the governance of environmental change, and treats the joint degradation of detection, margin and estimation near a threshold, and the turn to act-regarding restraint. The other papers of the series treat, in order: the object; the field and its three structures; coordination through change; heterogeneous knowledge; the event layer; criticality detection; observational capacity; representational requirements; routing and inquiry; forecasting from the stream; the knowledge commons; attribution and ownership; responsibility at transitions; and supervisory duties. Paper 12 is reserved and is not yet drafted. A related paper by the author, developed independently of this series, treats the temporal correspondence limitation of evidence-based policy making, being the divergence between the system state that evidence represents and the system state at the moment of decision. Its results on the partition of a state space into estimable and unobservable components, on the blindness of a recency-weighted evidence base to the approach to criticality, and on the allocation of a justificatory burden where temporal correspondence fails, were reached by a different route and bear on several papers of this series.

Suggested citation. Huang, W. Governing Near Environmental Criticality: The Triple Bind and the Turn to Act-Regarding Restraint. Working draft.

Discussion Paper Note

This paper is a preliminary discussion paper intended to share an evolving idea and invite further dialogue, criticism, revision, and independent development.

The author does not claim exclusive epistemic ownership over the viewpoints, concepts, or lines of reasoning presented here, nor does the author claim priority as their first originator. Similar or related ideas may have appeared previously in other intellectual, cultural, or disciplinary traditions. Any legal rights retained in this work are intended to support attribution, responsible use, and protection against exploitative or harmful appropriation, and not to restrict independent inquiry, criticism, revision, or further development.

The arguments in this paper should therefore be understood as provisional and historically situated rather than definitive. Readers are encouraged to question, revise, extend, reinterpret, or independently develop the ideas presented here. Where appropriate, acknowledgment of this paper as one point of encounter in the development of related ideas is appreciated, but such acknowledgment should not be understood as granting the author epistemic ownership over the ideas themselves.

Responsible Use and Rights Reservation

The author encourages good-faith discussion, criticism, independent development, and responsible use of the knowledge presented in this work. The author does not claim exclusive epistemic ownership over the ideas or viewpoints discussed herein, nor claim priority as their first originator.

However, the author expressly reserves all rights and remedies available under applicable law with respect to uses of this work or related materials that involve unlawful conduct, harmful or abusive exploitation, improper commercial appropriation, infringement of applicable intellectual property or other legal rights, or conduct contrary to applicable national, regional, or international law.

This reservation is intended to preserve the ability to respond to misuse and harmful appropriation. It should not be interpreted as restricting legitimate academic inquiry, criticism, independent reasoning, or the further development of related ideas.

1 Introduction

The paper on criticality detection in this series treats the approach to a threshold as something an arrangement may observe and grade. It expressly reserves what happens once the threshold is near, and this paper takes that up. The situation is worse than a continuation of the same difficulties, and the reason is structural.

An arrangement governing near a threshold depends on three things. It depends on detection, since it must know that it is near one. It depends on a margin, since any action it takes must not itself carry the system across. And it depends on estimation, since the amount of margin it has and the rate at which the system is approaching are quantities it must estimate to act at all. Each of these degrades as the threshold nears. That much is known in pieces and is scattered across literatures that do not cite one another.

One thing is not stated in those literatures and is stated here: the three are not independent. They are monotone functions of the same approach parameter, and a portfolio of three risks that are perfectly rank-correlated is not a portfolio. An arrangement cannot hedge the failure of its detection by acting more cautiously, because caution requires a margin that has shrunk by the same approach; it cannot hedge the shrinking margin by estimating more carefully, because estimation degrades by the same approach; and it cannot hedge estimation error by observing longer, because the observation the estimate needs takes time the approach is consuming.

Claim 1.1. *Detection reliability, the safe intervention margin, and the accuracy of estimation are monotone in a single approach parameter and therefore degrade together. They are not three risks admitting separate management, and no one of them can be traded against another, because the trades all require the quantity that is degrading. This is the triple bind, and its practical content is that the standard institutional response to uncertainty, which is to observe more closely, act more cautiously, and learn from the response, has all three of its legs cut at once and in the same direction.*

The paper's second position concerns the third leg, where the difficulty is sharper than imprecision. Estimation of a system's proximity to a threshold ordinarily proceeds through the persistence of its fluctuations, and estimation of persistence near the unit root is biased downward. Because the recovery rate is a logarithm of persistence, a small downward bias in persistence becomes a large overstatement of the recovery rate, which is to say of stability, and the overstatement grows as the threshold nears. §2 reports a simulation confirming this and quantifying it: the same bias that overstates stability by about five per cent far from a threshold overstates it by nearly a factor of three near one.

Claim 1.2. *The estimation error near a threshold is not symmetric. It runs toward apparent safety, and it grows as actual safety diminishes. An arrangement acting on estimates of its distance from a threshold is therefore systematically told that it has more room than it has, and is told so most emphatically when it has least.*

Claim 1.2 is the paper's most concrete result and the one that converts the triple bind from a statement about coincident difficulties into a statement about a directional failure. Difficulties that are merely large may be met by widening intervals; a bias that runs toward safety is not

Draft

met by widening intervals, because the interval is centred in the wrong place.

The paper's third position is constructive and follows from the first two. If every input that describes the *system* degrades as the threshold nears, an arrangement that governs near criticality must transfer weight to inputs that do not. The inputs that do not are properties of the *proposed act*: whether it can be undone, over what period, and at whose cost. The reversibility of an act is a property of the act and of the coupling it alters, and it does not become harder to determine because the system is near a threshold. §3 develops this into a regime distinction and argues that the burden criterion of the series' first paper, which keys on irreversible foreclosure, is not merely applicable near criticality but is the only governing input that survives it.

The paper proceeds as follows. §2 restates the carried positions and establishes the three degradations and their coincidence. §3 develops the turn to act-regarding restraint and the regime distinction. §4 surveys the owning literatures. §5 treats the objections and §6 the open questions, and §7 the declinations.

2 The Three Degradations and Their Coincidence

2.1 Positions Carried from the Companion Papers

Five results are used and restated so the paper stands alone.

The *object of governance* is the set of relational conditions under which trajectories are generated. The governing principle constrains, under a dischargeable burden of justification, those alterations of conditions that would irreversibly foreclose the capacity of trajectories other than the acting party's own to continue generating; the trigger is a structural property and there is no maximand [14].

The *detection position* holds that criticality detection is a governance function whose findings enter as graded standing levels with duties attached; that the indicator suites it runs on must be carried with the transition classes they do not precede; and that restraint is grade-independent, since a low grade means absence of detected approach and not established distance.

The *differential position* holds that a class of transitions is induced by the rate at which a system is forced, that the states traversed on such a path are contained in those traversed on a non-transitioning path, and that indicators keyed to the loss of stability of an occupied equilibrium do not precede them.

The *capacity position* establishes that a coupling of timescale τ requires a record of length proportional to τ before it can be established, so that the claims requiring the longest records are those concerning the slowest and least reversible couplings.

The *assembly position* holds that heterogeneous holdings are brought to bear with disagreement preserved, and that an assembly may terminate in the finding that the material does not warrant the action contemplated, which is a determinate result and not a failure.

Draft

2.2 The Approach Parameter

Let the governed system occupy a stable state whose local dynamics have a dominant restoring rate $\lambda > 0$, and let a slow forcing carry the system toward a threshold at which $\lambda \to 0$ and the occupied state merges with the boundary of its basin. Write μ for a normalised measure of approach, increasing toward the threshold, with λ decreasing in μ. This is the standard local picture near a fold and the paper claims no part of it; what follows are consequences for governance.

Two quantities are defined against this picture. The *margin* Δ is the distance from the occupied state to the boundary of its basin, which shrinks as the threshold nears. The *persistence* $\phi = e^{-\lambda \Delta t}$ is the lag-one autocorrelation of fluctuations at sampling interval Δt, which rises toward one as λ falls.

2.3 Detection Degrades

Early warning indicators operate on the rise in variance and autocorrelation that accompanies the fall in λ. Two properties of that estimation problem bear on governance.

The stationary variance of fluctuations is proportional to $1/(1 - \phi^2)$, which diverges as $\phi \to 1$, so the fluctuations against which a signal is read grow at the same time the signal does. And the effective number of independent observations in a record of fixed length falls as persistence rises, since successive observations become more nearly the same observation, so a record of a given calendar span carries less information about ϕ the closer ϕ is to one. The detection literature's own results on limits to detection follow from these properties and are conceded to it.

To these the detection paper adds two failures that do not attenuate with proximity: entire transition classes that the suites do not precede, including transitions induced by the rate of forcing, and the muting of classical signatures where human response is coupled to the system observed.

2.4 The Margin Shrinks

The second degradation requires no argument beyond the definition, and its governance content lies in what it does to the standard remedy for uncertainty. The margin Δ is the distance to the basin boundary, and its shrinking toward zero is what it is for a threshold to be near.

The consequence is that *caution is not a free hedge*. An arrangement uncertain about its position ordinarily responds by acting in smaller increments, and the increment that counts as small is judged against the system's ordinary variability, which is the quantity that has been growing. An intervention of a magnitude that was demonstrably safe at $\mu = 0.5$ may exceed the margin at $\mu = 0.9$ without any change in the intervention, and an arrangement calibrating its increments on past experience of safe action is calibrating on a margin that no longer exists.

Draft

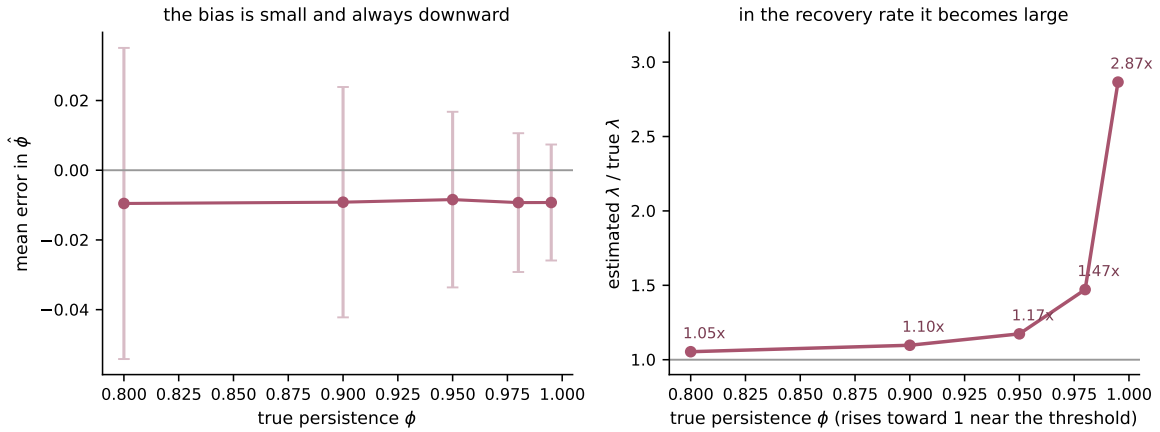

Figure 1: Estimation of persistence near the unit root, from four thousand simulated records of two hundred observations at each value. The left panel shows the mean error in $\hat{\phi}$, which is small, downward, and nearly constant across the range. The right panel shows the same error expressed in the recovery rate $\lambda = -\log \phi$, which is the quantity governance uses: the identical bias produces a five per cent overstatement of stability at $\phi = 0.80$ and a nearly threefold overstatement at $\phi = 0.995$. The annotations are computed by the simulation, not asserted.

2.5 Estimation Degrades Asymmetrically

The third degradation is the paper's own contribution and is worse than the loss of precision the first two would suggest.

Estimation of ϕ from a finite record by the ordinary least-squares estimator is biased downward, and the bias is a well-established property of autoregressive estimation and not a defect of any particular method [15, 17]. Its first-order form is known: for the least-squares estimator of a first-order autoregression from T observations, the bias is $-2\phi/T$ to order T^{-1} [22]. The paper's simulation confirms it across the range of interest, and agrees with the analytic form to within 0.0016 at every value tested, which is reported by the accompanying computation. Figure 1 shows the result: at every value of ϕ from 0.80 to 0.995, with records of two hundred observations, the mean estimate lies below the truth by approximately 0.009.

The bias in ϕ is small and, over this range, nearly constant, since $-2\phi/T$ varies little as ϕ runs from 0.80 to 1. Its consequence in the recovery rate is neither small nor constant, and this is the point.

Because $\lambda = -\log \phi$, and because the logarithm is steep near one, a fixed downward error in ϕ maps to an error in λ that grows without bound as $\phi \to 1$. The simulation quantifies it: the bias that overstates λ by a factor of 1.05 at $\phi = 0.80$ overstates it by a factor of 2.87 at $\phi = 0.995$. An arrangement reading its distance from a threshold off an estimated recovery rate is told the system recovers nearly three times faster than it does, at the moment when the truth is that it barely recovers at all.

Claim 1.2 follows. The error is directional, it runs toward apparent safety, and it grows as actual safety diminishes. Three consequences are worth stating.

The failure is invisible to the ordinary remedy. Widening a confidence interval addresses imprecision and not bias, and an interval correctly widened around a mis-centred estimate still excludes the truth at the rate the bias produces.

Draft

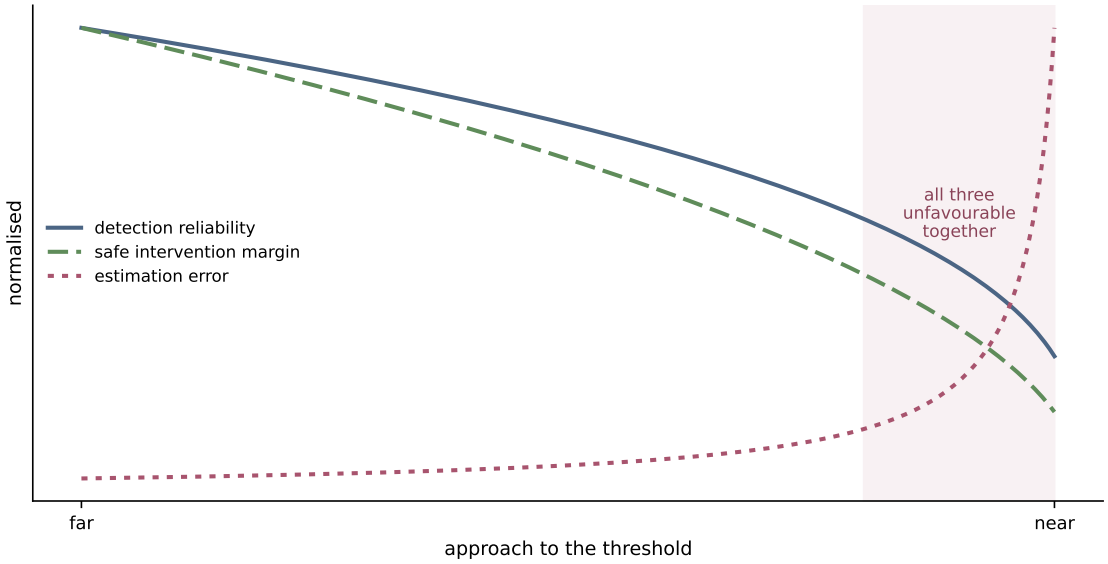

Figure 2: The three degradations against a single approach parameter. Detection reliability and the safe intervention margin fall; estimation error rises. The forms plotted are illustrative and the structural point is not: all three are monotone in μ, so their unfavourable values coincide, and the region in which each is worst is the same region.

The failure compounds with the first degradation, since the record needed to reduce the bias is longer exactly where the effective sample size is smallest.

And the failure is worst in the systems governance most cares about. The slowest couplings require the longest records, and slow couplings are the least reversible; so the estimates most likely to be badly biased concern the alterations that most need governing.

2.6 The Coincidence, and Why It Is Not a Portfolio

Figure 2 displays the coincidence. Its consequence for governance is stated as follows.

Claim 2.1. *Because the three degradations are monotone in the same parameter, no one of them may be hedged by another. Uncertainty about position cannot be met by acting in smaller increments, since the increment that would be small is judged against a margin that has shrunk. A shrunken margin cannot be met by estimating position more carefully, since estimation degrades with the same approach and is biased toward reporting more room than exists. And biased estimation cannot be met by observing longer, since the effective information in a record falls as persistence rises and the approach continues during the observation. The standard triad of adaptive governance, observe, act incrementally, and learn from the response, therefore fails as a whole and not in one of its parts.*

Claim 2.1 is stated at the level of the arrangement and not of the analyst, and the distinction matters. An analyst aware of all three difficulties may still produce the best available estimate with honest uncertainty attached. An arrangement that acts on that estimate faces the situation the claim describes: the honest uncertainty is wide, the bias inside it points one way, and the arrangement's usual devices for proceeding under wide uncertainty are the ones the bind has disabled.

Draft

3 The Turn to Act-Regarding Restraint

3.1 The Inputs That Survive the Approach

Every input examined in §2 is a description of the *system*: its distance from a threshold, its recovery rate, its margin, its detected grade. Each degrades as the threshold nears, and Claim 2.1 establishes that they cannot be traded against one another. If governance near criticality is possible at all, it must rest on inputs of another kind.

Claim 3.1. *Properties of a proposed act do not degrade as the system approaches a threshold. Whether an alteration can be undone, over what period, at whose cost, and whether the capacity it would remove could be reconstituted, are determined by the character of the act and of the coupling it alters, and their determination does not become harder because the system's recovery rate has fallen. An arrangement governing near criticality must therefore transfer weight from system-regarding inputs to act-regarding ones, and the transfer is not a counsel of caution but a consequence of which inputs remain available.*

Claim 3.1 requires two qualifications, both of which narrow it and neither of which defeats it.

The reversibility of an alteration is not wholly independent of the system, since whether a severed coupling can be restored depends on the system's own dynamics, and a system near a threshold may fail to recover from an alteration that it would have absorbed further away. This is true and it cuts the right way: proximity makes alterations *less* reversible, so an act-regarding assessment made without reference to proximity errs toward permitting, which is the same direction as the estimation bias and is a reason to treat act-regarding thresholds as tightening with grade, as the detection paper's grade structure provides.

And some act-regarding determinations are contestable, since whether an occupied habitat can be reoccupied or a discontinued record resumed admits argument. The claim is comparative and not absolute: these determinations are contestable in ways that do not worsen with μ, whereas the system-regarding determinations are contestable in ways that do.

3.2 Two Regimes

The consequence is a distinction between two governing regimes, and the paper states it as the practical form of the whole argument.

In the *learning regime*, an arrangement observes, acts in increments, and revises against the response. Its inputs are system-regarding, its instruments are incremental, and its central virtue is that it improves. This is the regime adaptive and experimentalist governance describe, and the paper takes no issue with it where it applies.

In the *restraint regime*, an arrangement's system-regarding inputs are unreliable in the specific ways §2 establishes. It continues to observe, since observation is what will eventually establish where the system is, and it stops treating its observations as licences. Its governing input becomes the character of proposed alterations, and the burden criterion carried from the series' first paper is the operative rule: an alteration that would irreversibly foreclose another trajectory's capacity to continue generating bears a burden of justification, and near critical-

Draft

Governing input	Requirements	Behaviour as $\mu \to 1$	Direction of error
Detected grade	Indicator suite, adequate record, spanning carriage	Reliability falls; classes escape entirely	Toward no detection
Distance to threshold	Estimated recovery rate	Bias amplifies in λ	Toward apparent safety
Safe increment	A margin against which small is judged	Margin shrinks; past calibration invalid	Toward permitting
Response to a trial	The system's return after perturbation	Return time diverges; trial may not be recoverable	Toward permitting
Reversibility of the act	The act's character and the coupling altered	Determination unaffected by μ	Contestable, not directional
Foreclosure of another's capacity	The coupling's structure and who depends on it	Determination unaffected by μ	Contestable, not directional

Table 1: Governing inputs near a threshold. The first four are system-regarding and degrade, three of them in the same direction, toward permitting. The last two are act-regarding, and it is their invariance under μ, and not any superior precision, that recommends them in this regime.

ity the class of alterations meeting that description is wider, because proximity makes more alterations irreversible.

Claim 3.2. *The transition between regimes is itself a governance determination, and it cannot be made on the system-regarding inputs whose failure defines the second regime. It must therefore be made on the grade structure of the detection arrangement, which is a public and contestable object, and on the grade-independent floor the detection paper establishes, under which a low grade means absence of detected approach and not established distance. An arrangement that waits for its system-regarding inputs to tell it that they have become unreliable will not be told.*

Claim 3.2 states the argument's least comfortable consequence. The bind's onset is not announced by the instruments the bind disables, so an arrangement must enter the restraint regime on grounds that are partly precautionary and partly structural: a grade transition, a class of transition its suite does not precede, a coupling whose alteration it cannot establish the reversibility of. This is unsatisfying and the paper prefers it to the alternative, which is a rule that enters the restraint regime when the estimates say so, and the estimates say so late by construction.

3.3 Three Misreadings Foreclosed

Three misreadings are worth foreclosing, since the position is easily overstated in each direction.

It is *not a prohibition*. The burden criterion is dischargeable, and an alteration whose proponent gives sufficient reasons proceeds. The regime changes what must be argued and by whom, and does not remove the possibility of acting.

It is *not a maximand*. The paper proposes no quantity to be minimised, no option value to be preserved, and no index of reversibility to be optimised. The series' first paper declines maximands for reasons that apply here with additional force, since any quantity aggregated

Draft

over a system near a threshold would be estimated by the apparatus §2 has just impeached.

And it is *not a suspension of observation*. An arrangement in the restraint regime observes more and not less, because the record it accumulates is what will eventually establish where the system is and because the capability and coupling events its observation generates are inputs to the routing and inquiry machinery of the series. The weight observation bears in licensing action changes, and the effort devoted to it does not.

3.4 A Worked Situation

A reservoir system supplies a city and an irrigated district and supports a downstream wetland whose vegetation community has an alternative low-water state from which recovery, where it has been observed elsewhere, has taken decades. Monitoring over fifteen years shows rising variance in the wetland's water-table series and a rising lag-one autocorrelation, presently estimated at 0.97.

The learning-regime response is available and is what an ordinary adaptive arrangement would do: estimate the distance to the threshold from the fitted recovery rate, licence an incremental increase in abstraction well within that distance, observe the response, and revise. Each step is defensible and the composite is what the triple bind describes. The estimated recovery rate is biased toward stability by a factor the simulation of §2 puts at roughly double in this range, so the estimated distance is an overestimate. The increment judged small is judged against a variability that has grown, so it is larger relative to the margin than it appears. And the trial's informativeness depends on observing the system's return, which now takes longer than the interval before the next decision.

The restraint-regime response differs in what it asks. The question becomes whether the proposed increase in abstraction would irreversibly foreclose the wetland's capacity, which is answerable from the character of the alteration and the comparative record of recovery elsewhere, and is not answerable more or less well according to how near the threshold is. If the answer is that recovery from the low-water state takes decades and is uncertain, the burden attaches, and it attaches without any estimate of the distance to the threshold. The proponent may discharge it, and what it must argue is about the alteration, which is the thing it knows most about.

The case also displays the paper's limits honestly. The comparative record of recovery elsewhere is itself an observational claim, drawn from other systems whose comparability is contestable. The restraint regime shifts the argument's ground and does not remove the need for evidence, and its advantage is that the evidence it needs is of a kind that does not degrade with the approach.

4 Prior Treatments of Action near Thresholds

Six literatures own parts of the position, and the paper's contribution is a composition and not any element of it.

Draft

4.1 Critical Slowing Down and the Limits of Detection

The early warning literature established that a class of transitions is preceded by rising variance and autocorrelation as a system's recovery slows [21, 6], and its critical branch established the limits: transitions occur without warning in chaotic and shock-driven regimes [12], detection performance depends on data quantities field records rarely supply, and false positives are common enough that multiple indicators must agree [5, 8].

This literature owns the first degradation entire, and the paper's use of it is application. The literature does not ask, because it is not its question, what an arrangement should do given that the indicator's reliability is a function of the same parameter as the danger it indicates.

4.2 Autoregressive Estimation near the Unit Root

The downward bias of least-squares estimation of autoregressive coefficients is classical, established for the first-order case and refined since [15, 17, 22], and the associated literature on unit-root inference establishes the non-standard distributions that hold as the coefficient approaches one [7, 18]. Bias-corrected estimators exist and their properties are known.

The concession is complete: the bias is the statistics literature's, the paper contributes nothing to it, and the simulation of §2 confirms a known result and discovers nothing. The paper's addition is the translation into the recovery rate and the governance reading. That $\lambda = -\log \phi$ converts a small constant bias into an unbounded one as $\phi \to 1$ is arithmetic, and appears not to have been stated as a property of early warning practice, where ϕ is ordinarily reported directly and the direction of its error is not remarked. Where such a statement exists in a literature not surveyed here, priority belongs to it.

4.3 Rate-Induced Transitions

The mathematical literature distinguishes transitions induced by the rate at which a parameter is moved from those induced by its value [2, 27]. The differential paper of this series draws the governance consequence, and it is carried here as the class that escapes the first degradation entirely, since a rate-induced transition need involve no loss of stability of the tracked state and therefore no slowing to detect.

4.4 Decision under Deep Uncertainty and Robustness

Robust decision making seeks strategies performing acceptably across a wide range of plausible futures in place of optimising against a distribution [16], and the associated field develops adaptive pathways and related methods [26, 11]. Info-gap analysis treats robustness to error in one's own model as the object [4].

This is the closest existing answer to the situation and the concession is substantial. Two differences bear. These methods evaluate strategies against a set of futures, and the set is generated by models whose parameters are estimated by the apparatus §2 impeaches, so a robustness analysis near a threshold inherits the bias in the generation of its own scenario

Draft

set unless the set is widened on grounds independent of the estimates. And the adaptive-pathways form depends on signposts, being observable indicators that trigger a switch of pathway, which is precisely the input the first degradation attacks; Claim 3.2 is the statement of that difficulty for this architecture.

4.5 Precaution and Irreversibility in Decision Theory and Law

The economics of irreversibility established that where an action forecloses options and information will arrive, there is a positive value to delay, quantified as quasi-option value [1, 13], and the real-options literature developed the apparatus [9]. The precautionary principle in environmental law and its analytical treatments address action under scientific uncertainty, with an extensive critical literature on its formulation and coherence [24, 20, 25].

These literatures own the proposition that irreversibility and uncertainty jointly warrant restraint, and the concession is unreserved. The paper's addition is a mechanism and a bound. The mechanism is the triple bind, which explains *why* the usual alternative to restraint, incremental action with learning, is unavailable here, where the option-value literature establishes that delay has value without establishing that the alternative fails. The bound is that quasi-option value is a quantity computed against a model, so it participates in the third degradation, and the position taken here declines the quantification for that reason, and keys on a structural property of the act. This is a weaker instrument than option value and it is one the bind does not disable.

4.6 Governance of Thresholds and Safe Operating Spaces

Planetary boundaries specify a region within which a system should be kept [19, 23], viability theory supplies the mathematics of keeping a system within constraints [3], and the recent tipping-point assessment reports thresholds near or crossed and calls for governance commensurate [10].

This family owns the region-based statement of the governing problem, and the relation to the present paper is complementary and clarifying. A boundary specifies where the system must stay and is stated in system-regarding terms; the present paper's claim is that near the boundary, the arrangement's ability to know where it stands relative to it degrades in the specific ways established here, which is a difficulty for the boundary framing and not an objection to it. The act-regarding turn is what remains available when the region-based input has become unreliable.

4.7 Residue of the Survey

Claim 4.1. *Consider an arrangement governing a system approaching a threshold, whose detection depends on statistics that degrade with the approach, whose safe increment is judged against a margin that shrinks with the approach, and whose estimate of position is biased toward stability by an amount that amplifies with the approach. The claims that these three are monotone in a single parameter and therefore constitute a bind admitting no internal hedge; that the estimation failure is directional toward*

Draft

apparent safety and amplifies in the recovery rate; that the standard triad of observation, incremental action and learning fails as a whole in consequence; and that an arrangement must therefore transfer weight to act-regarding inputs whose determination is invariant under the approach, are advanced by none of the surveyed literatures. The early warning literature owns detection and its limits; autoregressive statistics owns the bias; the rate-induced literature owns the escaping class; robust decision making owns strategy selection against a scenario set; the irreversibility economics owns the value of delay under arriving information; and the boundaries family owns the region-based statement. The composition, and the claim that act-regarding restraint is what survives the bind, and not merely what prudence recommends, is the residue claimed.

5 Objections and the Limits of the Position

5.1 The Paralysis Objection

The restraint regime attaches a burden to alterations whose reversibility cannot be shown, and near a threshold that class widens. The objection is that the position licenses inaction in exactly the situations that most need action, since a system near a threshold may require intervention to be brought back, and an arrangement that has entered the restraint regime will find every restorative intervention burdened along with every damaging one.

The answer distinguishes two things the objection conflates. The burden attaches to alterations that would irreversibly foreclose *another trajectory's* capacity to continue generating, and a restorative intervention ordinarily forecloses nothing; the criterion is not a bar on action but a test that most restorative action passes. Where a restorative intervention would itself foreclose, as with an engineered intervention whose failure mode is worse than the condition it addresses, the burden is correctly attached and is dischargeable by argument about the alteration.

The concession is that the criterion's application is not always clear, and that an arrangement disposed to inaction can find a foreclosure in most proposals. The paper's only answer is that the criterion is public and its application contestable, which is a procedural safeguard and not a guarantee, and that this exposure is the price of a criterion that does not rest on the estimates the bind disables.

5.2 The Objection from the Bias's Size

Claim 1.2 rests on a simulation whose bias in ϕ is about 0.009, which is small against the standard deviations reported in the same figure, and the objection is that a bias an order of magnitude below the sampling noise cannot bear the weight placed on it.

Four parts to the answer. The bias is not an artefact of the simulation: it agrees with the analytic first-order form $-2\phi/T$ established for this estimator, so the objection is to a known property of the method and not to the paper's computation. The comparison is misdirected: noise averages out across repeated estimates and bias does not, so an arrangement estimating repeatedly over time converges on a value that is wrong in a fixed direction. The magnitude in ϕ is the wrong scale on which to judge it, since the governing quantity is λ, in which the

Draft

same bias is a factor of nearly three at the values of interest, and Figure 1 exists to make that translation visible. And the simulation's parameters are favourable to the objection: records of two hundred observations are longer and cleaner than most environmental series, and shorter or irregularly sampled records exhibit larger bias.

The genuine limit is different and the paper states it: bias-corrected estimators exist, and an arrangement using them is not subject to the failure in the form described. The claim's force is therefore conditional on practice, and its practical import is a recommendation, that estimates of persistence used in governance be bias-corrected and that the correction be recorded, which is cheap and is not currently standard.

5.3 The Objection from Regime Determination

Claim 3.2 concedes that the bind's onset cannot be established by the inputs the bind disables, and proposes entry on grade transitions and structural grounds. The objection is that this makes the regime switch precautionary in the pejorative sense: an arrangement may declare the restraint regime whenever it wishes, and the declaration cannot be checked against the system's actual state.

The paper accepts that the switch is not verifiable against the system and denies that it is therefore arbitrary. The grounds for entry are public objects: a grade defined by stated indicator conditions in a versioned register, a documented class of transition the suite does not precede, an inability to establish a coupling's reversibility recorded with its basis. Each is contestable on its own terms, and a party disputing entry disputes a determinate claim. Adjudication by appeal to where the system actually is is unavailable, and the paper's position is that such an appeal was never available near a threshold and that arrangements which believe otherwise are relying on the estimates §2 impeaches.

5.4 The Objection from Symmetry

An arrangement in the restraint regime is subject to the same bind, and the objection is that the paper has exempted its own recommendation from its own analysis: if estimation is unreliable near a threshold, the determination that one is near a threshold is unreliable, and an arrangement may impose burdens on a false alarm as easily as it may permit on a false reassurance.

The answer is the asymmetry established in §2 and it is the paper's most important reply. The estimation failure is not symmetric: it runs toward apparent safety. An arrangement is therefore more likely to believe itself far from a threshold when near than the reverse, so the errors the bind produces are concentrated on the permitting side, and a rule that shifts weight toward restraint corrects in the direction the bias runs. This does not make false restraint impossible; it makes it the less likely of the two errors, which is the most the analysis supports.

5.5 Limits Internal to the Position

Four limits are recorded. The local picture of §2 is the standard one near a fold, and systems whose approach is not of that form, including those approaching a threshold through

Draft

a rate-induced mechanism, are not described by the approach parameter at all, so the bind's statement does not cover them and the detection paper's escape classes apply. The illustrative forms in Figure 2 are not measured, so the coincidence is established structurally and its shape in any field is unknown. The act-regarding turn presupposes that reversibility can be determined, which the capacity paper shows requires records the parties may lack, so the inputs the paper recommends are not free of the difficulties it attributes to the others, only free of their dependence on μ. And the paper says nothing about the regime near a threshold that has already been crossed, where the governing question is recovery and the analysis here does not apply.

A further limit is inherited rather than internal. The paper applies the series' principle to alterations described as irreversibly foreclosing, and the series' first paper supplies no method for establishing irreversibility, recording the absence as an open question. Where the property is stipulated by the case under discussion, nothing turns on this. Where it would be contested, the paper's conclusions are conditional on a determination the series has not yet made, and the seventh paper's establishment condition constrains what evidence such a determination would require without supplying it.

The first paper also records that a sequence of alterations, each satisfying the principle when it is assessed, may compose into a foreclosure on which the principle never fires. The approach parameter used here describes the state such a sequence produces, and the bind this paper identifies is what that state costs a governing arrangement. The paper does not, however, supply the rule governing the sequence that the first paper's question asks for: it characterises where accumulation leads without stating when along the way the principle should have fired. That gap remains open.

6 Questions Left Open for the Programme

6.1 Questions Concerning the Bind

Open Question 6.1. ***Empirical shape of the coincidence.*** *An answer would estimate, for a system with adequate records, how the three quantities actually covary with proximity, and would establish whether the coincidence is as tight in practice as its common dependence on the approach parameter implies, since a loose coincidence would leave room for the hedges Claim 2.1 denies.*

Open Question 6.2. ***Bias correction in governance practice.*** *§5 concedes that bias-corrected estimators defeat the failure in its stated form. An answer would establish whether corrected estimation is used in the practice that informs environmental governance, would quantify the residual bias of the corrections near the unit root, and would state what an arrangement should require of the estimates it relies on.*

Open Question 6.3. ***The bind under other approach mechanisms.*** *The bind is stated for approach to a fold. An answer would establish which of the three degradations hold for rate-induced approach, for noise-induced transitions, and for systems approaching thresholds of other types, since the paper's scope is narrower than its statement suggests.*

Draft

Open Question 6.4. ***Interaction with the observational capacity distribution.*** *The estimation failure is worst where records are shortest, and records are shortest where capacity is least. An answer would state how the bind's incidence is distributed across parties, which would make it a justice question as well as a methodological one.*

6.2 Questions Concerning the Turn

Open Question 6.5. ***Determination of reversibility.*** *The act-regarding turn presupposes that reversibility is determinable. An answer would state what evidence establishes it, how comparative evidence from other systems is to be assessed, and what an arrangement does where reversibility is itself contested, which the worked situation shows is the ordinary case.*

Open Question 6.6. ***Proximity-dependent tightening.*** *§3 concedes that proximity makes alterations less reversible, so act-regarding thresholds should tighten with grade. An answer would state the tightening, and would confront the difficulty that a tightening keyed to grade reintroduces dependence on the detection input the bind degrades.*

Open Question 6.7. ***Restorative intervention.*** *An answer would characterise the class of interventions that restore without foreclosing, would state what evidence establishes membership, and would address the case where a restorative intervention's failure mode is a foreclosure, which §5 identifies and does not resolve.*

Open Question 6.8. ***Active probing.*** *A body of work considers deliberate perturbation to establish a system's position. Whether probing is admissible near a threshold, and under what conditions, is a question the present paper does not treat, and an answer would need to establish what a probe reveals and at what risk when the margin is the quantity in doubt.*

6.3 Questions Concerning the Regime Switch

Open Question 6.9. ***Entry criteria.*** *An answer would state the criteria for entering the restraint regime as determinately as the detection paper states grade definitions, and would establish who may contest an entry and on what ground.*

Open Question 6.10. ***Exit.*** *The paper says nothing about leaving the restraint regime. An answer would state what warrants exit, and would confront the asymmetry the detection literature documents, that descending an alert level is institutionally harder than ascending, together with the fact that the inputs establishing exit are the ones the bind disabled.*

Open Question 6.11. ***Partial entry.*** *A field contains many objects at different proximities. An answer would state whether the regime is entered for a system, for a coupling, or for a class of alteration, since a whole-field entry is too blunt and a per-coupling entry requires per-coupling determinations the arrangement may not be able to make.*

Draft

Open Question 6.12. ***After crossing.*** *An answer would state what governs where a threshold has been crossed and the question is recovery, which §5 identifies as outside the paper's scope and which the tipping-point assessments indicate is already the situation for some elements.*

7 Declinations and Limits of the Position

The paper establishes no statistical result. The estimation bias is classical, the simulation confirms a known property, and the paper's contribution is its translation into the recovery rate and the governance reading placed on it.

The paper establishes no dynamical result. The local picture near a fold, the shrinking margin, and the rate-induced escape class belong to the literatures cited.

The paper claims no novelty in precaution or in the value of delay under irreversibility, and §4 concedes both, together with the observation that its own instrument is weaker than quasi-option value and is chosen because the bind does not disable it.

The paper measures nothing. Figure 2 is illustrative, no field is characterised, and the coincidence is established from the common dependence of the three quantities on the approach parameter and not from data.

The paper does not treat systems past a threshold, nor the admissibility of active probing, and records both as outside its scope.

References

[1] Arrow, K. J. and Fisher, A. C. (1974). Environmental preservation, uncertainty, and irreversibility. *Quarterly Journal of Economics*, 88(2), 312–319.

[2] Ashwin, P., Wieczorek, S., Vitolo, R., and Cox, P. (2012). Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. *Philosophical Transactions of the Royal Society A*, 370, 1166–1184. Correction: *Phil. Trans. R. Soc. A* 371, 20130098 (2013).

[3] Aubin, J.-P. (1991). *Viability Theory*. Birkhäuser.

[4] Ben-Haim, Y. (2006). *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*, 2nd edn. Academic Press.

[5] Boettiger, C. and Hastings, A. (2012). Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface*, 9(75), 2527–2539.

[6] Dakos, V., Carpenter, S. R., Brock, W. A., et al. (2012). Methods for detecting early warnings of critical transitions in time series. *PLoS ONE*, 7(7), e41010.

[7] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431.

[8] Ditlevsen, P. D. and Johnsen, S. J. (2010). Tipping points: early warning and wishful thinking. *Geophysical Research Letters*, 37, L19703.

[9] Dixit, A. K. and Pindyck, R. S. (1994). *Investment under Uncertainty*. Princeton University Press.

Draft

[10] Lenton, T. M., et al. (eds.) (2025). *The Global Tipping Points Report 2025*. University of Exeter.

[11] Haasnoot, M., Kwakkel, J. H., Walker, W. E., and ter Maat, J. (2013). Dynamic adaptive policy pathways: a method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, 23(2), 485–498.

[12] Hastings, A. and Wysham, D. B. (2010). Regime shifts in ecological systems can occur with no warning. *Ecology Letters*, 13(4), 464–472.

[13] Henry, C. (1974). Investment decisions under uncertainty: the irreversibility effect. *American Economic Review*, 64(6), 1006–1012.

[14] Huang, W. (2026). Irreversible foreclosure and the allocation of justificatory burden in governance beyond epistemic reach. *Knowledge Commons*. `https://doi.org/10.17613/6bkgx-zmk02`

[15] Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, 41(3–4), 403–404.

[16] Lempert, R. J., Popper, S. W., and Bankes, S. C. (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. RAND Corporation, MR-1626-RPC, Santa Monica CA. doi:10.7249/MR1626.

[17] Marriott, F. H. C. and Pope, J. A. (1954). Bias in the estimation of autocorrelations. *Biometrika*, 41(3–4), 390–402.

[18] Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, 55(2), 277–301.

[19] Rockström, J., Steffen, W., Noone, K., et al. (2009). A safe operating space for humanity. *Nature*, 461, 472–475.

[20] Sandin, P. (1999). Dimensions of the precautionary principle. *Human and Ecological Risk Assessment*, 5(5), 889–907.

[21] Scheffer, M., Bascompte, J., Brock, W. A., et al. (2009). Early-warning signals for critical transitions. *Nature*, 461, 53–59.

[22] Shaman, P. and Stine, R. A. (1988). The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association*, 83(403), 842–848.

[23] Steffen, W., Richardson, K., Rockström, J., et al. (2015). Planetary boundaries: guiding human development on a changing planet. *Science*, 347(6223), 1259855.

[24] Sunstein, C. R. (2005). *Laws of Fear: Beyond the Precautionary Principle*. Cambridge University Press.

[25] World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) (2005). *The Precautionary Principle*. UNESCO, Paris. UNESDOC 139578.

[26] Walker, W. E., Haasnoot, M., and Kwakkel, J. H. (2013). Adapt or perish: a review of planning approaches for adaptation under deep uncertainty. *Sustainability*, 5(3), 955–979.

[27] Wieczorek, S., Ashwin, P., Luke, C. M., and Cox, P. M. (2011). Excitability in ramped systems: the compost-bomb instability. *Proceedings of the Royal Society A*, 467, 1243–1269.

Draft