

When Correction Becomes Reproduction

Recursive Blame, Corrective Institutions, and the Conditions of Repair

Flyxion

Independent Researcher

October 2026

Abstract

Human societies repeatedly attribute collective harms to identifiable individuals, populations, institutions, or ideologies, and the institutions built to correct those harms are staffed, funded, and evaluated through the same acts of attribution. This essay asks when such correction lowers the visible prevalence of a harm while raising its capacity to recur. It separates the prevalence of a harm from its reproductive capacity, defines the latter at the level of a mechanism composed of an incentive structure and a classification practice, and states a testable proposition: a corrective intervention may be locally effective and globally counterproductive with respect to the specified harm when its suppression of observable harm is accompanied by changes in institutional incentives, classification practices, or adaptive capacities that raise the future production of that harm. The essay sets admissibility conditions that any claimed instance must satisfy, develops a minimal dynamic model with explicit conditions for delayed reversal and an extension to intervention that responds to observed harm, applies the framework, as an illustrative and unadjudicated case analysis, to a narrow mechanism in the political economy of concentrated technological power, and proposes design criteria for forms of repair that do not reproduce what they oppose. Its aim is a research program with stated conditions for being wrong, not a universal explanation.

1 Introduction: Perpetrator-Centered Explanation and Its Limits

Accounts of social dysfunction tend to settle on a culpable party. When a market fails, a movement radicalizes, a public institution loses trust, or a communications platform degrades public discussion, explanation usually proceeds by identifying who is responsible, and

the identification organizes both the immediate response and the next round of conflict. Attribution of this kind is not mistaken in principle. Individuals commit fraud, institutions conceal harm, and coalitions pursue advantage at the expense of others, and a theory unable to name such conduct would be useless for accountability. The difficulty lies in the inference from the existence of culpable actors to the sufficiency of an explanation that stops with them.

Perpetrator-centered explanation becomes inadequate where a harm persists through changes in the people who carry it. If the removal of a set of actors leaves the same behavior recurring among their successors, then something other than their individual dispositions is doing causal work: the conditions that reward the behavior, the classifications through which it is recognized and condemned, and the institutions charged with responding to it. Merton’s analysis of the unanticipated consequences of purposive social action established the general point that deliberate interventions produce effects their authors did not foresee [12], and later discussion has stressed that some such consequences are unintended without being unanticipated [4]. The present essay is concerned with a narrower and more demanding case, in which the corrective apparatus itself contributes to the recurrence it is meant to prevent.

The thesis is that a corrective intervention may be locally effective and globally counterproductive, in the sense made precise in Section 7, when its immediate suppression of observable harm is accompanied by changes in institutional incentives, classification practices, or adaptive capacities that raise the future production of that harm. The claim is stated so that it can fail. It does not assert that correction is generally futile, and it does not treat every policy failure as a pathology. It also separates the case from ordinary adaptive evasion, in which actors respond to a suppression by finding a way around it, by requiring that some part of the corrective apparatus’s own operation help maintain or regenerate the conditions responsible for the original harm.

Three terms are used in distinct senses. Mechanism reproduction means that correction raises the reproductive capacity of the mechanism that produces a harm. Pathological reversal is the narrower case in which that effect outweighs suppression, so that the long-run prevalence of the specified harm exceeds its no-intervention baseline, and this is the sense of locally effective and globally counterproductive used here, with respect to the specified harm and not to any wider measure of welfare. Feedback amplification refers to loops in which correction responds to the harm it addresses, and theoretical reflexivity refers to the application of the theory to its own categories and proposals. The word recursion is used for mechanism reproduction through correction, and the other senses are named explicitly.

The essay proceeds as follows. Sections 2 through 5 establish the concepts, Section 6 sets the conditions the theory must meet, Section 7 formalizes the argument, Sections 8 and 9

apply it to communication environments and to a single case, and Sections 10 and 11 turn to accountability and repair. In detail: Section 2 defines prevalence, correction, and reproductive capacity, and fixes the unit of reproduction at the level of mechanism. Section 3 examines how classification can reproduce the phenomena it names, and Section 4 distinguishes scapegoating from legitimate attribution. Section 5 turns to the political economy of corrective institutions. Section 6 states the conditions the theory must meet before any instance counts as an instance, including the requirement that the theory apply to itself, and Section 7 develops a minimal dynamic model. Section 8 considers communication environments as selection conditions, Section 9 applies the framework to a single mechanism in concentrated technological power as an illustrative case analysis and research design, not a completed test, and Sections 10 and 11 address accountability and the design of nonreproductive repair. Section 12 concludes.

2 Preliminaries: Prevalence, Reproductive Capacity, and Processes Without a Central Agent

Three quantities structure the analysis. Let H_t denote the true prevalence at time t of a specified harm, C_t the intensity of corrective intervention directed against it, and R_t the reproductive capacity of the pattern that produces it. Prevalence and correction are familiar. Reproductive capacity is the unusual quantity, and it is defined as latent, not observed. An observed recurrence rate is not the same thing as capacity: a pattern may have substantial capacity yet fail to recur within a given window because its enabling conditions are briefly absent, and a recurring pattern may reflect persistent external conditions instead of the reproduction of the mechanism itself. Capacity is therefore defined as the expected number of descendants of a mechanism over a stated horizon T , given the present state of the system and its environment,

$$R_t = \mathbb{E} [N_{\text{desc}}(t, t + T) \mid x_t, \mathcal{E}], \quad (1)$$

where x_t is the social configuration, $\mathcal{E}$ describes the relevant environmental conditions, and N_{desc} counts later instances attributable to the mechanism.

The horizon T must be chosen in relation to the decay rates a and ρ of the model in Section 7, so that baseline transients have vanished. For a stated tolerance ϵ it should satisfy $\max(a, \rho)^T \leq \epsilon$, that is, $T \geq \ln \epsilon / \ln \max(a, \rho)$. It must be fixed before the data are examined, and the sensitivity of conclusions to T should be reported, since a horizon that is too short makes capacity look small even when the mechanism is strongly reproductive, and an unreported choice of horizon could be used to make the thesis disappear.

The definition imposes an obligation, which is to say what qualifies as a descendant of a mechanism and not merely a superficially similar occurrence. The essay answers by locating the unit of reproduction at the level of mechanism. A mechanism is a pairing $M = (I, K)$ of an incentive structure I , meaning the pattern of rewards and penalties that make a behavior worth repeating, and a classification practice K , meaning the procedures by which the behavior and its carriers are recognized, labeled, and responded to. An instance is a descendant of M if it preserves M across changes in participants, names, and vocabulary. This choice guards against two opposite errors. If the unit were ideological content, a pattern would count as ended whenever its vocabulary changed. If the unit were any resemblance in outcome, every historical echo would count as continuity. A mechanism-level criterion excludes both, but it is demanding, because it requires showing that the incentive structure and the classification practice, and not merely the outcome, are carried forward.

A further distinction separates the recurrence of a mechanism from its transmission. Recurrence means that the same incentive structure and classification practice appear again, whereas transmission means that they appear because an earlier instance passed them on, through personnel, imitation, or institutional inheritance. Two institutions can independently develop the same mechanism under the same conditions, and that is recurrence without transmission. The thesis of the essay concerns the capacity of a pattern to recur as a consequence of the operation of the corrective apparatus, so the proxy must not credit every repeated pattern to a common lineage. Empirical work must therefore separate recurrence produced by persisting external conditions from recurrence produced through the apparatus, for example by comparing settings that share conditions but differ in whether the apparatus has operated.

Recurrence is what the proxy observes. Transmission is inferred only when the mechanism is shown to be carried by identifiable carriers such as personnel, documents, or institutional lineage, and where none is identified the essay reports recurrence and does not claim transmission. The institutional channel is typically a transmission channel, whereas the classification channel can operate through imitation or through independent re-invention under the same conditions. Mechanism identity must also be specified before recurrence is observed, because classifying cases as instances of the same mechanism on the ground that they produce similar harms would make the test circular. The reward contingencies, institutional roles, classification rules, and expected behavioral responses that constitute M should therefore be preregistered, and a later case counts as a descendant only if it preserves those prespecified relationships. A component counts as preserved if it is functionally equivalent: an incentive structure is preserved if it retains the rank-ordering of behaviors by reward, and a classification practice is preserved if it assigns the same cases to categories with the

same consequences, whatever the vocabulary. The definition requires both components, and the conjunction is deliberate, because an incentive structure that survives radical changes in classification, or a classification practice that reproduces harm under different incentives, is a related mechanism and not the same one. Such cases of partial survival are recorded and analyzed separately, and the criteria must be applied consistently across cases.

The framework also requires care about agency. An individual agent, an institution, a distributed process, and a self-reproducing pattern are different kinds of thing, and the argument concerns the last two. A distributed process can produce apparently coordinated outcomes without central direction, because variants that fail to secure attention, coalitions, or resources disappear while variants that succeed propagate. Apparent strategic intelligence of this kind is explained by differential reproduction and environmental selection, not by a hidden intelligent system, and the essay assumes no such system. Functional autonomy is attributed to a pattern only in a limited, testable sense: its persistence across turnover is better predicted by the mechanism and the environment than by the intentions of its current carriers.

Because R_t cannot be observed directly, empirical work requires a proxy, and the essay commits to one. The proxy is recurrence after turnover of personnel, labels, or institutions, measured over a horizon fixed in advance and compared across settings with similar environmental conditions $\mathcal{E}$. The proxy inherits the limitation noted above, since recurrence can be driven by conditions external to the mechanism, and the research design must therefore compare settings that differ in the corrective apparatus while holding $\mathcal{E}$ approximately constant. Without a stated proxy and a comparison design, the central claim of the essay would remain a metaphor.

3 Classification and Causation: Looping Effects and Unintended Consequences

Social categories are not passive descriptions. Hacking’s account of looping effects holds that classifications of people differ from classifications of natural objects because the people classified become aware of the classification and respond to it, and their response changes the kind itself [10]. The implication for corrective systems is direct. A category introduced to identify harmful conduct can become a stable identity, shaping the conduct of those assigned to it, the expectations of those who encounter it, and the way institutions organize their work. When this occurs the category no longer merely describes a prior condition, and it becomes a component of the process that maintains the condition. Related traditions in

the sociology of deviance and the theory of performativity reach similar conclusions through different routes, and the essay draws on them for the general claim that labeling is causally active.

Merton's 1936 article supplies the sociological foundation for the intervention problem. It identified several factors in unanticipated consequences, namely ignorance, error, the imperious immediacy of interest, and basic values, and it added a further circumstance peculiar to human conduct, that a public prediction becomes a new element in the situation and so tends to change the course it forecasts [12]. The last is especially relevant, since a public classification or forecast alters the behavior of those it concerns and can thereby defeat or confirm itself. The later distinction between consequences that are unintended and those that are unanticipated [4] sharpens the point for institutions, because a consequence that is foreseeable yet tolerated raises questions of responsibility that a purely cognitive account of unanticipated effects does not.

Girard's account of collective accusation supplies a mechanism by which classification can generate cohesion through victimization. On his account, communities in crisis converge on a victim whose punishment restores unity, and the process is concealed from participants by their belief in the victim's guilt [7, 8]. The essay does not adopt his claim that this mechanism was the origin of religion and culture, a claim that Riordan's review describes as speculative yet "nevertheless a hypothesis concerning human evolution," to be assessed against competing evolutionary accounts [16]. It uses the narrower and more defensible idea that accusation can bind a group and that the binding effect can give the practice an incentive to recur independent of the accused's actual contribution to the harm.

These strands support a proposition that the later sections develop. A classification can contribute to causal understanding, to the reproduction of the phenomenon it names, or to both at once, and the contributions can be distinguished empirically. A classification can accurately identify harmful conduct while also changing the conduct of those classified, and Hacking's own examples often combine the two. A classification contributes to understanding when it predicts conduct better than the antecedent conditions it replaced, when it can be revised in light of evidence about the classified, and when removing it would leave the underlying pattern unchanged. It contributes to reproduction when those it names adopt it as an identity, when institutions organize their budgets and procedures around it independent of the harm it was introduced to address, and when removing the category changes the rate at which the pattern recurs. The proposition does not make all categories pathological or all accusations false. Some classifications identify real harms and make effective intervention possible, and the research problem is to estimate, case by case, how much of each kind of contribution a given classification makes.

4 Scapegoating as Causal Misidentification and Its Legitimate Counterpart

Scapegoating is treated here as more than a moral error. It is a mechanism that converts a diffuse structural problem into a conflict with an identifiable party, and the conversion has two effects that bear on reproduction. It simplifies coordination, because a shared opponent is easier to organize around than a distributed condition, and it hides the generating conditions of the original problem, because attention and resources directed at the target are not directed at those conditions. Girard's account of collective accusation gives the first effect a central place, since on his view accusation produces cohesion and the cohesion rewards the practice independent of the accused's actual contribution to the crisis [7, 8]. Riordan's review examines the empirical basis of the strong evolutionary version, describes it as speculative with little direct empirical support for some of its premises, and finds commonalities between it and other evolutionary hypotheses, such as Wrangham's execution hypothesis on self-domestication [16], and, as in Section 3, the essay requires only the weaker proposition that accusation can bind a group and give the practice an incentive to recur.

4.1 Five Kinds of Attribution

The analysis requires distinguishing scapegoating from forms of attribution that are legitimate, and five kinds are separated. Causal attribution identifies a factor whose change would change the outcome. Individual accountability assigns responsibility to a person for particular conduct that the person undertook. Collective responsibility assigns responsibility to a group for its coordinated acts or for practices it has institutionalized. Stereotyping infers conduct from membership of a category without evidence about the individual. Scapegoating assigns blame for a harm to a party that is not its principal cause, in a way that is sustained by the cohesive or mobilizing benefit to those who accuse and not by the evidence about the target.

Two of three tests assess whether an attribution is valid, and the third assesses whether identifying the target explains the pattern. The first is the evidence standard: the attribution must rest on evidence that can be checked by parties who do not benefit from it. The second is proportionality: the severity of the response must scale with the target's actual contribution to the harm, and a response that stays constant as evidence about the contribution changes is a sign that the benefit to the accusers is doing the work. The third connects the analysis to the model and concerns explanatory sufficiency, not the validity of holding someone responsible. It asks whether removing or sanctioning the target changes the

generating conditions, so that the harm does not recur through successors who occupy the same position under the same incentives. In the terms of Section 2, a target is replaceable when the mechanism M is carried forward by its successors. A replaceable target can still be correctly held responsible for particular conduct, so failing the third test does not weaken individual accountability. It shows only that the identification of a culpable party is not an explanation of the pattern, and an attribution that fails it should not be presented as one.

4.2 How Scapegoating Can Be Reproductive

In the model of Section 7, scapegoating can raise reproductive capacity through two of the channels of the transmission coefficient. Through the classification channel γ_K , it hardens the category assigned to the target into an identity that organizes the conduct of those assigned to it and the expectations of those who encounter them, in the manner described by Hacking [10]. Through the institutional channel γ_I , it gives institutions a defined population on which to build budgets, procedures, and jurisdiction. It can also reduce the extinguishing channel γ_{ext} relative to what would otherwise occur, since attention and resources committed to the accused are not committed to removing the enabling conditions. The benefits and costs are distributed unevenly, because the cohesion gains accrue to the accusers while the costs fall on the target and on the accuracy of the diagnosis, and this asymmetry is part of the incentive structure I of the mechanism. None of these channels operates automatically. Whether scapegoating raises capacity in a given case is an empirical question to be answered by the estimation strategy of Section 7.

The reproductive effects do not depend on the accusation being false. A well-founded campaign against a genuinely culpable party can also harden adversarial classifications or create a standing bureaucracy, and a groundless accusation need not do so. Scapegoating, which concerns the misidentification of cause, and reproductive correction, which concerns effects on capacity, are therefore analytically independent, and the framework is not a criticism of blame or accountability as such.

4.3 The Recursive Case

Movements organized against scapegoating face a particular risk. A campaign that identifies a class of persons as scapegoaters, and organizes opposition against them, repeats the structure it opposes with a new target, and it acquires the same incentives to sustain the category. The three tests apply to such a campaign as they apply to any other attribution. An anti-scapegoating practice passes if it directs its attention to practices and mechanisms, such as the procedures by which collective accusation is organized, and not to persons, if its

claims rest on evidence that can be checked by parties who do not benefit from them, and if it gives those it criticizes a means of leaving the category by changing their conduct. A practice that fails these tests is itself an instance of the pattern, and the framework offers no exemption to those who adopt its vocabulary.

5 Corrective Institutions and the Political Economy of Management

Institutions created to resolve a problem can develop incentives to preserve their own authority, budgets, or jurisdiction. This does not mean that corrective institutions manufacture the problems they address, and the people who staff them may sincerely want those problems solved. It means that organizational survival can become partly decoupled from successful problem resolution, so that the institution’s standing depends on the persistence of the conditions that justify it. In the model of Section 7, developed below, this is the institutional channel γ_I , and the present section describes the mechanisms through which the channel operates and the indicators by which it can be detected.

5.1 Mechanisms

Five mechanisms from the literature on organizations and regulation bear on the channel. Principal-agent problems arise because those who fund a corrective institution cannot directly observe its effort or its effect on the underlying harm, and so reward proxies that the institution can influence. Regulatory capture, in the sense developed by Stigler, describes the tendency of a regulatory body to be shaped over time by the interests of those it regulates or by other concentrated interests [19], and the concept extends to any corrective body dependent on a narrow set of constituencies. Bureaucratic self-preservation describes the tendency of organizations to seek to maintain and extend their mandates, staff, and budgets. Co-optation, in Selznick’s sense, describes how an organization absorbs elements of its environment in order to preserve its stability, and in doing so alters its goals [17]. Finally, Goodhart effects arise when a measure of corrective success becomes a target for optimization, so that the measure ceases to track what it was meant to track. Goodhart’s original observation concerned monetary policy, where a statistical regularity tends to break down once it is used for control [9], and the broader formulation in terms of measures and targets, together with an ethnographic account of audit in universities, is Strathern’s [20]. Each mechanism is a way in which the institution’s reward structure I can come apart from the reduction of H or R . Wilson’s study of bureaucracy cautions against reading persistence

as pathology, since organizational tasks, constraints, and environments shape behavior [26], while Power’s account of the audit society [15] and Campbell’s analysis of the corruption of quantitative indicators [3] describe how verification practices and indicators come to organize the institutions they evaluate.

5.2 The Reward Structure

The central question is whether a corrective institution is rewarded for eliminating the conditions that justify its existence or for managing them indefinitely. An institution whose resources fall when recurrence falls has an incentive to reduce capacity, which corresponds to a large extinguishing channel γ_{ext} . An institution whose resources scale with its caseload, with the number of violations it identifies, or with the activity it can document has the opposite incentive, which corresponds to a large institutional channel γ_I . The model does not require that institutions behave cynically. It requires only that behavior is shaped, on average and over time, by what the institution is rewarded for, and that those rewards can be observed and measured.

An institution’s survival is not by itself reproduction of the harm. The institutional channel requires a demonstrated causal sequence in which correction produces institutional reinforcement, institutional reinforcement raises reproductive capacity, and reproductive capacity raises future harm. The middle step is not automatic. An institution might preserve its funding without preserving the harm, and might even become redundant while the harm stays suppressed, which is self-preservation without pathological recursion. The question is therefore not only whether an institution benefits from the persistence of a problem but whether its response causally increases the capacity of the mechanism that produces it.

5.3 Indicators

The channel can be assessed through indicators that do not depend on knowing anyone’s motives. The first is whether funding scales with caseload or documented activity instead of with measured recurrence after turnover. The second is whether the institution’s metrics of success count activity, such as cases processed or violations found, instead of outcomes. The third is whether the institution’s jurisdiction has expanded while reproductive capacity, as measured by the proxy of Section 2, has stayed flat or risen. The fourth is whether the institution has ever been wound down, narrowed, or redefined on the ground that its targets were met, since the absence of any such event over a long period is informative. The fifth concerns detection: if an institution is rewarded for discovering violations, the detection responsiveness η of Section 7 is positive, and measured prevalence will rise with

its activity whatever the underlying trend, so its own statistics cannot be used to evaluate it. Together these indicators allow the institutional channel to be estimated separately from the classification channel, and they do not presuppose that any particular institution is pathological.

5.4 Recurrent Demand and Induced Demand

Recurrence of an institution's activity is not by itself evidence that the harm has not been extinguished or that the institution reproduces it. Some harms are generated continuously by external conditions that no institution can eliminate. A fire department does not become pathological because fires continue to occur, and an epidemiological surveillance system does not fail because new infections arise. The framework therefore distinguishes recurrent demand for correction, which persists because external conditions generate the harm, from correction-induced reproduction of demand, in which the operation of the corrective apparatus itself raises later demand. Only the second supports the thesis, and in the model it is the coefficient g , estimated net of any external source of harm. An institution whose caseload is stable or rising is a case of recurrent demand unless the indicators above, together with the mediation test of Section 7, show that its own operation contributes to the demand.

5.5 What the Argument Does Not Claim

The argument does not claim that corrective institutions generally raise the capacity of the harms they address, that institutional self-interest is the dominant force in their behavior, or that institutions cannot be designed to be rewarded for resolution. Some are plausibly designed in that way, for example through outcome-based funding or mandatory sunset provisions, and the framework predicts that such institutions have a smaller institutional channel and a larger extinguishing channel. The claim is conditional and comparative: where the reward structure favors management over elimination, the transmission coefficient of Section 7 is more likely to be positive, and the institutions can be compared by the indicators above. Section 11 returns to the question of how reward structures can be designed to point toward resolution.

6 Reflexivity and Admissibility Conditions

A theory that says attempts to identify culprits reproduce the pathology they target must be able to apply that claim to itself. Otherwise it names a universal source of harm, calls it

a disease, and invites collective mobilization against it, which is the structure the theory describes. This section precedes the formal model because it fixes what the model is permitted to claim. It states the conditions an instance must satisfy before it counts as pathological recursion, sorts the possible outcomes of correction into five classes with distinguishing observations, and sets rules intended to prevent the essay’s own category from hardening into a label for a type of person or movement.

6.1 The Theory Applied to Itself

The risk is one of self-sealing. Strong absorption theses hold that opposition to a system is itself incorporated into the system and so strengthens it, and one influential reading of Baudrillard’s account of simulation [2] takes this form, although that account is broader than any single reading of it and the point here concerns the logical form of the thesis, not a verdict on its author. A thesis of that form can accommodate any outcome. If a protest succeeds it has been absorbed, if it fails it has been irrelevant, and if it is ignored it has been neutralized, so no observation could count against the claim and the claim says nothing about the world. The present essay claims something much weaker and much more exposed: that some corrective interventions, under conditions that can be stated and measured, raise the capacity of the pattern they oppose to recur. The weaker claim earns its content by being refutable, and the remainder of this section specifies how.

The reflexive requirement extends to the essay’s own proposals. The forms of repair discussed in Section 11 are interventions, and they are subject to the same model. A body created to verify claims, to audit institutions, or to detect recursive correction can develop institutional dependency, can harden the categories it applies, and can distort detection, so its proposal must be accompanied by an account of the channels through which it could raise capacity and the observations that would show it had done so. The theory receives no exemption on the ground that it is the one pointing out the problem.

6.2 Admissibility Conditions for an Instance

A claimed instance of pathological recursion is admissible only if it supplies four things. First, it identifies a specific mechanism $M = (I, K)$ in the sense of Section 2, naming the incentive structure and the classification practice that are said to persist across turnover. A general reference to “the system” or “the culture” does not qualify. Second, it identifies a measurable outcome, in the form of a harm variable with a detection method that is independent of the corrective apparatus, for the reasons given in Section 7. Third, it specifies a plausible counterfactual, meaning the path the harm would have followed without the intervention,

together with the evidence on which that path is estimated. Fourth, it states observations that would disconfirm the attribution, and states them before the data are examined. The five refutation conditions of Section 7 serve as the default list, and a particular case may add further conditions of its own.

These requirements mirror the model. The mechanism corresponds to the unit of reproduction, the outcome to H and its observed counterpart Y , the counterfactual to the baseline from which deviations are measured, and the disconfirming observations to the refutation criteria. An instance that cannot be stated in these terms is not rejected as false but is excluded as not yet testable, and the essay does not assert that such cases are free of recursion.

6.3 Five Outcomes and Their Signatures

Not every failure of correction is pathological recursion, and the essay must show that recursive reproduction is analytically distinct from the other outcomes instead of renaming them. The outcomes of an intervention are sorted into five classes, each with a signature in the quantities of the model, and the classes are presented as signatures under the model’s assumptions and not as an exhaustive classification of real interventions. A successful repair reduces harm and does not raise capacity, and an unsuccessful repair has no detectable effect on either. Adaptive evasion reduces harm temporarily and, in its simplest form, returns it toward baseline as actors adapt. Displacement reduces harm in the targeted domain while raising a second harm. Recursive reproduction means that correction raises capacity, $g > 0$, and it has a narrower subcase, pathological reversal, in which the reproductive effect outweighs suppression and harm eventually exceeds baseline. A remedy can have a genuine reproductive side effect while remaining beneficial overall, and the framework treats that mixed case as reproduction and not as pathological reversal. Table 1 states the signatures.

The classes are ideal types and can co-occur. An intervention may displace some harm, provoke some evasion, and also raise capacity, and displacement can itself create new institutional incentives and so become reproductive. Adaptive evasion in its simplest form leaves harm at or below baseline, but actors may also evade by adopting methods more harmful than the original ones, and displacement need not move a comparable amount of harm, since a few episodes may be displaced into a much more damaging domain. In such cases the task is to estimate the contribution of each, not to assign a single label. The distinctive claim of the essay is confined to the reproductive pathway, and it holds only where the mediating pathway through capacity can be shown to operate. The central discriminator is that later excess harm or recurrence is mediated by recurrence after turnover and survives an indepen-

Outcome	Harm H	Capacity R	Distinguishing observation
Successful repair	Falls and stays below baseline	Unchanged or falls	Recurrence after turnover is lower than in comparison settings
Unsuccessful repair	Unchanged (b near zero)	Unchanged	No detectable effect of C on H or R
Adaptive evasion	Falls, then returns toward baseline (in the simplest model, never above it)	Unchanged	Effectiveness of C erodes as actors adapt, with no mediation by R
Displacement	Falls in the targeted domain	Unchanged	A second harm H' rises, by an amount to be weighed against the fall in H
Recursive reproduction	Falls, then may exceed baseline (pathological reversal if it does)	Rises ($g > 0$)	Later harm or recurrence is mediated by capacity, as set out in the text

Table 1: Outcomes of correction and their signatures in the model.

dent detection measure. A capacity increase may also leave no observed recurrence within the measurement horizon, which is neither evidence that the increase occurred nor evidence that it did not, and the design must distinguish a theoretically possible latent effect from an empirically supported one.

6.4 Rules Against Hardening the Category

Hacking’s account of looping effects applies directly to the essay’s own vocabulary [10]. A category such as recursive blame, once available, could be applied to persons or movements as a label, and the label could then be used as an accusation, which would reproduce the structure the essay describes. Three rules are adopted to limit this. The first is that the category applies to specified practices, mechanisms, and institutional arrangements. Responsibility for establishing or maintaining those arrangements may be attributed to identifiable actors, but such attribution does not license an inference from conduct to an immutable kind of person, so that a claim of the form “this corrective practice raises the capacity of the harm it targets” is admissible while a claim of the form “these people are recursive” is not. The second is that every application states, in advance, the observation that would count against it. The third is that no application is permitted to function as a verdict of culpability, since the framework concerns the conditions under which conduct recurs, while questions of individual responsibility are treated separately in Section 10. These rules do not remove the risk, but they make misuse of the category visible as a departure from the

framework.

A worked pair illustrates the rules as a decision procedure. An admissible application is the claim that a funding rule scaling with caseload raises the capacity of the harm it targets by giving the corrective body an incentive to sustain the activity, since it names a mechanism, a measurable outcome, and observations that would count against it. An inadmissible application is the claim that the officials of an agency are recursive blamers, since it assigns the category to a kind of person and states no disconfirming observation. A borderline case is the claim that the network around a particular investor functions as a mechanism. It is admissible only if the network is specified through roles, reward contingencies, and classification rules, with preregistered outcomes and disconfirming observations, and it is inadmissible if it is used to infer the character of the people in it from the outcome.

7 A Formal Model of Pathological Recursion

This section develops a deliberately minimal dynamic model whose purpose is to establish one modest result and to say exactly when it fails. The result is that, under explicitly stated conditions, a corrective intervention reduces harm over a short horizon and increases it over a longer one, and that this pattern is determined by the relative strength of three effects: the immediate suppression of harm, the effect of correction on reproductive capacity, and the persistence of that capacity. The section states its assumptions, specifies the channel through which correction can affect capacity, proves two propositions, presents counterexamples, separates actual from observed prevalence, distinguishes the pattern from adaptive evasion, displacement, and ordinary failure, and closes with the observations that would refute the model in a given case. The structure belongs partly to the system-dynamics tradition of delayed feedback, including Forrester’s work on counterintuitive policy effects [6] and Senge’s archetype of fixes that fail [18], and the contribution here is not the observation that corrective feedback can be delayed but its connection to classification, institutional incentives, and mechanism-level reproduction.

7.1 State Variables and Assumptions

All quantities are measured as deviations from the path the system would follow in the absence of intervention. Let H_t be the deviation of actual harm prevalence from its no-intervention baseline, R_t the corresponding deviation of reproductive capacity as defined in Section 2, and $C_t \geq 0$ the intensity of corrective intervention. The true prevalence of Section 2 therefore equals a baseline level h_0 plus H_t , and reproductive capacity enters as a

scalar state variable that stands in, in reduced form, for the latent quantity defined there, so that R_t denotes a level in Section 2 and its deviation from baseline here. The state equation for capacity is an approximation adopted for tractability and is not derived from that definition, whose expected descendant count depends on the horizon, the environment, and the reproduction process, and the persistence law it assumes is itself an empirical claim. Because the system is linear, deviations from baseline behave exactly like responses to an impulse, and the baseline itself drops out of the analysis. The dynamics are

$$H_{t+1} = aH_t - bC_t + dR_t, \quad R_{t+1} = \rho R_t + gC_t. \quad (2)$$

The model rests on five assumptions, each of which is a substantive claim that can fail. **A1 (Persistence)**. Harm and capacity decay when left alone, so that $0 \leq a < 1$ and $0 \leq \rho < 1$. **A2 (Suppression)**. Correction reduces harm directly in the following period, so that $b > 0$. **A3 (Yield)**. Reproductive capacity produces harm, so that $d > 0$. **A4 (Linearity and stationarity)**. The parameters are constant over the horizon and the effects of interventions add. **A5 (Exogenous intervention)**. Intervention intensity is set independently of current harm, so that any association between C_t and later harm is not produced by correction responding to harm. The coefficient g is deliberately left unrestricted in sign, for reasons given next. Assumptions A4 and A5 are strong, and the closing subsection returns to the question of how they are relaxed and tested.

7.2 The Transmission Channel from Correction to Capacity

The central refinement of the model is that an increase in intervention is not assumed to raise later harm. Whether it does so depends on the sign and size of g , which in turn depends on a transmission mechanism that must be specified and defended case by case. The coefficient is decomposed as

$$g = \gamma_I + \gamma_K - \gamma_{\text{ext}}, \quad (3)$$

where $\gamma_I \geq 0$ is the institutional channel, through which intervention raises the revenue, jurisdiction, or standing of the corrective apparatus and thereby its incentive and ability to sustain the activity that regenerates the conditions of the harm, $\gamma_K \geq 0$ is the classification channel, through which intervention hardens the category assigned to those involved into a stable adversarial identity, and $\gamma_{\text{ext}} \geq 0$ is the extinguishing channel, through which intervention removes enabling conditions and lowers capacity. Correction raises capacity only if $\gamma_I + \gamma_K > \gamma_{\text{ext}}$. The decomposition is not a result. It is a requirement that anyone asserting pathological recursion in a particular case identify which channels operate, estimate their

relative weights, and show that the first two outweigh the third.

The decomposition becomes a measurement scheme only if each channel has observable counterparts. Table 2 lists candidates. No single observable identifies a channel, and the weights must be estimated jointly, but the table shows what data would distinguish the institutional channel from the classification channel and from extinguishing. The extinguishing channel is the hardest to observe, because it concerns what would have removed the enabling conditions, which is a counterfactual, so its observables are proxies for the removal of conditions and not positive indicators of it.

Channel	Mechanism	Candidate observables
γ_I	Institutional reinforcement	Elasticity of institutional resources with respect to caseload; expansion of mandate or jurisdiction after intervention; metrics that count activity instead of recurrence
γ_K	Classification hardening	Rate at which those classified adopt the category as an identity; persistence of category assignment; organization of procedures around the category
γ_{ext}	Removal of enabling conditions	Uptake of remediation after turnover; changes in the enabling rules or rewards; exits from the category through changed conduct

Table 2: Candidate observables for the three channels of the transmission coefficient.

7.3 Delayed Reversal Under a Single Intervention

Consider a one-period intervention of unit intensity at time t_0 , with $C_t = 0$ at every other time and $H_{t_0} = R_{t_0} = 0$. Write $h_n = H_{t_0+n}$ for the response of harm n periods after the intervention.

Proposition 1 (Delayed reversal). *Under A1 through A4, the response of harm to a unit pulse satisfies $h_1 = -b < 0$ and $h_{n+1} = ah_n + dg\rho^{n-1}$ for $n \geq 1$, and:*

- (i) $h_2 = dg - ab$, so harm exceeds its baseline two periods after the intervention if and only if $dg > ab$;
- (ii) $h_n \rightarrow 0$ as $n \rightarrow \infty$, so the effect of a single pulse is temporary;
- (iii) the cumulative effect of the pulse is

$$S = \sum_{n \geq 1} h_n = \frac{\frac{dg}{1-\rho} - b}{1-a}, \quad (4)$$

so cumulative harm over the whole horizon rises above baseline if and only if $dg > b(1 - \rho)$.

Proof. At t_0 the intervention enters the harm equation with coefficient $-b$, giving $h_1 = -b$. It enters the capacity equation with coefficient g , so $R_{t_0+k} = g\rho^{k-1}$ for $k \geq 1$, since no further intervention occurs. For $n \geq 1$ the harm equation then gives $h_{n+1} = ah_n + dR_{t_0+n} = ah_n + dg\rho^{n-1}$. Setting $n = 1$ gives $h_2 = -ab + dg$, which is positive exactly when $dg > ab$, proving (i). The recursion is a stable linear recursion with $a < 1$ driven by a forcing term that decays geometrically at rate $\rho < 1$, so $h_n \rightarrow 0$, proving (ii). For (iii), sum the recursion over $n \geq 1$. Writing $S = \sum_{n \geq 1} h_n$, the identity $h_1 = -b$ and $h_{n+1} = ah_n + dg\rho^{n-1}$ give

$$S = h_1 + \sum_{n \geq 1} h_{n+1} = -b + aS + dg \sum_{n \geq 1} \rho^{n-1} = -b + aS + \frac{dg}{1 - \rho},$$

so $S(1 - a) = dg/(1 - \rho) - b$, which is (4). Because $1 - a > 0$, $S > 0$ if and only if $dg/(1 - \rho) > b$, that is, $dg > b(1 - \rho)$. $\square$

Corollary 1. *If $g \leq 0$, then $h_n < 0$ for every $n \geq 1$. A positive net effect of correction on capacity is therefore necessary for any reversal at any lag.*

Proof. With $g \leq 0$ and $d > 0$, the forcing term $dg\rho^{n-1}$ is nonpositive at every step, and $h_1 = -b < 0$, so by induction every h_n is negative. $\square$

7.4 Three Thresholds

The second-period condition of Proposition 1 is one of three distinct thresholds, and a fuller statement of the impulse response shows how they differ.

Proposition 2 (Three thresholds). *Under A1 through A4 with $g > 0$, the response to a unit pulse is, for $n \geq 1$,*

$$h_n = -ba^{n-1} + dg \sum_{j=0}^{n-2} a^{n-2-j} \rho^j, \quad (5)$$

which equals $-ba^{n-1} + dg(a^{n-1} - \rho^{n-1})/(a - \rho)$ for $a \neq \rho$, and $-ba^{n-1} + dg(n-1)a^{n-2}$ for $a = \rho$ and $n \geq 2$. Three thresholds follow. (i) Harm exceeds baseline at the second period if and only if $dg > ab$. (ii) Harm exceeds baseline for all sufficiently large n if and only if $\rho \geq a$ or $dg > b(a - \rho)$. (iii) Cumulative harm exceeds baseline if and only if $dg > b(1 - \rho)$. Each of (i) and (iii) implies (ii), and (i) and (iii) are independent of one another.

Proof. The closed form follows by induction from $h_1 = -b$ and $h_{n+1} = ah_n + dg\rho^{n-1}$, and claims (i) and (iii) are Proposition 1. For (ii), when $\rho > a$ the term in ρ^{n-1} has the positive

coefficient $dg/(\rho - a)$ and dominates the geometric decay at rate a . When $\rho = a$ the term $(n - 1)a^{n-2}$ dominates. When $\rho < a$ the term in a^{n-1} has coefficient $-b + dg/(a - \rho)$, which is positive if and only if $dg > b(a - \rho)$. If $dg > ab$ then $dg > b(a - \rho)$ because $a - \rho \leq a$, and if $dg > b(1 - \rho)$ then $dg > b(a - \rho)$ because $a < 1$, so each of (i) and (iii) implies (ii). The two are independent because $ab < b(1 - \rho)$ exactly when $a + \rho < 1$, and either order is possible. In Case C of Table 3, (i) holds and (iii) fails, while with $a = 0.9$, $\rho = 0.5$, $b = 1$, and $dg = 0.7$, (iii) holds and (i) fails. $\square$

The conditions answer different questions: whether a rebound occurs soon, whether one eventually occurs, and whether the net effect is harmful. With $a = 0.5$, $\rho = 0.9$, $b = d = 1$, and $g = 0.3$, the responses begin $h_1 = -1$, $h_2 = -0.2$, and $h_3 = 0.17$, so that correction suppresses harm, remains beneficial at the second period, and produces an excess only later. The second-period condition is therefore not a general threshold for whether reversal occurs at all.

7.5 Continuing Policy

A single pulse and a continuing policy are different objects, and the distinction matters for the thesis, because correction in practice is usually continuing. Suppose intervention is held at a constant intensity $c > 0$ from t_0 onward.

Proposition 3 (Sustained correction). *Under A1 through A4 with $C_t = c$ for all $t \geq t_0$, harm falls by bc in the first period and converges to the deviation $H^* = cS$, with S as in (4). The sustained policy therefore suppresses harm initially and raises its long-run prevalence above the no-intervention baseline if and only if $dg > b(1 - \rho)$.*

Proof. Capacity converges to the fixed point of $R = \rho R + gc$, namely $R^* = gc/(1 - \rho)$. Harm converges to the fixed point of $H = aH - bc + dR^*$, namely $H^* = (-bc + dgc/(1 - \rho))/(1 - a) = cS$. Convergence follows from $a, \rho < 1$. The first-period change is $-bc$ by the harm equation. Since $1 - a > 0$, $H^* > 0$ if and only if $dg > b(1 - \rho)$. $\square$

The comparison with Proposition 1 is informative. The steady-state effect of a continuing policy of intensity c equals c times the cumulative effect of a single pulse, so the sign condition that governs one-time interventions also governs sustained ones. The first proposition describes a transient, while the second describes what a society that keeps correcting at a fixed intensity should expect to observe in the long run. That the steady-state response equals the cumulative impulse response is a standard property of linear time-invariant systems, and the contribution here lies in the interpretation of the state variables and channels. The thesis of the essay takes a precise form in the model, for the specified harm and not

for welfare in general: correction suppresses harm at first and raises its long-run level above the no-intervention baseline when $g > 0$ and $dg > b(1 - \rho)$, which is the case of pathological reversal, while the broader phenomenon of mechanism reproduction requires only $g > 0$.

7.6 Counterexamples and Boundary Cases

The model also says when the pattern does not occur. Table 3 reports the response h_n to a unit pulse for three parameter sets, computed directly from the recursion in Proposition 1.

Case	(a, ρ, b, d, g)	h_1	h_2	h_3	h_4	h_5	S
A: no reversal	(0.5, 0.5, 1, 0.2, 0.2)	-1	-0.460	-0.210	-0.095	-0.043	-1.84
B: reversal	(0.5, 0.9, 1, 1, 0.6)	-1	0.100	0.590	0.781	0.828	10.0
C: rebound, net benefit	(0.8, 0.1, 1, 1, 0.85)	-1	0.050	0.125	0.109	0.088	-0.278

Table 3: Response of harm to a unit pulse of correction, in deviations from the no-intervention baseline.

In Case A the product dg is small relative to both ab and $b(1 - \rho)$, so suppression dominates at every lag and the intervention is simply beneficial. In Case B capacity is both strongly raised by correction and highly persistent, so harm first falls, then exceeds its baseline from the second period onward, and the cumulative effect is strongly positive. Case C shows that the two conditions in Proposition 1 are distinct. Here $dg > ab$, so harm rebounds above baseline after the first period, yet $dg < b(1 - \rho)$, so the cumulative effect is negative and the intervention is net beneficial despite the rebound. A temporary rebound is therefore not evidence of pathological recursion, and the essay's thesis concerns the cumulative and sustained conditions, not the appearance of any rebound.

7.7 Actual and Observed Prevalence

Harm is seldom observed directly, and intervention changes how much of it is detected, reported, and classified. Let observed prevalence be

$$Y_t = (1 + \eta C_t)(h_0 + H_t), \quad (6)$$

where $h_0 > 0$ is baseline harm, so that $L_t = h_0 + H_t$ is the true level of harm recovered from its deviation and Y_t is the observed level, and η measures the responsiveness of detection to correction, positive when intervention increases reporting and negative when it suppresses it. Under a sustained policy of intensity c , the observed change relative to baseline is

$$\Delta Y^* = cS(1 + \eta c) + \eta c h_0. \quad (7)$$

If correction truly reduces long-run harm, so that $S < 0$, observed prevalence nevertheless rises whenever $\eta h_0 > (1 + \eta c)|S|$, so improved detection can make a successful intervention look unsuccessful. If $\eta < 0$ and $S > 0$, concealment can make a counterproductive intervention look successful. Because classification is among the mechanisms under study, the second error is the more dangerous one. The measurement problem is therefore not a technical afterthought, and any empirical test must rely on a detection measure that is independent of the corrective apparatus.

7.8 Pathological Recursion and Its Neighbors

A model that merely redescribed familiar phenomena would establish nothing, so the section states what distinguishes pathological recursion from its neighbors within the model. Ordinary intervention failure corresponds to $b \approx 0$ with $g \approx 0$, in which case harm is unchanged and no reversal occurs. Adaptive evasion corresponds to an effectiveness coefficient that erodes as actors adapt, which can be written b_t decreasing in cumulative intervention, with $g = 0$. In that case harm may return to its baseline but cannot exceed it, because nothing in the system raises harm above the level it would have reached without intervention. Displacement corresponds to harm in the targeted domain falling while a second harm variable H' rises, and it is detected by measuring H' as well as H .

The discriminating signature of correction-induced reproduction is a positive pathway from correction to capacity, $g > 0$, that mediates later harm. By Corollary 1, any overshoot of the baseline requires $g > 0$, and by Proposition 2 overshoot at some lag requires that pathway to be strong enough relative to suppression and persistence. In the simplest model, overshoot of the no-intervention baseline therefore indicates a capacity-mediated pathway, since adaptive evasion with $g = 0$ cannot produce it. Overshoot is not necessary for reproduction, because when suppression dominates, correction can raise capacity and remain net beneficial, a mixed case that the framework still classifies as reproduction. Nor is overshoot alone sufficient in general, because evasion by more harmful methods can also raise harm above baseline, so mediation through the proxy for capacity must be shown in either case. This yields a testable difference. Under adaptive evasion, later harm should be explained by changes in the effectiveness of correction, while under reproduction later harm or recurrence above its counterfactual level should be mediated by the proxy for capacity, namely recurrence after turnover.

7.9 Extension: Intervention That Responds to Harm

Assumption A5 treats intervention as set independently of current harm, but many corrective regimes scale their intensity with the harm they observe. Consider the proportional rule $C_t = \phi L_t$ with gain $\phi \geq 0$, where $L_t = h_0 + H_t$ is the level of harm. Substituting into (2) gives a linear system in (H_t, R_t) ,

$$\begin{pmatrix} H_{t+1} \\ R_{t+1} \end{pmatrix} = A \begin{pmatrix} H_t \\ R_t \end{pmatrix} + \phi h_0 \begin{pmatrix} -b \\ g \end{pmatrix}, \quad A = \begin{pmatrix} a - b\phi & d \\ g\phi & \rho \end{pmatrix}. \quad (8)$$

Proposition 4 (Closed-loop equilibrium and stability). *Under A1 through A4 and the rule $C_t = \phi L_t$ with $\phi \geq 0$, let S be defined by (4), and let $T = \text{tr } A$ and $D = \det A$. Then*

$$\det(I - A) = 1 - T + D = (1 - a)(1 - \rho)(1 - \phi S). \quad (9)$$

If $\phi S \neq 1$, the system has a unique algebraic equilibrium, with level of harm $L^ = h_0/(1 - \phi S)$. For $h_0 > 0$ this level is positive if and only if $\phi S < 1$. At $\phi S = 1$ no finite equilibrium exists when $g, \phi, h_0 > 0$. If $\phi S > 1$, the algebraic equilibrium has a negative level of harm, which indicates that no nonnegative bounded equilibrium exists under the assumed feedback law, and A has a real eigenvalue greater than one. The equilibrium is asymptotically stable if and only if $1 - T + D > 0$, $1 + T + D > 0$, and $|D| < 1$, and the first of these conditions is $\phi S < 1$, so that $\phi S < 1$ is necessary but not sufficient for stability.*

Proof. Direct expansion gives $1 - T + D = (1 - a + b\phi)(1 - \rho) - dg\phi = (1 - a)(1 - \rho) + \phi[b(1 - \rho) - dg]$, and since $S = (dg - b(1 - \rho))/((1 - a)(1 - \rho))$ this equals $(1 - a)(1 - \rho)(1 - \phi S)$. When the determinant is nonzero, the equilibrium equations $(I - A)x^* = \phi h_0(-b, g)^\top$ have a unique solution, and eliminating $R^* = g\phi L^*/(1 - \rho)$ from $(1 - a)(L^* - h_0) + b\phi L^* = dR^*$ gives $L^* = h_0/(1 - \phi S)$. When $\phi S = 1$ the matrix $I - A$ is singular with left null vector $w = (g\phi, 1 - a + b\phi)$, and $w^\top \phi h_0(-b, g)^\top = (1 - a)g\phi h_0 \neq 0$, so the forcing is not in its range and no equilibrium exists. For $\phi S > 1$, the characteristic polynomial $p(\lambda) = \det(\lambda I - A)$ satisfies $p(1) = \det(I - A) < 0$ and $p(\lambda) \rightarrow \infty$ as $\lambda \rightarrow \infty$, so p has a real root greater than one. The stability conditions for a two-by-two real matrix are the Schur conditions stated. $\square$

Three properties must be kept apart: the existence of an algebraic equilibrium, its admissibility as a nonnegative level of harm, and its dynamic stability. If $S < 0$, so that correction is beneficial in the open loop, feedback lowers the equilibrium level below the baseline h_0 . If $S > 0$, feedback raises it above baseline for any positive gain, and the level diverges as the gain approaches the critical value $\phi^* = 1/S$ from below. At ϕ^* no finite equilibrium exists,

and beyond it the algebraic equilibrium is negative and unstable. For Case B of Table 3, $S = 10$ and $\phi^* = 0.1$. A gain of 0.05 doubles equilibrium harm, a gain of 0.09 raises it tenfold, and a gain of 0.1 or more admits no bounded positive equilibrium. Direct simulation of the recursion agrees with these values.

Divergence in the linear model indicates the failure of a bounded equilibrium under the assumed feedback law. It does not imply that actual harm can grow without physical, institutional, or resource constraints, and the linear approximation should not be read outside the range in which levels remain nonnegative.

Stability can also fail at high gain for the opposite reason, since the second and third Schur conditions bound the gain from above through overcorrection. In Case C, where $S < 0$, a gain of 2 gives eigenvalues of approximately -2.007 and 0.907 . The unstable mode is negative, so the divergence is oscillatory, which illustrates the difference between destabilization by positive reinforcement and destabilization by excessive negative feedback.

A second loop arises when the regime responds to measured prevalence. With observed prevalence $Y_t = (1 + \eta C_t)L_t$ and contemporaneous detection, the rule $C_t = \phi Y_t$ gives the algebraic solution $C = \phi L / (1 - \phi \eta L)$ for a given level of harm L . This is finite and nonnegative only when $\phi \eta L < 1$, and at or beyond that threshold the algebraic rule admits no finite nonnegative solution, which is a singularity and not yet a dynamic result. A dynamic statement requires an explicit update law. With $C_{t+1} = \phi Y_t = \phi(1 + \eta C_t)L$ and L held fixed, intensity follows $C_{t+1} = \phi L + \phi \eta L C_t$, which converges to $\phi L / (1 - \phi \eta L)$ if $\phi \eta L < 1$ and grows without bound if $\phi \eta L \geq 1$. The detection loop therefore has its own threshold, independent of S , because more correction produces more recorded harm, which produces more correction. This is the formal counterpart of the fifth indicator of Section 5, and it shows why an institution rewarded for the violations it finds cannot be evaluated by its own statistics. The two loops are analyzed separately, the detection loop with L held fixed, so the separate thresholds are approximations that hold when one loop operates much faster than the other. The joint system has its own stability region, which is not in general the intersection of the two conditions, and its analysis is a further extension. Whether a given regime lies in the stable range, and what its gain is, are empirical questions that require modeling the response rule alongside the dynamics.

7.10 Falsification Criteria

A case is inconsistent with the model if any of the following is established, and each condition must be assessed with its uncertainty. First, if the estimated effect of correction on capacity, g , is nonpositive with an interval that excludes meaningful positive values, then

by Corollary 1 no reversal is possible and the case is not an instance of reproduction. Second, if the estimated parameters satisfy $dg < b(1 - \rho)$ with an interval that excludes the boundary, then sustained correction reduces long-run harm and the case is not an instance of pathological reversal, though it may still be an instance of reproduction if $g > 0$. Third, if no overshoot of the counterfactual baseline appears over a horizon fixed in advance, the case does not support reversal, and this is decisive only if the estimated parameters imply that overshoot should have occurred within that horizon, since a persistence ρ close to one can delay effects beyond a short window. Fourth, if the apparent reversal disappears when harm is measured by a detection method independent of the corrective apparatus, it was an artifact of measurement. Fifth, if the same intervention lowers harm in the targeted domain while raising a second harm, the case is at least partly displacement, which does not by itself exclude a capacity-mediated effect, because the two can operate together.

The language of refutation should distinguish three levels. Failure to establish an individual instance means that a particular case did not satisfy the admissibility conditions or the criteria above, and it says nothing about other cases. Rejection of a proposed mechanism means that, for a stated mechanism, the estimated channel coefficients or the mediating pathway fail across the cases in which it was proposed. Evidence against the broader research program arises only when carefully selected cases that independently satisfy the proposed causal conditions repeatedly fail to exhibit the predicted outcomes. Observing a case in which the conditions are absent does not refute the general possibility, and finding successful corrective interventions is compatible with the program, which asserts that recursion occurs under particular conditions and not generally. An estimate whose interval contains both moderately negative and moderately positive values does not refute a mechanism either. It fails to establish the sign of the effect.

7.11 Identification and Limitations

Estimation is an identification problem, and fitting the two equations directly will not in general recover the parameters. The coefficients of Equation (2) are theoretical quantities, and what data can identify are their empirical counterparts under stated assumptions. Three difficulties arise. Capacity is latent and its proxy is recurrence after turnover, so if recurrence also enters the measured harm outcome, the explanatory variable and the outcome are not independent, and the capacity proxy requires its own measurement model, distinct from the harm measure. Turnover of leadership is often endogenous, since it may occur because harm has risen, because a corrective program failed, or because political conditions changed, and turnover used as an instrument or a comparison must be shown to be independent of those

causes. And assumption A5 requires variation in intervention that is independent of current harm, which candidate designs such as staggered adoption across comparable units, program termination and sunset provisions, and externally imposed changes in leadership approximate to different degrees. The two measures must differ in kind. The harm measure is a level, such as a concentration share or a count of episodes in a period, while the capacity proxy is a persistence rate conditional on turnover, such as the probability that an episode of the pattern follows the replacement of the people or institutions that carried the previous one. The proxy for capacity is the recurrence rate after turnover defined in Section 2, with horizons and decision thresholds preregistered, and the parameters are identified only to the extent that these conditions are met. A credible mediation design must also address treatment assignment, unmeasured confounding, the temporal order of mediator and outcome, and measurement error in the mediator [13]. Measurement error is represented minimally by a measurement equation $Z_t = \lambda R_t + \xi_t$ relating the observed proxy Z_t to latent capacity, with ξ_t capturing error and environmental influences not represented in R_t . The empirical target is the causal contrast

$$\Delta_R(T) = \mathbb{E}[N_{\text{desc}} \mid \text{do}(C = c), \mathcal{E}] - \mathbb{E}[N_{\text{desc}} \mid \text{do}(C = 0), \mathcal{E}], \quad (10)$$

which is a conceptual estimand whose identification requires additional assumptions or a credible experimental or quasi-experimental design. Turnover must be either plausibly exogenous or modeled as part of the causal process.

Two limitations should be stated. The linear, constant-parameter model illustrates conditions and does not demonstrate that they hold, so the empirical content lies entirely in estimating a , b , d , g , and ρ . The baseline model also treats intervention as exogenous, and the preceding extension analyzes only a proportional response rule, so that richer rules, nonlinear dynamics, and agent-based settings remain for later work and should be attempted only after the linear model has been confronted with data. The linear model also cannot represent saturation, in which correction meets diminishing returns and institutions respond by broadening their target category. That form of mission creep would be modeled by an effectiveness coefficient b_t that declines in cumulative intervention together with growth in the scope of C_t , and it is the nonlinear extension most likely to matter.

8 Communication Environments as Selection Conditions

This section asks whether particular properties of communication environments alter the reproductive capacity of destructive social patterns, independent of the intentions of the people who use them. It does not offer a history of media or a general theory of media evolution. Its task is narrower: to show how communication environments modify the mechanisms already established, by shifting the parameters of the model of Section 7, and to derive a small set of predictions that can be tested and that would count against the framework if they failed.

8.1 Affordances and Incentives

Two sets of variables must be kept apart. The first consists of a medium’s technical affordances: the speed at which content travels, the breadth of its reach, its persistence and retrievability, the ease with which it can be copied, the presence and design of recommendation systems, and the composition of the audience, including whether it consists of people who know the speaker. The second consists of the incentives that govern use of the medium: the objective a recommendation system optimizes, the revenue model of the platform, the rules for moderation, and the identity requirements imposed on participants. The two sets are distinct, and conflating them produces the claim that a technology is inherently antagonistic. A system that permits rapid distribution does not necessarily reward antagonism, and a recommendation system need not optimize engagement at the expense of accuracy. Whether an environment amplifies a destructive pattern depends on the combination of what it permits and what it rewards.

The environment enters the model through its parameters and does not require new state variables. Let θ denote the vector of affordances and incentives that characterize an environment. The coefficients of Section 7 then become functions of θ , so that the transmission coefficient is

$$g(\theta) = \gamma_I(\theta) + \gamma_K(\theta) - \gamma_{\text{ext}}(\theta), \quad (11)$$

and the persistence of capacity $\rho(\theta)$ and the detection responsiveness $\eta(\theta)$ depend on the environment as well. The claim of the section is that specific components of θ move specific coefficients, and that the direction of each movement is conditional and can be tested.

8.2 Channels Through Which Environments Act

Persistence and retrievability act on the persistence of capacity and on the extinguishing channel. When accusations and classifications remain permanently retrievable, a pattern that would otherwise fade can be reactivated, which raises ρ , and the revision or forgetting that allows a person or practice to leave a category becomes harder, which lowers γ_{ext} . Permanent records can also preserve evidence of misconduct and prevent harmful mechanisms from reappearing unnoticed, so the effect is conditional on what is preserved, who can retrieve it, and how it is interpreted. Replicability and reach act on the classification channel. A category that can be copied and distributed to large audiences at little cost reaches many people who encounter it without the context that would qualify it, and this strengthens the identity-forming effect that Hacking describes [10], so that γ_K rises.

Recommendation systems act on the institutional channel, because the platform is itself an institution with a reward structure. A platform whose revenue scales with engagement has an incentive to surface content that sustains repeated identification of threats, since such content holds attention, and this is the structure the model describes as an institution rewarded for management over resolution, with γ_I rising accordingly. This requires the further premise that antagonistic material generates the rewarded engagement more effectively than the alternatives, because engagement optimization alone does not entail antagonistic amplification. Audience composition acts on both channels at once. An audience of strangers has no stake in the individual accused and no means of checking the claims against its own knowledge, so it lowers the evidence standard of Section 4 and raises the benefit to those who accuse.

Environments can also act in the opposite direction. Provenance requirements, rights of reply, and visible context raise γ_{ext} by making it easier to correct a false attribution, and moderation that is accountable to the people it affects can slow the hardening of categories. The effects are therefore conditional mechanisms, and the section does not attribute any of them to a technology as such.

8.3 Reproduction and Visibility

Greater circulation of accusations can increase their observed frequency without increasing the underlying harmful conduct, and the section therefore carries forward the distinction between actual and measured prevalence drawn in Section 7. Environments that record, index, and display conduct increase detection, so that $\eta(\theta)$ is larger, and an increase in measured episodes after the introduction of such an environment is not by itself evidence of an increase in harm. The reverse also holds. Where an environment makes accusation

cheap, a rise in the number of accusations may measure the cost of accusing and not the rate of wrongdoing. The tests below therefore rely on measures of conduct and recurrence that are independent of the environment’s own records.

8.4 Counterexamples and Limits

The framework must accommodate cases in which rapid communication improves outcomes. Wide and fast circulation can expose wrongdoing that was previously concealed, extend accountability to powerful actors, and accelerate effective correction. In the terms of the model, such cases correspond to an environment that raises γ_{ext} or that raises detection in a way that reveals real harm and thereby enables intervention that lowers H over the longer horizon. A rise in measured prevalence immediately after the introduction of such an environment is the expected signature of better detection, and it is to be distinguished from a rise in actual harm by the independent measures just described.

The section also declines to treat small audiences, slower circulation, or unrecorded discussion as inherently superior. These arrangements can reduce performative incentives and lower the cost of tentative speech, but they can also conceal misconduct, inhibit external scrutiny, and protect entrenched authority. Their value depends on the mechanism and on the outcome being measured, and the same environment can lower γ_K while lowering detection and thereby raising the persistence of an undetected harm.

8.5 Predictions

Four conditional predictions follow, each stated with the observation that would count against it. First, under engagement-maximizing recommendation, increases in algorithmic amplification should be associated with higher recurrence after turnover of antagonistic episodes, as measured by the proxy of Section 2, and not only with higher measured frequency. If amplification raises only measured frequency while recurrence is unchanged, the effect is one of visibility. Second, under recommendation that prioritizes verified information or reduces repeated exposure to antagonistic content, amplification should not raise recurrence, and a finding that it does would count against the claim that the ranking objective and not reach is the operative variable. Third, environments in which accusations are permanently retrievable should show a longer persistence of accusation-driven episodes, corresponding to a larger ρ , than comparable environments in which they expire or are revisable, and equal persistence in the two would count against the persistence channel. Fourth, when detection tools improve, measured prevalence should rise before any independent measure of conduct does, and where the independent measure rises as well the improvement

in detection cannot explain the increase. These predictions are conditional on the stated ranking objectives and design features, and they are intended as inputs to the case analysis of Section 9 and to later empirical work, not as conclusions established here. Each prediction must also specify whether its outcome is the frequency of accusations, the frequency of harmful conduct, or the persistence of the mechanism, since these are different outcomes and the measurement argument depends on keeping them separate.

9 Illustrative Case Analysis: Procurement, Concentration, and Dependency in Defense Technology

The preceding section treated communication environments as general selection conditions. This section turns to a different domain, in which the framework is applied to a further mechanism. The widespread claim that a small number of technology billionaires are taking over the world is a perpetrator-centered explanation, and it is therefore the most demanding test of the framework. It is also too broad to serve as a single empirical test, because it spans corporate governance, political influence, defense procurement, surveillance, platform competition, and ideological coordination, and different causal accounts may be correct in different domains. This section selects one mechanism, the relationship between defense procurement, technological concentration, and institutional dependency, states what each of two competing readings predicts about it, and specifies what each class of available evidence can and cannot show. The section is an illustrative case analysis and empirical research design and not a completed test of the model, because the primary procurement records needed to estimate its parameters have not yet been assembled.

The section addresses two questions that should be kept apart. The first concerns the origins and persistence of a harmful arrangement, and it is the contest between coordinated intent and convergent incentives. The second concerns the unintended consequences of attempts to correct the arrangement, and it is the question to which the formal model principally speaks. Much of the section examines the first question, because it determines what a correction would have to address, and the research design is organized around the second, as a design for testing whether specified corrective interventions raise the persistence of dependency.

9.1 Two Readings and Three Classes of Evidence

The first reading is coordinated intent, on which a small group deliberately steers outcomes. The second is convergent incentives, on which many actors facing the same selection pres-

asures, procurement rules, and sources of capital produce similar outcomes without a plan. The readings are not exclusive. Powerful actors can coordinate deliberately inside an environment that independently rewards the same conduct, and the empirical task is to estimate the relative contribution of each and not to choose between them. The two readings also do not exhaust the possibilities. Institutional inertia, technical compatibility requirements, network effects, accumulated expertise, and political constraints can contribute to persistence without fitting either, and the category of structural incentives must be specified through particular procurement rules and dependencies, since it would otherwise explain nearly everything.

Three classes of evidence bear on the question, and they discriminate between the readings to different degrees. The first class is public statements of belief and strategy, which show what actors say they aim to do. The second is documented coordination, such as communications, joint planning, shared financing, or concerted action recorded in primary documents. The third is structural incentives, meaning the rules, contracts, funding flows, and switching costs that reward particular behavior whatever the actors believe. Statements of belief are the weakest discriminator, because actors under convergent incentives also hold and voice beliefs, which may rationalize the conduct the incentives reward. Documented coordination can support the intent reading only to the extent that it is shown to alter outcomes beyond what the incentives already produce. Structural incentives can support the convergent reading only to the extent that similar conduct is shown among actors with no documented contact and with differing stated aims.

The organizing counterfactual of the section is whether the procurement arrangements, financial incentives, technological dependencies, and competitive pressures would remain if a particular executive disappeared. Two propositions govern its interpretation. The persistence of a harmful arrangement after changes in its personnel is consistent with structural reproduction, but it does not establish that individual intentions were causally irrelevant, since intentions may have created an arrangement that then persists without them. Conversely, documented coordination establishes intentional participation without explaining why similar arrangements recur across different participants. Both can hold at once, and the case analysis therefore asks what share of the observed persistence each accounts for.

9.2 What Thiel's 2009 Essay Establishes

The primary document used here is Thiel's essay in *Cato Unbound* [21], which states a position in his own words. The essay says that he no longer believes freedom and democracy are compatible, and it argues that libertarians should seek escape from politics through

new technological spaces, naming cyberspace, outer space, and seasteading as frontiers. It describes a deadly race between politics and technology and suggests that the outcome may depend on the effort of a single person who builds the machinery of freedom. It attributes part of the difficulty to the expansion of the franchise and of welfare beneficiaries. In remarks reproduced beneath the essay from a later response in the same forum [22], Thiel states that it would be absurd to suggest that women’s votes would be taken away, that he does not think any class of people should be disenfranchised, and that he has little hope that voting will make things better.

These statements establish what one prominent investor publicly argued at a particular time. They are evidence of the first class and nothing more. The essay does not address procurement, does not describe coordination among technology executives, and does not show that any outcome followed from its argument. A stated strategy of exit from democratic politics is compatible with both readings, since it may describe a coordinated project or may be one among many beliefs held by actors whose conduct is driven by their incentives. The essay is therefore useful for stating the intent reading in its strongest documented form and for fixing what its proponents would need to show, and it cannot by itself decide between the readings.

9.3 Operationalizing the Mechanism

The mechanism is mapped onto the model of Section 7. Harm H_t is a measure of avoidable institutional dependency in a defined segment of defense technology, for example the cost and operational disruption of replacing a supplier or the number of qualified alternative bidders, with concentration treated as one potential predictor and not as harm itself, since concentration can arise from scale economies, security requirements, or technical specialization. Correction C_t is the intensity of a corrective intervention, such as a competition reform, an interoperability mandate, or a new compliance regime. The proxy for reproductive capacity is the persistence of dependency after turnover of firm leadership and of individual firms, as defined in Section 2. Data come from contract award records, corporate disclosures, and documented communications, which are to be assembled before any claim is made.

Three questions structure the test, and Table 4 states the prediction of each reading. The first asks whether the pattern persists when leadership changes. The second asks whether contracts and procurement rules generate convergent behavior independent of ideological agreement. The third asks whether proposed remedies reduce concentration or unintentionally reinforce incumbent advantages.

Question	If coordinated intent dominates	If convergent incentives dominate
Does the pattern persist when leadership changes?	The pattern weakens or changes direction when key individuals leave	The pattern persists under successors facing the same rules and sources of capital
Do procurement rules generate convergent behavior independent of joint planning agreement?	Conduct aligns beyond what the rules require, with documented joint planning	Similar conduct appears among actors with no documented contact and differing stated aims, tracking changes in the rules
Do remedies reduce concentration or reinforce incumbents?	Remedies are evaded or neutralized through concerted response	Compliance costs and switching costs raise barriers to entry, and concentration persists or rises whatever anyone plans

Table 4: Predictions of the two readings for three questions about defense procurement and technological concentration.

9.4 A Design for the Corrective-Intervention Question

The formal model concerns the unintended consequences of corrective intervention, so the design that bears on it follows a corrective intervention and asks whether it raised the persistence of dependency. The causal sequence to be tested is procurement reform, then higher entry or switching costs, then greater incumbent persistence. The sequence is plausible, since a regulation intended to increase competition can impose compliance costs that fall more heavily on smaller firms, but it has not been demonstrated in this domain. The design would select corrective interventions adopted at different times across comparable segments, such as a competition reform, an interoperability mandate, or a compliance regime, and compare segments before and after adoption with segments not yet subject to the intervention. The outcome measures are indicators of avoidable dependency, with the persistence of dependency after supplier or leadership turnover as the proxy for capacity. Because reforms often follow scandal, adoption is endogenous, and the design must use variation in timing that is plausibly unrelated to current dependency or model the response rule, as in the closed-loop extension of Section 7.

9.5 Relative Contributions and the Limits of the Test

Because the readings are not exclusive, the aim is to estimate how much of the observed persistence is attributable to each. Cases in which segments with similar procurement rules differ in documented coordination offer the clearest comparisons, since they separate the contribution of coordination from that of the rules. A finding that persistence is high in segments with no documented coordination would support a large structural component, while a finding that persistence drops sharply when documented coordinators exit would

support a large intent component. The expected outcome is a mixed one, and the section commits to reporting it as mixed if the evidence so indicates.

The limitations should be stated. Documented coordination is observed only where records exist, so its absence from the data does not show its absence from practice, and the test therefore cannot establish a negative about intent. Corrective intensity in this setting also probably responds to harm, since procurement reform tends to follow scandal, so the case is closer to the closed-loop regime than assumption A5 allows and its estimates are correspondingly less clean. In addition, no world without defense procurement is observed, so a clean no-intervention baseline cannot be specified, and in settings without one the test degrades from asking whether overshoot above baseline occurs to asking whether the capacity-mediated component is positive, which is a weaker claim than the headline of the model. The test is also restricted to a single mechanism in a single domain, and its conclusions do not extend to corporate governance, platform competition, or political influence without separate analysis.

9.6 Two Book-Length Accounts as Competing Emphases

Two 2026 books address the same field with different emphases, and they serve here both as secondary sources and as examples of explanatory method. Klein and Taylor’s *End Times Fascism* [11] organizes its argument around apocalyptic and messianic ideas, the unequal distribution of protection against anticipated catastrophe, and the political uses of scapegoating, and it also draws on structural analysis, citing an account of an “authoritarian stack” that embeds itself through personnel pipelines, procurement loops, and technical architectures. Durán’s *The Nerd Reich* [5] gives greater weight to identifiable figures, their networks, and their ideological projects, and several of its chapters are organized around themes such as exit, apocalypse, and the Antichrist. The question for the framework is not whether either author is right about the dangers identified. It is whether the explanation identifies the conditions that reproduce concentrated technological power or mainly the actors who currently exercise it. On the portions examined, both books do both, and neither is a pure case of perpetrator-centered explanation. The first attends explicitly to scapegoating and to material conditions even as its chapters foreground figures and ideologies, and the second documents contractual and personnel relationships that can be read structurally even though its organizing frame is the network and its ideology. Klein and Taylor’s own description of the object of study is a hybrid. They describe an alliance of very different groups, each moving toward a different but related apocalyptic vision, that is fractious and divided on many questions yet united by an increasingly open disdain for democracy and by

justifications that differ from group to group. That description contains coordination, in the form of an alliance, and convergence, in the form of different justifications arriving at similar conduct, and so it already poses the question the case analysis asks. The comparison does not assess whether the authors' political judgments are correct. An explanation centered on culpable individuals may be accurate about particular decisions while incomplete about institutional reproduction, and a structural explanation can become evasive if it discounts documented coordination, misconduct, or personal responsibility, and the section tests for both failures.

To judge how far the books depart from a perpetrator-centered account, a benchmark is needed. An ideal-typical perpetrator-centered account explains outcomes by the biography and intentions of identifiable individuals, treats intent as a sufficient explanation, and offers no institutional analysis, so that removing the individuals would be expected to end the pattern. On the portions examined, both books depart from this type in identifiable ways. Klein and Taylor describe an alliance of groups with different aims and justifications, cite an analysis of procurement loops and technical architectures, and treat scapegoating as a tool of the movements they study. Durán documents contracts and personnel relationships, including Palantir's government contracting under the preceding administration, and attributes outcomes partly to networks and institutions as well as to named individuals, although his frame gives more weight to individuals than the first book does. Neither is a pure structural account either, since both attribute substantial causal weight to ideology and intent. The verdict that each does both is therefore a claim about degree, and the test of this section is intended to estimate that degree empirically.

The second book also reports items that bear on the three questions of Table 4, citing press accounts for each. It reports that Palantir had secured about \$2.5 billion in government contracts since 2009, including \$531 million in 2024 under the preceding administration, so that the relationship predates the present one. It reports that in mid-2025 the Pentagon awarded contracts, described as worth up to \$200 million each in most cases, to several artificial intelligence firms. That the firms' stated aims differ is the essay's own characterization and is not reported by the book. It reports that the Army announced plans to consolidate seventy-five technology contracts, worth as much as \$10 billion, into a single contract with Palantir, which was already prime contractor on fifteen of them and a subcontractor on the other sixty. The Army's own announcement of July 31, 2025 [23] confirms that it awarded Palantir an enterprise agreement of up to ten years that merges seventy-five contracts, fifteen prime and sixty related, into one, with a ceiling of \$10 billion that the Army describes as the maximum potential value of the contract and not any specific obligation or commitment. And it reports that executives from three firms were inducted into an Army Reserve

detachment in June 2025, and that a news report counted sixteen men connected to one investor and his companies in the administration. The first book cites an analyst's claim that each new contract deepens a dependency that becomes irreversible once a supplier is indispensable. Both books rely on the same analysis of the authoritarian stack, so they are not independent confirmations of it. The analyst's authoritarian stack, built through personnel pipelines, procurement loops, and technical architectures, can be read as a candidate realization of the institutional channel γ_I of Section 7 under the environment-dependent form $\gamma_I(\theta)$ of Section 8, with procurement rules and switching costs among the parameters in θ , but that reading is a hypothesis to be estimated and not a result.

These items do not decide between the readings, but they show what the test would measure. Continuity of a firm's government business across administrations fits the structural reading on the question of leadership change, though it concerns one firm and total contract value and not concentration relative to a counterfactual. Awards to several firms fit convergent behavior if their stated aims do differ, which is to be documented, but the awards share a single buyer, and a common buyer is itself the kind of structural condition the reading invokes. The consolidation of seventy-five contracts into one enterprise agreement is a candidate instance of a procurement decision that raises concentration, and it may be the clearest example of the dependency mechanism. The Army's announcement establishes that the agreement was awarded, but its ceiling is the maximum potential value and not obligated or expended funds, and the minimum guarantee, the terms, and the realized value are not established from the sources examined, so they must be before the consolidation is treated as an instance. Concentration of awards is also distinct from demonstrated institutional dependency, which requires evidence of switching costs and reduced contestability. Administrative convenience is only one possible explanation, however, and the contracting rationale and the effects on concentration and switching costs must be established independently. Personnel placements and reserve induction show channels of access and are evidence of the second class, but access is not shown to have produced the outcomes. The claim of irreversibility is itself a testable statement about reproductive capacity, since it predicts that dependency should persist after changes of government and of leadership, and the proxy of Section 2 can be applied to it directly. All of these items are reported at second hand, and the underlying award notices and contract documents are to be obtained before any of them is relied on. The final analysis must also distinguish contract ceilings, obligations, actual expenditures, and cumulative award values, since a contract worth up to a stated amount does not establish that the amount has been spent, nor does a consolidated contract imply the same amount of new procurement.

9.7 A Structured Test of the Consolidation

The consolidation of seventy-five contracts into one enterprise agreement is the case’s best candidate for an instance of the institutional channel, and the admissibility conditions of Section 6 require a structured test of it. The competing explanations are administrative convenience, in which the buyer consolidates to reduce transaction costs and contract overhead, and dependency raising, in which consolidation raises switching costs and reduces contestability so that incumbent persistence rises. The mechanism is specified as the procurement rules and incentives that reward consolidation together with the classification of a supplier as indispensable or enterprise-standard. The measurable outcome is avoidable dependency, assessed by the number of qualified bidders for follow-on work, the cost and time of replacing the supplier, the terms on data portability and interface openness, and lifecycle cost. The counterfactual is an award structured to several prime contractors or with enforceable portability terms. The documents that would discriminate are the acquisition strategy and market research, the justification for the chosen contract structure, the contract’s terms on data rights, exit, and interoperability, and later competition records. Under administrative convenience one expects documented transaction-cost savings, preserved portability and exit terms, and continued competition for follow-on work. Under dependency raising one expects weakened exit terms, a falling number of qualified bidders after consolidation, and rising switching costs. Preserved contestability together with documented savings would count against the dependency reading. The Army’s stated rationale is cost: the removal of reseller pass-through fees, volume-based discounts, and management of a single contract in place of many. An Army contracting official also told a trade outlet that the Army is considering ten to fifteen further enterprise agreements for other commercial software [25], which, if documented, would indicate a procurement model and not an arrangement specific to one supplier, although a model of this kind can itself raise switching costs for each supplier it covers.

9.8 A First Look at the Federal Award Data

The federal award data are the public records on USAspending.gov [24], queried through its programming interface on October 31, 2026 for contract awards to three Palantir entities, Palantir Technologies Inc., Palantir USA Inc., and Palantir USG Inc. The figures are obligations by fiscal year, running from October to September, including modifications to existing awards, and they are not disbursements. Subcontracts are not captured. Fiscal 2026 and, to a lesser degree, fiscal 2025 may be incomplete or revised because of reporting lags. Cumulative obligations through fiscal 2024 were about \$2.58 billion, consistent with

the \$2.5 billion relayed by Durán, and fiscal 2024 obligations were \$541 million against his \$531 million, a difference that plausibly reflects calendar-year and fiscal-year definitions.

Fiscal year	Obligations (\$M)	Defense share	Fiscal year	Obligations (\$M)	Defense share
2014	75	38%	2021	229	33%
2015	95	53%	2022	377	51%
2016	132	65%	2023	348	44%
2017	132	56%	2024	541	58%
2018	117	62%	2025	1,021	70%
2019	163	52%	2026	1,992	67%
2020	269	51%			

Table 5: Annual contract obligations to three Palantir entities, with the share awarded by the Department of Defense. Fiscal 2026 may be incomplete or revised.

Obligations rose from about \$75 million in fiscal 2014 to \$541 million in fiscal 2024, so the growth predates the present administration. Assigning months to administrations approximately, with January assigned to the outgoing one, annual averages were about \$52 million under the Obama administration, \$179 million under the first Trump administration, and \$419 million under the Biden administration, and about \$1.7 billion annualized over the twenty months of the second Trump administration to date, a figure that is sensitive to the observation window and to the timing of large awards. That series establishes continuity and growth across administrations, which is consistent with both repeat purchasing and lock-in and cannot distinguish them. The defense share rose from about half in fiscal 2019 to about two thirds in fiscal 2025 and 2026.

Among the 120 largest awards, which account for about 89 percent of the obligations in the set, 50 percent of obligations by value carry a not-competed, only-one-source coding in the latest record, and 46 percent carry full and open competition, mostly orders under multiple-award vehicles with fair opportunity. Counting awards coded not competed under simplified acquisition procedures as well, the share is 53 percent by value, and 34 percent by count, so the value share reflects a few very large awards. By the era in which the award began, the not-competed share on that basis was about 53 percent for the Obama era, 15 percent for the first Trump administration, 55 percent for the Biden administration, and 60 percent in the second Trump administration to date. The last figure falls to 9 percent when the fifteen orders under the enterprise agreement are excluded, because they inherit its single-award, one-offer coding, and the share for the 120 largest awards falls from 53 to 39 percent by value, so the high recent figure is largely a feature of that one vehicle and not independent evidence. The two figures answer different questions, since the first describes the composition of recorded obligations and the second the portfolio that remains once the dominant vehicle is removed, so the comparison is a sensitivity analysis

and not a replacement estimate. Fifteen orders inheriting the coding of one agreement reflect one structural decision, not fifteen independent instances of noncompetitive selection, even though each order is a genuine financial commitment. Competition for the original vehicle, competition for a particular order, and the practical availability of substitutes are three different things, and the coding captures the first two imperfectly and the third not at all. A sole-source coding for proprietary commercial software is common and is not by itself evidence of dependency, the justification documents were not examined, and the coding comes from the latest transaction record, which can differ from the original.

The enterprise agreement, contract W519TC25D0039, is recorded as a single-award indefinite-delivery contract signed on July 31, 2025, with a ceiling of \$10 billion and no obligations on the vehicle itself, coded not competed, with only one source and one offer. By October 31, 2026 it had forty orders totaling \$1.23 billion in obligations, about 12 percent of the ceiling, of which \$55 million fell in fiscal 2025 and \$1,172 million in fiscal 2026, and reported outlays of \$11.5 million, a figure drawn from a different reporting stream with its own timing and coverage, so that the two should not be read as delayed execution. One order, a \$638 million obligation for a Defense Department Maven Smart System task order, accounts for 52 percent of the obligations under the agreement, and the three largest orders for 73 percent, so the forty orders are not forty independent instances of demand. The agreement carried about 59 percent of Palantir's fiscal 2026 obligations, and the remaining \$820 million outside it still exceeded the whole of fiscal 2024, so the jump in fiscal 2026 is not only a relabeling of existing purchases through a new vehicle. One order from a civilian agency appears among the orders, which should be checked against the Army's description of use by defense agencies.

The data do not measure switching costs, proprietary integration requirements, interoperability, lifecycle cost, or the practical availability of alternatives, and they cannot distinguish supplier lock-in from ordinary long-term demand. Annual obligations also do not align with administrations, because fiscal years end on September 30 and administrations change on January 20. Four propositions that are often conflated are supported to different degrees. The continuity of Palantir's relationship with the government across administrations is established. The consolidation of its purchasing mechanisms into one agreement is established. The concentration of financial commitments is supported, since three orders carry 73 percent of the obligations under the agreement, although concentration of suppliers across the market is not measured and would require data on competing suppliers, market boundaries, and substitution. Institutional dependency arising from restricted alternatives is not established and cannot be inferred from the first three. The evidence is mixed against the dependency criteria. The single-award, sole-source coding of the agreement is compatible

with dependency, but it is equally compatible with intellectual-property exclusivity, security requirements, specialized demand, and incumbent technical advantages, and compatibility is weaker than discriminating evidence. Obligations of about 12 percent of the ceiling after about fifteen months, the substantial share of obligations placed through multiple-award fair-opportunity vehicles, and the volume discounts the Army describes, which have not been verified, point the other way. The strongest remaining test is whether the agreement permits an agency to recover its data, migrate operational workflows, preserve interoperability, and replace the supplier without prohibitive cost. The sole-source justification may explain the procurement decision, but the exit and data-rights provisions bear more directly on whether dependency persists. Those terms are not in the data and mark the evidentiary boundary of this draft.

9.9 Crisis as an Economic Resource

Klein and Taylor argue that powerful corporate players have developed a symbiotic relationship with catastrophe, one that no longer only profits from short-lived shocks but invests in a continuing state of emergency, and they present Palantir as emblematic. In discussing a company executive's hostile public rhetoric, they note that it is partly ideological and partly financially rewarding, since prolonged conflict benefits a supplier of tools for managing it. In the model of Section 7 this is a candidate mechanism for the institutional channel. If the standing of the institutions positioned to manage a harm grows with its intensity or persistence, then worsening conditions strengthen their position and they have a reduced incentive to extinguish them.

Three distinctions keep the claim testable. First, benefiting from a crisis is not causing it, and evidence that a firm profits from emergency does not show that any actor raised the crisis's capacity to recur. Second, the mechanism requires that revenue respond to the intensity of emergency, which can be measured as the elasticity of contract value with respect to indicators of emergency or conflict, controlling for general budget growth. Third, the model asks whether a specific intervention, such as a procurement reform or an enforcement action, raises capacity through this channel, which requires comparing periods or jurisdictions with and without the intervention. If the mechanism operates, supplier revenue should rise with emergency declarations and persist after the emergencies lapse, a pattern that is distinguishable from demand that simply follows need, and the absence of such an elasticity would count against the mechanism. The discriminating observation is whether the elasticity persists after emergency indicators lapse, relative to a comparison group of suppliers whose business does not depend on the emergency and to a benchmark in which demand simply

follows measured need. A positive elasticity alone is compatible with ordinary demand following need. The mechanism is supported if revenue remains elevated, or falls substantially less than the indicators, after the emergency lapses, and it is not supported if revenue tracks the indicators down. When a threat is represented as existential or irreversible, the urgency of preventing it may itself become a justification for exceptional authority, creating a further pathway through which the proposed remedy strengthens the institutional dependencies it was intended to overcome.

9.10 The Books' Own Classifications and Remedies

The framework applies to the books' vocabulary as it applies to any other. Terms such as end times fascism and nerd reich are not merely descriptive labels, since they organize heterogeneous events and actors into recognizable political categories. By the criteria of Section 3, such a category improves understanding if it predicts conduct better than the antecedent conditions it replaces, such as procurement rules and sources of capital, and it becomes a stabilizing component if those labeled adopt it as an identity or institutions organize around it independent of the harm. Klein and Taylor's insistence that the alliance is fractious and composed of groups with different aims supports discrimination within the category, though whether the category as a whole predicts conduct better than the conditions is an empirical question that this section does not settle. The rules of Section 6 also apply: a claim that a practice raises the capacity of a harm is admissible, while a claim that a kind of person is the source of the harm is not.

The same holds for the remedies the books propose. Klein and Taylor offer what they call Hereness, a politics of neighborliness and attachment to place that they present as an alternative to exclusionary survivalism and illustrate with movements against data centers and a municipal electoral campaign. Under the repair criteria of Section 11, the question is whether such arrangements reduce the conditions that generate antagonism, working through the extinguishing channel by changing siting rules, public procurement, or the dependencies described above, or whether they mainly create another coalition opposed to an existing one, working through the classification channel by hardening an opposing identity. This is an empirical question and not a presumed contradiction. The relevant measures are whether concentration and dependency fall after turnover, as measured independently of the movements concerned, whether the categories they use are revisable and offer an exit, and whether their claims rest on evidence that parties who do not benefit can check.

9.11 Remedies Aimed at Conditions

The framework favors remedies aimed at the conditions that reward the pattern, such as competition and procurement rules, transparency duties, and interoperability and data portability requirements, over remedies aimed at the persons who currently occupy dominant positions. Such remedies are interventions and are subject to the model. A procurement reform that raises compliance costs can favor incumbents who can bear them, which is a form of the capture that Stigler described [19], and a transparency regime can be satisfied formally while leaving concentration unchanged. The case is therefore also a test of the repair criteria of Section 11: remedies qualify as nonreproductive only if concentration and dependency fall after turnover, as measured independently of the bodies administering the remedies. Accountability for individual conduct is preserved by the paired remedy principle of Section 10, so that structural explanation does not reduce a whole institutional configuration to the intentions of a few prominent people and does not excuse any of them.

9.12 Status of the Sources

Thiel’s essay has been read in full as a primary document. The two books were examined through their front matter, tables of contents, introductions, and the passages concerning scapegoating, procurement, and contracts, and their wider evidence has not been evaluated. Their claims are treated as claims requiring evaluation against records, and the contract and personnel items above are reported at second hand from the press accounts the books cite. One of the files examined appears to be a pre-publication proof, with placeholders in its front matter, so page references and wording are to be checked against the published edition. The Army’s announcement of its enterprise agreement with Palantir was checked directly, and Palantir’s federal contract obligations, the competition coding of its largest awards, and the orders under the enterprise agreement were taken from the federal award data. The other contract and personnel figures, further award data, corporate disclosures, and documented communications remain to be assembled, and no empirical conclusion is drawn in this draft. Commentary that explains the pattern by the character of the individuals involved is excluded as evidence, because it is an instance of the character-type explanation that the framework examines.

10 Accountability Without Recursive Scapegoating

Two commitments pull against each other. Structural explanation must not be allowed to erase responsibility for harmful acts, since an account that attributes every outcome to

incentives leaves no room for the judgment that a particular person or institution should have acted otherwise. Conversely, legitimate accountability must not require attributing an entire social pathology to a morally contaminated population, since that is the move the essay identifies as scapegoating. This section argues that the two commitments are compatible once the objects of responsibility are distinguished, and it sets out the features of an accountability system that does not reproduce the pattern it addresses.

10.1 Two Objects of Responsibility

Responsibility attaches to two different objects. The first is conduct: the particular acts of individuals or the coordinated acts of groups and institutions, which can be investigated, attributed, and sanctioned. The second is the conditions under which that conduct becomes likely, rewarded, or protected, which are the incentive structure I and the classification practice K of the mechanism defined in Section 2. Those who design, maintain, or benefit from such conditions bear a distinct responsibility for them, and that responsibility is not discharged by punishing the conduct it produces.

The distinction corresponds to the central finding of the model. A system that addresses only conduct is exposed to replaceability, because the conditions that rewarded the conduct remain and the mechanism is carried forward by successors, which is the failure the third test of Section 4 detects. A system that addresses only conditions erases culpability and removes the deterrent that individual accountability supplies. Neither alone is adequate, and an accountability system should be evaluated by whether it addresses both and by whether recurrence after turnover falls.

10.2 Features of a Nonrecursive Accountability System

A system that preserves accountability without assigning permanent blame has several features. It preserves evidence, so that attributions can be checked by parties who do not benefit from them, which satisfies the evidence standard of Section 4. It provides due process, so that those accused can contest the attribution and so that errors are identified and corrected, which raises the extinguishing channel γ_{ext} . It applies proportionality, so that the severity of a response tracks the demonstrated contribution of the party to the harm. It provides restitution where harm has occurred, which treats the injured as the object of the remedy and not the accused as the object of the proceeding. It includes institutional reform, which addresses the conditions that made the conduct likely. And it avoids converting culpability into a permanent identity assigned to a population, by attaching findings to conduct and to roles instead of to persons as kinds, and by providing a route by which those found responsi-

ble can leave the category through changed conduct, which limits the classification channel γ_K .

10.3 Structural Explanation Is Not Exculpation

Explaining why conduct was rewarded does not reduce the responsibility of the actor who engaged in it, and the existence of culpable actors does not make a remedy aimed at them sufficient. The two statements are consistent, and holding them together yields a practical rule that is called here the paired remedy principle. Where harmful conduct is recurrent or structurally rewarded, sanctions should be accompanied by examination of the conditions that reproduce it, and where institutional reforms respond to documented misconduct, the reforms should not substitute for independent assessment of responsibility for that conduct. The principle is proportional to evidence of mechanism-level recurrence and is not a universal procedural requirement, since many routine sanctions concern isolated conduct for which there is little reason to suspect a reproductive mechanism, and some structural reforms address conditions that do not involve culpable conduct. The pairing guards against the two characteristic failures, the scapegoating that stops at the accused and the exculpation that stops at the structure, and the two halves should be reported separately so that neither can be silently dropped.

There are cases in which the principal cause of a harm is the conduct of identifiable actors and the conditions are not responsible for its recurrence, and in these cases the third test of Section 4 is passed, since removal of the actors does change the outcome. Targeted sanction is then the appropriate instrument, and the framework does not recommend structural reform in its place. The aim is not to reduce accountability but to match the object of the response to the source of the harm.

11 Nonreproductive Repair: Design Criteria

The constructive question is what arrangements reduce the reproductive capacity of a harmful pattern while retaining the ability to respond to real harms. The model supplies the criterion. An arrangement is nonreproductive in the strong sense if its net effect on capacity is nonpositive, $g \leq 0$, and in the weaker long-run sense if $dg \leq b(1 - \rho)$, so that sustained use does not raise harm above baseline. Design principles can therefore be stated in terms of the coefficients: raise effective suppression b , lower the institutional and classification channels γ_I and γ_K , raise the extinguishing channel γ_{ext} , and lower the persistence ρ where persistence is not needed for accountability.

11.1 Principles by Channel

For the institutional channel, funding should be tied to resolution in place of caseload, and mandates should include sunset provisions whose renewal requires demonstrated outcomes, so that the institution gains from reducing the conditions that justify it. Detection should be separated from funding, so that the body that finds violations is not paid in proportion to the number it finds. For the classification channel, categories should be revisable and time-limited, should carry a route of exit tied to verified change in conduct, and should name practices and roles in preference to persons. For the extinguishing channel, remedies should aim at conditions, decisions should be reversible where harm from error would be severe, and claims should be subject to distributed verification and adversarial cooperation, meaning that independent parties with opposed priors check one another’s findings. Pluralistic deliberation belongs here as well, since it exposes the assumptions on which a classification rests.

Several of these principles need qualification. Automatic expiration can destroy institutional knowledge and undermine long-term enforcement, so the more general principle is periodic mandate review, in which continued existence must remain justified by independently assessed outcomes and risks. Outcome-based funding requires risk adjustment as well as independent detection, because an institution penalized for outcomes caused by external conditions may avoid difficult cases, exclude high-risk populations, or redefine the scope of its responsibility. Durable records and revisable classifications are different operations, since an institution can preserve an accurate historical record while changing the current status assigned to a person or practice, so that accountability and category exit can coexist without treating forgetting as the principal means of reconciliation.

Where records of accusation persist indefinitely, a bounded expiry or a right of revision can lower ρ , but only where the interest in accountability can be met by other means, since a record that is erased cannot support later correction of an institution’s behavior. Settings that do not reward performance for an audience, such as small, private, or one-to-one forms of review, can lower the cost of tentative speech and reduce the incentive to accuse for effect. As Section 8 argued, such settings can also conceal misconduct, and they should be paired with external scrutiny and not substituted for it.

11.2 Automated Allocation as a Remedy

A further difficulty arises when the proposed remedy is to replace discretionary human administration with automated allocation. Systems designed to distribute resources according to explicit rules may reduce some forms of favoritism, rent extraction, and arbitrary decision-making, but they do not eliminate the institutional conditions under which allocation occurs.

The selection of objectives, the definition of legitimate claims, the ownership of computational infrastructure, and the authority to revise decisions remain consequential exercises of power. An algorithmic system can therefore suppress particular abuses while reproducing the concentration of authority that made them possible, especially when those affected cannot inspect its criteria, contest its classifications, or alter its governing objectives. The relevant test is not whether allocation is performed by humans or machines, nor whether the system is described as public or private, but whether its operation reduces dependency and the reproductive capacity of the harmful mechanism. Automated governance is consequently subject to the same requirements of independent verification, contestability, and institutional revision as any other corrective arrangement. More generally, replacing the occupants of a position of power is not equivalent to changing the mechanism through which that power reproduces itself.

11.3 Alternative Framings of the Corrective Question

The choice of corrective arrangement depends on how the problem is framed, and the first framing to hand is not the only one. A problem of conflict between groups can be framed as one of detection and force, which invites surveillance and coercive responses, or as one of language, contact, and material conditions, which invites responses such as language and cultural exchange and the provision of infrastructure and public services. The essay does not claim that the second framing yields solutions. It claims only that the framings identify different mechanisms, that each is an intervention subject to the same model, and that each should be held to the same criterion, namely whether recurrence after turnover falls under independent measurement. Identification and monitoring systems extended before their error rates and exclusion effects are understood can be tested against that criterion as well, since a system that misclassifies or excludes some of those it covers may raise and not lower the capacity of the harm it addresses. The gentler framing is also conditional. The contact hypothesis holds that exchange between groups reduces prejudice under conditions such as equal status, shared goals, and institutional support [1, 14], and contact without those conditions can have the opposite effect.

11.4 Failure Modes and Safeguards

Each arrangement is an intervention and is subject to the same model, as Section 6 requires. Table 6 lists the arrangements discussed, the channel each addresses, the main way each can fail, and the safeguard that the framework suggests. The central lesson is that several of the most natural reforms contain a route to the failure they are meant to prevent. Outcome-

based funding, for example, rewards an institution for recorded reductions in recurrence, which gives it an incentive to detect less, so that the apparent success is an artifact of the detection responsiveness η . Sunset clauses invite the manufacture of outcomes to meet the renewal test, which is a Goodhart effect of the kind described in Section 5.

The reflexive consequence is that an institution created to detect recursive correction can itself acquire the incentive structure the theory describes. An auditor funded in proportion to the number of cases of recursion it identifies has an institutional incentive and a classification incentive of exactly this kind, so the theory must apply to the mechanisms used to enforce its own repair criteria. This does not entail an infinite regress of auditing institutions. It requires that evaluation have a stopping rule, independent outcome measures, and limits on institutional accumulation.

Arrangement	Channel addressed	ad-	Main failure mode	Safeguard
Outcome-based funding	Institutional down, γ_{ext} up)	$(\gamma_I$	Rewards concealment, since recorded recurrence falls when detection falls	Detection by a body independent of the funded institution
Sunset clauses with conditional renewal	Institutional		Outcomes manufactured to meet the renewal test	Preregistered measures with external audit
Revisable categories with exit	Classification (γ_K down)		Exit granted pro forma and loses meaning	Exit tied to verified change, with independent review
Distributed verification and adversarial cooperation	Extinguishing (γ_{ext} up)		Verifiers converge on shared interests or are captured	Opposed priors, rotation of verifiers, public methods
Reversible decisions and rights of reply	Extinguishing; persistence		Delays accountability for genuine harm	Time limits on review; interim measures that preserve evidence
Remedies paired with sanctions	Institutional and classification		Sanction omitted in favor of reform, erasing responsibility	Separate reporting on each half of the pairing

Table 6: Repair arrangements, the channels they address, their failure modes, and suggested safeguards.

11.5 A Criterion of Success

Three properties are distinguished, because none entails the others without further assumptions. An intervention is capacity-reducing if it lowers reproductive capacity relative to the counterfactual, capacity-neutral if it does not materially change capacity, and net beneficial if its total effect on the specified harm is negative over the evaluation horizon. An arrangement is nonreproductive in the strong sense if it is capacity-neutral or capacity-reducing,

$g \leq 0$, and in the weaker long-run sense if it is net beneficial with $dg \leq b(1 - \rho)$ even though capacity rises. The measurable criterion for the strong sense is that recurrence after turnover does not rise, and for a capacity-reducing arrangement that it falls, relative to comparison settings, as measured by a detection method independent of the arrangement itself, over a horizon and with thresholds fixed in advance. A fall in visible prevalence alone does not qualify, for the reasons given in Section 7, and a capacity-neutral arrangement that reduces harm through suppression qualifies as nonreproductive even though recurrence is unchanged. Nor is a continuing demand for correction by itself a failure, since some harms are generated by external conditions that no arrangement can remove, and the criterion concerns whether the arrangement raises later demand.

The criterion can fail. If recurrence rises, if harm reappears after the arrangement is withdrawn, or if a fall in measured prevalence disappears under independent detection, the arrangement is not shown to be nonreproductive, and if recurrence is unchanged it has not been shown to reduce capacity, though it may still be capacity-neutral. The essay does not claim that any of the arrangements discussed here will meet the standard in practice, and it offers them as hypotheses to be tested under the same discipline as the harms they address.

12 Conclusion

The essay began from the distinction between the visible perpetrators of a harm and the conditions that repeatedly produce it. Naming culpable parties is sometimes necessary and often just, but it does not explain why a harm survives changes in the people who carry it, and it can become part of the process that sustains the harm when classification, mobilization, and institutional investment form around the accused. The contribution is to state this possibility precisely. Separating the prevalence of a harm from the reproductive capacity of its mechanism, and locating the unit of reproduction in an incentive structure paired with a classification practice, distinguishes mechanism reproduction, in which correction raises capacity, from the narrower case of pathological reversal, in which that effect outweighs suppression.

The formal results are conditional. In the minimal model of Section 7, three thresholds govern a single intervention: overshoot at the second period when $dg > ab$, eventual overshoot under a weaker condition, and positive cumulative harm when $dg > b(1 - \rho)$, which also governs the long-run effect of a continuing policy. Reversal requires a positive pathway from correction to capacity, and a rebound alone is not evidence of recursion. Under a proportional feedback rule the same quantity governs the equilibrium level of harm, $h_0/(1 - \phi S)$, with a critical gain $1/S$ beyond which no bounded positive equilibrium exists

when $S > 0$, although overcorrection can destabilize beneficial regimes at high gain. Applied to institutions, communication environments, and a narrow case in concentrated technological power, the framework yields predictions and indicators in place of verdicts, and it locates the decisive question in whether recurrence after turnover changes because of correction.

The claims in the essay have different evidentiary status, and the conclusion preserves the difference. The mathematical propositions are established under their stated assumptions and say nothing about whether those assumptions hold in a given setting. The sociological mechanisms are theoretically motivated by an established literature but have not been estimated. The procurement application is an untested research design. Its evidence establishes an increasingly consequential contractual relationship, but not that the relationship is involuntary, irreversible, or institutionally pathological, which are separate propositions requiring separate tests, and most of the figures it uses are second-hand, with the Army's consolidation checked against an official announcement and Palantir's obligations and competition coding taken from federal award data. The repair criteria are proposals awaiting evaluation, and the essay does not claim that any of them works.

The essay holds itself to its own standard. The admissibility conditions of Section 6 apply to its categories and to the repairs it proposes, which are interventions subject to the same model. A given case would fail to count as an instance if the estimated effect of correction on capacity were nonpositive, if the long-run condition failed with an interval that excluded the boundary, if no overshoot appeared over a horizon in which the parameters imply it should have, if an apparent reversal disappeared under independent detection, or if the harm proved to have been displaced to another domain. A proposed mechanism would be rejected if its channel coefficients or mediating pathway failed across the cases in which it was proposed, and the program as a whole would be weakened only if carefully selected cases that independently satisfy its causal conditions repeatedly failed to exhibit the predicted outcomes. Several tasks remain. The parameters must be identified and not merely fitted, which requires a measurement model for the capacity proxy, a justification or relaxation of the assumption that intervention is independent of current harm, and the contract records that the case analysis has not yet assembled. The central unresolved question is how reproductive capacity can be measured, and the framework should be tested first where recurrence after turnover can be observed over long horizons and where institutions have been wound down or restructured.

References

- [1] Allport, G. W. (1954). *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.

- [2] Baudrillard, J. (1994). *Simulacra and Simulation*. Translated by S. F. Glaser. Ann Arbor: University of Michigan Press. (Original work published 1981.)
- [3] Campbell, D. T. (1979). Assessing the Impact of Planned Social Change. *Evaluation and Program Planning*, 2(1), 67–90.
- [4] de Zwart, F. (2015). Unintended but Not Unanticipated Consequences. *Theory and Society*, 44(3), 283–297.
- [5] Durán, G. (2026). *The Nerd Reich: Silicon Valley Fascism and the War on Democracy*. New York: Avid Reader Press.
- [6] Forrester, J. W. (1969). *Urban Dynamics*. Cambridge, MA: MIT Press.
- [7] Girard, R. (1977). *Violence and the Sacred*. Translated by P. Gregory. Baltimore: Johns Hopkins University Press. (Original work published 1972.)
- [8] Girard, R. (1986). *The Scapegoat*. Translated by Y. Freccero. Baltimore: Johns Hopkins University Press. (Original work published 1982.)
- [9] Goodhart, C. A. E. (1984). Problems of Monetary Management: The U.K. Experience. In *Monetary Theory and Practice: The U.K. Experience*. London: Macmillan. (Originally presented in 1975 and published in *Papers in Monetary Economics*, vol. I, Sydney: Reserve Bank of Australia.)
- [10] Hacking, I. (1995). The Looping Effects of Human Kinds. In D. Sperber, D. Premack, and A. J. Premack (Eds.), *Causal Cognition: A Multidisciplinary Debate* (pp. 351–394). Oxford: Oxford University Press.
- [11] Klein, N., and Taylor, A. (2026). *End Times Fascism: And the Fight for the Living World*. New York: Farrar, Straus and Giroux.
- [12] Merton, R. K. (1936). The Unanticipated Consequences of Purposive Social Action. *American Sociological Review*, 1(6), 894–904.
- [13] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge: Cambridge University Press.
- [14] Pettigrew, T. F., and Tropp, L. R. (2006). A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology*, 90(5), 751–783.
- [15] Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

- [16] Riordan, D. V. (2021). The Scapegoat Mechanism in Human Evolution: An Analysis of René Girard’s Hypothesis on the Process of Hominization. *Biological Theory*, 16(4), 242–256.
- [17] Selznick, P. (1949). *TVA and the Grass Roots: A Study in the Sociology of Formal Organization*. Berkeley: University of California Press.
- [18] Senge, P. M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday.
- [19] Stigler, G. J. (1971). The Theory of Economic Regulation. *Bell Journal of Economics and Management Science*, 2(1), 3–21.
- [20] Strathern, M. (1997). “Improving Ratings”: Audit in the British University System. *European Review*, 5(3), 305–321.
- [21] Thiel, P. (2009a). The Education of a Libertarian. *Cato Unbound*, April 13, 2009. <https://www.cato-unbound.org/2009/04/13/peter-thiel/education-libertarian/>
- [22] Thiel, P. (2009b). Your Suffrage Isn’t in Danger. Your Other Rights Are. *Cato Unbound*, May 1, 2009. <https://www.cato-unbound.org/2009/05/01/peter-thiel/suffrage-isnt-danger-other-rights-are>
- [23] U.S. Army (2025). U.S. Army Awards Enterprise Service Agreement to Enhance Military Readiness and Drive Operational Efficiency. Army.mil, July 31, 2025. <https://www.army.mil/article/287506/>
- [24] USAspending.gov (2026). Federal contract award data for Palantir Technologies Inc., Palantir USA Inc., and Palantir USG Inc., queried October 31, 2026. U.S. Department of the Treasury. <https://www.usaspending.gov/>
- [25] Welch, C. (2025). Army Consolidates Dozens of Palantir Software Contracts into One Deal Worth Up to \$10 Billion. *Breaking Defense*, August 1, 2025. <https://breakingdefense.com/2025/08/army-consolidates-dozens-of-palantir-software-contracts-into-one-deal-worth-up-to-10-billion/>
- [26] Wilson, J. Q. (1989). *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books.