

Evidence-to-Action near Criticality: A Qualification for EBPM

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Evidence-based policymaking (EBPM) uses systematic evidence to inform public decisions, yet the move from evidential support to intervention also depends on whether the response conditions under which the evidence was generated remain sufficiently stable for action. This paper develops a conditional qualification for settings near a plausible critical transition. In such settings, evidence can remain locally informative while intervention, policy announcement, or induced expectation change alters the feedback structure, coupling pattern, attractor accessibility, or recovery regime that made the evidential relation decision-relevant. The paper terms this possibility *critical-regime instability of evidence-to-action relations*. The qualification concerns action warrant rather than the value of evidence itself: a locally credible estimate can continue to support inference about a current response relation even as forward transportability weakens, reversibility contracts, or the set of supportable interventions narrows. To make this distinction explicit, the paper develops a local action-admissibility framework in which evidential validity, dynamical reversibility, and normative warrant are treated as related but separable conditions. It then examines regime transport, reflexive evidence, and *evidence-restraint inversion*, where improved knowledge of leverage or susceptibility can support bounded, staged, monitored, or reversible action rather than greater intervention intensity. The constructive implication is an adaptive evidence-informed policy architecture in which evidence updates both expected policy effects and the estimated boundaries of admissible action. Near criticality remains an empirically defeasible, domain-specific hypothesis, and the paper concludes with a research program designed to distinguish critical-regime instability from ordinary nonstationarity, contextual variation, and structural change. The resulting qualification expands the role of evidence from estimating effects alone toward characterizing the response conditions, transition risks, and action boundaries relevant to policy choice.

Keywords: evidence-based policymaking; critical-regime instability; regime transport; reversibility; action warrant

Discussion Paper Note

This position paper develops a focused qualification for evidence-based policymaking (EBPM) under conditions of plausible near criticality. It asks how far a locally supported response relation can be transported into action when intervention can alter the regime in which that relation was estimated. The term *near criticality* is used conditionally for configurations with transition-relevant changes in susceptibility, feedback, attractor accessibility, or recovery structure; crisis rhetoric, volatility, or disagreement alone provide insufficient grounds for the attribution. The paper is a focused companion to *Governance Near Criticality: Rethinking Intervention under Critical Transition Uncertainty* and engages directly with EBPM and evidence-informed policy scholarship.

Responsible Use and Rights Reservation

The paper develops a conditional framework for reasoning about evidence and intervention under uncertain response regimes. It is intended to support careful inquiry into policy sensitivity, reversibility, and justification. Applications should preserve empirical humility, transparent uncertainty, contestability, affected-party standing, and the distinction between descriptive leverage and normative warrant.

Except where otherwise indicated, copyright 2026 Wanhong Huang. This work is made available under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). Reuse remains independent from authorial endorsement. Copyright exceptions and limitations remain available where applicable.

Notices

Status. This release-candidate draft presents a complete focused position paper on evidence-to-action relations near criticality. Near criticality is treated as a conditional, empirically defeasible description of a response configuration whose attribution requires transition-relevant evidence beyond political crisis, volatility, or disagreement.

Statement on the use of language models. The exploratory discussions and preparation of this paper involved OpenAI's ChatGPT. ChatGPT supported exploratory dialogue, source discovery followed by website verification, conceptual reconstruction, argumentative criticism, and LaTeX drafting. The author selected the research problem, directed and approved theoretical commitments, and bears sole responsibility for the manuscript, including its definitions, claims, interpretations, and errors. Authorship credit remains with the human author.

1 Introduction

This section positions the paper within evidence-based policymaking (EBPM) and isolates its narrow qualification. The central distinction separates the epistemic value of evidence from the additional warrant required for intervention when action can alter the response regime in which the evidence was generated. Existing EBPM scholarship already treats evidence use as a problem of judgment, context, transport, and governance. Near criticality adds a dynamical possibility: intervention can change the response field to which otherwise credible local evidence applies.

Evidence-based policymaking is understood here as a family of efforts to improve public decisions through systematic use of research evidence, with multiple institutional forms of translation from inquiry to action. The literature already contains substantial caution against a mechanical picture in which a study result directly selects a policy. Head describes rigorous research as an important input into policy

development while emphasizing the institutional and political conditions under which evidence is interpreted (Head 2010). Cairney’s account similarly places evidence inside policy processes characterized by bounded attention, ambiguity, institutional competition, and selective use of information (Cairney 2016). Parkhurst develops the related idea of the good governance of evidence, distinguishing technically valid evidence from the procedural and political conditions governing how evidence is used (Parkhurst 2017). Together, these arguments supply the starting point for the present qualification: evidence enters policy through judgment, institutions, and political procedure.

A second established concern is transport. Evidence about an intervention in one setting carries forward only to the extent that the target setting preserves the conditions relevant to the inference. Cartwright and Hardie emphasize the importance of the support factors that allow a causal relation to operate in a target context (Cartwright and Hardie 2012). This problem is usually expressed across populations, institutions, places, or implementation conditions. The present paper extends the same discipline in a dynamical direction. The relevant target context can change through time even within the same jurisdiction, organization, or policy domain. More strongly, policy action can contribute to that change. Evidence may therefore be credible for a local response relation while becoming less informative about the consequences of acting on that relation once the intervention modifies the system’s feedback architecture.

The qualification becomes especially important under a plausible critical transition. Critical-transition research studies systems in which changes in state, forcing, coupling, or recovery structure can produce shifts between contrasting dynamical regimes. The literature also emphasizes the difficulty of predicting such transitions and the conditional value of early-warning indicators (Scheffer, Bascompte, et al. 2009). The present paper uses those results only as conditional analytical resources for political analysis. Political systems contain strategic actors, contested interpretations, institutions, normative orders, and reflexive responses that differ from the physical and ecological systems in which many critical-transition models were developed. The term *near criticality* is therefore used here as a domain-specific hypothesis about a response configuration. Its application requires evidence that susceptibility, feedback, recovery, basin accessibility, or another transition-relevant property is changing in the policy domain under study.

This cautious use of criticality creates a precise evidence-to-action problem. Suppose evidence supports a local estimate of how a policy variable affects a current outcome. Under a sufficiently stable response regime, the policy can be implemented, its consequences observed, and the evidential relation updated without assuming that the act of intervention has transformed the basic field of response. Near a critical regime, the same sequence can become less secure. The intervention can alter expectations, strategic behavior, institutional couplings, reinforcing feedback, or the accessibility of alternative states. Subsequent observations can then be generated under a materially changed response field. The original evidence may remain methodologically sound and historically informative while providing an incomplete basis for judging the new trajectory.

The paper calls this possibility *critical-regime instability of evidence-to-action relations*. The phrase identifies a relation between three objects: an evidential basis, a response regime, and an intervention capable of changing that regime. The concept applies only where intervention has a credible channel to regime change; uncertainty or nonlinearity alone provides insufficient support. The analytical burden is more specific. A study invoking this qualification should identify the response property that appears sensitive, the intervention channel that can affect it, the time horizon over which the regime may change, and the observable pattern that would count against the criticality interpretation.

This framing preserves the value of local evidence. A well-estimated response relation can remain useful for diagnosis, monitoring, comparison, and short-run prediction. The qualification enters at the transition from evidence to action. An intervention decision asks more than whether the estimated response is credible. It also asks whether the response relation is likely to remain relevant after action, whether adverse movement remains reversible, whether future options are preserved, whether affected actors have standing in the decision, and whether the resulting distribution of risk is acceptable. These

considerations can move independently of evidential confidence.

Working Claim 1.1 (Evidence-to-action qualification). Near a plausible critical transition, stronger local evidential support can coincide with a smaller set of adequately supported interventions. Contraction may reflect reduced reversibility, regime sensitivity, distributional exposure, or weaker normative warrant while the underlying evidence remains locally informative.

The working claim is deliberately conditional. A sensitive system can still contain interventions that are weakly coupled to the transition mechanism, highly reversible, institutionally authorized, or buffered against persistent harm. In such cases, better evidence can straightforwardly enlarge practical confidence. The qualification becomes relevant when greater precision reveals greater leverage over trajectory formation and therefore a more demanding decision problem.

This possibility motivates the term *qualification* in the title. Near criticality is presented as a qualification within EBPM rather than as a refutation of it. Existing EBPM scholarship already argues that evidence quality, contextual fit, judgment, political values, and institutional procedure require separate attention (Head 2010; Cairney 2016; Parkhurst 2017; Cartwright and Hardie 2012). The contribution proposed here is narrower: dynamical regime instability can create an additional boundary between valid evidence and warranted action. That boundary is especially important where intervention changes the conditions under which the evidence would otherwise be transported forward in time.

The qualification also differs from a generic appeal to complexity. Complexity can justify humility without identifying the structure of the decision problem. The present framework requires a more discriminating account. Evidence should be evaluated relative to a stated inferential task; near criticality should be supported by transition-relevant response evidence; reversibility should be assessed over a declared horizon; and normative warrant should remain visible as a distinct requirement. These distinctions allow the framework to fail in an informative way. A policy domain may display uncertainty, strategic behavior, or rapid change while still lacking evidence of a critical transition. In that case, ordinary problems of evidence quality, transport, implementation, and political judgment remain the more appropriate explanation.

A further implication concerns the meaning of stronger evidence. In many policy settings, better estimation of policy effects supports more confident selection among interventions. Near criticality, stronger evidence about susceptibility or leverage can also reveal that small interventions have unusually large or persistent consequences. The result can be an *evidence-restraint inversion*: the informational gain increases the responsibility to bound the intervention, preserve fallback capacity, strengthen monitoring, or delay irreversible commitment until the relevant response structure is better understood. The inversion leaves the direction of action open. Delay and non-action can themselves carry irreversible consequences. Its role is to show that evidential confidence and intervention intensity can move independently.

The paper therefore treats action, restraint, and continued observation within a common evidential framework. A regime-sensitive use of evidence asks how stable the supported response relation is over the intended horizon, how intervention can reconfigure that relation, and which consequences remain revisable. The informational task thereby extends from estimating an expected effect toward characterizing the conditions under which the estimate remains decision-relevant.

The analysis begins from ordinary evidence-to-action reasoning and then follows what changes when intervention can modify the regime it enters. Throughout, evidential strength and action warrant remain analytically distinct.

The intended contribution is focused in scope. Near criticality enlarges the evidential task: effect estimates are interpreted together with the response regime, transition sensitivity, reversibility, and the conditions under which intervention remains admissible. Keeping these tasks distinct preserves a strongly evidence-informed policy process while treating local epistemic validity as one component of action warrant.

2 Evidence-to-Action Relations

This section establishes the baseline evidence-to-action architecture against which critical-regime instability becomes visible. Evidence can support a target inference, while public action additionally depends on transport, timing, reversibility, institutional authority, distribution, and acceptable exposure. Keeping these functions separate preserves the epistemic importance of evidence while locating the additional judgment introduced by intervention.

Evidence-based policymaking gains much of its force from a simple and defensible intuition: public decisions should respond to disciplined inquiry into how the world works. The difficulty begins when this intuition is compressed into a single relation between “what the evidence says” and “what policy should do.” Evidence is generated for particular inferential tasks. Policy action is selected under legal, institutional, temporal, distributive, and strategic conditions that generally exceed those tasks. Existing EBPM scholarship already treats this translation as a problem of judgment and governance, with evidence entering through institutional interpretation (Head 2010; Cairney 2016; Parkhurst 2017). The present paper retains that understanding and isolates one additional source of separation: the dynamical regime in which an evidenced relation is expected to remain decision relevant.

2.1 Target-Relative Evidence

Evidence acquires meaning relative to a question. A randomized evaluation can support an estimate of an average treatment effect under stated conditions. Administrative records can support a claim about observed uptake or distribution. Qualitative inquiry can illuminate mechanisms, interpretations, implementation conditions, or forms of exclusion that are poorly represented by a single quantitative indicator. Historical evidence can identify sequences, institutional dependencies, and path-specific constraints. Each evidential form contributes through the target inference for which it is being used; action requires further translation beyond that target claim.

Definition 2.1 (Target-relative evidential support). *Target-relative evidential support* is the degree to which an evidential basis justifies a specified inferential claim under stated assumptions, measurement conditions, comparison structures, and scope conditions. The target claim can concern direction, magnitude, mechanism, ranking, distribution, implementation, or another policy-relevant property.

The definition makes two forms of discipline explicit. First, the target claim should be stated before evidence is treated as decisive. Evidence supporting one question can remain weak for another. A study estimating a short-run average effect may provide limited information about long-run institutional adaptation. Evidence that one intervention outperforms another on a measured outcome may provide little information about the distribution of costs across affected groups. Evidence about implementation fidelity can remain compatible with uncertainty about causal effect. Second, evidential strength should remain tied to the assumptions that support the inference. Measurement validity, comparison design, missing data, model specification, and contextual fit can matter differently across target claims.

This target-relative view is compatible with the plural conception of evidence already present in EBPM. Head emphasizes that policy knowledge combines research, political judgment, and practical experience across a plural evidential architecture (Head 2010). Cairney similarly describes policy processes in which attention, framing, institutional rules, and actor strategies shape which evidence becomes usable (Cairney 2016). These accounts preserve the epistemic role of evidence while showing why its inferential task must be specified before its policy role can be evaluated.

A useful consequence follows. Strong evidence for a narrow claim can coexist with substantial uncertainty about the broader decision problem. This configuration can contain methodologically strong evidence for its declared target together with uncertainty located elsewhere in the decision architecture.

Remark 2.2 (Local strength). The phrase *locally valid evidence* in this paper refers to evidence that remains informative for a declared target inference within the response conditions for which the infer-

ence is supported. The term *local* denotes scope and regime dependence; it carries no implication of geographical narrowness or methodological weakness.

This distinction becomes important later because near criticality can preserve local evidential strength while weakening the forward relevance of the inferred relation for intervention. The evidential result need not be reclassified as erroneous merely because the system subsequently enters a different response regime.

2.2 Inferential Support

The transition from evidence to policy begins with an inferential step. A policymaker, analyst, or institution uses observations to support a claim about a response relation. The claim can be causal, predictive, descriptive, comparative, mechanistic, or diagnostic. Different claims impose different empirical burdens. A predictive model can perform well while leaving the manipulable causal mechanism unidentified. A causal estimate can be internally credible while transport to another context remains uncertain. A mechanism can be plausible while its aggregate policy effect remains poorly identified.

The paper therefore uses *response relation* as a deliberately broad term. A response relation is a target-specific association between a perturbation, intervention, condition, or state and a policy-relevant consequence. The relation can be estimated quantitatively, reconstructed qualitatively, or supported through mixed evidence. What matters for the present argument is that the relation has a stated scope.

Definition 2.3 (Response relation). A *response relation* is an empirically supported relation between a specified change and a specified consequence under declared contextual and temporal conditions. Its evidential meaning includes the inferential task, the population or system of reference, the relevant horizon, and the assumptions required for interpretation.

This definition places “effect” alongside other policy-relevant response objects. Some decisions depend on threshold behavior, implementation bottlenecks, distributional exposure, compliance, institutional substitution, network diffusion, or changing expectations. The evidential task is therefore to support the particular response relation that matters for the decision.

Evidence can also support several response relations at once. An intervention may increase an average outcome, produce heterogeneous effects across groups, change administrative burden, alter participation, and generate feedback into later behavior. These relations can be jointly relevant without being reducible to one scalar objective. The evidence-to-action problem consequently begins before criticality: policy choice already requires a way to combine multiple supported claims whose practical implications can differ.

2.3 Decision Translation

Decision translation refers to the institutional and analytical process through which supported claims are converted into a comparison among available actions. Translation introduces information beyond the target estimate. It requires a specification of feasible interventions, decision horizon, implementation constraints, affected populations, legal authority, resources, and relevant losses or values.

Cartwright and Hardie’s account of evidence for use emphasizes that a causal result becomes policy relevant only when the factors required for the causal relation are available in the target setting (Cartwright and Hardie 2012). The same discipline applies within a decision. A response relation can be credible while the intervention needed to exploit it is administratively infeasible, politically unauthorized, poorly timed, or dependent on support factors that are absent. The translation from evidence to policy therefore includes a feasibility structure in addition to an evidential structure.

Decision translation also introduces alternatives. Evidence rarely evaluates an intervention against every feasible option. A positive effect relative to a baseline leaves comparative appraisal among alterna-

tive, staged, smaller, delayed, targeted, or observational strategies open. The action set matters because a policy recommendation is comparative even when the underlying study is not.

Definition 2.4 (Decision translation). *Decision translation* is the process that maps target-relative evidential claims into an appraisal of feasible policy options under stated temporal, institutional, distributive, and implementation conditions.

The definition keeps translation distinct from evidence generation. This functional separation remains compatible with close institutional interaction between research and policy. Analysts can anticipate decision needs, policymakers can shape research questions, and implementation can generate new evidence. The distinction is functional. It identifies the point at which empirical support for a claim becomes one input into a choice among actions.

The distinction also makes disagreement more legible. Two analysts can agree about an estimated response while disagreeing about the preferred action because they differ on feasibility, horizon, distribution, legal authority, reversibility, or acceptable risk. Treating that disagreement as a dispute about evidence can obscure the actual source of difference. Empirically answerable questions retain their status even where value disagreement remains. A disciplined evidence-to-action process should preserve both boundaries.

2.4 Action Warrant

The paper uses *action warrant* for the broader justification required to select or authorize an intervention. Action warrant includes the relevant evidential support but also requires a defensible relation between the intervention and the conditions under which it will operate. Near criticality will later add a specific dynamical qualification to that relation. At this stage, the point is more general: evidence supporting a response claim and warrant supporting an intervention are analytically distinct objects.

Definition 2.5 (Action warrant). *Action warrant* is the context-specific justification for selecting, authorizing, or continuing an intervention given the available evidence, feasible alternatives, expected consequences, implementation conditions, reversibility, distributional effects, institutional authority, and relevant normative commitments.

This definition is intentionally broader than expected-effect maximization. Public policy often operates under rights constraints, procedural requirements, asymmetric harms, irreversibility, and contested objectives. Parkhurst's account of evidence governance is relevant here because the legitimate use of evidence depends partly on the institutions and procedures through which evidence enters collective choice (Parkhurst 2017). A technically credible estimate supplies limited guidance about whose interests count, which losses are acceptable, and which authority may impose them.

The distinction can be stated as a working proposition.

Proposition 2.6 (Inferential support and action warrant). *Strong target-relative evidential support supplies one component of well-grounded action warrant. Full warrant also depends on the intervention, decision horizon, institutional setting, affected parties, and persistence of the response conditions to which the evidence applies.*

This proposition extends evidence-intensive policymaking into additional components of action warrant. Many components of action warrant are themselves empirically investigable. Reversibility can sometimes be estimated; distributional exposure can be measured; implementation failure can be modeled; institutional capacity can be studied; and alternative actions can be compared. The point is that these questions should be asked rather than silently absorbed into the original effect estimate.

A second consequence concerns evidence aggregation. More studies of the same local response can reduce uncertainty about that response while leaving other components of action warrant largely

unchanged. If the principal decision uncertainty concerns irreversibility, institutional legitimacy, or the possibility that intervention changes the governing regime, further precision on the local effect may have diminishing decision value. Evidence accumulation and decision confidence can therefore diverge even before the stronger near criticality claim is introduced.

2.5 Sequential Updating

Evidence-to-action relations are often iterative. An intervention is selected, implemented, observed, and revised. This sequence can support adaptive policy because later observations reveal whether the earlier response relation transported successfully into practice. Monitoring can identify implementation failure, heterogeneous effects, unintended consequences, or changing context. Under favorable conditions, policy becomes a learning process rather than a one-time application of prior evidence.

Sequential updating depends on comparability across time. Later evidence can remain informative under changing conditions when the relevant differences are identified and modeled. Policy learning routinely accommodates changing populations, institutions, technologies, and behavior. What matters is whether the analyst can identify which differences are relevant to the response relation and update the model accordingly. Ordinary nonstationarity can therefore remain compatible with evidence-informed adaptation.

The stronger difficulty appears when intervention changes the response architecture in a way that weakens the continuity required for ordinary updating. An intervention can alter strategic expectations, institutional dependencies, feedback loops, or the accessibility of future states. The post-intervention observations are then informative, but they may describe a response regime that differs materially from the regime under which the earlier evidence was generated. The evidence-to-action loop remains possible, yet the meaning of transport across iterations becomes a substantive question.

This possibility clarifies the role of time in the paper. Evidence is never timeless. Even apparently stable estimates are interpreted over a decision horizon. A short-run relation can remain valid while longer-run adaptation changes the sign, magnitude, or persistence of the effect. A policy can also produce small immediate changes while altering conditions that matter only after several rounds of response. The action warrant therefore depends partly on the horizon over which the relation is expected to remain relevant.

Remark 2.7 (Updating under comparable response conditions). Sequential evidence-informed policy is easiest to interpret when intervention changes outcomes while leaving the response conditions sufficiently comparable for later observations to update the earlier relation. The relevant standard is task-specific comparability across historical change.

Critical-regime instability begins when comparability itself becomes fragile because intervention can materially alter the response field. All policy environments change; the analytical question is whether enough response structure survives for evidence to remain transportable across the intended action horizon.

2.6 Regime Dependence

The evidence-to-action relation can now be stated in its minimal form. Evidence supports a target inference about a response relation. Decision translation places that inference among feasible alternatives and contextual constraints. Action warrant evaluates whether an intervention is sufficiently justified. Sequential updating then tests how the relation behaves in practice. Each stage can remain strongly evidence-informed while addressing a different question.

The remaining issue is the stability of the response regime connecting these stages. Standard evidence-to-action reasoning often allows substantial contextual change because transport can be investigated and models can be updated. Near criticality introduces a more specific concern: an action

can help move the system into a region in which the earlier mapping between perturbation and consequence becomes materially altered. Evidence that was valid for the pre-intervention relation can remain historically correct while its forward decision relevance contracts.

The paper therefore treats regime dependence as a qualification on temporal transport within empirical inquiry. This framing also supplies a falsifiable boundary. If the response relation remains sufficiently stable under intervention, ordinary evidence-to-action updating is adequate. If the relation changes gradually and predictably, standard adaptive methods can remain adequate. The stronger framework becomes relevant only when there is credible reason to investigate transition-sensitive changes in feedback, recovery, coupling, attractor accessibility, or another regime-defining property.

The stronger case begins with critical-regime instability: a condition in which the evidence-to-action relation becomes unstable at the regime level. The analysis distinguishes this condition from ordinary uncertainty, implementation variation, and nonstationarity.

3 Critical-Regime Instability

This section defines the paper’s central dynamical qualification: evidence can remain informative about a local response relation while intervention changes the response regime that gives the relation forward decision relevance. Critical-regime attribution is reserved for cases in which a transition-sensitive mechanism is supported beyond ordinary uncertainty, gradual nonstationarity, implementation variation, or generic complexity.

A policy relation is estimated within an effective response architecture composed of feedback, institutional constraints, expectations, material conditions, strategic responses, and state-dependent mechanisms. Many of these elements can change without defeating ordinary evidence-to-action reasoning. The stronger problem arises when intervention can materially change the architecture itself during the horizon for which the original evidence is expected to guide action. Critical-transition research supplies one family of formal distinctions concerning recovery, susceptibility, and regime change (Scheffer, Bascompte, et al. 2009); their policy use here remains conditional on independent evidence about the response property under study.

3.1 Response Regime

The paper uses *response regime* for the effective configuration that organizes how a policy-relevant system responds to perturbation over a stated horizon. A response regime can be analytically useful even when it lacks the sharply bounded form of a physical phase. It can refer more modestly to a sufficiently persistent configuration of feedback, coupling, recovery, institutional constraint, and strategic expectation within which a response relation can be estimated and transported.

Definition 3.1 (Response regime). A *response regime* is a horizon-specific configuration of transition rules, feedback relations, couplings, constraints, and recovery properties that jointly determine how policy-relevant perturbations are translated into subsequent states and outcomes.

The horizon qualification is important. A regime can be sufficiently stable for a one-week operational decision while changing too rapidly to support a five-year institutional commitment. Likewise, the relevant regime depends on the target inference. A response architecture can remain stable for estimating service uptake while changing in ways that matter for distributional incidence, strategic adaptation, or long-run institutional dependence. Regime identity is therefore analytical and task-relative rather than a universal label attached to the whole political system.

A minimal representation is given in Equation 1.

$$x_{t+1} = F_{R_t}(x_t, u_t, \eta_t), \quad R_{t+1} = G(R_t, x_t, u_t, \eta_t). \quad (1)$$

Here, x_t denotes the policy-relevant state, u_t an intervention, η_t unresolved fluctuation or disturbance, and R_t the effective response regime. The first relation represents state evolution conditional on the current regime. The second makes the regime itself historically variable. The analytical issue arises when u_t contributes materially to the second relation. Policy then affects both the immediate state and the future rule through which later perturbations are translated.

Equation 1 is intentionally generic and permits R_t to remain latent or only partially identified. The regime can be inferred through partial indicators such as changes in recovery time, interaction structure, institutional response, strategic coordination, persistence, or the accessibility of alternative states. The representation is useful because it separates two questions that can otherwise be conflated: what an intervention does *within* a response regime, and what an intervention does *to* that regime.

3.2 Intervention-Coupled Regime Change

An intervention can influence the response regime through several channels. A policy announcement can change expectations before implementation. A subsidy or tax can alter investment patterns that later change the elasticity of response. An emergency measure can create administrative capacity that persists after the initial event. A prohibition can reorganize networks and strategic incentives. A large infrastructure commitment can change which later options are materially or politically reachable. These mechanisms illustrate how policy can participate in constructing the field through which future effects are generated; criticality attribution requires additional evidence of transition sensitivity.

The stronger near-critical case appears when relatively limited perturbations can contribute to a disproportionate change in persistence, feedback, accessibility, or recovery. The scale of the intervention should therefore be separated from the scale of the resulting trajectory change. A small action can remain small in its immediate observable effect while shifting the system toward a different basin of attraction or reinforcing configuration. Conversely, a large intervention can produce a large but temporary response without changing the effective regime. Magnitude alone leaves the underlying response structure underdetermined.

Definition 3.2 (Critical-regime instability of evidence-to-action relations). *Critical-regime instability of evidence-to-action relations* is a condition in which evidence remains informative about a local or current response relation while action based on that relation can materially alter the feedback structure, coupling pattern, recovery properties, attractor accessibility, or other regime-defining conditions in which the evidential relation was estimated.

The instability lies in the relation among evidence, action, and regime. The definition preserves the validity of evidence for the response relation in which it was generated. Evidence can remain methodologically strong for the pre-action relation. The difficulty concerns forward use: the intervention can move the system into a configuration for which the original response estimate has weaker transportability.

Figure 1 contrasts two conceptual structures. The left panel represents an iterative evidence-to-action loop in which policy changes outcomes while the response regime remains sufficiently comparable for ordinary updating. The right panel adds a regime-changing path through expectations, couplings, or feedback. Later evidence remains valuable, but it is generated under a field that may differ materially from the field in which the initial relation was learned.

3.3 Action Relevance under Regime Change

Critical-regime instability separates historical validity from forward action relevance. A study can remain methodologically sound for the response relation observed under R_t while supporting weaker intervention claims after the system has moved toward R_{t+k} . The change lies in transportability, not in retroactive invalidation of the earlier evidence.

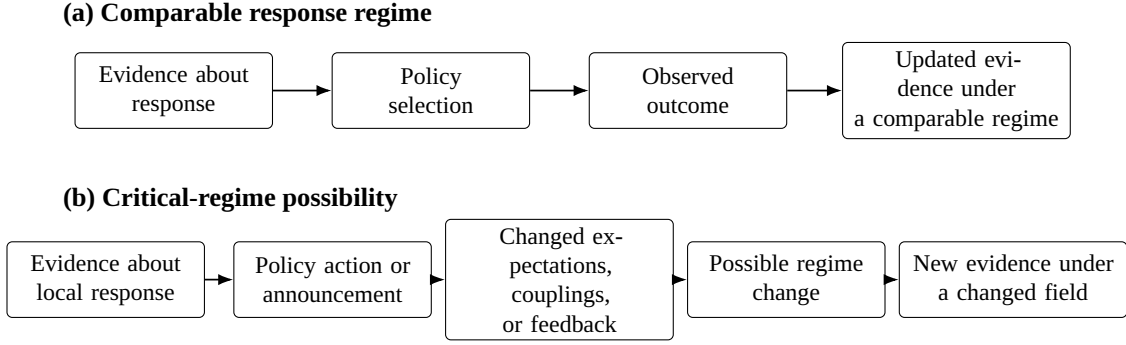


Figure 1: Two evidence-to-action structures. Panel (a) represents iterative updating under a sufficiently comparable response regime. Panel (b) represents the additional possibility that action changes expectations, coupling, feedback, or another regime-defining property, so later evidence is generated under a materially changed field. The diagram is conceptual; transition attribution remains conditional on evidence about the response mechanism.

Proposition 3.3 (Local validity under regime change). *Evidence may remain valid for the response relation and conditions under which it was generated even as support for intervention weakens after material regime change. Historical evidential value and forward transportability are therefore distinct.*

More precise estimation of the local response can therefore leave the main decision uncertainty unresolved. A narrower confidence interval around F_{R_t} in Equation 1 improves knowledge of the current relation while uncertainty about how action affects R_{t+1} can remain large. Research priorities may consequently shift from further local effect estimation toward recovery, coupling, reinforcing feedback, path dependence, or other regime-defining properties.

Near-critical relevance also requires more than volatility or a large observed effect. The response property must be specified. The issue is whether intervention plausibly affects recovery, reinforcing feedback, coupling, or accessibility of persistent trajectories in a way that changes the forward evidence-to-action relation.

Working Claim 3.4 (Transition-sensitive action relevance). Criticality language becomes relevant to evidence-to-action reasoning when there is credible evidence that intervention can affect a transition-relevant property of the response architecture, such as recovery, reinforcing feedback, coupling, or accessibility of persistent trajectories. Short-run effect size or volatility provides only partial evidence for this classification.

3.4 Classification Discipline

Critical-regime instability should therefore be treated as a hypothesis about a particular evidence-to-action relation rather than as a diagnosis of an entire society. The unit of analysis can be a policy subsystem, market, administrative process, ecological-social interface, infrastructure network, or another bounded response configuration. Different subsystems can occupy different regimes at the same time, and an intervention can be transition-relevant for one target while remaining ordinary for another.

Three distinctions are especially important. First, regime instability differs from uncertainty about the current state. A system can be difficult to measure while its response architecture remains stable. Second, regime instability differs from nonstationarity. Gradual and modelable change can often be handled through updating without invoking a critical transition. Third, regime instability differs from policy controversy. Disagreement over values, authority, or distribution can be intense in a dynamically ordinary setting. The present qualification concerns the additional case in which the response architecture itself is plausibly susceptible to intervention-induced change.

Research Obligation 3.5 (Critical-regime attribution). An analysis invoking critical-regime instability should specify the target response relation, the regime-defining property thought to be changing, the intervention channel capable of affecting that property, the relevant decision horizon, plausible non-critical alternatives, and evidence that would count against the critical-regime interpretation.

The framework should be withdrawn when ordinary transport, gradual updating, or a simpler explanation accounts for the observed evidence-to-action relation. When the stronger condition remains plausible, the relevant transport problem concerns whether the response architecture remains sufficiently comparable for forward use of the evidence.

4 Regime Transport

Regime transport isolates the temporal comparability problem created by Section 3. Evidence can remain sound for the source configuration while providing weaker support for action when the target configuration no longer preserves the response architecture relevant to the inference. The issue can arise across jurisdictions, institutions, or periods, and also within the same site after a material regime shift.

Evidence-based policy routinely transports findings beyond the exact conditions under which they were observed. A causal relation estimated in one population can inform another population; a program evaluated in one institution can guide implementation elsewhere; a relation learned in an earlier period can be used for a later decision. Such transport requires attention to the conditions that support the response in the target setting. Cartwright and Hardie emphasize that causal effectiveness depends on the presence of support factors that allow a policy to work in the intended context (Cartwright and Hardie 2012). The present paper extends this concern temporally. The target setting can retain the same name, jurisdiction, administrative boundary, and nominal policy instrument while the response regime governing policy consequences has changed.

The resulting problem is easy to obscure because ordinary indicators of contextual continuity can remain visible. The same ministry may administer the same program. The same legal authority may act on the same population. The same measured variables may be available. Yet strategic expectations, institutional couplings, recovery properties, network structure, or reinforcing feedback can have shifted enough to alter the response relation relevant to action. Regime transport therefore concerns continuity in the effective response architecture, not continuity in labels alone.

4.1 Regime-Indexed Transport

The paper uses *regime transport* for the use of evidence generated under one response regime to support an inference or intervention under another. The concept is target-relative. A regime change can matter for one policy outcome while leaving another response relation sufficiently stable. The relevant comparison is therefore between the response properties needed for the declared decision and the response properties preserved across the transport step.

Definition 4.1 (Regime transport). *Regime transport* is the use of evidence generated under response regime R_a to support a target inference or action under response regime R_b , where the adequacy of transport depends on whether the regime-defining properties relevant to that target remain sufficiently comparable over the stated decision horizon.

Global identity between R_a and R_b is unnecessary. A policy relation can transport despite substantial social change when the mechanisms relevant to the target inference remain stable or their differences are explicitly modeled. Conversely, seemingly modest contextual change can be transport-relevant when it alters a feedback loop, recovery process, strategic response, or institutional dependency that materially shapes the intervention's consequences.

Table 1: Analytical forms of evidence transport relevant to policy action.

Transport form	Primary change	Central transport question	Characteristic failure
Population transport	Composition, exposure, or baseline characteristics	Does the estimated response apply to the target population?	Effect modification or missing support conditions
Institutional transport	Administrative capacity, rules, implementation, or authority	Can the intervention be realized with comparable operative conditions?	Implementation or institutional mismatch
Temporal-context transport	Technology, behavior, environment, or background conditions over time	Has contextual change altered the response enough to require updating?	Stale or weakly comparable evidence
Regime transport	Feedback, coupling, recovery, strategic expectation, or persistent trajectory accessibility	Does the target configuration preserve the response architecture relevant to action?	Evidence remains locally valid while forward intervention relevance changes

Table 1 distinguishes regime transport from neighboring forms of transport. The distinctions are analytical and can overlap in a single policy decision. A single policy decision can require several forms at once.

The last row identifies the distinctive concern of this paper. Regime transport can fail even when the observed population remains similar, the institution remains functional, and the policy mechanism appears recognizable. The failure arises because the response architecture has changed in a way that affects how the same perturbation propagates or persists.

4.2 Same-Site Regime Shift

Regime transport is especially important when spatial or institutional continuity creates an appearance of evidential continuity. Many policy applications use historical evidence from the same jurisdiction, agency, sector, or population. Same-site evidence often deserves substantial weight because it avoids several familiar external-validity problems. Same-site continuity can nonetheless coexist with substantial dynamical change.

Consider a policy instrument whose short-run behavioral effect has been measured repeatedly within one jurisdiction. Over time, repeated use of the instrument can alter expectations, induce organizational adaptation, encourage avoidance, change investment, or generate political counter-mobilization. The measured population remains the same in an administrative sense, while the system in which the intervention operates has accumulated history. The target of transport is therefore a later configuration with different response conditions.

The same structure can arise through background historical change even when the policy has not been repeatedly used. A financial support measure can be evaluated before a market becomes highly interdependent; an emergency communication protocol can be evaluated before public expectations become strongly conditioned by official announcements; a resource allocation rule can be studied before institutions reorganize around that rule. The earlier evidence can remain a valid description of the earlier response. The policy question concerns how much of that response relation survives into the later regime.

Proposition 4.2 (Same-site transport). *Geographic, institutional, or population continuity provides incomplete support for temporal transport when the response architecture relevant to the target inference can change endogenously. Same-site historical evidence therefore requires a regime-comparability as-*

assessment whenever feedback, coupling, recovery, or strategic adaptation plausibly alters the policy response.

The proposition is deliberately conditional. Many same-site applications do preserve enough response structure for historical evidence to remain highly informative. The purpose is to prevent administrative sameness from functioning as an unexamined proxy for dynamical comparability.

4.3 Support Structure

Transport depends on more than the persistence of an average effect. A response relation is generated through a support structure: institutions implement the policy, actors interpret it, complementary resources are available, network positions shape transmission, legal arrangements stabilize commitments, and feedback processes determine persistence. Some components can be directly observed, while others remain latent or only partially reconstructed.

The support structure matters because two regimes can produce similar local effects for different reasons. An intervention may initially work through administrative compliance and later produce a similar short-run outcome through strategic anticipation. The measured effect can appear stable while the underlying response architecture changes. That difference becomes important when policy intensity, duration, or surrounding conditions change, because the two architectures can amplify, saturate, reverse, or recover differently.

A regime-transport assessment should therefore identify the support features that are decision-relevant, keeping reconstruction proportional to the target inference. The required granularity depends on the claim. If the policy decision concerns a modest short-run adjustment, local comparability can be sufficient. If it concerns a large, durable, or difficult-to-reverse intervention, support features associated with persistence and transition become more important.

Definition 4.3 (Transport-relevant support structure). The *transport-relevant support structure* for target inference q is the subset of institutional, behavioral, material, relational, and feedback conditions whose preservation or modeled variation is required for evidence from one response regime to remain adequately informative for q in another.

This definition keeps transport disciplined. The task is to identify the regime differences capable of changing the target relation and determine whether those differences are sufficiently understood for the intended action.

The definition also clarifies the relation to ordinary contextual reasoning in EBPM. Context remains central. Regime transport adds a more specific concern: some support conditions are themselves changed by policy, by anticipation of policy, or by cumulative interaction among policies and actors. The target context is therefore partly endogenous to the evidence-to-action sequence.

4.4 Transport Horizon

Transportability varies with the horizon over which evidence is expected to guide action. A relation can transport over hours or weeks while becoming weak over years. Short-horizon transport can remain reliable because the relevant feedback structure changes slowly relative to the decision. Long-horizon transport can fail because adaptation, accumulation, institutional reorganization, or basin movement becomes material.

This horizon dependence is particularly important near criticality. A small perturbation can produce a modest immediate effect while contributing to a later change in recovery or persistent trajectory accessibility. Evidence about the immediate response can therefore remain accurate at the same time that the longer-horizon policy consequence becomes regime-dependent. The inferential object should specify the horizon so that “the policy effect” remains temporally indexed.

Horizon also interacts with reversibility. A short experiment can generate useful evidence while remaining easy to terminate. A formally similar intervention implemented at scale or maintained for a longer period can alter infrastructure, expectations, investment, or institutional commitments that reduce the feasibility of return. Evidence that transports for the experimental horizon can therefore provide weaker support for indefinite continuation.

Working Claim 4.4 (Horizon-dependent transport). Evidence can transport adequately for a short decision horizon while providing weaker support for a longer commitment when regime-defining feedback, adaptation, or persistence becomes material over time. Transport claims should therefore specify the horizon for which comparability is being asserted.

This claim is compatible with staged policy. When uncertainty concerns the longer-horizon regime response, sequential implementation can sometimes create new information while limiting exposure. The value of staging depends on whether early steps remain sufficiently reversible and whether monitoring can detect the relevant regime changes. These conditions are developed explicitly in the later analysis of reversibility and adaptive policy.

Transport failure and evidence failure refer to different inferential problems. Evidence failure concerns inadequate support for the claimed inference in the source conditions. Transport fails when that earlier inference remains credible but the target configuration has changed enough to weaken the new application. Divergence between historical estimates and later outcomes can therefore provide evidence about regime dependence while the earlier methods remain sound for their source conditions.

Remark 4.5 (Historical validity). A finding can remain historically valid for the regime and horizon in which it was generated while losing action relevance after a material regime shift. Later evidence can revise the domain of applicability while preserving the earlier result within its source regime and horizon.

Replication across changing regimes can accordingly be interpreted as a comparison of response architectures when the studies preserve enough context to identify changing support conditions. This makes temporal heterogeneity an explicit object of explanation.

4.5 Regime-Indexed Updating

Evidence-informed policy near changing regimes requires updating both the estimated response and the assessment of transportability. The first task asks what happened under the observed configuration. The second asks whether the configuration relevant to the next action remains comparable in the dimensions that matter for the target inference. Treating these tasks separately prevents precision in the first from being mistaken for certainty in the second.

A regime-indexed evidence record should therefore retain enough contextual information to reconstruct the response architecture relevant to later use. The required record can include intervention intensity, timing, implementation conditions, institutional constraints, strategic responses, recovery patterns, network changes, and other target-relevant support features. The objective is selective comparability at the granularity required by the target inference.

This approach changes the interpretation of evidence accumulation. More evidence from one regime can strengthen knowledge of that regime while contributing little to uncertainty about a transition into another. Conversely, a smaller amount of evidence spanning distinct response configurations can reveal that an apparently stable relation is regime-dependent. Evidence quantity and transport coverage therefore describe distinct dimensions of the evidence base.

Research Obligation 4.6 (Regime-transport discipline). A policy analysis transporting evidence across time should state the target response relation, the source regime, the decision horizon, the transport-relevant support structure, evidence concerning changes in that structure, and the conditions under which the earlier relation would be revised, restricted, or withheld from forward action.

Regime change can arise from background historical processes, while policy evidence and policy communication can also alter expectations and behavior before or during implementation. That reflexive pathway is examined separately in Section 5.

5 Reflexive Evidence

Reflexive evidence adds a second route through which the evidence-to-action relation can move: publication, measurement, warning, or anticipated intervention can become events inside the system being described. The analysis distinguishes ordinary social reflexivity from the stronger near-critical concern of this paper and reserves critical-regime relevance for cases in which the informational event plausibly changes a transition-relevant response property.

Evidence used in public policy is often observed by more actors than the analyst and decision-maker. Forecasts are published, risk classifications are reported, evaluations circulate among affected organizations, and announced policy intentions alter expectations before implementation is complete. Once these informational events enter the social field, they can affect the very behavior, coordination, and institutional response through which later outcomes are generated. Evidence can therefore operate simultaneously as information about a response relation and as an event inside that relation.

This possibility has long-standing precedents across the social sciences. Merton's account of self-fulfilling prophecy describes situations in which a publicly held definition of a situation changes conduct in ways that contribute to making the definition consequential (Merton 1948). Lucas's critique of econometric policy evaluation similarly highlights the danger of assuming that historically estimated behavioral relations remain invariant when policy rules change the expectations and decisions that generated those relations (Lucas 1976). More recent work on reactivity shows how public measures can reorganize expectations and institutional behavior (Espeland and Sauder 2007), while studies of performativity in financial markets demonstrate that theories and calculative devices can participate in reshaping the environments to which they are applied (MacKenzie and Millo 2003). The present argument draws on this family of problems while retaining a narrower focus on evidence-to-action relations near plausible critical transitions.

5.1 Evidence as a Public Event

Evidence can enter the policy system through several channels. Publication can change beliefs before formal action occurs. Official interpretation can signal which outcomes institutions consider important. A forecast can alter timing, investment, or precaution. A risk classification can change access to resources or insurance. An evaluation can induce organizations to redesign practice in anticipation of future assessment. These effects can leave the underlying evidence valid while changing the subsequent configuration in which the evidence is used.

The distinction is clearest when the same evidential statement has different consequences under different disclosure conditions. A confidential estimate used by a small decision team can leave most actor expectations unchanged. The same estimate released publicly can alter strategic behavior, mobilize support or opposition, and change the baseline against which policy effects are later measured. The content of the evidence is unchanged; its position in the causal and relational architecture has changed.

Definition 5.1 (Reflexive evidence). *Reflexive evidence* is evidence whose production, disclosure, interpretation, or anticipated use can alter actor expectations, behavior, coordination, institutional response, or other conditions that materially enter the response relation to which the evidence is subsequently applied.

The definition is deliberately broader than strategic gaming. Actors can respond in good faith to new information. A credible warning can induce precaution. A published evaluation can motivate genuine

Table 2: Reflexive pathways connecting evidence to subsequent response conditions.

Pathway	Informational event	Possible behavioral or institutional response	Evidence-to-action concern
Expectation update	Forecast, risk estimate, or public interpretation	Anticipation, precaution, delay, acceleration, or repricing	The pre-disclosure response relation can change before intervention
Measurement reactivity	Ranking, indicator, target, or evaluation criterion	Effort reallocation, category adaptation, strategic reporting, or organizational redesign	The measured construct and the behavior producing it can co-evolve
Policy anticipation	Announced rule, likely intervention, or implementation signal	Pre-emptive compliance, avoidance, investment, resistance, or coalition adjustment	Part of the policy effect occurs before nominal implementation
Performative use	Model, classification, or calculative device becomes embedded in practice	Coordination around the representation changes routines or market structure	The model can contribute to conditions that later make it more or less accurate
Feedback through interpretation	Evidence is contested, amplified, or institutionally endorsed	Mobilization, trust revision, legitimacy change, or altered enforcement	Evidential interpretation becomes part of the response architecture

improvement. A forecast can support coordination. Reflexivity concerns endogeneity between information and the field of action; moral evaluation remains a separate question.

Table 2 distinguishes several pathways through which evidence can become dynamically active. The pathways can overlap in a single policy episode.

5.2 Expectation-Mediated Response

Expectations provide the most direct bridge between evidence and altered response. A policy estimate can influence behavior before the policy is enacted if affected actors believe the estimate will guide future action. Anticipatory responses can then change the state on which implementation operates. This is especially important when the intervention depends on timing, coordination, or strategic interaction.

The policy analyst can therefore face two different inferential objects. The first is the response that would occur if the evidence remained observationally passive. The second is the response after relevant actors have learned, interpreted, or reacted to the evidence. In some domains the difference is small. In others it is part of the policy mechanism itself. Public-health warnings, financial guidance, environmental risk communication, enforcement priorities, and resource-allocation announcements can all create anticipatory behavior that belongs to the intervention environment rather than to a separate background context.

Lucas's critique is relevant here as a methodological analogy for one expectation-mediated mechanism of policy reflexivity. If behavior depends on expectations about policy, estimated relations under one policy rule can provide weak guidance after the rule or the expectation structure changes (Lucas 1976). The broader lesson for this paper is that the evidential relation should be indexed to the expectation regime when expectations materially shape response. A coefficient estimated before disclosure can remain statistically well estimated while providing incomplete guidance after disclosure has altered the behavioral field.

Proposition 5.2 (Expectation-conditioned transport). *When actor expectations materially enter the policy response, transport from pre-disclosure evidence to post-disclosure action requires an assessment of*

whether the informational event has changed the expectation structure relevant to the target outcome. Post-disclosure transportability therefore depends on the expectation structure in addition to the precision of the original estimate.

Public disclosure can remain appropriate and may be required for public reason, accountability, contestability, and coordination. The analytical consequence is narrower: disclosure can be part of the treatment environment and should be represented accordingly when it changes the response architecture.

5.3 Measurement Reactivity

Reflexivity also appears when evidence is generated through measures that become targets of organizational attention. Espeland and Sauder use the concept of reactivity to analyze how public rankings reshape expectations and institutional behavior, including changes induced by self-fulfilling processes and commensuration (Espeland and Sauder 2007). The relevance for EBPM extends across rankings, indicators, thresholds, targets, eligibility criteria, and related evaluative devices. Indicators, thresholds, targets, eligibility criteria, risk scores, and performance metrics can all change behavior once actors learn how they are used.

Measurement reactivity creates a distinction between improving the measured quantity and improving the policy-relevant condition the measure was intended to represent. The two can remain aligned, partially diverge, or become coupled in ways that change over time. A target can induce genuine improvement while also encouraging substitution toward more measurable activities. A risk score can improve triage while generating category migration around a threshold. An evaluation criterion can strengthen accountability while narrowing attention to what is formally assessed.

These possibilities matter for evidence-to-action inference because repeated use of a measure can change its diagnostic environment. Historical validation of an indicator can remain informative about an earlier period while later strategic or institutional adaptation changes the mapping between the indicator and the latent condition of interest. The problem therefore combines reflexivity with the regime-transport issue developed in Section 4.

Working Claim 5.3 (Reactive diagnosticity). When a policy-relevant measure becomes consequential for the actors it classifies or evaluates, its evidential meaning can become historically conditional on the behavioral and institutional responses induced by that use. Longitudinal validation should therefore examine whether the measure–target relation changes as the measure becomes embedded in decision practice.

This claim leaves room for stable indicators. Many measures remain useful after becoming policy-relevant, especially when manipulation is difficult, the target is materially anchored, or the measure is one component of a broader assessment. The point is to treat reactivity as an empirical possibility whose presence and direction require evidence.

5.4 Performative Embedding

A stronger form of reflexivity occurs when representations become integrated into the practices that organize the domain. MacKenzie and Millo’s historical study of options markets shows how economic theory and pricing practice could contribute to market changes that made the theoretical representation more accurate in important respects (MacKenzie and Millo 2003). The example is domain-specific, yet it illustrates a general analytical possibility: a model can participate in producing the regularities against which it is later evaluated.

For policy analysis, performative embedding can arise when classifications, benchmarks, scenario models, or risk frameworks become institutional operating rules. The representation then has at least two roles. It summarizes an observed field, and it helps coordinate future action within that field. Subsequent

fit between model and observation can partly reflect this institutional embedding. Likewise, later model failure can arise because embedding produces counteracting adaptation, saturation, or resistance.

This dual role complicates simple notions of external validation. A model that changes its environment can still be useful and empirically disciplined. The relevant question becomes whether the model's use preserves, strengthens, or changes the response relation for which policy inference is required. Evidence of performativity therefore belongs inside the regime description as a mechanism shaping the response relation.

5.5 Reflexive Amplification near Criticality

Reflexivity alone supplies limited evidence about near criticality. Expectations can change, metrics can induce adaptation, and theories can become performative in systems that remain dynamically stable. The near-critical qualification enters only when these reflexive effects couple to a configuration with elevated susceptibility, weak recovery, competing self-reinforcing trajectories, or transition-sensitive feedback. Under those conditions, an informational event that would ordinarily produce a modest adjustment can contribute to a larger change in trajectory accessibility or persistence.

Amplification provides a useful way to express the distinction. In a stable regime, publication of a risk estimate can update behavior while the governing response structure remains comparable. In a transition-sensitive regime, the same update can alter coordination, expectations, or institutional commitments enough to change which future responses are accessible. The evidence can then participate in shifting the regime to which it is subsequently applied.

Critical-transition research motivates attention to changing recovery and susceptibility near some tipping processes (Scheffer, Bascompte, et al. 2009). The present paper treats that dynamical lesson as conditional. Where empirical evidence supports a transition-sensitive social response architecture, reflexive informational events should be treated as possible perturbations to that architecture.

Working Claim 5.4 (Reflexive amplification). If a policy domain is already supported by independent evidence of elevated transition sensitivity, evidence disclosure or policy signaling can require analysis as a perturbation to the response regime itself. The warrant for this analysis should be grounded in demonstrated susceptibility and coupling structure, with reflexivity treated as the perturbation channel.

This claim keeps two classifications separate. Public evidence can enter the response process endogenously, while anticipation, reactivity, or performativity remain distinct from evidence of a critical transition unless the relevant susceptibility mechanism is independently supported. Reflexivity and near criticality are distinct properties whose conjunction creates the stronger evidence-to-action problem.

5.6 Reflexive Classification Discipline

A practical analysis should identify the channel through which evidence enters the system and the time at which the channel becomes active. Did actors observe the result? Was it institutionally endorsed? Did they expect policy to follow? Could they change behavior before implementation? Did the measure become a target? Did organizational routines incorporate the model? These questions help separate reflexive mechanisms from ordinary contextual drift.

Counterfactual comparison is also important. The explanatory contribution of the informational event weakens when similar changes plausibly arise under non-publication, non-measurement, or absent policy anticipation. Conversely, a clear temporal sequence from disclosure to expectation change, behavioral adjustment, and altered response structure strengthens the case for reflexive interpretation. Natural experiments, staggered disclosure, confidential-versus-public evidence comparisons, announcement windows, and institutional process tracing can all contribute to such identification when appropriate.

Research Obligation 5.5 (Reflexive evidence attribution). A claim that evidence has reflexively altered an evidence-to-action relation should specify the informational event, the actors exposed to it, the behavioral or institutional response channel, the temporal sequence, plausible alternative explanations, and the response property that changed. Near-critical amplification should be attributed only when transition sensitivity is independently supported.

Reflexivity therefore adds a distinct pathway of regime movement: evidence and anticipated policy can become perturbations within the process being evaluated. The remaining question is how to separate the persistence of local evidential validity from changes in reversibility and warrant.

6 Local Action Admissibility

This section formalizes the paper’s core separation among evidential validity, reversibility, and normative warrant. A neighborhood representation is used only to express target-relative local support: the region in which evidence remains informative can differ from the region in which intervention remains recoverable or justified. The representation is provisional and can be implemented through discrete alternatives, network states, policy portfolios, or other domain-appropriate structures.

6.1 Target-Relative Local Validity

Evidence is locally valid when it remains sufficiently informative for a specified inferential task within a delimited neighborhood of states, interventions, or contexts. The task can concern sign, magnitude, ranking, mechanism, risk, or another explicitly stated target. Local validity therefore depends on what the evidence is being asked to support. A study can remain informative about an immediate directional effect while offering much weaker support for a long-horizon trajectory claim.

This target-relative formulation aligns with broader arguments in evidence-based policy that evidence must be relevant to the causal role a policy is expected to play and to the support factors present in the target setting (Cartwright and Hardie 2012). The present framework adds a dynamical qualification. Even where those support factors are presently adequate, intervention can modify them. Local validity should therefore be indexed to the response regime in which the evidence is interpreted and to the horizon over which that interpretation is intended to remain useful.

Definition 6.1 (Local evidential validity). Let q denote a declared inferential target and let x_t denote the present configuration. Evidence has *local evidential validity* for q over a neighborhood of x_t when the available evidence remains sufficiently informative for q across that neighborhood under a stated adequacy criterion.

The definition deliberately leaves the adequacy criterion open. Different applications may use prediction error, posterior uncertainty, qualitative direction, rank stability, calibration, or another criterion. The important requirement is explicitness. A neighborhood adequate for estimating a short-run average response may provide weak coverage for forecasting the persistence of a new institutional trajectory.

6.2 Evidence-Validity Neighborhood

Let $N_E(x_t; q)$ denote the set of nearby configurations or candidate interventions for which the current evidential representation remains sufficiently informative for target q . The neighborhood is epistemic; normative conditions enter separately through N_W . Membership says that the evidence continues to support the stated inference to the required standard. Reversibility, legitimacy, distributional acceptability, and institutional authorization remain separate conditions.

Several mechanisms can contract N_E . Support factors can disappear, composition can change, measurement relations can shift, actors can adapt strategically, or the intervention itself can alter the response

regime. Conversely, new measurements, improved models, or additional observations can expand the neighborhood by strengthening discrimination among plausible responses. The resulting movement is empirical and can proceed in either direction.

Regime transport from Section 4 can be restated through this object. Historical evidence may remain valid for the regime in which it was obtained while the present target configuration moves outside the neighborhood in which that evidence transports adequately. This is an epistemic contraction. The stronger claim developed below concerns a different possibility: N_E can remain comparatively broad while the set of interventions that preserve reversibility or satisfy stated warrant conditions becomes narrower.

Remark 6.2 (Evidence retention). A contraction of the action-admissible region can occur while N_E remains broad. High-quality evidence may retain descriptive and inferential value even as reversibility and warrant support fewer interventions.

6.3 Reversibility Neighborhood

Let $N_R(x_t; h)$ denote the set of candidate interventions expected to remain sufficiently reversible over decision horizon h relative to stated loss, recovery, and option-preservation criteria. Reversibility here is broader than literal restoration of an earlier state. It can include the practical possibility of withdrawing an intervention, recovering institutional capacity, restoring previously accessible alternatives, or preventing a temporary perturbation from becoming a persistent trajectory commitment.

The horizon parameter matters. An intervention can appear reversible over weeks while changing investments, expectations, legal commitments, organizational routines, or coalition structures in ways that are difficult to reverse over years. The same intervention can therefore belong to N_R for one horizon and fall outside it for another. Section 7 develops this problem in detail.

Critical-transition research supplies a formal motivation for separating local response from recovery properties. In some dynamical systems, declining recovery capacity and increasing sensitivity can precede regime shifts (Scheffer, Bascompte, et al. 2009). The present argument uses this distinction only conditionally for political systems: where an empirical policy domain exhibits credible evidence of changing recovery, persistence, or basin accessibility, reversibility deserves independent analysis alongside local effect estimation.

Definition 6.3 (Dynamical reversibility neighborhood). The *dynamical reversibility neighborhood* $N_R(x_t; h)$ is the set of candidate interventions for which available evidence and model uncertainty support a reasonable expectation that relevant losses, commitments, or trajectory changes remain recoverable within horizon h under the stated reversibility criteria.

Uncertainty is part of the neighborhood specification. A policy can be potentially reversible in principle while the decision maker lacks enough information to rely on that reversibility. The relevant object is therefore the supportable expectation of recoverability under stated criteria.

6.4 Warrant Neighborhood

Evidence-to-action reasoning also depends on criteria beyond effect estimation and dynamical recoverability. Let $N_W(x_t; q, h)$ denote the set of candidate interventions for which there is adequate action warrant under the stated normative, legal, procedural, distributional, and authority conditions relevant to the decision. The subscript W refers specifically to warrant. The substantive content of this neighborhood is supplied by the applicable legal, institutional, and ethical framework.

This separation is important for neutrality. A policy can be well supported empirically and dynamically reversible while remaining unauthorized, procedurally defective, rights-infringing under the applicable framework, or unjustified given the distribution of burdens and benefits. Conversely, an emergency

mandate or other legitimate authorization can alter the warrant conditions without changing the underlying evidence-validity neighborhood. The determinants of N_W are therefore partly independent of the determinants of N_E and N_R .

The distinction is consistent with evidence-governance scholarship that treats the quality of evidence and the quality of evidence use as related but separate questions (Parkhurst 2017). In the present framework, normative warrant is represented as a distinct condition on intervention admissibility whose content should be specified by the relevant legal, institutional, and ethical framework.

Definition 6.4 (Warrant neighborhood). The *warrant neighborhood* $N_W(x_t; q, h)$ is the set of candidate interventions that satisfy the stated normative, legal, procedural, distributional, and authority conditions for acting on target inference q over horizon h .

The definition leaves ranking among competing normative frameworks to the applicable decision context. Its analytical purpose is to keep evidential strength, political or ethical authorization, and empirical support visibly distinct.

6.5 Action-Admissible Region

The paper’s central distinction can now be represented compactly. The locally action-admissible region is defined by Equation 2.

$$N_A(x_t; q, h) = N_E(x_t; q) \cap N_R(x_t; h) \cap N_W(x_t; q, h). \quad (2)$$

In Equation 2, N_A contains candidate interventions that satisfy all three conditions simultaneously. The equation is a conjunction; optimization requires a separate decision rule. It leaves policy ranking, common units, and relative estimation precision to separate analysis. It states only that evidential validity, supportable reversibility, and warrant are analytically distinct conditions for the action class considered here.

Proposition 6.5 (Conjunctive admissibility). *For evidence-to-action reasoning under possible regime change, expansion of N_A depends on the joint movement of the neighborhoods in Equation 2. An expanded evidence-validity neighborhood can coexist with a contracted reversibility or warrant neighborhood.*

The proposition blocks a monotonic inference from “better evidence” to “more intervention.” Better evidence can expand the set of claims that are supportable while simultaneously revealing that intervention has greater leverage, persistence, distributional consequence, or irreversibility than previously understood. In that case, knowledge improves while admissible action contracts.

6.6 Neighborhood Divergence

Near criticality, the three neighborhoods can diverge because their boundaries depend on different properties of the system. N_E depends primarily on inferential adequacy for the target task. N_R depends on recovery, persistence, hysteresis-like effects, and option preservation over a stated horizon. N_W depends on the normative and institutional conditions under which intervention is authorized and justified.

The relevant near-critical possibility is therefore asymmetric. A system can remain locally legible while becoming dynamically harder to intervene in safely. Improved estimation of a local response can preserve or enlarge N_E while evidence of increasing susceptibility, slow recovery, or path dependence contracts N_R . Likewise, widening distributional stakes or procedural requirements can contract N_W while empirical measurement remains unchanged.

Proposition 6.6 (Neighborhood divergence near criticality). *As transition sensitivity becomes more plausible, an epistemically credible local neighborhood can remain broader than the region in which intervention is supportably reversible or warranted. Local knowledge can therefore survive a contraction of the action-admissible region.*

This proposition is conditional. Near criticality can leave one or more neighborhoods broad. Some transitions remain reversible, some interventions are weakly coupled to the transition mechanism, and some decision settings provide strong prior authorization for action. The claim is that divergence becomes an empirical possibility requiring direct testing even when evidential confidence is high.

The distinction also clarifies why criticality language should remain scale-specific. A subsystem can exhibit elevated susceptibility while the broader institutional environment supplies buffers that preserve reversibility. The reverse can also occur: a locally stable subsystem can be embedded in a larger configuration whose transition changes the meaning of local evidence. Neighborhood estimation should therefore specify the scale at which evidence, reversibility, and warrant are being assessed.

6.7 Horizon Dependence

The three neighborhoods can move differently with the decision horizon. Short-horizon effect evidence may be strong because the response relation changes slowly over the first few periods. Reversibility can nevertheless decline with horizon as commitments accumulate. Normative warrant can also change as affected populations, opportunity costs, or institutional mandates evolve.

Letting the horizon remain explicit prevents one temporal scale from dominating the evidence-to-action judgment. A policy supported for immediate stabilization can require a different evidential and warrant structure if maintained as a long-term institutional settlement. Likewise, an intervention that is difficult to justify as a permanent redesign may be admissible as a bounded trial when exit conditions, monitoring, and recovery capacity are credible.

This temporal structure also provides a bridge to adaptive policy. An intervention can be evaluated as a sequence of commitments with distinct review points. Staged commitments, review points, reversible pilots, and pre-specified stopping conditions can change the shape of N_R and, through procedural safeguards, sometimes the shape of N_W . Whether they do so is empirical and institutional. The framework therefore directs attention toward how action architecture changes admissibility, alongside the substantive option selected.

Research Obligation 6.7 (Neighborhood specification). An analysis that invokes local validity near criticality should specify the inferential target, evidence-validity criterion, decision horizon, reversibility criterion, warrant framework, relevant scale, and principal uncertainties affecting the boundaries of N_E , N_R , and N_W . A claim of neighborhood divergence should identify which boundary moved and what evidence supports that movement.

Local validity therefore yields the paper’s central qualification in formal form. Evidence can remain informative within a comparatively broad N_E while the intersection in Equation 2 contracts. This contraction places reversibility at the center of the subsequent analysis, especially where persistent transition and trajectory closure can make recovery progressively more difficult.

7 Reversibility

Reversibility specifies the second condition in the action-admissibility framework of Section 6. The relevant question is whether target-relevant capacities, options, or conditions remain recoverable over the stated horizon when intervention can alter the response regime. This target-relative conception separates formal withdrawal from recovery and leaves room for the costs of excessive flexibility where policy effectiveness depends on durable expectations.

Table 3: Target-relative dimensions of reversibility.

Dimension	Recovery object	Illustrative failure mode
Instrument exit	Ability to terminate or modify the formal intervention	A measure can be legally withdrawn but implementation infrastructure remains difficult to dismantle.
State recovery	Return of target-relevant system properties toward a pre-intervention or acceptable region	The measured outcome recovers slowly or settles in a different persistent configuration.
Institutional rollback	Restoration of procedures, authorities, organizational capacities, or governance relations	Temporary institutional arrangements become embedded through routines, contracts, or authority shifts.
Option preservation	Retention or recovery of meaningful future alternatives	One intervention closes pathways that later evidence would have made preferable.
Distributional repair	Recoverability of losses or burdens borne by affected groups	Aggregate conditions recover while concentrated harms remain uncompensated or practically irreversible.

Reversibility matters in this paper because near criticality can increase the distance between local effect knowledge and recoverability. An intervention may have a well-estimated immediate effect while changing investments, expectations, institutions, coalition structure, or accessible alternatives in ways that persist after the intervention is withdrawn. The evidential question and the recovery question therefore remain distinct. The former asks what the intervention is expected to do. The latter asks what remains possible after it has done it.

7.1 Operational Reversibility

Literal restoration of a previous state is often unavailable in social and institutional settings. Information has been revealed, resources have moved, legal acts have occurred, organizations have adapted, and affected actors have changed their expectations. Reversibility is therefore target-relative: consequences relevant to the stated decision problem may be recoverable, repairable, or reopenable within an acceptable horizon and cost even when the earlier state cannot be recreated literally.

Definition 7.1 (Operational reversibility). An intervention is *operationally reversible* relative to target q and horizon h when withdrawal, modification, or compensating action can recover the target-relevant capacities, options, or conditions required by the stated reversibility criterion within horizon h , with uncertainty and recovery costs remaining within the accepted bounds.

Operational reversibility can concern several different objects. A policy instrument may be easy to cancel while the institutional structure created around it remains durable. A subsidy can expire while investments induced by the subsidy remain sunk. A temporary emergency rule can lapse while expectations, organizational routines, or litigation alter later governance. A market intervention can be withdrawn while distributional losses remain concentrated among actors with limited recovery capacity. The reversibility analysis should therefore specify what is supposed to be recoverable.

Table 3 separates five dimensions that can otherwise be conflated.

The dimensions can diverge. A policy can score highly on instrument exit and poorly on option preservation. Another can be institutionally persistent while producing easily repairable material effects. Reversibility should therefore be reported as a profile relative to the target decision rather than as a single scalar property.

7.2 Recovery Path

Reversibility is a path property. Ending an intervention can leave the return path materially different from the departure path. In dynamical systems, hysteresis can arise when the state reached after a transition persists even after the forcing variable returns toward an earlier level. Ecological research on alternative stable states provides a well-known formal and empirical illustration of such path asymmetry (Scheffer, Carpenter, et al. 2001). The present paper uses the concept conditionally for political systems to identify a policy-relevant possibility: institutional dismantling, sunk investment, changed expectations, legal entrenchment, network reorganization, or altered coalition structure can make recovery follow a different path from intervention.

Proposition 7.2 (Withdrawal-recovery separation). *Formal withdrawal and target-relevant recovery are distinct. Reversibility analysis should therefore evaluate the recovery path separately from the exit mechanism, including restoration of relevant state, institutional capacity, distribution, or option set.*

This separation becomes important near a plausible transition because small differences in timing can change the recovery environment. An intervention withdrawn before expectations and investments reorganize can be easier to reverse than the same intervention withdrawn after complementary commitments accumulate. The relevant unit is therefore the intervention history, including duration, sequencing, feedback, and accumulated commitment.

The concept of hysteresis should remain empirically disciplined. Persistence alone provides limited evidence for hysteresis, and historical path dependence can arise through mechanisms outside bifurcation or alternative-stable-state models. Where direct dynamical identification is unavailable, the paper will use terms such as *recovery asymmetry*, *path-dependent persistence*, or *hysteresis-like persistence*. This keeps the analogy visible without converting metaphor into diagnosis.

7.3 Option Preservation

A second dimension concerns the future actions that remain available. An intervention can be locally effective and still reduce future choice by creating sunk costs, specialized infrastructure, legal commitments, dependency relations, or expectations that make alternative trajectories increasingly expensive. Reversibility therefore includes an option-space component.

The economic literature on irreversibility under uncertainty supplies a useful analytical precedent. Arrow and Fisher showed that uncertainty can give additional value to preserving an irreversible environmental option (Arrow and Fisher 1974). Henry developed the irreversibility effect for investment decisions under uncertainty (Henry 1974), while Dixit and Pindyck systematized the option value of waiting when investment is partly irreversible and new information can arrive (Dixit and Pindyck 1994). These models address specific economic structures. Their narrower lesson is relevant here: under uncertainty, retaining the capacity to act later can itself have decision value when present commitment is difficult to undo.

Definition 7.3 (Option preservation). *Option preservation* is the retention of materially distinct future actions or trajectories whose accessibility could matter after additional information, institutional learning, or regime change.

Option preservation is target-relative. Keeping every imaginable future open is neither feasible nor necessarily desirable. Some options conflict with rights, legal obligations, or collective commitments; some options impose continuing costs merely by being maintained. The analytical task is to identify which future alternatives have decision relevance and whether the intervention closes them before the uncertainty that distinguishes among them is likely to resolve.

This distinction also prevents the reversibility argument from collapsing into generalized precaution. The framework treats delay as one option whose value depends on timing, uncertainty, and transition

risk. Delay can impose large costs, allow harmful dynamics to deepen, or eliminate an opportunity for effective intervention. The relevant comparison is between the value of current action and the value of retaining future response capacity under the actual timing, uncertainty, and transition risks of the domain.

7.4 Learning through Staged Commitment

Where uncertainty can be reduced through observation, staged commitment can preserve recovery capacity while generating information before larger or more persistent commitments are made. The real-options literature motivates the value of retaining choice before irreversible investment (Dixit and Pindyck 1994); policy staging extends the intuition to interventions whose early phases can reveal response properties. The analogy remains conditional because pilots can alter expectations, create constituencies, or change institutions and therefore function as interventions as well as observations.

Proposition 7.4 (Staged-learning value). *When an intervention can be decomposed into meaningful stages and early stages generate information relevant to later commitment, a staged architecture can increase the informational value of action while reducing the amount of irreversible exposure incurred before that information arrives.*

The proposition has important boundary conditions. The early stage must be informative about the later decision; scale effects must remain sufficiently understood; actors must have credible capacity to stop or modify the program; and the delay created by staging must remain within acceptable loss bounds. Where these conditions fail, staging can create false confidence or merely postpone unavoidable commitment.

Reversible pilots also raise a transport problem. Evidence from a limited trial can fail to transport to full implementation if the scaled intervention changes expectations, strategic response, or network structure. Reversibility can therefore improve the learning process while the regime-transport problem developed in Section 4 remains active.

7.5 Commitment Trade-offs

Reversibility and commitment serve different policy functions. Some interventions work partly because affected actors expect them to persist. Long-lived infrastructure, legal guarantees, institutional independence, insurance arrangements, and other policy structures can derive part of their effect from credible duration. Economic work on time consistency illustrates how expectations about future policy can alter present behavior even when policymakers possess accurate information about current conditions (Kydland and Prescott 1977).

This creates a trade-off. Greater exit flexibility can preserve options and reduce lock-in, while reduced commitment can weaken coordination, investment, compliance, or the credibility of the intervention itself. A policy designed for maximal reversibility can therefore become less effective or generate different behavior from the policy whose evidence base is being used to justify it.

Remark 7.5 (Reversibility neutrality). A larger reversibility neighborhood is treated as one decision property within the action-admissibility framework; its value remains context dependent. Some decisions can warrant durable commitment when the relevant evidence, authority, distributional conditions, and risks support that structure.

The trade-off is particularly important near criticality because commitment can itself reshape the response regime. Durable commitments can stabilize expectations and reduce susceptibility in one setting while deepening lock-in in another. Empirical analysis should therefore examine how commitment changes the relevant feedback structure and evaluate flexibility and permanence relative to the target decision.

7.6 Reversibility Architecture

Reversibility can be partly designed. The relevant architecture includes exit procedures, modularity, monitoring, review thresholds, reserve capacity, fallback institutions, legal rollback mechanisms, compensation arrangements, and resources for recovery. These features can change the practical boundary of N_R even when the substantive policy objective remains the same.

A useful distinction is between *intrinsic reversibility* and *supported reversibility*. Intrinsic reversibility arises from the intervention's own technical or institutional properties. Supported reversibility arises because additional capacities make recovery feasible: redundant infrastructure, financial reserves, interoperable systems, retained expertise, legal review, data retention, or pre-negotiated compensation. Supported reversibility can therefore require investment before intervention.

This point connects reversibility to governance capacity. An intervention that is safely reversible for a well-resourced institution can be effectively irreversible for another institution lacking monitoring, fiscal space, administrative continuity, or recovery infrastructure. Reversibility is therefore partly relational: it depends on the actor attempting recovery and the capacities available at the relevant time.

Near criticality, these capacities can themselves degrade. A transition can remove personnel, funding, authority, or institutional alternatives that had made reversal possible at the moment of adoption. Reversibility assessment should therefore ask whether the recovery mechanism is robust to the same transition risk that motivates caution about the intervention.

Research Obligation 7.6 (Reversibility specification). An evidence-to-action analysis that relies on reversibility should specify the recovery object, decision horizon, acceptable recovery loss, expected recovery path, principal sources of path dependence, relevant future options, institutional recovery capacity, and the conditions under which commitment itself contributes to policy effectiveness. Claims of reversibility should distinguish formal withdrawal from target-relevant recovery.

Reversibility qualifies the move from evidence to action while leaving the direction and timing of action open. It asks whether intervention can be unwound, repaired, or redirected when later evidence changes the preferred trajectory.

8 Evidence-Restraint Inversion

Evidence-restraint inversion captures a conditional divergence between informational confidence and intervention intensity. Improved evidence can reveal unusually high susceptibility, greater policy leverage, or a narrower recovery path, thereby strengthening the case for bounded exposure, staging, monitoring, or restraint. The concept remains conditional: stronger evidence can also support stronger intervention when delay or under-action carries the larger irreversible cost.

The central issue is the direction in which uncertainty reduction changes the action problem. Suppose early evidence leaves the magnitude of policy leverage unclear. Additional evidence can show that the intervention is weakly coupled to the relevant transition mechanism, in which case greater confidence can support implementation. The same process of learning may show that a small intervention materially shifts feedback, expectations, or basin accessibility. The informational gain is similar in form: uncertainty about the response relation has decreased. Its action implication differs because the estimated consequences of exercising that leverage have changed.

8.1 Intervention Intensity under Greater Confidence

Evidence-to-action reasoning can implicitly couple two quantities that should remain distinct: confidence in an estimated response relation and the intensity of intervention justified by that relation. In ordinary settings the two can move together. Better evidence that a treatment is effective can support broader

adoption; stronger evidence that an enforcement mechanism changes behavior can support more confident deployment; improved estimates of infrastructure demand can support larger investment. These examples create a familiar pattern in which improved evidence expands the region of action that appears supportable.

Near criticality, improved estimation can reveal a different structure. High susceptibility means that relatively small perturbations can produce large or persistent changes. Evidence of such leverage can improve prediction of the immediate response while simultaneously increasing uncertainty about the post-intervention regime, the recovery path, or the distribution of downstream effects. The relevant action question therefore becomes partly architectural: how much exposure should be created before the system's response and recoverability are better understood?

Definition 8.1 (Evidence-restraint inversion). An *evidence-restraint inversion* occurs when credible new evidence increases confidence that an intervention has substantial leverage, transition sensitivity, or persistent consequences, and the informational gain strengthens the case for bounded exposure, preserved reversibility, closer monitoring, staged commitment, or delayed escalation.

The term *inversion* refers to the relation between evidential confidence and intervention intensity. Additional evidence remains valuable precisely because it clarifies a response property that changes the appropriate intervention architecture. Restraint can therefore be evidence-responsive.

8.2 Susceptibility as Decision-Relevant Information

Critical-transition research motivates one route to the inversion. Changes in recovery rate, variance, autocorrelation, or related indicators can under some conditions provide information about declining resilience or increasing susceptibility (Scheffer, Bascompte, et al. 2009). These indicators remain fallible and domain dependent. When converging evidence nevertheless supports a credible increase in susceptibility, the policy implication depends on how the contemplated intervention couples to the transition mechanism.

If the intervention is weakly coupled, improved evidence about susceptibility can have little effect on action admissibility. If the intervention is strongly coupled, the same evidence can increase the expected consequence of a small policy error. A policy that appeared incremental under an earlier response regime can become dynamically consequential under the revised estimate. The relevant evidential gain therefore concerns both expected effect and the sensitivity of the surrounding trajectory to intervention.

This distinction also helps separate evidence-restraint inversion from generalized precaution. The inversion requires a specific evidential development: stronger information about leverage, transition sensitivity, persistence, or recovery difficulty. Uncertainty, novelty, controversy, or political disagreement supplies no specific evidence for the pattern without a leverage or persistence mechanism. The restraint must be connected to a response property that the evidence has made more credible.

Proposition 8.2 (Susceptibility-sensitive restraint). When credible evidence indicates that a contemplated intervention is strongly coupled to a transition-sensitive response mechanism and that recovery capacity is limited or uncertain, improved estimation of intervention leverage can increase the evidential case for limiting initial exposure even while confidence in the local response relation rises.

The costs of delay and under-action condition this proposition. A rapidly worsening process can make restraint costly; a narrow intervention window can make early commitment necessary; and some transition-sensitive systems may require prompt action precisely because later recovery becomes harder. Evidence-restraint inversion therefore concerns the structure of the decision problem and leaves the preferred direction of action open.

Stronger evidence of leverage can also increase the value of exit capacity because the cost of a mistaken trajectory shift has become more credible. Section 7 develops the underlying logic of option preservation and staged commitment; here the point is narrower. Evidence-restraint inversion can change the

Table 4: Illustrative relations between evidential improvement and intervention architecture.

Evidential development	Response interpretation	Potential action implication
More precise local effect estimate	Stable response regime remains plausible	Greater confidence in ordinary implementation or comparison among interventions
Stronger evidence of high susceptibility	Small perturbations can produce larger or more persistent changes	Smaller initial exposure, stronger monitoring, or staged commitment
Improved evidence of difficult recovery	Post-intervention restoration appears costly or path dependent	Greater value assigned to fallback capacity and option preservation
Evidence of weak intervention coupling	Policy is unlikely to alter the transition mechanism	Broader action region can remain supportable despite system-level sensitivity
Evidence of rapidly worsening inaction costs	Delay materially reduces later options or recovery capacity	Earlier or stronger action can remain warranted despite transition sensitivity

scale at which learning remains recoverable. High estimated leverage can favor smaller initial exposure or stronger fallback capacity, while evidence of weak coupling can support broader experimentation.

8.3 Stopping Architecture

Evidence-restraint inversion also changes the role of monitoring. Under a stable response regime, monitoring commonly measures implementation fidelity and outcome performance. Under transition sensitivity, monitoring must additionally detect whether the response architecture itself is changing. Relevant observations can include recovery time, coupling strength, concentration of effects, expectation shifts, option loss, institutional dependence, or other indicators tied to the hypothesized transition mechanism.

This creates a distinction between *performance monitoring* and *regime monitoring*. Performance monitoring asks whether the intervention produces the expected outcome. Regime monitoring asks whether the conditions that made the original evidence-to-action relation informative remain sufficiently comparable. Both can be necessary when intervention is capable of altering the response field.

Stopping conditions should follow the same logic. A threshold can be tied to evidence that reversibility is degrading, that the intervention is generating an unanticipated feedback loop, or that the response regime is moving outside the neighborhood in which the original evidence remains transportable. Such thresholds can support adaptive restraint under uncertainty about whether a critical transition is imminent.

Table 4 summarizes several evidence patterns that can produce different action implications.

The table emphasizes that improved evidence can support several intervention architectures, with restraint as one conditional response. The same criticality-sensitive framework can justify stronger action when evidence shows that inaction closes options faster than intervention does.

8.4 Restraint under Competing Harms

Evidence-restraint inversion must be evaluated against the baseline trajectory. Intervention can produce persistent change, while non-intervention can allow an existing dynamic to deepen; waiting can therefore alter expectations, investment, institutional capacity, and reachable futures as surely as overt action.

This comparison is particularly important when the system is already moving toward a harmful transition. Stronger evidence of critical slowing down, lock-in, or loss of recovery capacity could increase the case for timely intervention. The qualification developed here concerns intervention intensity and architecture while leaving action versus inaction to the substantive decision context. A bounded intervention can be more active than the status quo while still being restrained relative to a larger commitment.

The commitment trade-offs developed in Section 7 also remain relevant. Some interventions require credible duration to change expectations or coordinate behavior. If restraint undermines the mechanism through which the intervention works, staged or easily reversible designs can generate misleading evidence. A criticality-sensitive EBPM framework therefore needs to assess whether the proposed form of restraint preserves the policy mechanism being evaluated.

8.5 Decision Discipline

Evidence-restraint inversion is most useful as a decision discipline. It asks analysts to state whether new evidence changes only the estimated effect of an intervention or also the estimated sensitivity of the surrounding response regime. Where leverage rises, the analysis should ask whether the intervention remains reversible, whether the decision horizon changes, which alternatives remain accessible, and which monitoring signals would indicate that the evidential regime has shifted.

Research Obligation 8.3 (Evidence-restraint attribution). A claim of evidence-restraint inversion should identify the evidential improvement, the response property whose estimated leverage or persistence has increased, the mechanism connecting that property to intervention risk, the relevant reversibility or option-preservation concern, the costs of delay or under-action, and the conditions under which stronger intervention would remain warranted. Restraint should be specified through the evidence-responsive architecture and the mechanism that warrants it.

The qualification is therefore directional: better evidence can support stronger intervention, bounded intervention, or restraint depending on what it reveals about susceptibility, recoverability, and competing trajectories.

9 Adaptive Evidence-Informed Policy

Adaptive evidence-informed policy provides the constructive implication of the preceding analysis. When intervention, observation, and regime evolution are coupled, evidence should update both expected policy performance and the boundary within which those estimates remain action-relevant. This requirement complements established work on adaptive policy, robust decision making, and dynamic adaptive pathways (Walker, Rahman, et al. 2001; Walker, Marchau, et al. 2010; Lempert et al. 2003; Haasnoot et al. 2013).

Adaptive policy has long been proposed as a response to uncertainty that unfolds through time. Walker, Rahman, and Cave formulate adaptive policy as an approach in which adjustment is anticipated from the outset and built into policy design (Walker, Rahman, et al. 2001). Later work on deep uncertainty emphasizes policies that combine near-term action with explicit monitoring and adaptation as information changes (Walker, Marchau, et al. 2010). Robust decision making likewise emphasizes strategies that remain acceptable across multiple plausible futures (Lempert et al. 2003). Dynamic Adaptive Policy Pathways extends this family of approaches by linking actions to adaptation pathways, signposts, and conditions under which a pathway should change (Haasnoot et al. 2013).

The present paper draws on these traditions while distinguishing near criticality from deep uncertainty. A system can be deeply uncertain while dynamically insensitive to a contemplated intervention, and a transition-sensitive system can contain well-estimated local relations. The distinctive issue developed here is that action can alter the regime in which the evidence-to-action mapping was informative. Adaptive policy therefore becomes relevant through a specific mechanism: the policy must be capable of responding when evidence indicates that the response architecture itself has moved.

Definition 9.1 (Adaptive evidence-informed policy). An *adaptive evidence-informed policy* is a policy architecture in which initial action, monitoring, review, and revision are designed jointly so that accu-

mutating evidence updates both expected policy effects and the current action-admissible region through renewed assessment of transportability, reversibility, and warrant.

The definition emphasizes the architecture through which flexibility becomes evidentially governed. A policy can be frequently changed while remaining poorly adaptive if revisions respond only to short-run performance and ignore changes in the response regime. Conversely, a relatively stable policy can be adaptively governed when it contains credible monitoring, review, contingency, and revision capacity that can be activated when relevant conditions change.

9.1 Adaptive Architecture

Adaptive architecture begins in the design of initial implementation. The initial policy choice should identify the evidential relation being used, the conditions under which that relation is expected to remain transportable, and the types of change that would justify reconsideration. This makes adaptation part of the evidential design and a planned response to changing conditions.

For the framework developed in Sections 3–8, the architecture has two linked tasks. The first is ordinary policy learning: estimate whether outcomes are developing as expected. The second is regime learning: assess whether the response field remains sufficiently comparable for the original evidence-to-action relation to retain its relevance. The two tasks can move differently. Outcomes can improve while reversibility deteriorates, or short-run performance can disappoint while evidence indicates that the underlying regime remains stable and recoverable.

This dual perspective changes the meaning of policy success. A policy can achieve its immediate target while creating feedback, dependency, or option loss that narrows later choices. Another policy can produce modest near-term gains while preserving information, fallback capacity, and alternative trajectories. Their relative evaluation depends on explicit horizons, targets, and normative criteria.

Working Claim 9.2 (Dual updating). Under plausible regime sensitivity, policy learning should update two objects: the estimated consequences of available interventions and the estimated boundary of the action-admissible region. The two updates are analytically independent.

Dual updating provides a bridge between the local-validity framework in Section 6 and adaptive policy practice. The evidence-validity neighborhood can expand as estimates improve, while new information about persistence, recovery, or affected-party consequences contracts another part of the admissible region. An adaptive architecture must therefore preserve the possibility of changing policy even when confidence in a local effect estimate is increasing.

9.2 Signposts

Adaptation requires observations that are tied to the mechanism of concern. Adaptive-policy and dynamic-pathway approaches commonly use signposts or monitoring variables to indicate when assumptions are weakening or when a policy pathway should be reconsidered (Walker, Rahman, et al. 2001; Haasnoot et al. 2013). A criticality-sensitive version should distinguish ordinary outcome indicators from *regime signposts*.

Outcome indicators concern the policy’s stated effects: service uptake, emissions, employment, health outcomes, costs, or other target measures. Regime signposts concern the response architecture that gives those outcomes their dynamic meaning. Depending on the domain, these can include recovery time, concentration of dependence, coupling among subsystems, expectation shifts, fallback capacity, path accessibility, or indicators of self-reinforcing feedback. The relevant signpost must be justified by the hypothesized transition mechanism; volatility or statistical unusualness alone provides weak guidance.

A signpost is therefore more than a variable to be monitored. It has an evidential role within an explicit causal or dynamical account. If the mechanism concerns institutional dependence, a measure

of short-run performance may provide little information about reversibility. If the mechanism concerns expectation coordination, administrative implementation data may miss the response channel that matters. A regime-sensitive architecture should explain why each signpost is informative about the continuation or deterioration of the evidential conditions supporting action.

Signposts also need uncertainty ranges. Early-warning research shows that candidate indicators can generate false alarms or fail before some transitions (Scheffer, Bascompte, et al. 2009). The adaptive architecture should therefore avoid treating a single threshold crossing as proof of criticality. Converging indicators, mechanism-specific interpretation, and competing explanations remain necessary.

9.3 Triggered Revision

Signposts become operational when they inform review and revision. A trigger can specify a condition under which a policy is reconsidered, narrowed, expanded, paused, or shifted to another pathway. The attraction of triggers is procedural clarity: they reduce the need to reinvent the decision process after every new observation. Their limitation is that the trigger itself can become outdated when the response regime changes.

Criticality-sensitive triggers should therefore be treated as revisable decision rules with an explicit interpretation protocol. A pre-specified threshold can be useful when the monitored relation is stable enough to support it. Where reflexivity is strong, announcing the threshold can alter behavior before it is reached. Where measurement is noisy, a rigid rule can create repeated switching or strategic manipulation. Trigger design should consequently include an interpretation protocol: what evidence counts, who reviews it, which alternative explanations are considered, and what degree of action follows.

The architecture should also distinguish *review triggers* from *action triggers*. A review trigger initiates structured reassessment. An action trigger commits the institution to a specified response. Near uncertain critical transitions, review triggers can be preferable when the evidential basis for a fixed response remains weak. In other settings, an action trigger can provide valuable credibility and reduce strategic delay. The choice depends on the policy mechanism and institutional setting, with criticality language adding no independent warrant.

9.4 Portfolio Design

Adaptive policy can also be expressed through a portfolio of actions with different degrees of commitment. Robust-decision approaches emphasize strategies that perform acceptably across multiple plausible futures (Lempert et al. 2003). Dynamic pathways similarly allow sequences of actions to change as conditions evolve (Haasnoot et al. 2013). The near criticality qualification adds attention to how each action changes later option space.

A portfolio can include a core action, monitoring capacity, reversible experiments, reserve capacity, compensatory mechanisms, and prepared alternatives. These elements serve different evidential functions. Reversible experiments can reveal local response. Reserve capacity can support recovery. Prepared alternatives reduce the institutional delay involved in switching pathways. Compensation can address harms that become visible during implementation. The appropriate combination is domain-specific.

Portfolio design also prevents adaptation from being reduced to incrementalism. A policy can contain a large or durable core commitment while preserving adaptive capacity around implementation, compensation, monitoring, or complementary measures. Conversely, a sequence of small interventions can cumulatively generate lock-in. The relevant question is how the portfolio shapes future reachability, recoverability, and learning capacity over the decision horizon.

Table 5 summarizes the main components of the proposed architecture.

Table 5: Components of an adaptive evidence-informed policy architecture.

Component	Ordinary evidential role	Criticality-sensitive extension
Initial action	Acts on the best available evidence for the declared target	Limits exposure when leverage, persistence, or recovery remains uncertain
Outcome monitoring	Tracks implementation and target performance	Remains separate from evidence about regime comparability
Regime signposts	Usually implicit or secondary	Tracks feedback, coupling, recovery, expectation, or option-space changes tied to the transition hypothesis
Review triggers	Initiate reassessment when performance or assumptions change	Reassess transportability, reversibility, and warrant before expanding commitment
Prepared alternatives	Provide contingency options	Preserve reachable pathways when the current regime deteriorates or assumptions fail
Fallback capacity	Supports correction after policy error	Maintains supported reversibility when intrinsic recovery is weak
Institutional review	Updates interpretation and implementation	Re-evaluates the admissible-action boundary and the validity of the trigger architecture itself

9.5 Embedded Learning

The architecture developed here treats implementation as a source of evidence while bounding the exposure created by learning through action. Policy action can reveal response properties that become identifiable only after intervention, especially when feedback appears endogenously through action. Yet learning through intervention imposes consequences on affected actors and can alter future states. Near criticality, the cost of learning can rise precisely because the intervention used to generate information can change the regime.

This creates a need for *bounded learning*. The scale and duration of an intervention used partly for learning should reflect the expected informational gain, the reversibility of exposure, the distribution of risk, and the urgency of the policy problem. Small trials are useful when effects scale smoothly and local learning transports. They can be misleading when system effects arise only after large-scale adoption or when commitment is part of the mechanism. The design should therefore explain what can reasonably be learned from the chosen scale.

Learning also requires preservation of negative and ambiguous evidence. Adaptive systems can become biased toward escalation if successful short-run observations are recorded while near misses, recovery costs, or failed triggers receive less attention. A regime-sensitive EBPM architecture should retain evidence about boundary conditions, failed assumptions, and interventions that were deliberately withheld because the estimated action region had contracted.

9.6 Institutional Discipline

Adaptation creates discretion, and discretion creates governance questions. A policy that can change when conditions change also creates opportunities for opportunistic reinterpretation, shifting objectives, selective invocation of uncertainty, or strategic postponement. Adaptive architecture therefore requires institutional discipline alongside technical monitoring.

At minimum, the institution should specify who can interpret signposts, who can authorize escalation or withdrawal, how affected parties can contest the interpretation, how changes are documented, and how earlier commitments constrain later discretion. Existing work on the governance of evidence already

emphasizes that technical evidence must be embedded within legitimate processes of use (Parkhurst 2017). Near-critical adaptation makes this especially important because the evidential object includes both the policy-effect estimate and the evolving boundary of warranted action.

Precommitment can sometimes strengthen adaptation by stabilizing procedures for future revision. Institutions can precommit to review dates, disclosure requirements, consultation procedures, or compensation rules while leaving the substantive policy response contingent on future evidence. This separates procedural stability from policy rigidity. It also allows actors to form expectations about how adaptation will occur while allowing the substantive future state to remain uncertain.

The same discipline should apply to claims of emergency or criticality. A criticality-sensitive framework can be misused to justify exceptional discretion, accelerated intervention, or indefinite postponement. The empirical burden developed throughout this paper remains intact. Adaptation should be triggered by specified evidence about changing response conditions and accompanied by a transparent account of which part of the admissible-action region is thought to have changed.

Research Obligation 9.3 (Adaptive policy specification). An adaptive evidence-informed policy should identify the target inference supporting initial action, the monitoring variables and regime signposts, the conditions that initiate review, the authority and procedure for revision, the available fallback or alternative pathways, the expected limits of learning from implementation, and the evidence that would justify expanding, narrowing, or terminating the intervention. The architecture should distinguish changes in expected policy performance from changes in the response regime that supports action warrant.

The framework remains compatible with EBPM while changing the temporal structure of evidence use. Evidence informs initial action, monitoring, interpretation of changing response conditions, and revision. Its policy role therefore extends from intervention selection to the evolving relation between action and the response regime. Critical-regime instability remains a defeasible hypothesis suitable for direct empirical testing.

10 Empirical Research Program

This section makes the qualification empirically defeasible. The research program distinguishes critical-regime instability from ordinary contextual variation, nonstationarity, implementation failure, and exogenous shocks, while allowing the proposed mechanism to be absent or too weak to alter ordinary evidence-to-action reasoning. The central task is to connect observable evidence-action episodes to competing explanations and explicit boundary tests.

The program begins from a demanding evidential posture. Early-warning research in other dynamical domains shows that candidate indicators can generate false alarms, miss transitions, or depend strongly on the assumed mechanism and observation window (Scheffer, Bascompte, et al. 2009; Boettiger and Hastings 2012; Dakos et al. 2015). Structural-break and regime-switching methods likewise represent changing data-generating processes while leaving the near-critical interpretation to separate mechanism-specific evidence (Hamilton 1989; Bai and Perron 2003). Nonlinear forecasting can sometimes discriminate structured dynamics from measurement error or externally induced noise, yet its success depends on data quality and model adequacy (Sugihara and May 1990). These literatures support methodological pluralism and caution against reading criticality directly from one statistic.

10.1 Unit of Analysis

The primary empirical unit is an *evidence–action episode*: a temporally bounded configuration in which a specified body of evidence supports a target inference, an intervention is contemplated or implemented, and the response conditions that connect evidence to action can be observed before and after the decision.

The episode can occur at the level of a program, organization, jurisdiction, policy subsystem, market, or other analytically justified scale.

This unit keeps the paper's claim narrower than a general diagnosis of political instability. The empirical question concerns whether the evidential relation used to justify action remains sufficiently transportable after the system responds to the action, announcement, or surrounding interpretation. Dramatic crisis alone falls outside the central claim unless the evidence-to-action relation changes through the proposed mechanism. A comparatively quiet episode can remain relevant when small interventions, warnings, or expectation shifts materially change the response field.

Each episode should therefore identify at least five elements: the target inference supported by the evidence; the intervention or policy signal under consideration; the response variables through which the target is evaluated; the features of the response regime hypothesized to matter for transportability; and the decision horizon over which reversibility and warrant are assessed. The same episode can support several target inferences, but their evidential status should be estimated separately.

Research Obligation 10.1 (Episode specification). An empirical study of critical-regime instability should state the evidence–action episode, the target inference, the intervention or announcement, the response-regime features under examination, the relevant scale, and the decision horizon. These elements should be specified before interpreting volatility, structural change, or policy disagreement as evidence of the mechanism proposed in this paper.

10.2 Construct Discipline

Critical-regime instability is a relational construct. It concerns the relation among evidence, intervention, and an evolving response regime. Measurement should therefore avoid collapsing the framework into a single “criticality score.” Several construct families can provide partial evidence: changing response sensitivity, altered recovery dynamics, feedback reinforcement, expectation shifts, coupling changes, reversal cost, and loss of accessible alternatives. Their empirical meaning depends on the hypothesized mechanism.

Table 6 summarizes the main constructs developed in this paper and the empirical burden attached to each. The table separates observable implications from alternative explanations because the same pattern can arise through ordinary nonstationarity, exogenous shocks, strategic adaptation, measurement change, or institutional replacement. A near-critical interpretation gains support when multiple observations fit one mechanism while competing explanations become less adequate.

Construct validation should use converging evidence where feasible. For example, a claim of deteriorating reversibility can draw on recovery time, rollback cost, dependence measures, institutional exit barriers, and observed failures of attempted reversal. A claim of changing coupling can combine network measures, synchronized responses, transmission sequences, and qualitative process evidence. No single indicator is required across all domains.

The same discipline applies to candidate early-warning measures. Rising variance, autocorrelation, slowing recovery, and related patterns can be informative under specific transition mechanisms, yet their error rates can be substantial and some transitions provide weak or absent signals (Boettiger and Hastings 2012; Dakos et al. 2015). The empirical program should therefore treat these measures as components of a model-comparison exercise whose interpretation depends on the proposed mechanism.

10.3 Regime Identification

Regime identification asks whether a response architecture relevant to the target inference has materially changed. Several statistical architectures can assist, while substantive interpretation remains mechanism-dependent. Regime-switching models can represent observations generated under latent states with different parameter structures (Hamilton 1989). Structural-break methods can estimate changes in coefficients or break dates within specified models (Bai and Perron 2003). Nonlinear forecasting can test

Table 6: Empirical architecture for evidence-to-action research near criticality. The entries identify research targets and discriminating tasks for a plural measurement strategy.

Construct	Illustrative empirical signature	Primary competing explanation or test
Regime transport	Previously supported response relation loses predictive or decision relevance after response conditions change	ordinary contextual heterogeneity; population shift; measurement drift
Reflexive evidence	Publication, warning, or policy anticipation changes expectations or behavior before direct implementation	common information shock; concurrent intervention; selection into exposure
Neighborhood divergence	Local effect estimation remains stable while recovery cost, option loss, or warrant constraints change	estimation error; ordinary risk aversion; unrelated institutional constraint
Reversibility	Return path becomes slower, costlier, or materially different after intervention	temporary adjustment cost; one-off shock; administrative delay
Evidence-restraint inversion	Better evidence of susceptibility or leverage supports narrower exposure or stronger monitoring	generic precaution; political opposition; resource scarcity
Adaptive architecture	Signposts and review triggers detect changes in response conditions early enough to alter policy pathway	routine performance management; exogenous policy change; discretionary drift

whether local state-space structure improves prediction relative to models dominated by measurement error or unstructured noise (Sugihara and May 1990). Each approach addresses a different empirical problem.

For the present framework, the key criterion is substantive: the inferred regime difference must matter for the target evidence-to-action relation. A statistically detectable break in an unrelated parameter can be irrelevant to policy transport. Conversely, a small change in an expectation channel or institutional coupling can be highly consequential when it changes which intervention responses remain accessible or reversible.

Empirical studies should therefore pair statistical regime identification with mechanism reconstruction. Relevant evidence can include temporal ordering, institutional changes, communication records, network transmission, actor expectations, rollback attempts, and shifts in response to comparable perturbations. Historical and qualitative material can be especially important when the data-generating process is itself partly constituted by interpretation and institutional change.

Hypothesis 10.2 (Regime-transport degradation). *When a response regime relevant to a target inference changes materially, evidence generated under the earlier regime will exhibit weaker out-of-regime performance for that target than comparable evidence evaluated within the regime in which it was generated, conditional on ordinary population and measurement differences.*

This hypothesis is intentionally narrower than a prediction of universal forecast failure. Some relations can remain invariant across regime changes. Evidence of stable transport across well-identified response shifts would weaken the claim for the target relation and domain under study.

10.4 Transport Designs

Regime transport can be studied through temporal, comparative, and matched designs. A temporal design follows the same policy domain across a suspected response-regime change and estimates whether a previously supported response relation retains predictive or decision relevance. A comparative design examines units exposed to similar interventions under different response architectures. A matched design seeks episodes with similar measured populations and institutional settings but different feedback or expectation structures.

The central advantage of same-site temporal comparison is that it limits some familiar transport differences while making the regime question more visible. Its limitation is historical confounding: many conditions can change together. Comparative designs can help by identifying units where selected regime features vary more sharply than other contextual factors. Process tracing can then test whether the hypothesized feedback or coupling pathway actually connects the regime difference to the observed change in policy response.

The decision horizon should be varied explicitly. Evidence can transport well over short horizons while losing relevance for longer-run trajectories. Studies should therefore report horizon-specific performance where feasible. The result can be reported as a horizon-specific transport profile.

Transport failure should be reported separately from evidential invalidity. A finding can continue to describe the earlier regime accurately while becoming less useful for action in the later one. Empirical reporting should preserve this distinction so historical evidence remains indexed to the regime in which it was valid.

10.5 Intervention-Coupled Identification

The most demanding empirical problem concerns intervention-coupled regime change. The framework predicts that policy implementation, announcement, or evidence disclosure can sometimes affect the response architecture itself. Identification therefore requires attention to timing and mechanism.

Useful designs can exploit phased implementation, staggered announcements, threshold rules, naturally occurring differences in information exposure, or institutional boundaries that produce variation in anticipatory response. When experimental manipulation is ethically and politically appropriate, vignette or laboratory designs can study expectation and coordination channels before field application. Field experiments can be informative for local mechanisms while remaining limited when large-scale adoption or public commitment is itself part of the causal pathway.

Reflexive evidence requires a particularly careful temporal decomposition. Researchers should distinguish the informational event from the material intervention and measure responses in the interval between them where possible. Pre-implementation expectation changes, investment shifts, mobilization, strategic compliance, or withdrawal can then be assessed as candidate reflexive pathways. The Lucas critique and the performativity literature supply established reasons to expect policy rules and models to alter behavior in some domains (Lucas 1976; MacKenzie and Millo 2003).

Hypothesis 10.3 (Empirical reflexive amplification). *When evidence disclosure or policy anticipation changes actor expectations through a mechanism coupled to a transition-sensitive response field, the resulting pre-implementation response will be larger or more persistent than under otherwise comparable conditions with weaker susceptibility along that mechanism.*

A finding that reflexive responses are stable in magnitude across substantially different susceptibility conditions would weaken this specific amplification claim. Reflexivity can remain present under other mechanisms.

10.6 Neighborhood Tests

The neighborhood framework in Section 6 generates a direct empirical program. The evidence-validity neighborhood, reversibility neighborhood, and warrant neighborhood should be estimated through different observables because their determinants differ.

The evidence-validity neighborhood can be studied through predictive performance, effect stability, directional accuracy, or another target-specific loss function across nearby states and interventions. The reversibility neighborhood can be studied through recovery time, rollback cost, option preservation, dependence, institutional persistence, and the success of attempted reversal. The warrant neighborhood requires explicit normative and procedural criteria; its empirical components can include authority,

affected-party incidence, procedural compliance, distributional exposure, and stated legal or institutional constraints.

The framework predicts that these neighborhoods can move differently. Better measurement can expand the evidence-validity neighborhood while evidence of rising reversal cost contracts the reversibility neighborhood. New authorization can expand the warrant neighborhood while regime uncertainty leaves the other two unchanged.

Hypothesis 10.4 (Empirical neighborhood divergence). *In some transition-sensitive evidence–action episodes, improvements in target-relative local estimation will coexist with deterioration in recovery, option-preservation, or warrant indicators, producing an action-admissible region that contracts despite improved evidential precision.*

This hypothesis should be tested against simpler explanations. A contraction caused solely by budget pressure or unrelated legal change supports an alternative explanation to critical-regime instability. Likewise, improved evidence followed by ordinary political caution provides limited support unless the caution is linked to the response mechanism specified in the study.

10.7 Adaptive Policy Evaluation

Section 9 proposed adaptive evidence-informed policy as an architecture for dual updating. Its evaluation should therefore extend beyond conventional outcome comparison. Studies should ask whether the architecture detects changes in response conditions, whether triggers are activated at useful times, whether alternative pathways remain available, and whether revision occurs before avoidable option loss or irreversible exposure.

Comparative evaluations can examine policies with similar substantive objectives but different monitoring and revision architectures. Relevant outcomes include target performance, time to detect regime change, frequency of false review triggers, recovery after pathway change, retained fallback capacity, implementation burden, and distribution of adaptation costs. Simulation can supplement field evidence by stress-testing trigger rules across multiple plausible regime evolutions, with its evidential role limited to mechanism plausibility and design comparison.

Hypothesis 10.5 (Boundary-sensitive adaptation). *Where regime change is empirically detectable and policy pathways are genuinely revisable, architectures that monitor response conditions in addition to target outcomes will identify some loss of transportability or reversibility earlier than architectures that monitor performance outcomes alone.*

The hypothesis has clear failure conditions. Regime signposts can be too noisy, delayed, manipulable, or poorly linked to the mechanism. Adaptive architectures can also impose decision costs that exceed their value. Evidence that regime-sensitive monitoring adds no reliable information beyond ordinary outcome tracking would narrow the practical relevance of the qualification developed here.

10.8 Boundary Tests

The empirical program should actively search for cases in which the paper’s qualification adds little. Many policy domains can remain sufficiently stable for ordinary evidence transport. Some interventions are weakly coupled to response-regime dynamics. Some systems recover rapidly after perturbation. Some apparent transitions are better represented as exogenous shocks, gradual trends, administrative changes, or ordinary strategic adjustment.

At least four findings would materially weaken a critical-regime interpretation in a given domain. First, response relations transport reliably across suspected regime changes after ordinary contextual differences are controlled. Second, intervention and announcement leave the response architecture relevant to the target measurably stable. Third, recovery and option-preservation indicators remain stable as

susceptibility is claimed to increase. Fourth, regime-sensitive monitoring provides little discriminating value beyond standard performance measures.

Research Obligation 10.6 (Competing explanations). An empirical study should compare the critical-regime account with plausible alternatives, specify which observations discriminate among them, and report evidence that counts against the proposed mechanism. Mechanism-specific identification should carry the primary inferential burden beyond model flexibility, retrospective narrative, or indicator multiplicity.

This research program preserves the constructive posture of the paper. Evidence remains central to policy analysis, including evidence that the near-critical qualification is unnecessary for a particular case. The empirical task is to determine when locally valid evidence travels safely into action, when action changes the conditions of its own evidential relevance, and when adaptive architecture materially improves that relation.

11 Conclusion

This paper develops a focused qualification to the move from evidence to policy action. The qualification applies when evidence remains locally informative while intervention can alter the response regime that made the evidential relation decision-relevant. In such settings, evidential confidence and action warrant can diverge. Policy reasoning must therefore consider both whether the response estimate is credible and whether its relation remains transportable, recoverable, and supportable over the intended decision horizon.

The concept of *critical-regime instability of evidence-to-action relations* identifies this mechanism while reserving criticality attribution for evidence beyond volatility, crisis, or generic nonlinearity. A critical-regime interpretation requires a transition-sensitive response property, a plausible intervention channel, and observable patterns capable of discriminating the proposed mechanism from ordinary non-stationarity, exogenous shocks, strategic adaptation, or implementation failure. Historical evidence can therefore remain valid for the regime in which it was generated while losing some forward intervention relevance after the response architecture changes.

The paper's local-validity framework separates three conditions. The evidence-validity neighborhood concerns where a response relation remains sufficiently informative for a declared inferential task. The reversibility neighborhood concerns where target-relevant recovery remains supportable over a specified horizon. The warrant neighborhood concerns the legal, procedural, distributional, and normative conditions governing action. Their intersection defines the locally admissible action region. Near a plausible critical transition, the evidential neighborhood can remain broad while reversibility or warrant contracts.

This divergence motivates *evidence-restraint inversion*. Stronger evidence of leverage, susceptibility, or persistent transition can justify bounded exposure, staged commitment, monitoring, fallback capacity, or delayed irreversible commitment. The inversion is conditional. Evidence of worsening delay, shrinking options under inaction, or a well-buffered intervention can support earlier or stronger action. The contribution is therefore the separation of informational confidence from intervention intensity and from the architecture through which intervention is implemented.

Reversibility is central because formal withdrawal can leave institutional, material, distributional, or expectation effects that persist. Conversely, reserve capacity, fallback institutions, retained expertise, compensation mechanisms, or staged implementation can preserve recovery paths even under durable policy commitments. Reversibility is thus target-relative, horizon-sensitive, and partly institutionally produced.

The constructive implication is an adaptive evidence-informed architecture in which evidence updates both expected policy effects and the estimated boundary of admissible action. Ordinary outcome

monitoring can be complemented by regime signposts, review triggers, fallback pathways, and explicit authority for revision. Such arrangements preserve the strengths of EBPM while making regime sensitivity and recovery conditions visible within the decision process.

The core qualification stated in Working Claim 1.1 can therefore be read as a conditional separation between policy knowledge and intervention scope. Where transition-sensitive response dynamics are credible, stronger local evidence can improve policy knowledge while the admissible action region remains stable, expands, or contracts according to regime transport, reversibility, normative authorization, and the decision horizon.

The empirical program keeps this claim defeasible. Domains in which response relations transport reliably, intervention leaves the relevant regime sufficiently comparable, and recovery remains broad provide evidence against the need for the qualification. Where the mechanism is supported, the role of evidence expands beyond effect estimation to characterize response sensitivity, regime comparability, recovery capacity, accessible alternatives, and the conditions under which stronger commitment remains warranted.

References

- Arrow, Kenneth J. and Anthony C. Fisher (1974). "Environmental Preservation, Uncertainty, and Irreversibility". In: *The Quarterly Journal of Economics* 88.2, pp. 312–319. DOI: 10.2307/1883074. URL: <https://doi.org/10.2307/1883074>.
- Bai, Jushan and Pierre Perron (2003). "Computation and Analysis of Multiple Structural Change Models". In: *Journal of Applied Econometrics* 18.1, pp. 1–22. DOI: 10.1002/jae.659. URL: <https://doi.org/10.1002/jae.659>.
- Boettiger, Carl and Alan Hastings (2012). "Quantifying Limits to Detection of Early Warning for Critical Transitions". In: *Journal of the Royal Society Interface* 9.75, pp. 2527–2539. DOI: 10.1098/rsif.2012.0125. URL: <https://doi.org/10.1098/rsif.2012.0125>.
- Cairney, Paul (2016). *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan. DOI: 10.1057/978-1-137-51781-4. URL: <https://doi.org/10.1057/978-1-137-51781-4>.
- Cartwright, Nancy and Jeremy Hardie (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press. ISBN: 9780199841608. DOI: 10.1093/acprof:osobl/9780199841608.001.0001. URL: <https://doi.org/10.1093/acprof:osobl/9780199841608.001.0001>.
- Dakos, Vasilis, Stephen R. Carpenter, Egbert H. van Nes, and Marten Scheffer (2015). "Resilience Indicators: Prospects and Limitations for Early Warnings of Regime Shifts". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 370.1659, p. 20130263. DOI: 10.1098/rstb.2013.0263. URL: <https://doi.org/10.1098/rstb.2013.0263>.
- Dixit, Avinash K. and Robert S. Pindyck (1994). *Investment under Uncertainty*. Princeton, NJ: Princeton University Press. ISBN: 9780691034102. URL: <https://www.jstor.org/stable/j.ctt7sncv>.
- Espeland, Wendy Nelson and Michael Sauder (2007). "Rankings and Reactivity: How Public Measures Recreate Social Worlds". In: *American Journal of Sociology* 113.1, pp. 1–40. DOI: 10.1086/517897. URL: <https://doi.org/10.1086/517897>.
- Haasnoot, Marjolijn, Jan H. Kwakkel, Warren E. Walker, and Judith ter Maat (2013). "Dynamic Adaptive Policy Pathways: A Method for Crafting Robust Decisions for a Deeply Uncertain World". In: *Global Environmental Change* 23.2, pp. 485–498. DOI: 10.1016/j.gloenvcha.2012.12.006. URL: <https://doi.org/10.1016/j.gloenvcha.2012.12.006>.
- Hamilton, James D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle". In: *Econometrica* 57.2, pp. 357–384. DOI: 10.2307/1912559. URL: <https://doi.org/10.2307/1912559>.
- Head, Brian W. (2010). "Reconsidering Evidence-Based Policy: Key Issues and Challenges". In: *Policy and Society* 29.2, pp. 77–94. DOI: 10.1016/j.polsoc.2010.03.001. URL: <https://doi.org/10.1016/j.polsoc.2010.03.001>.
- Henry, Claude (1974). "Investment Decisions Under Uncertainty: The 'Irreversibility Effect'". In: *The American Economic Review* 64.6, pp. 1006–1012. URL: <https://www.jstor.org/stable/1815248>.
- Kydland, Finn E. and Edward C. Prescott (1977). "Rules Rather than Discretion: The Inconsistency of Optimal Plans". In: *Journal of Political Economy* 85.3, pp. 473–491. DOI: 10.1086/260580. URL: <https://doi.org/10.1086/260580>.
- Lempert, Robert J., Steven W. Popper, and Steven C. Bankes (2003). *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. Santa Monica, CA: RAND Corporation. ISBN: 9780833032751. URL: https://www.rand.org/pubs/monograph_reports/MR1626.html.
- Lucas Robert E., Jr. (1976). "Econometric Policy Evaluation: A Critique". In: *Carnegie-Rochester Conference Series on Public Policy* 1, pp. 19–46. DOI: 10.1016/S0167-2231(76)80003-6. URL: [https://doi.org/10.1016/S0167-2231\(76\)80003-6](https://doi.org/10.1016/S0167-2231(76)80003-6).
- MacKenzie, Donald and Yuval Millo (2003). "Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange". In: *American Journal of Sociology* 109.1, pp. 107–145. DOI: 10.1086/374404. URL: <https://doi.org/10.1086/374404>.

- Merton, Robert K. (1948). "The Self-Fulfilling Prophecy". In: *The Antioch Review* 8.2, pp. 193–210. DOI: 10.2307/4609267. URL: <https://doi.org/10.2307/4609267>.
- Parkhurst, Justin O. (2017). *The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence*. Abingdon: Routledge. ISBN: 9781138939400. DOI: 10.4324/97811315675008. URL: <https://doi.org/10.4324/97811315675008>.
- Scheffer, Marten, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk, and George Sugihara (2009). "Early-Warning Signals for Critical Transitions". In: *Nature* 461, pp. 53–59. DOI: 10.1038/nature08227. URL: <https://doi.org/10.1038/nature08227>.
- Scheffer, Marten, Steve Carpenter, Jonathan A. Foley, Carl Folke, and Brian Walker (2001). "Catastrophic Shifts in Ecosystems". In: *Nature* 413, pp. 591–596. DOI: 10.1038/35098000. URL: <https://doi.org/10.1038/35098000>.
- Sugihara, George and Robert M. May (1990). "Nonlinear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series". In: *Nature* 344.6268, pp. 734–741. DOI: 10.1038/344734a0. URL: <https://doi.org/10.1038/344734a0>.
- Walker, Warren E., Vincent A. W. J. Marchau, and Darren Swanson (2010). "Addressing Deep Uncertainty Using Adaptive Policies: Introduction to Section 2". In: *Technological Forecasting and Social Change* 77.6, pp. 917–923. DOI: 10.1016/j.techfore.2010.04.004. URL: <https://doi.org/10.1016/j.techfore.2010.04.004>.
- Walker, Warren E., S. Adnan Rahman, and Jonathan Cave (2001). "Adaptive Policies, Policy Analysis, and Policy-Making". In: *European Journal of Operational Research* 128.2, pp. 282–289. DOI: 10.1016/S0377-2217(00)00071-0. URL: [https://doi.org/10.1016/S0377-2217\(00\)00071-0](https://doi.org/10.1016/S0377-2217(00)00071-0).

DRAFT