

Relational Trust Dynamics: Self-Reinforcement, Attractor Criteria, and Poetic Composition

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Prior research already documents reciprocity, spiraling cooperation, temporal development, self-reinforcing relations, and trust-attractor language. This paper therefore makes a narrower claim. It asks when trust-dependent interaction changes the conditions of later trust formation strongly enough to justify a dynamical description. The language of *trust attractors* is reserved for configurations showing more than persistence: relevant evidence would include return after perturbation, basin dependence, thresholds, multistability, hysteresis, or comparable signatures. Relational self-reinforcement denotes the broader mechanism class from which such dynamics may, but need not, arise.

The paper also develops the *poetic structure of trust*. Context, framing, affect, ritual, embodiment, and symbolic cues are already known to matter for trust. The proposed contribution is a compositional hypothesis: heterogeneous, indirect, and temporally dispersed elements can acquire trust-relevant significance through sequence, interaction, memory, embodiment, and relational history in ways not adequately represented by fixed context-independent cue weights. A small event can therefore gain large relational leverage without becoming sufficient evidence of trustworthiness.

These claims meet in a mechanism-explicit account of recursive relational formation. Trust can alter disclosure, cooperation, interpretation, monitoring, vulnerability, institutional assurance, and the evidence generated by later interaction. The same recursion can remain corrigible, support repair, or produce closure, dependence, and manufactured confidence. Distrust can develop parallel dynamics, while rupture can change the pathways through which later repair is interpreted.

The resulting governance problem is not trust maximization. It concerns the conditions under which reliance can emerge, be tested, contested, revised, repaired, refused, or exited, together with the distribution of vulnerability and power over trust-generating environments. The framework links dynamics, composition, governance, justice, and empirical identification while preserving the distinctions among experienced trust, trustworthiness, dynamical stability, and justice.

Keywords: trust; relational self-reinforcement; trust attractors; poetic composition; path dependence; vulnerability; trust governance; relational justice

Discussion Note

Purpose, Recursion, Scope

This paper develops one connected problem from three directions: trust can be relationally generated; established trust can change the conditions of later interaction; and heterogeneous events can acquire trust-relevant significance through composition within a relational history. The argument therefore asks not only whether trust is present, but how trust participates in producing the subsequent relation from which later trust, distrust, verification, or refusal will emerge.

The language of *trust attractors* is used cautiously and without a priority claim. Trust evolution, reciprocity, spirals, self-reinforcing relations, and explicit trust-attractor language already have substantial precedent in trust research. The stronger attractor claim concerns recurrent organization under perturbation: return, basin dependence, threshold behavior, multistability, hysteresis, or comparable signatures. Persistence alone is not sufficient evidence. Where those stronger signatures are absent, self-reinforcement or historical dependence is the better description.

A minimal recursive representation is

$$A_t = \mathcal{A}(R_t, T_t, \xi_t), \quad (1)$$

$$E_{t+1} = \mathcal{E}(R_t, A_t, \eta_t), \quad (2)$$

$$R_{t+1} = \mathcal{R}(R_t, A_t, E_{t+1}), \quad (3)$$

$$T_{t+1} = \mathcal{T}(R_{t+1}, H_{t+1}), \quad (4)$$

where R_t denotes the relevant relational configuration, T_t a local trust-related orientation, A_t action, E_{t+1} newly generated evidence or experience, and H_{t+1} accumulated history. The notation is illustrative. Its purpose is to make one recursive fact explicit: present trust can alter action; action can alter the relation; and the altered relation can change the evidence and interpretation available to later trust formation.

Composition, History, Meaning

The phrase *poetic structure* concerns composition rather than identity. Trust is not equated with poetry. The concept marks cases in which heterogeneous, partial, indirect, temporally dispersed, or non-propositional elements acquire significance through relation, sequence, recurrence, embodiment, memory, and context. A gesture, remembered act, silence, shared difficulty, institutional symbol, spatial setting, or rhythm of response may carry little force in isolation yet become consequential inside a particular history.

The empirical burden is correspondingly strict. Poetic-composition language is useful only where order, interaction, context, or retrospective integration adds explanatory value beyond simpler additive, incentive-based, or evidential accounts. The same restraint applies to attractor language: evocative concepts must restrict inquiry rather than merely redescribe events after the fact.

Reinforcement, Vulnerability, Governance

Trust can change disclosure, delegation, cooperation, interpretation, monitoring, and willingness to accept vulnerability. Those changes can generate new evidence and new dependencies. Reinforcement can therefore remain informative and revisable, or become epistemically closed through reduced monitoring, narrowing alternatives, interpretive shielding, or rising switching costs.

The paper preserves several separations throughout. Experienced trust need not track trustworthiness. Dynamical stability need not indicate justice. Cooperation can occur without trust, and trust can coexist with monitoring, disagreement, or bounded distrust. These distinctions prevent recursive stability from becoming self-validation.

They also change the governance question. Guarantees, procedures, interfaces, architecture, rituals, reputational systems, narratives, and shared experiences can alter the conditions under which reliance emerges. The relevant objective is therefore not trust maximization. It is the governance of conditions

under which trust can be tested, revised, refused, repaired, or exited, with attention to vulnerability, evidence access, contestability, and power over the trust-generating field.

Distrust, Repair, Corrigibility

Distrust receives independent treatment because it need not be a simple low value of trust. It can organize withdrawal, defensive monitoring, restricted disclosure, and hostile interpretation, thereby generating its own recursive dynamics. Rupture can also change the later meaning of assurances and repair attempts. A repaired relation consequently need not return to the configuration that existed before the rupture; the violation and the repair become part of the relation's history.

The framework is intentionally corrigible. Evidence for self-reinforcement should identify a feedback channel. Evidence for a trust attractor should go beyond persistence. Evidence for poetic composition should test sequence, context, and interaction. Governance analysis should separate experienced trust from protection, trustworthiness, vulnerability, voice, and exit. Where simpler models explain the observations with equal adequacy, the additional relational machinery should be narrowed or discarded.

Responsible Use and Rights Reservation

This section separates requested scholarly conduct from the legal permissions stated on the following page. It records an ethical request for responsible use and then defines the narrower scope of retained legal rights.

The author encourages good-faith discussion, criticism, independent inquiry, and responsible use of the material in this work. Separately from the licence's terms, the author asks users to consider foreseeable harms when adapting or applying the proposed framework. This ethical request leaves the licence's permissions and legally authorized uses unchanged.

The author retains the rights preserved under CC BY-NC 4.0 and may pursue remedies to which the author is legally entitled for breach of the licence or violation of the author's independently applicable rights. Reuse remains independent from authorial endorsement. Third-party rights require authorization from their respective holders where applicable. Copyright exceptions and limitations, including applicable forms of fair use or fair dealing, remain fully available.

Notices

This page consolidates the manuscript’s publication status, licence, development disclosure, research-programme relation, and suggested citation.

Status. This manuscript is a complete working-paper version of the argument. It is circulated for scholarly discussion and remains open to correction, empirical testing, specialist review, and later publication-stage revision. The formal constructions are conceptual and illustrative unless a stronger empirical status is stated explicitly.

Licence. Except where otherwise indicated, copyright 2026 Wanhong Huang. This work is made available under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). Subject to its terms, the licence permits sharing and adaptation for noncommercial purposes with appropriate attribution, a link to the licence, an indication of changes, and attribution that preserves the licensor’s independence from the reuse. Reuse is governed solely by that licence; the responsible-use request on the preceding page remains separate from its terms. The licence deed and legal-code link are available at <https://creativecommons.org/licenses/by-nc/4.0/>. The licence governs in case of conflict with this summary. Third-party material remains subject to the rights held by its respective rights holders.

Statement on the use of language models. The exploratory discussions and preparation of this paper involved OpenAI’s ChatGPT. ChatGPT supported exploratory dialogue, formal reconstruction, source discovery followed by website verification, exact calculation, argumentative criticism, and drafting in $\text{\LaTeX}$. The author selected the research questions, directed and approved the theoretical commitments and epistemic status of the claims, and bears sole responsibility for the manuscript, including its definitions, formal constructions, taxonomy, calculations, arguments, conclusions, and errors. Authorship credit remains with the human author. The author remains responsible for source selection, citation accuracy, and verification against the cited works.

Suggested citation. Huang, Wanhong. “Trust Attractors: Relational Self-Reinforcement and the Poetic Structure of Trust.” Working paper, 2026.

1 Trust, Relation, Recursion

This section establishes the problem that organizes the paper. Its purpose is not to replace existing theories of trust with a single relational definition. The literature already contains influential accounts centered on vulnerability, risk, expectation, affect, social complexity, institutional assurance, trustworthiness, and the development of working relationships [5, 34, 46, 50, 64, 29, 42]. The present inquiry begins from a narrower observation that cuts across several of these traditions: once trust has emerged, subsequent interaction often occurs under conditions partly produced by that prior trust. What is trusted at time t can affect what becomes disclosed, attempted, observed, remembered, and interpreted at time $t + 1$. The relation therefore does not merely carry a judgment about an independently fixed object. It can participate in generating the later conditions under which further trust, distrust, verification, or withdrawal become possible.

This recursive feature motivates the central problem of the paper. Trust may be both historically generated and historically generative. It can arise from a relational history and then modify that history's continuation. A trusted collaborator may receive information that a distrusted collaborator would never see. A trusted institution may receive voluntary compliance that generates records of successful coordination unavailable under universal refusal. A trusted friend may be given an opportunity to repair a mistake that would end a more fragile relation. Conversely, a history of distrust can restrict communication, increase defensive interpretation, and prevent experiences that might otherwise revise the relation. In each case, present trust affects the production of future relational evidence.

The paper uses the existing term *trust attractor* for part of this problem, but it does not claim priority for the term or for the general use of dynamical-systems language in trust research. Levick, Woog, and Knox used trust-and-goodwill attractor language in a complexity-informed inquiry, and Long and Sitkin later developed an explicit dynamical-systems approach under the title "Trust Attractors" [40, 45]. The present paper takes a narrower methodological position. It distinguishes self-reinforcement from attraction and reserves stronger attractor claims for cases in which a specified state representation, perturbation class, and observation scheme support recurrent or basin-like behavior. The term therefore does not presuppose that every trust relation possesses a low-dimensional phase space, that trust is a scalar state variable, or that persistence establishes a mathematical attractor.

A second problem enters at the level of composition. Trust is often affected by elements that are individually weak, indirect, heterogeneous, or temporally dispersed. A remembered act, a recurring gesture, a silence, a shared hardship, an institutional seal, a familiar room, a pattern of response, or an unexpected act under pressure may acquire significance only within a particular relational history. The paper uses *poetic structure* for this compositional problem. The phrase does not identify trust with poetry. It points to the possibility that relational meaning depends on arrangement, sequence, recurrence, omission, resonance, and context, so that the force of an element cannot always be inferred from its isolated informational content. This compositional dimension becomes especially important once trust is studied through time, because the same event can have different consequences in different relational histories.

The third problem is normative. Conditions that make trust easier to generate can also be designed. Guarantees, procedures, rituals, interfaces, architecture, reputational systems, narratives, social proof, and controlled experiences can alter whether a participant feels able to rely on another actor or institution. Such interventions can support valuable cooperation, especially where strangers must coordinate under uncertainty. They can also generate confidence that outpaces the trustworthiness of the underlying relation. A theory concerned only with increasing trust therefore risks confusing a produced relational state with the conditions that would make reliance warranted. The paper treats this divergence as a central problem for trust governance.

1.1 Vulnerability, Uncertainty, Exposure

Many influential approaches converge on the point that trust matters where an actor cannot fully secure an outcome in advance. Baier's philosophical account places vulnerability near the center of the phenomenon, drawing attention to the moral significance of accepted dependence and the possibility of betrayal [5].

Mayer, Davis, and Schoorman likewise distinguish trust from related constructs by connecting it to a willingness to accept vulnerability under conditions in which another party's actions matter [50]. Rousseau and colleagues' cross-disciplinary synthesis similarly emphasizes positive expectations together with the willingness to accept vulnerability under risk and interdependence [64]. These formulations differ in theoretical purpose, but they share a useful boundary condition: where everything can be guaranteed, verified, and controlled without remainder, the distinctive work performed by trust becomes harder to identify.

The present paper retains that boundary while changing the object of attention. Instead of asking only why an actor accepts vulnerability at a particular moment, it asks what accepting vulnerability does to the subsequent relation. The act may expose the trusting party to loss, but it may also create a new interaction, disclose new information, produce an opportunity for reciprocity, and alter how future ambiguity is interpreted. Vulnerability is therefore not only a condition under which trust is expressed. In some settings it is also a mechanism through which relational history becomes newly generative.

This point should not be romanticized. Exposure can produce learning, mutuality, and cooperation, but it can also produce exploitation. A person who discloses more because of trust may enable deeper coordination or may hand an adversarial actor precisely the information needed for manipulation. A person who grants a second chance may enable repair or may extend a harmful relation. The same structural opening can therefore support sharply different trajectories. The analysis cannot infer normative value from relational intensity or persistence alone.

Claim 1.1 (Recursive exposure). Where trust changes the actions available to participants, the consequences of those actions can modify the evidence, expectations, dependencies, and interpretive conditions relevant to later trust. The trust relation can therefore become one causal component of its own subsequent development.

The claim is intentionally weaker than a universal law of positive feedback. Trust does not invariably generate more trust. An act enabled by trust can produce disconfirmation, betrayal, disappointment, or a demand for stronger verification. The important point is recursive participation: prior trust can change the pathway through which later trust is updated.

1.2 Judgment, Relation, Duration

A static representation can be useful for many purposes. One may ask whether a trustor believes a trustee to be reliable, whether available evidence supports that belief, whether the trustee possesses relevant abilities and motives, or whether institutional safeguards reduce expected risk. Yet repeated relations create a further question. The object assessed at a later time may no longer be relationally identical to the object encountered at the beginning, because the history of interaction has altered expectations, commitments, information, roles, and available actions.

Luhmann's account famously connects trust to problems of social complexity and the need to proceed despite futures that cannot be exhaustively known [46]. Hardin develops a different line, emphasizing encapsulated interest and the relational grounds on which one party may have reason to take another's interests into account [29]. Organizational research likewise distinguishes antecedents, forms, levels, and developmental processes of trust rather than treating it as a context-free quantity [50, 42, 52]. These literatures provide different answers to what trust consists in, but they make clear that trust is embedded in conditions of uncertainty, interdependence, and history.

The present framework adds a recursive emphasis. Suppose two relations yield the same currently observable cooperative act. One arrived there after years of reliable interaction and successful repair; the other arose yesterday under a strong external guarantee. The present act may look similar, yet the two relations need not react similarly to a missed obligation tomorrow. Their histories can alter both interpretation and recovery. A single snapshot can therefore underdescribe the relational state relevant to future behavior.

This is one reason the paper does not begin by defining a universal scalar $T \in [0, 1]$ and interpreting every change in trust through movement along one axis. A single number may be useful within a particular

empirical model, but the conceptual framework must initially remain broad enough to represent domain specificity, heterogeneous mechanisms, and historical dependence. A person may rely on another's technical competence while withholding confidence in their motives. An institution may be trusted to preserve records but not to interpret them impartially. A relation may sustain practical cooperation while retaining substantial affective suspicion. Trust and distrust can therefore coexist in different domains rather than functioning only as opposite points on one line.

Remark 1.2 (History belongs to the object of analysis). If earlier interactions change later interpretation, available evidence, or relational options, history cannot be treated solely as background description. At least part of that history belongs to the state needed to explain subsequent relational evolution.

This remark becomes important when the attractor vocabulary is introduced more formally later in the paper. A dynamical representation that ignores the history responsible for hysteresis, repair, or cumulative suspicion can create a false impression of Markovian simplicity. One possible response is to enlarge the state representation; another is to use explicitly history-dependent operators. The correct choice is an empirical and formal question rather than a matter to settle by terminology.

1.3 Recursion, Feedback, Reproduction

Existing theories already contain resources for studying reciprocity, relationship development, institutional support, affect, cognition, and the management of vulnerability. The contribution proposed here should therefore not be stated through an exaggerated claim that relational or temporal aspects have been absent from trust research. They have not. The narrower gap concerns how to organize several familiar observations around one recursive question: *how does generated trust alter the conditions under which later trust is generated?*

This question has substantial precedent. Jones and George theorize the evolution of organizational trust through interaction among values, attitudes, moods, and emotions; Serva, Fuller, and Mayer identify reciprocal longitudinal relations among trust, perceived trustworthiness, and risk taking; Ferrin, Bligh, and Kohles model spirals of cooperation and perceived trustworthiness; Huang and Wilkinson develop a complex-adaptive account of business relationships; Vanneste, Puranam, and Kretschmer show that relationship duration has only a small average association with trust and that multiple countervailing mechanisms matter; and Møllering and Sydow explicitly analyze self-reinforcing inter-organizational trust and the possibility of a trust trap [33, 68, 21, 32, 77, 57]. The novelty claim of this paper is therefore not that trust can evolve, reciprocate, spiral, become path-dependent, or be discussed through attractor language. It is narrower: to separate mechanism-level self-reinforcement from stronger attractor claims, to specify what empirical signatures would warrant those stronger claims, and to connect that recursive dynamics to a compositional account of trust-relevant meaning.

That question has at least five components.

First, trust can alter *action*. A participant may cooperate, disclose, delegate, wait, forgive, invest, or refrain from monitoring because the relation supports those choices.

Second, altered action can change *evidence production*. Some evidence of reliability appears only after another actor has been given an opportunity to act. Refusal can protect against exploitation while simultaneously preventing the generation of certain kinds of evidence.

Third, trust can alter *interpretation*. The same delay, silence, mistake, or incomplete message can receive different meanings within different histories. Karen Jones's account of trust emphasizes an affective attitude involving optimism toward another's goodwill and competence, illustrating why trust cannot always be reduced to detached belief updating [34]. The present paper does not adopt that account as an exhaustive theory, but it takes seriously the possibility that trust changes how evidence becomes salient and intelligible.

Fourth, trust can alter *relational structure*. Repeated cooperation may create obligations, routines, shared resources, reputational links, mutual knowledge, switching costs, or third-party expectations. Later interaction then occurs in a field partly constituted by earlier interaction.

Fifth, these changes can affect *future trust*. New evidence, altered interpretation, and transformed structure may reinforce trust, weaken it, split it across domains, or generate a transition toward distrust.

Recursive formation therefore does not imply monotonic growth. It identifies a feedback architecture within which multiple trajectories remain possible.

A compact conceptual representation can make this architecture explicit. Let R_t denote the relational configuration relevant at time t , H_t the relational history carried into that time, T_t the presently realized trust configuration, A_t the action generated under those conditions, and E_{t+1} the evidence or experience generated through the interaction. Then a minimal scheme is

$$A_t = \mathcal{A}(R_t, H_t, T_t, \xi_t), \quad (5)$$

$$E_{t+1} = \mathcal{E}(R_t, A_t, \eta_t), \quad (6)$$

$$R_{t+1} = \mathcal{R}(R_t, A_t, E_{t+1}), \quad (7)$$

$$H_{t+1} = \mathcal{H}(H_t, R_t, A_t, E_{t+1}), \quad (8)$$

$$T_{t+1} = \mathcal{T}(R_{t+1}, H_{t+1}, \zeta_{t+1}), \quad (9)$$

where ξ_t , η_t , and ζ_t collect contingencies not represented explicitly. No particular functional form is assumed here. The scheme simply prevents one conceptual omission: T_t can influence A_t , and the action can participate in producing the later conditions on which T_{t+1} depends.

Definition 1.3 (Relational self-reinforcement). A trust process exhibits *relational self-reinforcement* over a specified range of conditions when trust-dependent interaction modifies the subsequent relational configuration in ways that increase the accessibility, resilience, or reproducibility of trust relative to an otherwise comparable trajectory.

The definition contains several restrictions. The comparison class must be specified; otherwise “reinforcement” has no counterfactual content. The range of conditions must also be specified because a process can be robust to small perturbations and fragile to a qualitatively different shock. Finally, the definition refers to accessibility, resilience, or reproducibility rather than to moral desirability. A relation can be self-reinforcing and harmful.

1.4 Attractors, Basins, Dynamics

The attractor vocabulary becomes useful only after the recursive problem is clear. Positive feedback by itself does not establish an attractor. Nor does persistence. The stronger idea concerns a relational region in which multiple nearby trajectories tend, under specified dynamics, to return toward or remain within a recognizable configuration despite some class of perturbations. The empirical signatures might include recovery after disturbance, threshold behavior, hysteresis, path-dependent transitions, or persistent differences between relations that currently appear similar.

A familiar interpersonal example is successful repair. Two people with a long history of reliable co-operation may survive a minor failure because the event is interpreted within a broader pattern, because channels for explanation already exist, and because both parties have experience of previous repair. Successful repair then becomes part of the history available during a later conflict. The relation has not merely returned to its earlier condition; it may have acquired a new repair pathway. In dynamical language, the interaction has potentially altered the landscape through which later disturbances evolve.

The reverse process is equally important. Repeated opacity can create a configuration in which new assurances are discounted, requests for evidence are read defensively, and ordinary ambiguities accumulate toward suspicion. The problem is not simply that “trust is low.” The relation may have entered a history-dependent region in which the same intervention that would have worked earlier no longer produces the same effect. This possibility motivates later discussion of basin deformation, metastability, transition, and attractor escape.

Claim 1.4 (Landscape deformation). In recurrent trust relations, interaction can change not only the current position of the relation but also the pathways, thresholds, and interpretive weights governing later transitions. A fixed-landscape model can therefore be insufficient even when attractor language is otherwise useful.

This claim gives the present project a distinctive direction. The paper is not mainly interested in locating an actor on a pre-given trust scale. It is interested in how relational interaction can modify the effective landscape on which later trust trajectories unfold.

1.5 Composition, Significance, Indirection

The recursive account still leaves open a question about what enters the relation. Trust is often discussed through reasons, signals, beliefs, and expectations. These are indispensable categories, yet lived trust can also accumulate through elements whose contribution is distributed across form and history. A gesture can matter because it repeats an earlier gesture under new conditions. An omission can matter because something expected did not occur. A shared place can acquire significance because multiple consequential episodes have been organized around it. A phrase can become reassuring because its meaning has been altered by prior use between particular participants.

The term *poetic structure* is introduced to preserve this compositional problem without making a claim of equivalence between trust and poetry. The relevant similarity concerns how significance can emerge from relations among heterogeneous elements. Juxtaposition, recurrence, timing, silence, rhythm, contrast, memory, and context can matter to the whole without any single element functioning as deductively sufficient evidence.

This point has two consequences. The first is epistemic. An observer who extracts individual cues from the history and evaluates them independently may fail to reproduce the significance they had within the relation. The second is causal. A small event can redirect a trajectory when it occurs near a threshold or resonates with an already structured history. Its effect need not be proportional to its isolated magnitude.

Example 1.5 (A small act in two histories). Consider the same act of delayed assistance in two relations. In the first, previous delays have repeatedly been followed by explanation, repair, and reliable support. In the second, previous delays have accompanied abandonment and concealed conflict. The observable delay can be identical while its relational significance differs sharply. The difference lies neither in the isolated event nor in an arbitrary subjective reaction; it lies partly in the history through which the event becomes interpretable.

The example also clarifies why poetic composition should not be confused with irrationality. Context-sensitive interpretation can be evidentially appropriate when histories genuinely differ. The research problem is to determine when such composition tracks relevant structure and when it produces projection, bias, or manipulable confidence.

1.6 Trust Production, Trustworthiness, Governance

Once the generative conditions of trust become an explicit object of analysis, a governance problem follows immediately. Actors with sufficient control over the relational environment may be able to influence the production of trust. Professional credentials, institutional guarantees, interface conventions, spatial design, social endorsements, repeated rituals, selective transparency, and curated narratives can all affect whether reliance feels reasonable or normal. Some of these devices increase genuine accountability. Others may only increase experienced confidence.

The distinction between trust and trustworthiness is therefore indispensable. Hardin's work explicitly separates the two in title and analysis [29]; organizational accounts likewise distinguish trust from antecedent characteristics associated with the trustee [50]. The present framework extends this distinction to *trust production*: an intervention can successfully increase trust while leaving the underlying trustworthiness of the relation unchanged or even worsened.

Definition 1.6 (Trust-production divergence). *Trust-production divergence* occurs when a process that increases experienced or enacted trust does not generate a corresponding improvement in the properties that would warrant reliance under the relevant conditions.

This divergence is not exceptional to digital manipulation or deceptive marketing. It can arise wherever one actor has disproportionate control over the conditions through which another actor interprets reliability. The ethical question therefore concerns more than whether a statement is literally true. It concerns who can structure the field of encounter, which vulnerabilities are made easier to accept, what counterevidence remains visible, and whether the relation remains contestable and reversible.

The normative goal of this paper is consequently not the maximization of trust. High trust can co-exist with exploitation, dependency, epistemic closure, and unequal exit costs. Low trust can sometimes express justified caution. A better research objective is to identify relational conditions under which trust can be generated, tested, revised, repaired, and withdrawn without systematically separating experienced confidence from warranted reliance.

1.7 Scope, Contributions, Structure

The paper makes three core contributions. Governance and justice follow downstream from them.

First, it gives *recursive relational formation* a mechanism-explicit form. The contribution is not the discovery of reciprocal or self-reinforcing trust, which is already present in prior research. It is the separation of current trust, the mediator through which trust changes later interaction, the evidence generated through that interaction, and the later trust update. This separation makes reinforcement empirically contestable rather than circular.

Second, building on prior trust-attractor work, it develops stricter criteria for when attractor terminology is warranted. Stability, duration, positive feedback, and path dependence are insufficient on their own. Attractor language should be retained only when a specified model and observation scheme support signatures such as return under perturbation, basin dependence, threshold transitions, multistability, or hysteresis.

Third, it develops a compositional hypothesis under the term *poetic structure of trust*. The claim is not that context, framing, affect, embodiment, ritual, or symbols matter to trust; each has antecedents in existing research. The claim is that heterogeneous cues can interact historically and sequentially so that their joint trust-relevant significance is not well represented by fixed, context-independent additive contributions.

The governance and justice analyses follow from these three claims. If trust can recursively alter evidence and if trust-relevant meaning can be shaped by compositional environments, then design power over those environments requires analysis of corrigibility, vulnerability, contestability, third-party effects, and exit.

These contributions remain preliminary. The paper does not claim a universal model of trust, does not infer mathematical attractors from metaphor, and does not treat stability as evidence of justice. Its aim is to construct a framework within which empirical, formal, historical, and normative investigations can be posed more precisely.

The sections that follow proceed from conceptual differentiation to dynamics and then to governance. The next section situates the argument more carefully within major traditions of trust research. Subsequent sections distinguish the objects of analysis, develop relational self-reinforcement, examine attractor-like dynamics and historical dependence, and then turn to poetic composition, contingency, memory, and interpretation. The later parts of the paper examine manufactured trust, vulnerability, contestability, repair, and institutional design. The conclusion returns to a modest proposition: trust is not only something relations contain. Under some conditions, trust participates in changing the relations from which later trust will be generated.

2 Vulnerability, Uncertainty, Interpretation

The purpose of this section is to locate the present argument within the existing trust literature without forcing heterogeneous traditions into a single definition. Trust has been studied in philosophy, sociology, psychology, organization studies, economics, political theory, and adjacent fields. These literatures do not merely offer different answers to one common question. They frequently select different objects for

explanation. Some ask what makes reliance rational or justified; some examine willingness to accept vulnerability; some distinguish cognitive and affective bases of interpersonal trust; some study the social and institutional production of expectations; and some investigate the consequences of trust for coordination, organizing, and cooperation [5, 34, 50, 64, 46, 44, 81, 53].

The present paper depends on several of these traditions, but its central object is slightly different. It asks what happens when trust is not only an outcome of prior relations but also one of the conditions that reorganize later relations. This question requires attention to feedback: trust can change communication, disclosure, monitoring, cooperation, interpretation, memory, and willingness to remain exposed; those changes can then alter the evidence and relational possibilities from which later trust is formed. Existing work contains many components of this picture. The contribution sought here is not the discovery that trust is relational, emotional, institutional, temporal, or consequential. It is the more specific attempt to place recursive relational formation at the center of analysis and to connect it with attractor-like stability, heterogeneous composition, and the governance of trust-generating conditions.

Remark 2.1 (A cumulative rather than replacement strategy). The argument developed here should be read cumulatively. It does not require a claim that earlier theories are static, individualistic, or insensitive to history. Many are explicitly relational and processual. The narrower claim is that the feedback by which prior trust modifies the conditions of later trust can be isolated more sharply and studied in its own right.

2.1 Vulnerability, Exposure

A first family of approaches makes vulnerability indispensable to the problem of trust. Baier draws attention to relations in which one party entrusts something of value to another and thereby becomes open to forms of harm that mere prediction does not adequately capture [5]. In organization studies, Mayer, Davis, and Schoorman similarly distinguish trust from judgments about a trustee's characteristics by connecting trust to a willingness to become vulnerable to another party's actions [50]. Rousseau and colleagues, working toward a cross-disciplinary formulation, likewise emphasize positive expectations in conditions involving vulnerability and interdependence [64].

This literature matters for the present framework because vulnerability is not only a passive condition surrounding trust. Once enacted, it can transform the relation. Disclosure gives another party information that did not previously exist within the relation. Delegation creates an opportunity for performance or betrayal. Reliance can redistribute responsibility and expose hidden competence, negligence, care, or opportunism. Forgoing verification can alter the interaction itself. A trusting act therefore sometimes produces the very events from which later judgments will be made.

The recursive problem begins here. Suppose actor *A* discloses information to actor *B* because *A* already has some trust in *B*. If *B* responds with care, *A* obtains evidence that could not have been generated without the initial exposure. If *B* exploits the disclosure, the same opening generates contrary evidence. In either case, trust has affected the experimental conditions of the relationship. The later evidential field is partly endogenous to the earlier relational state.

This endogeneity complicates a simple separation between “evidence first” and “trust second.” In many repeated relations, evidence and trust can enter a feedback sequence. Prior evidence contributes to trust; trust changes action; action generates new evidence; and the resulting evidence can reinforce, redirect, or rupture the relation. The framework developed later in this paper will use this loop to supply one of the minimal structures from which a trust attractor may emerge.

2.2 Expectation, Interpretation, Uncertainty

A second tradition emphasizes the relation between trust and uncertain futures. Luhmann links trust to the reduction of social complexity: action must proceed even though future states and other actors cannot be exhaustively known [46]. Möllering later reconstructs a major line of trust research around reason, routine, reflexivity, and what he calls suspension. In his account, interpretive grounds do not mechanically

determine a trustful expectation. Trust involves a movement beyond what available interpretation can conclusively establish [55, 56].

This point is especially important for a relational-dynamical account. If trust followed deductively from complete information, the problem of self-reinforcement would be much thinner. New evidence could simply update a fully specified calculation. In actual trust relations, however, action occurs while some uncertainty remains unresolved. Participants proceed, refrain from checking, disclose, wait, delegate, cooperate, or remain present despite an incomplete evidential situation. Those actions can then change the situation that later must be interpreted.

The presence of an irreducible remainder also helps explain why identical observable events can produce different trust updates. An unanswered message may indicate benign delay within one relationship and impending abandonment within another. A procedural failure may be absorbed into ordinary noise where an institution has a long history of repair, while the same failure may confirm a pre-existing suspicion elsewhere. The relevant difference cannot always be found inside the isolated event. It can reside in the relational history that organizes interpretation.

For the present paper, this does not justify arbitrary belief. The distinction between experienced trust, justified trust, and trustworthiness remains essential. Rather, the point is that interpretation itself has a history. Trust can help constitute the background against which later signals become salient, forgivable, suspicious, ordinary, or exceptional. A recursive account must therefore model not only changing expectations but also changing interpretive conditions.

Claim 2.2 (Historical interpretation). When relational history changes the meaning assigned to later events, trust updates cannot in general be reconstructed from the current event alone. At minimum, the analysis requires information about the state or history through which the event is interpreted.

This claim is modest but consequential. It does not yet specify whether the relevant memory can be compressed into a finite state vector or whether the full trajectory matters. That question belongs to the later formal analysis. The conceptual point is simply that temporal ordering and relational history may carry explanatory information.

2.3 Cognition, Affect, Embodiment

A third line of research resists reducing trust to detached calculation. Jones develops an account centered on affective attitude, while McAllister's organizational research distinguishes cognition-based and affect-based forms of interpersonal trust [34, 52]. Lewis and Weigert likewise argue for a sociological understanding in which cognitive and emotional dimensions coexist within a broader social reality of trust [44]. These approaches differ substantially, but they jointly make it difficult to reduce every trust relation to the output of one homogeneous evidential calculus.

This heterogeneity provides one foundation for the paper's later account of poetic structure. A trust-relevant configuration can contain explicit reasons, affective responses, memories, habits, social expectations, spatial cues, embodied reactions, and symbolic elements at the same time. Their effects need not be additive or independently intelligible. A gesture may matter because it recalls an earlier event. A routine may become reassuring because it has survived repeated disturbances. A place may carry the memory of prior repair. A formal assurance may become persuasive only because it arrives from an institution already situated within a trusted network.

The important implication is not that affect replaces reason or that symbolic meaning overrides evidence. Such substitutions would reproduce the very reductionism this paper seeks to avoid. The implication is compositional. The relational effect of an element can depend on the configuration in which it appears. Consequently, a theory of trust formation may sometimes require more than a list of independent antecedents. It may need to represent interactions among elements, their temporal order, and the history through which their meaning has become available.

This also clarifies why the term *poetic* is introduced cautiously. The term identifies a structural problem of composition, not an invitation to suspend epistemic standards. The later argument will ask how

heterogeneous and partially indirect elements become significant together, while preserving the possibility that such significance can mislead, manipulate, or diverge from trustworthiness.

2.4 Institutions, Social Production

Trust is also studied beyond the dyad. Lewis and Weigert locate trust within a sociological account of social reality rather than reducing it to an isolated individual disposition [44]. Zucker's historical and organizational analysis examines institutional sources through which trust is produced, shifting attention toward characteristic-based, process-based, and institutionally supported conditions of economic coordination [81]. Later work on institutional trust building continues to examine how formal and informal institutional arrangements can shape trust between organizations [4].

This literature is indispensable for any governance-oriented theory of trust. Participants rarely encounter one another in an empty field. Professional roles, legal forms, reputational infrastructures, certification systems, organizational histories, social categories, technical standards, insurance, contracts, and public institutions can all modify the initial relational conditions. Such structures may lower the cost of first cooperation or make otherwise unavailable forms of reliance possible.

From a generative perspective, however, institutional support introduces a second recursive layer. Institutions can help generate trust, while trust can in turn alter the operation of institutions. High trust can reduce monitoring, increase voluntary compliance, widen discretion, and enable informal coordination. Low trust can increase documentation, verification, defensive procedure, and enforcement. These changes can then affect how institutions are experienced and what evidence they produce. A trusted procedure may generate smooth participation that appears to confirm its legitimacy; a distrusted procedure may generate resistance and surveillance that further changes the relationship between institution and participant.

McEvily, Perrone, and Zaheer provide an especially relevant bridge by showing how trust can influence organizing through structuring and mobilizing pathways [53]. Dirks and Ferrin similarly distinguish direct effects from processes through which trust changes the influence of other determinants on organizational outcomes [14]. These contributions reinforce a key premise of the present paper: trust does not merely register a relation; it can alter what participants subsequently do inside that relation.

The additional step proposed here is to feed these consequences back into the formation of later trust. If trust changes organizing, and changed organizing changes future interaction, then the consequences of trust can become new antecedents of trust. The resulting loop is not guaranteed to be positive. It can stabilize cooperation, conceal deterioration, produce dependence, expose contradictions, or generate a sudden transition toward distrust. What matters is that the loop itself becomes analytically visible.

2.5 Development, Coexistence, Multiplicity

Process-oriented work already rejects the idea that trust appears fully formed. Lewicki and Bunker describe developmental changes in work relationships and connect different relational conditions to different bases for trust [42]. Möllering emphasizes the process through which trust forms, including interpretation and suspension [55]. Institutional accounts likewise investigate trust-building processes rather than only terminal states [4]. These lines of research make clear that the present paper's concern with temporal formation has substantial precedents.

At the same time, development should not be narrated on a single ladder running from low trust to high trust. Lewicki, McAllister, and Bies argue that trust and distrust can coexist rather than forming opposite endpoints of one dimension [43]. This is highly compatible with a relational-landscape view. A person may trust another actor's technical competence while distrusting their motives. An organization may be trusted to execute a routine operation but distrusted when incentives change. A public institution may receive procedural reliance together with deep suspicion of its political objectives. Relations can therefore contain multiple domains, multiple evaluative dimensions, and partially conflicting histories.

Such multiplicity matters for attractor language. If trust and distrust can coexist, it is hazardous to represent the whole relation by a single scalar without empirical justification. A trajectory may approach stability in one domain while becoming unstable in another. A perturbation may leave competence trust

intact while destroying confidence in benevolence or integrity. Repair may restore one component but not the others. The relevant state space may therefore need to be multidimensional, domain-specific, or layered across scales.

Research Obligation 2.3 (Preserve relational heterogeneity). Any later formalization that compresses trust into a small number of variables must state what distinctions are being discarded. In particular, a scalar representation should not be taken to be ontologically complete merely because it is mathematically convenient.

The same caution applies to scale. Interpersonal, interorganizational, institutional, and political trust may share structural features while relying on different mechanisms. The recurrence of terms such as vulnerability, expectation, history, and institutional support does not by itself license a single universal model.

2.6 Trust, Distrust, Evidential Environments

The preceding traditions can now be brought together around a narrower question. Vulnerability creates consequential openings. Uncertainty requires participants to proceed without complete determination. Cognitive and affective processes shape interpretation. Institutions modify the conditions under which reliance becomes possible. Trust affects organizing and cooperation. Relations develop through time, and trust can coexist with distrust across domains. Taken together, these insights imply that a trust relation can modify the environment from which its later confirmation or disconfirmation is drawn.

Consider two simple trajectories. In the first, modest initial trust increases communication. Better communication makes coordination easier. Successful coordination produces shared experience and lowers the cost of future communication. Ambiguous failures are investigated before participants infer betrayal. The relation thereby becomes increasingly resilient to small perturbations. In the second trajectory, modest distrust increases monitoring and defensive withholding. Withholding reduces the information available for coordination. Coordination becomes more fragile, failures become more common, and each failure is interpreted against an already suspicious background. The result can be an increasingly closed distrust trajectory.

Neither sequence is inevitable, and neither should be taken for a universal psychological law. External shocks, asymmetric power, new information, institutional intervention, strategic manipulation, or deliberate reflection can interrupt either loop. The examples instead illustrate a structural possibility: trust and distrust can participate in producing the future conditions under which they are updated.

This possibility also creates an epistemic difficulty. A self-reinforcing relation can generate evidence that appears to validate the relation because alternative interactions are no longer sampled. A trusted actor receives more opportunities to demonstrate reliability; a distrusted actor receives fewer opportunities to do so. A trusted institution receives voluntary cooperation that makes its procedures work more smoothly; a distrusted institution may encounter resistance that degrades performance. Observed outcomes may therefore reflect both underlying trustworthiness and the relational pathway through which opportunities were distributed.

The problem resembles neither simple bias nor simple accuracy. A trusting relation may reveal genuine properties that would otherwise remain hidden, and a distrusting relation may protect participants from exploitation. Yet both can also become partially self-confirming. The analytical task is to distinguish cases in which feedback improves access to relevant evidence from cases in which feedback narrows the evidential environment and shields the relation from correction.

Definition 2.4 (Recursive trust formation). A trust process exhibits *recursive formation* when a trust-relevant state at one time changes actions, interactions, or interpretive conditions in ways that affect the production or evaluation of trust-relevant information at a later time.

The definition does not require positive feedback. Recursive formation includes reinforcement, weakening, oscillation, repair, and abrupt reversal. It also does not require conscious intention. Participants may alter future trust conditions simply by acting under the practical consequences of current trust or distrust.

2.7 Conceptual Opening, Attractors

The literature reviewed here therefore leaves the present paper with a precise opening rather than an empty field. Existing theories already explain many pieces of trust: accepted vulnerability, positive expectation, affect, interpretation, institutional production, developmental processes, multidimensionality, and organizational consequences. Dynamic work also goes further than a static antecedent–outcome picture: trust has been modeled through evolution, reciprocity, spirals, complex-adaptive interaction, temporal mechanisms, and self-reinforcing lock-in [33, 68, 21, 32, 77, 57]. The question pursued in the remainder of the paper is therefore not whether trust changes over time. It concerns which feedback mechanisms change the conditions of later trust, when those mechanisms justify stronger dynamical claims, and how historically composed meaning enters the same recursion.

This shift changes the unit of analysis. The object is no longer only the trustor’s current judgment, the trustee’s current properties, or the current institutional guarantee. It is also the evolving configuration that links state, action, environment, interpretation, and history. A useful notation will therefore need to distinguish at least the relational state at time t , the actions enabled or inhibited by that state, the observations generated by those actions, and the update process through which later relational states emerge.

The attractor vocabulary becomes relevant only after this recursive structure is established. A persistent high level of reported trust is not sufficient. An attractor claim concerns the behavior of trajectories under perturbation: whether a set of relational configurations tends to draw nearby trajectories toward it, whether multiple stable regions coexist, whether thresholds separate qualitatively different futures, and whether the landscape itself changes with history. These questions will be developed progressively rather than assumed in advance.

The same recursive structure prepares the later discussion of poetic composition. If the relational state determines which elements become salient, and if heterogeneous elements acquire significance through historical configuration, then trust formation cannot always be represented by fixed, context-independent weights assigned to isolated cues. Composition can affect transition. A small event may have little force in one region of the relational landscape and redirect the trajectory in another.

Finally, the recursive view sharpens the governance problem. Designing for trust is not merely a matter of sending better signals. An intervention can alter the future field of interaction, the distribution of vulnerability, the availability of verification, and the costs of exit. Once those effects feed back into the relation, governance concerns the evolution of a relational system rather than the optimization of a momentary attitude.

Claim 2.5 (Conceptual bridge). The trust-attractor hypothesis becomes theoretically meaningful when three conditions are considered together: trust affects subsequent interaction; subsequent interaction affects the production or interpretation of later trust-relevant evidence; and the resulting updates exhibit some degree of recurrence, path dependence, resilience, or threshold behavior. The next sections develop these conditions without presuming that every trust relation satisfies them.

This formulation also marks the paper’s principal restraint. The aim is not to rename all trust processes with dynamical terminology. The aim is to identify which trust processes display sufficiently strong recursive organization that a dynamical description clarifies rather than obscures them. That distinction will guide the formal constructions that follow.

3 State, Feedback, Attractors

The recursive claim developed above needs a more explicit dynamical vocabulary. Trust can alter disclosure, action, observation, memory, and interpretation; those changes can then enter later trust formation. Dynamical and attractor language already appears in prior trust research [40, 45, 32]. The task here is therefore not to introduce that language, but to specify what must be represented before terms such as *feedback*, *basin*, *perturbation*, and *attractor* carry sufficiently restrictive analytical content for the present relational framework.

The evolving object is therefore not assumed to be a universal scalar trust score. It is a relational configuration whose coordinates may include expectations, affect, memory, vulnerability, coordination opportunities, power over verification or exit, and institutional conditions. Actions and observations arise inside that configuration and can subsequently transform it. Only after this recursive structure is explicit does the section distinguish a strict dynamical attractor from a weaker *trust-attractor candidate* for empirical work. The distinction matters because attraction concerns trajectories from a neighborhood of states, not mere persistence [71]. Positive feedback and path dependence likewise show how historical events can alter later possibilities without implying that every reinforcing process converges to one fixed outcome [3].

3.1 Relational Configuration, State

Let $\mathcal{X}$ denote a state space appropriate to a specified trust domain. A relational configuration at time t is written

$$\mathbf{x}_t \in \mathcal{X}. \quad (10)$$

The notation deliberately leaves the dimension and geometry of $\mathcal{X}$ open. Different empirical settings may require different state spaces. An interpersonal relation, an interorganizational partnership, a humanitarian institution, and an online platform need not share the same variables merely because each contains something called trust.

For exposition, one admissible decomposition is

$$\mathbf{x}_t = (B_t, A_t, M_t, V_t, C_t, P_t, I_t), \quad (11)$$

where B_t collects trust-relevant beliefs or expectations, A_t affective or embodied orientations, M_t relational memory and interpretive history, V_t vulnerability and dependence, C_t available modes of coordination and contact, P_t power over monitoring, sanction, refusal, and exit, and I_t institutional or technical conditions. Equation (11) is an illustrative coordinate system, not an ontological decomposition of trust. A study can merge, omit, or replace its components when theory and evidence justify doing so.

Definition 3.1 (Relational configuration). A *relational configuration* is the collection of trust-relevant conditions that jointly constrain what participants can do, encounter, interpret, verify, contest, remember, and revise at a specified time and in a specified domain.

The definition places action possibilities inside the object of inquiry. This is important because two actors can report similar expectations while inhabiting very different relations. One may possess inexpensive exit, independent verification, and multiple alternative partners; another may depend on a single institution that controls both information and redress. Treating those cases as identical because a survey reports the same numerical trust score would erase conditions that can strongly affect subsequent trajectories.

The same point applies to memory. M_t should not be reduced automatically to a list of past observations. Relational memory can include which events were recorded, which were forgotten, which explanations became conventional, which breaches were repaired, and which experiences remain available for later reinterpretation. The history that matters for trust is therefore partly a history of relations among events, interpretations, and present possibilities.

Research Obligation 3.2 (Domain-specific state construction). Before estimating a trust dynamic, a study should state which relational components are represented, which are treated as external conditions, and which are omitted. Mathematical tractability does not by itself justify the claim that the selected variables exhaust the relation.

3.2 Action, Observation, Updating

A relational configuration matters dynamically when it affects what happens next. Let $\mathbf{a}_t$ denote actions selected by participants, $\mathbf{y}_t$ the trust-relevant observations generated during the interaction, and $\mathbf{u}_t$ exogenous interventions or shocks that are not produced by the relation under study. A minimal discrete-time

representation is

$$\mathbf{a}_t = \Pi_t(\mathbf{x}_t, \mathbf{d}_t), \quad (12)$$

$$\mathbf{y}_t = G_t(\mathbf{x}_t, \mathbf{a}_t, \varepsilon_t), \quad (13)$$

$$\mathbf{x}_{t+1} = \Phi_t(\mathbf{x}_t, \mathbf{a}_t, \mathbf{y}_t, \mathbf{u}_t; \mathcal{H}_t), \quad (14)$$

where $\mathbf{d}_t$ represents the practical demand faced at time t , ε_t represents unresolved variability, and $\mathcal{H}_t$ denotes the relational history available to the update process. No assumption is made here that participants consciously compute Π_t , G_t , or Φ_t . These maps can represent learned habits, institutional procedures, affective responses, strategic decisions, or mixed mechanisms.

Equations (12)–(14) separate three operations that static trust measures often combine. First, the current relation changes which actions become likely or even thinkable. Second, those actions help determine which evidence can be generated. Third, the produced evidence is interpreted within a history rather than added mechanically to an unchanging ledger. Recursive trust formation occurs when these operations feed back into later relational configurations.

A familiar example is disclosure. Suppose a participant discloses sensitive information because the present relation permits vulnerability. The disclosure creates an opportunity for the other participant to protect, exploit, ignore, or reciprocate. Whatever follows becomes evidence that would not have existed without the prior relational opening. The next trust-relevant state is therefore conditioned partly by an event whose possibility depended on the earlier state. The relation participates in constructing its own later evidential environment.

This recursive structure is compatible with both reinforcement and correction. If disclosure is protected, the relation may become more open. If disclosure is exploited, trust may collapse. If the event is ambiguous, the existing history may influence whether the participants investigate, excuse, contest, or withdraw. The same architecture can therefore produce convergence, oscillation, repair, polarization, or escape.

Claim 3.3 (Endogenous evidence production). When a relational configuration changes which interactions occur, it also changes the distribution of trust-relevant evidence that can subsequently be observed. Later evidence is then partly endogenous to the preceding relation.

This claim has an immediate epistemic consequence. Evidence produced inside a trust relation cannot always be treated as though it were sampled from an interaction process independent of trust. A trusted participant may receive more chances to demonstrate reliability. A distrusted participant may be excluded before such opportunities arise. Conversely, strong distrust may induce monitoring that reveals misconduct which trusting interaction would have missed. The relation affects not only how evidence is weighted but sometimes which evidence becomes possible.

3.3 Mechanism, Self-Reinforcement

The phrase *self-reinforcing trust* can become empty if it merely restates persistence. A relation is not self-reinforcing simply because trust at one time correlates with trust later. Reinforcement requires an endogenous mechanism through which the current relational configuration contributes to conditions that make related future configurations more likely or more resilient.

Several mechanisms are possible. Trust can increase communication and thereby improve coordination. It can increase disclosure and permit richer mutual models. It can reduce the immediate cost of requesting clarification after an ambiguous event. It can create reputational or role commitments that make reliable action more valuable. It can also produce shared routines that reduce the friction of future cooperation. None of these mechanisms is guaranteed, and several can operate in the opposite direction. Greater disclosure can increase exploitability; routine can conceal emerging problems; reputational commitment can protect a relation from justified criticism.

Definition 3.4 (Relational self-reinforcement). A trust-relevant relational process is *self-reinforcing* over a specified interval when consequences endogenously generated by its present configuration increase the persistence, recurrence, or local recoverability of a related class of future configurations.

Three qualifications are built into the definition. First, reinforcement is mechanism-relative: a study should identify what consequence feeds back into what later condition. Second, reinforcement is local and temporal rather than necessarily permanent. A relation can reinforce itself for years and still collapse after a sufficiently large breach. Third, reinforcement need not make the relation normatively desirable. Coercive dependence, manipulation, and closed information environments can also produce persistence.

A useful empirical design would compare the observed process with a counterfactual in which a specified feedback channel is interrupted. Let K denote a transition kernel for the observed relational process and $K^{(-r)}$ a corresponding kernel under intervention on reinforcement channel r . For a trust-supporting region $\mathcal{S}_T \subseteq \mathcal{X}$, one possible operational signature at horizon n is

$$K^n(\mathbf{x}, \mathcal{S}_T) > (K^{(-r)})^n(\mathbf{x}, \mathcal{S}_T), \quad (15)$$

for specified starting states $\mathbf{x}$ and a theoretically justified intervention. Equation (15) is not proposed as a universal estimator. It illustrates the stronger evidential burden implied by the term *self-reinforcement*: one should be able to identify a feedback channel whose removal changes later persistence or return.

3.4 Attractors, Basins, Multistability

The attractor concept can now be stated more carefully. Consider first an autonomous deterministic system

$$\mathbf{x}_{t+1} = F(\mathbf{x}_t). \quad (16)$$

For a compact invariant set $\mathcal{A} \subseteq \mathcal{X}$, a standard form of attraction requires a neighborhood U such that trajectories beginning in U approach $\mathcal{A}$ under repeated application of F :

$$\text{dist}(F^n(\mathbf{x}), \mathcal{A}) \longrightarrow 0 \quad \text{for } \mathbf{x} \in U \text{ as } n \longrightarrow \infty. \quad (17)$$

The set of initial conditions that approach $\mathcal{A}$ constitutes its basin of attraction under the chosen definition. Fixed points, periodic sets, and more complicated invariant sets can all participate in attractor theory; the relevant object need not be a single equilibrium [71].

This strict formulation immediately reveals why ordinary persistence is insufficient. If trust stays high only because no perturbation occurs, one has not yet demonstrated attraction. If a relation remains stable only while an external subsidy is continuously applied, the relevant attractor may belong to a larger system containing that subsidy rather than to the relation considered in isolation. If every nearby perturbation causes permanent departure, the observed state may be long-lived without being locally attracting.

Most real trust processes will not initially satisfy the assumptions of Equation (16). They may be stochastic, partially observed, history-dependent, strategically reflexive, and institutionally nonstationary. The present paper therefore distinguishes strict mathematical attractors from *trust-attractor candidates*.

Definition 3.5 (Trust-attractor candidate). A *trust-attractor candidate* is a set of relational configurations for which evidence suggests recurrent or persistent organization, a nontrivial region of approach or return, and some capacity to recover after bounded perturbation, while the exact dynamical representation remains an empirical and formal question.

This weaker term permits disciplined inquiry without claiming more mathematics than the evidence supports. A candidate becomes more persuasive when one can identify: (i) a state description with empirical meaning; (ii) repeated trajectories rather than a single narrative; (iii) local perturbations and observable recovery or escape; (iv) a basin-like dependence on initial or historical conditions; and (v) a mechanism connecting current relational conditions to later transition probabilities.

Remark 3.6. The adjective *trust* identifies the relational region under study; it does not imply that every coordinate of the attractor corresponds to positive trust. A stable trust relation may contain vigilance, verification, limited distrust, formal safeguards, or domain-specific refusal. Robust trust can therefore occupy a mixed relational configuration rather than a state of undifferentiated confidence.

Attractor language becomes especially useful when several long-run regimes are possible. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ denote two attracting sets with basins $\mathcal{B}(\mathcal{A}_1)$ and $\mathcal{B}(\mathcal{A}_2)$. Two relations exposed to similar present events can then diverge because they occupy different regions of the state space. A minor delay may be absorbed inside one relation and interpreted as evidence of betrayal inside another. The difference need not be located in the delay itself. It can arise from the historical configuration through which the event is encountered.

This offers a precise way to describe one form of path dependence. Present conditions may not uniquely determine future development if historically produced differences place otherwise similar relations in different basins or near different basin boundaries. Positive-feedback research has long shown how small historical events can become consequential when subsequent processes reinforce a selected trajectory [3]. In trust research, the analogous question is not whether every early event creates lock-in. It is which relational mechanisms amplify some differences, damp others, and alter the cost of moving between regimes.

A multidimensional state space also allows trust and distrust to coexist. Instead of placing them on one line, one can imagine regions in which technical reliance is high while motive-related suspicion remains high, or institutional procedural confidence is stable while confidence in leadership is low. A basin can therefore contain structured mixtures rather than a single global level of trust. Transitions between such regions may occur component by component.

Claim 3.7 (History-sensitive response). If trust-relevant trajectories occupy different basin regions, the same external event can generate different subsequent trajectories without requiring that the event itself possess different intrinsic informational content. Relational history can change the event’s dynamical consequence.

The claim provides one bridge to the later discussion of poetic structure. A small gesture can be dynamically weak in one relational region and decisive in another. Its consequence depends on composition with memory, vulnerability, expectation, and current proximity to a transition boundary. The significance of the event therefore cannot always be read from the event in isolation.

3.5 History, Perturbation, Landscape

A fixed attractor landscape is still too restrictive for many generative relations. Repeated interaction can change not only the current state but also the rules through which later states are produced. Participants learn new ways to interpret one another, institutions change procedures, reputational systems are redesigned, communication channels are opened or closed, and power over exit can shift. The effective transition structure can therefore evolve with the trajectory.

To represent this possibility, let θ_t collect parameters or rules governing relational updating. Instead of a fixed map F , consider

$$\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t, \mathbf{a}_t, \mathbf{y}_t; \theta_t), \quad (18)$$

$$\theta_{t+1} = \Psi(\theta_t, \mathbf{x}_t, \mathbf{a}_t, \mathbf{y}_t, \mathbf{u}_t). \quad (19)$$

Equation (18) describes movement under the presently available relational rules. Equation (19) describes change in those rules. The second equation is essential for relations in which history alters sensitivity, interpretive conventions, institutional safeguards, or the availability of future actions.

Definition 3.8 (Relational landscape deformation). *Relational landscape deformation* occurs when interaction history changes the transition structure governing later relational dynamics, so that future movement cannot be represented adequately by a single time-invariant update rule on the original state space.

This distinction separates two forms of change that are easily conflated. A relation can move to a new point within an otherwise fixed landscape, or the landscape itself can be reorganized. Successful repair after a breach may do both. The immediate relation can return toward cooperation, while the repair process creates a new norm of disclosure, a new verification procedure, or a new shared interpretation of what counts as warning. Later perturbations are then encountered under different transition rules.

The same mechanism can deepen harmful dependence. A manipulative relation can change the participant's information access, available alternatives, or criteria for interpreting warning signs. Persistence then reflects not merely movement toward a stable state but transformation of the conditions under which escape could occur. This is one reason stability cannot be treated as a proxy for legitimacy.

Proposition 3.9 (State stability does not imply rule stability). *An observed relational state can remain within a narrow region while the rules that govern future transitions change substantially. Consequently, a stable series of trust reports does not establish that the underlying relational landscape is stable.*

Proof. Let $\mathbf{x}_t$ remain in a neighborhood U for an interval while θ_t changes according to Equation (19). If $\Phi(\cdot; \theta_t)$ differs across that interval, then identical later perturbations can generate different responses even though the observed state remained inside U . Stability of $\mathbf{x}_t$ therefore does not entail stability of the transition rule. $\square$

The proposition is elementary, but it has substantive implications for trust measurement. Cross-sectional stability can coexist with growing fragility, increasing dependence, or improving resilience. To distinguish these cases one must study response properties and evolving relational mechanisms rather than only state levels.

Attractor claims become empirically informative when relations are observed under perturbation. A perturbation can be a missed obligation, contradictory information, leadership change, communication delay, institutional failure, third-party accusation, sudden dependency, or deliberate stress test. The relevant question is not simply whether trust falls. It is what the trajectory does afterward.

Several response patterns should be distinguished. A relation can return close to its prior region; it can settle into a different but still cooperative configuration; it can oscillate for an extended period; it can cross into a distrust-dominated regime; or it can dissolve. Repair therefore need not mean restoration of the pre-breach state. A repaired relation may have different verification practices, boundaries, memories, and distributions of vulnerability. In the coevolutionary representation above, repair can change both $\mathbf{x}_t$ and θ_t .

For a strict autonomous system with attractor $\mathcal{A}$, one can define a return-time observable after perturbation. If a perturbed state $\tilde{\mathbf{x}}_0$ remains inside the basin, then for tolerance $\epsilon > 0$ let

$$\tau_\epsilon = \inf \{n \geq 0 : \text{dist}(F^n(\tilde{\mathbf{x}}_0), \mathcal{A}) < \epsilon\}. \quad (20)$$

Shorter or longer return times can then become one component of an empirical resilience analysis. In social applications, however, the meaning of return must be specified carefully. A quick return to reported confidence can reflect genuine repair, suppression of dissent, renewed dependence, or measurement insensitivity.

Work on critical transitions in complex systems shows why recovery behavior, multistability, and regime shifts can carry information about changing stability [67]. That literature should not be transferred mechanically to trust. It does, however, motivate testable questions. Does recovery from comparable minor breaches become slower over time? Does the variance of relational responses increase near a transition? Does the relation begin to switch repeatedly between cooperative and defensive regimes? Are there hysteresis-like effects in which restoring the original external condition fails to restore the prior relation? These are hypotheses to be tested, not generic properties of trust.

Research Obligation 3.10 (Perturbation evidence). A strong empirical claim about a trust attractor should include evidence about responses to perturbation or naturally occurring disturbances. Persistence without perturbation provides weaker evidence because it does not reveal local recovery, basin boundaries, or escape behavior.

Escape deserves equal attention. A relation that repeatedly returns to a trust-supporting region after benign perturbations may nevertheless prevent exit after harmful ones. Conversely, the possibility of low-cost exit can make trust more ethically acceptable even when it reduces persistence. A governance framework therefore needs at least two distinct questions: whether the relation is dynamically resilient and whether participants retain meaningful capacity to revise or leave it.

3.6 Identification, Stability, Justice

The framework can now be translated into an empirical research programme. No single indicator establishes an attractor. The case becomes stronger when several forms of evidence converge.

First, repeated trajectories should reveal recurrence or persistent organization in a relational state space whose coordinates have substantive meaning. Second, nearby starting configurations should display some tendency to approach similar regions under comparable conditions. Third, bounded perturbations should permit observation of recovery, reorganization, or basin escape. Fourth, historical differences should help explain divergent responses to similar present events. Fifth, at least one endogenous feedback mechanism should connect current trust-relevant conditions to later opportunities, evidence, or interpretation. Sixth, where the landscape itself appears to change, the analysis should model or document that deformation rather than forcing the process into a stationary representation.

Candidate empirical signatures therefore include path-dependent response, asymmetric recovery after comparable breaches, hysteresis-like patterns, changes in return time, regime switching, structured resistance to small perturbations, and discontinuous change after threshold-crossing events. None of these is unique to trust, and none proves a specific mechanism on its own. Their value lies in converting a broad metaphor into a sequence of falsifiable questions.

Claim 3.11 (Empirical content). The trust-attractor hypothesis becomes empirically meaningful only when it restricts expected trajectories. It should identify relational states, feedback mechanisms, perturbations, and response patterns whose observation could strengthen or weaken the attractor interpretation.

This requirement also protects the framework from retrospective storytelling. If every persistence, collapse, repair, or betrayal can be redescribed after the fact as an attractor phenomenon, the concept predicts nothing. The model must expose itself to failure by stating which trajectories would be inconsistent with the proposed landscape or mechanism.

The section closes with a distinction that will govern the rest of the paper. Dynamical stability is a property of trajectories under specified transition rules. Trustworthiness concerns whether reliance is warranted by relevant features of the trustee, institution, or relation. Justice concerns additional questions about power, vulnerability, voice, revision, distribution, and exit. These properties can vary independently.

A trustworthy relation may remain dynamically fragile because participants have little shared history. An untrustworthy institution can occupy a highly stable region because information is controlled and alternatives are costly. A just institutional arrangement can deliberately preserve contestation and easy exit, thereby reducing lock-in. A durable relation can therefore be good or bad, warranted or unwarranted, emancipatory or coercive. Stability supplies no normative verdict.

Proposition 3.12 (Non-equivalence of stability and warrant). *The existence of a stable or attracting trust-supporting regime does not entail that the underlying object of trust is trustworthy, nor that the relation is just.*

The proposition is not a technical theorem about a particular empirical model; it records a categorical separation among three questions that the later analysis must preserve. The next stage of the paper can therefore study how trust-supporting regimes are generated and reinforced without treating their persistence as evidence of moral legitimacy. This separation will become especially important when the analysis turns to heterogeneous cues, poetic composition, engineered environments, and the possibility that relational conditions can be deliberately arranged to produce confidence.

4 Composition, Salience, Meaning

Dynamical language clarifies persistence, perturbation, and historical changes in relational trajectories, but it does not by itself explain why individually weak, indirect, or ambiguous events can acquire large trust-relevant significance within a particular history. A remembered act can change the meaning of a later

silence; a familiar rhythm of response can make a short message reassuring; a minor inconsistency can become decisive when it resonates with an older pattern of concealment.

This section develops *poetic relational composition* for that problem. The concept concerns structure rather than an identity between trust and poetry: heterogeneous and partially indeterminate elements can acquire meaning through relation, sequence, recurrence, embodiment, memory, and context. The component phenomena are not new to trust research. Framing can alter inferred intentions and trust judgments; embodied verbal and non-verbal conduct can combine in the construction of clinical trust; affect and symbolic interaction enter accounts of trust development; and observable ritual conformity can increase trust under experimentally controlled conditions [35, 10, 33, 73].

The narrower proposal concerns the *composition* of such elements. Its intellectual neighbors include Ricoeur's account of semantic innovation, Goodman's plural symbolic functions, Goffman's analysis of frames, and Böhme's work on atmospheres [63, 25, 24, 7]. The claim earns separate terminology only when sequence, interaction, historical embedding, or cross-modal relations add explanatory value beyond a model in which each cue contributes a fixed independent weight.

4.1 Composition, Heterogeneity, Condensation

The expression *poetic structure* can easily become too broad. If every context-dependent interpretation is called poetic, the term loses analytic value. If the term is restricted to verse, metaphor, or artistic production, it cannot describe the relational phenomenon at issue here. A narrower middle position is therefore required.

Definition 4.1 (Poetic relational composition). A configuration exhibits *poetic relational composition* when trust-relevant significance is generated through the historically situated composition of heterogeneous, partial, or indirect elements, such that the significance of the configuration cannot be adequately reconstructed by assigning each element a fixed context-independent contribution and simply adding those contributions together.

Five features are central to this definition. First, the relevant elements can be heterogeneous: propositions, gestures, silences, spatial arrangements, embodied responses, institutional symbols, temporal rhythms, remembered acts, third-party guarantees, or material objects can participate in the same configuration. Second, their meaning is historically indexed. A cue can carry a different significance after betrayal, repair, shared difficulty, or long cooperation. Third, the composition is interactive. One element can amplify, qualify, neutralize, or reverse another. Fourth, the result remains underdetermined with respect to trustworthiness. A configuration can produce confidence without proving that confidence warranted. Fifth, the composition has trajectory relevance: it changes the probabilities of later disclosure, reliance, verification, withdrawal, interpretation, or repair.

The adjective *poetic* therefore refers to a family of compositional properties rather than to a literary identity. The same structural properties can occur in ordinary interpersonal life, organizational routines, diplomacy, religious practice, interface design, architecture, public ceremony, or clinical interaction. Some configurations may be intentionally designed; others emerge without an author. Some may be experienced as beautiful; others may be ordinary, uncomfortable, or barely noticed.

Remark 4.2 (No aesthetic warrant). Calling a trust-generating configuration poetic does not confer epistemic or moral authority on it. Poetic composition can support warranted trust, misplaced trust, distrust, manipulation, repair, or resistance. The term characterizes a mode of relational meaning-formation, not the truth or justice of the resulting judgment.

This qualification is especially important because trust often contains a gap between what is explicitly known and what is practically accepted. Möllering's account of interpretation and suspension already shows why trust cannot be reduced to a completed calculation of evidence [55]. The present proposal adds a further question: how are the elements that enter such interpretation organized before, during, and after the act of trusting? The answer may involve more than evidence quantity. It can depend on form, sequence, salience, embodiment, and relation-specific history.

Let $\mathcal{C}_t = \{c_{1t}, \dots, c_{nt}\}$ denote the set of trust-relevant cues available or salient at time t . A cue need not be linguistic. It can be an observed action, tone of voice, institutional seal, remembered promise, physical arrangement, delay, bodily response, testimony from another person, or absence where an action was expected. Let H_t denote the relation's history, $\mathbf{x}_t$ its current relational state, and Γ_t a broader context containing relevant institutional, material, and cultural conditions.

A simple additive representation would assign each cue a fixed weight and calculate a trust-relevant quantity by summation. Such models can be useful for some purposes, but they exclude in advance the phenomenon under examination. The same silence can signify comfort in one relation, indifference in another, and danger in a third. A promise can be reassuring when accompanied by costly commitment and suspicious when embedded in a history of strategic reassurance. An institutional logo can stabilize expectation for one person and activate a history of exclusion for another.

The present framework therefore introduces an interpretive-compositional operator

$$m_t = \mathcal{P}(\mathcal{C}_{\leq t}, H_t, \mathbf{x}_t, \Gamma_t), \quad (21)$$

where m_t denotes the trust-relevant meaning generated from the available configuration. The notation does not assume that meaning is one-dimensional. It can be a vector containing, for example, perceived sincerity, danger, familiarity, coherence, recognition, reliability, or motive-related concern. Nor does Equation (21) imply that an individual consciously computes $\mathcal{P}$. The operator is a placeholder for the process through which a relational configuration becomes meaningful enough to affect later action and interpretation.

The relational dynamics from Section 3 can then be extended schematically:

$$\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t, \mathbf{a}_t, \mathbf{o}_t, m_t; \boldsymbol{\theta}_t). \quad (22)$$

The point of the additional term is not to privilege interpretation over material action. It is to represent the fact that actions and observations do not enter relational updating without organization. Their consequences depend partly on what they become within the relation.

Claim 4.3 (Context dependence). For a trust-relevant cue c , no general theory should assume that its relational effect is invariant across histories and configurations. If $H_t^{(1)}$, $\mathbf{x}_t^{(1)}$, and $\Gamma_t^{(1)}$ differ materially from $H_t^{(2)}$, $\mathbf{x}_t^{(2)}$, and $\Gamma_t^{(2)}$, the same observable cue can acquire different meanings and produce different later trajectories.

This claim is intentionally modest. It does not deny regularities. Some cues may have robust average effects across populations or institutions. The claim only rejects the stronger assumption that the cue carries a fixed relational meaning independently of the configuration in which it appears.

One reason poetic composition matters is that the observable size of an event and its relational consequence need not be proportional. A short sentence can reorganize a long conflict. A small breach can dissolve a fragile repair. A minor act can become memorable because it occurs at a point of unusual vulnerability. These cases should not be romanticized as mysterious. Their force can be analyzed through the amount of relational history that the event activates.

Definition 4.4 (Relational condensation). *Relational condensation* occurs when a comparatively small present event acquires substantial trust-relevant significance because it activates, reorganizes, or makes salient a larger distributed structure of memory, expectation, vulnerability, symbolism, or prior interaction.

Condensation gives a more precise interpretation to the intuition that a single gesture can “mean a great deal.” The event need not contain all of that meaning intrinsically. It can function as an access point into a wider relational configuration. A familiar phrase may recall a prior episode of care. A changed routine may make a previously invisible dependency visible. A public defense by a colleague may condense years of uncertainty about whether support will remain when it becomes costly.

This mechanism also explains why outsiders can misjudge trust relations. An observer who sees only the present event may regard the response as excessive or irrational because the observer lacks the history

through which the event is being interpreted. Conversely, insiders can overread a small cue because a strong existing narrative causes them to assimilate ambiguous evidence too easily. Relational condensation is therefore compatible with both insight and error.

A useful analytic distinction is between *event magnitude* and *relational leverage*. Event magnitude concerns the direct observable change produced by an occurrence. Relational leverage concerns how strongly that occurrence changes subsequent interpretation or transition probabilities within a historically formed configuration. A low-magnitude event can have high relational leverage when the system is highly susceptible to the pattern it activates.

Example 4.5 (A remembered act). Suppose two collaborators have worked together for several years. During an earlier crisis, one collaborator quietly accepted a personal cost to protect the other's work without public recognition. Years later, a brief ambiguous message arrives during a new conflict. The literal content of the message is weak evidence about future conduct. Yet its wording recalls the earlier event and activates a history in which costly support has already been observed. The later message can therefore have greater relational leverage than its propositional content alone would suggest. This increased leverage still does not prove that reliance is warranted in the new situation.

The example illustrates why a poetic account is not a theory of irrationality. The historical association may contain genuine information. The analytic point is that the information is carried and reorganized through a configuration whose force is not exhausted by the surface content of the immediate cue.

4.2 Sequence, Rhythm, Embodiment, Atmosphere

Trust is not generated only by what happens. It can also depend on when, how often, in what sequence, and with what interruption events occur. Temporal form therefore belongs to the relational configuration.

Consider response timing. A single delayed reply may have little significance. Repeated delays at moments of vulnerability can form a pattern. A previously reliable rhythm that suddenly breaks can become salient before the reason for the break is known. Conversely, a relation characterized by variable timing may not assign special significance to the same delay. The trust-relevant unit is not necessarily the event in isolation but the event's position within a temporal pattern.

Silence has the same ambiguity. Silence can indicate comfort, respect, withdrawal, concealment, punishment, or simply absence of communicative need. Its significance depends on expectation and history. This makes silence an important example of why a cue need not be a positive observable object. What is missing can become meaningful when the relation contains a sufficiently specific expectation of presence.

Let Δt_i denote intervals between a recurrent class of interactions. The pattern $\{\Delta t_i\}$ can itself become part of H_t . A disturbance in that pattern can then affect m_t even when no explicit proposition changes. The same applies to ordering. Apology followed by disclosure may be interpreted differently from disclosure that occurs only after detection. A guarantee provided before vulnerability is accepted can have a different meaning from an identical guarantee offered after dependence has already been created.

Proposition 4.6 (Temporal non-commutativity). *Where trust-relevant meaning depends on relational history, two otherwise similar events need not have the same effect when their order is reversed. Consequently, the composition of trust-relevant events can be temporally non-commutative.*

Proof. Let c_a and c_b be events whose meanings depend on the state produced by prior events. If processing c_a first changes the relational state or interpretive context in which c_b is encountered, then the meaning assigned to c_b under the sequence (c_a, c_b) can differ from its meaning under (c_b, c_a) . The resulting updates of $\mathbf{x}_t$ therefore need not be identical. Equality requires additional assumptions, such as history-independent cue effects, that the present framework does not impose. $\square$

The proposition is simple but important for empirical design. Studies that collapse relational events into unordered counts can erase a source of causally relevant structure. Ten positive interactions do not necessarily have the same consequence when they occur before rather than after a severe breach. Likewise,

a single disclosure can have different consequences depending on whether it was volunteered, requested, compelled, or discovered.

The compositional account also permits trust formation to include embodied and spatial conditions without reducing trust to atmosphere. Böhme's work is useful here because it treats atmospheres as aesthetically and spatially organized conditions of experience and explicitly connects them to architecture, design, staging, and commodity environments [7]. For the present argument, the important point is more limited: trust judgments and trustful conduct take place somewhere, in bodies, through sensory and material arrangements that can affect attention, exposure, comfort, vigilance, and perceived formality.

A meeting behind a security barrier, a conversation at a shared table, a hospital room, a courtroom, a private home, and an encrypted digital interface do not provide interchangeable experiential conditions. Lighting, distance, access control, seating, surveillance, uniforms, sound, and the visibility of exit can contribute to how vulnerability is felt and interpreted. These features may be instrumentally relevant as well: a locked door changes not only atmosphere but actual exit conditions.

The framework should therefore resist two opposite errors. The first is cognitive reductionism, which treats every trust-relevant contribution as an explicit belief. The second is atmospheric determinism, which treats felt comfort or solemnity as sufficient to generate justified trust. An elegant building, warm tone, familiar scent, or carefully designed interface can make a relation easier to enter without making its commitments more reliable.

Claim 4.7 (Embodied contribution without evidential sufficiency). Embodied, spatial, and atmospheric conditions can participate in the formation of experienced trust while remaining insufficient evidence of the trustworthiness of the person or institution that arranged those conditions.

This separation will become normatively important later. The more effectively a setting can shape experienced safety, the more important it becomes to ask who controls that setting, what information it makes salient, which forms of vulnerability it conceals, and whether participants can revise their judgment outside the staged environment.

4.3 Frames, Salience, Perturbation

A relation is exposed to more information than can be treated as equally important. Trust formation therefore requires selection. Certain events become salient; others remain background. Certain inconsistencies are interpreted as noise; others become diagnostic. Goffman's work on frames is relevant because it emphasizes that experience is organized through structures that affect what participants take an occurrence to be [24]. The present framework extends this insight into a historical-relational question: how do relations themselves help produce the frames through which later evidence is selected and interpreted?

Repeated cooperation can create a generous frame in which ambiguity is resolved charitably. Repeated deception can create a defensive frame in which the same ambiguity is treated as a warning. Institutions also provide frames: professional roles, legal categories, credentials, ceremonies, and procedural scripts can tell participants what kind of interaction is occurring and what counts as normal evidence within it.

The selection problem has an aesthetic dimension in a broad sense. Among many available cues, some appear fitting, discordant, sincere, artificial, coherent, or striking before the participant can fully articulate why. The paper does not assume that such judgments are infallible or reducible to beauty. Their importance lies in their capacity to reorganize attention. A small discordance can suddenly make a previously stable narrative visible as a narrative rather than a fact.

This provides one route by which contingency enters trust formation. Not every available cue becomes consequential. A particular cue can become salient at a particular moment because the current relational configuration is susceptible to it. Another observer, or the same observer at another time, may select a different feature. Such contingency need not imply arbitrariness. It can be conditioned by history, current vulnerability, learned categories, cultural forms, and the relational position of the interpreter.

Definition 4.8 (Relational salience). *Relational salience* is the context-dependent prominence acquired by an event, cue, absence, or pattern within a relational configuration, measured not merely by perceptual notice but by its capacity to reorganize trust-relevant interpretation or action.

Relational salience can itself become self-reinforcing. Once a participant begins attending to signs of reliability, future confirming cues may be noticed more readily. Once suspicion becomes dominant, anomalies may receive greater weight. This feedback can stabilize a trust or distrust trajectory, but it can also produce confirmation bias. The attractor framework therefore gains an interpretive mechanism: a basin can partly persist because the relation changes which evidence is noticed and how ambiguous evidence is classified.

The connection between poetic composition and attractor dynamics becomes clearest near a basin boundary. Suppose a relation is located well inside a stable trust-supporting region. Many small ambiguous events may be absorbed without substantial change. If the relation approaches a boundary because of accumulated strain, however, an event that would previously have been minor can become trajectory-changing.

Let d_t denote a local measure of distance from a relevant basin boundary and let δm_t denote the change in trust-relevant meaning produced by a newly salient composition. The paper does not assume a universal function between these quantities. The qualitative hypothesis is that, other conditions held sufficiently similar, the same δm_t can have greater probability of producing regime change when d_t is small than when the relation lies far inside a stable region.

Definition 4.9 (Poetic perturbation). A *poetic perturbation* is a relationally consequential change whose force arises substantially from the composition, timing, or historical resonance of its elements rather than from the independently measured magnitude of any single element.

The definition does not imply that the perturbation is pleasant or artistic. A discovered inconsistency can be poetic in this technical sense if its force comes from how it reorganizes several previously disconnected memories. A repair gesture can have the same structure when it connects apology, acknowledgment, material restitution, and a remembered norm into a newly coherent configuration.

Claim 4.10 (Boundary-sensitive amplification). Where a relation is near a transition boundary, a small event with high relational leverage can redirect the trajectory even when the event would not have done so in a more stable region. Large effects from small events therefore need not imply mysterious causation; they can arise from the conjunction of compositional meaning and dynamical susceptibility.

This claim links two parts of the paper that could otherwise remain separate. Poetic composition explains how an event acquires relational leverage. Attractor dynamics explains why that leverage can matter differently across regions of the relational landscape. The combination offers a hypothesis for why apparently disproportionate transitions sometimes occur without asserting that every abrupt change has this structure.

4.4 Polysemy, Ambiguity, Over-Reading

Poetic composition increases explanatory reach only if ambiguity is retained. The same cue can support several interpretations, and the relation can select among them incorrectly. A warm gesture can be sincere, strategic, habitual, or misread. A formal ritual can express genuine commitment while also functioning as impression management. A shared hardship can create solidarity among some participants and coercive dependence among others.

The framework therefore distinguishes *experienced coherence* from *epistemic adequacy*. A configuration can feel coherent because many cues fit one another under a particular frame. That coherence can be informative, but it can also be manufactured or selectively constructed. The fact that a story explains many observations does not establish that it is the only or best explanation of those observations.

This problem is especially acute under asymmetric power. A more powerful actor may control which cues are available, which memories are publicly validated, which interpretations appear legitimate, and which spatial or institutional settings frame the interaction. The actor can therefore influence not only the content of evidence but the compositional environment through which evidence becomes meaningful.

Proposition 4.11 (Compositional efficacy does not imply epistemic adequacy). *If a configuration effectively increases experienced trust, it does not follow that the configuration provides sufficient evidence for the trustworthiness of the trusted object.*

Proof. Experienced trust and trustworthiness are distinct variables. A configuration can alter interpretation, affect, salience, or perceived safety while leaving the underlying reliability, motives, competence, or institutional incentives unchanged. Therefore a positive effect on experienced trust does not logically establish a corresponding change in trustworthiness. $\square$

This proposition complements the previous section's separation of dynamical stability from warrant. Together, the two distinctions prevent the paper from treating successful trust production as a normative success.

4.5 Design, Manipulation, Identification

Once trust formation is understood compositionally, deliberate design becomes an unavoidable topic. Organizations routinely arrange interfaces, rooms, uniforms, procedures, narratives, symbols, response scripts, guarantees, and rituals. Such arrangements are not inherently manipulative. Many are necessary for intelligibility, coordination, accessibility, or reassurance. The harder question concerns when the design of trust-producing conditions becomes normatively problematic.

A useful distinction is between designs that make relevant features of a relation more legible and designs that create confidence while obscuring those features. A visible audit trail can increase trust by making accountability easier to inspect. A ceremonial display can increase trust while providing no additional accountability. A calm interface can help users understand a risky process, or it can reduce vigilance toward risks that remain unchanged. The surface technique does not determine the ethical category; the relation among experience, evidence, power, and available revision does.

Böhme's discussion of staging and aesthetic economy is useful because it shows that atmospheres can be deliberately produced within architecture, commerce, and public life [7]. In a trust framework, this suggests a research programme rather than an accusation. Which components of a designed environment affect experienced trust? Which of those components also track actual trustworthiness? Which make counterevidence harder to notice? Which preserve meaningful exit and independent verification?

Research Obligation 4.12 (Separate production from warrant). Empirical studies of designed trust environments should measure, where possible, both their effects on experienced trust and their relation to independent indicators of trustworthiness, accountability, contestability, and exit. An intervention that raises confidence alone is normatively underdetermined.

The threshold of manipulation cannot be defined by intentionality alone. Unintended environments can still distort judgment, while intentional reassurance can be legitimate. Later sections will therefore need a richer account involving disclosure, vulnerability, asymmetry, reversibility, and the capacity of participants to inspect or contest the trust-generating conditions.

The language of poetic structure risks becoming unfalsifiable unless translated into observable questions. Several strategies can make the proposal more empirically disciplined.

First, researchers can compare the same cue across different relational histories. If a cue's effect changes systematically with prior betrayal, cooperation, institutional identity, or vulnerability, this supports a context-dependent account. Second, sequence can be manipulated or observed: identical informational components can be presented or encountered in different orders to test whether temporal composition changes trust-relevant responses. Third, multimodal environments can be decomposed. Researchers can vary textual guarantees, spatial arrangements, visual symbols, interaction rhythm, and verification opportunities separately and jointly to estimate interaction effects. Fourth, longitudinal designs can examine whether certain cues acquire or lose salience as a relation develops. Fifth, perturbation studies can test whether highly condensed events are especially consequential near periods of relational instability.

Qualitative methods remain essential because the relevant meaning cannot always be specified in advance. Process tracing, diaries, interviews, conversation analysis, ethnography, archival reconstruction,

and multimodal analysis can help identify which events participants experienced as trajectory-changing and how those events became connected to prior history. These methods should not simply accept retrospective narratives at face value. They can compare narratives with contemporaneous records, behavioral changes, and the accounts of differently positioned participants.

Formal modeling can investigate candidate mechanisms without replacing the interpretive work. One model can assign context-dependent weights to cues. Another can allow interactions among cues to change with history. A third can couple attentional selection to the relational state, allowing trust to change which evidence becomes salient. The appropriate formalism will depend on the empirical problem.

Research Obligation 4.13 (Compositional evidence). A strong empirical claim about poetic relational composition should identify at least: (i) the heterogeneous elements involved; (ii) the historical or contextual conditions under which they acquire significance; (iii) evidence that their joint effect differs from a fixed additive interpretation; and (iv) a consequential change in trust-relevant interpretation, action, or trajectory.

This obligation protects the concept from decorative use. It is not enough to describe an event vividly and call it poetic. The analysis must show what was composed, how context changed its significance, and what difference that composition made.

4.6 Governance, Limits, Implications

The section can now return to the paper's broader argument. Trust-supporting relations are not produced only by explicit reasons or isolated acts. They can be sustained by histories, rhythms, symbols, absences, embodied experience, spaces, institutional frames, and contingent moments of salience. These elements can form configurations whose significance is generated through composition. When such configurations feed back into later attention, action, and evidence production, they can contribute to the formation and maintenance of trust attractors.

Three separations remain indispensable. Poetic composition is not identical to poetry. Experienced coherence is not identical to truth. Trust production is not identical to trustworthiness. These distinctions allow the concept to do analytic work without becoming romantic or normatively permissive.

The resulting governance problem is sharper than a generic question about how to increase trust. If trust can be shaped through compositional environments, then governance must ask who can arrange those environments, whose memories and symbols they privilege, what forms of attention they cultivate, which vulnerabilities they expose or conceal, and whether participants can reinterpret or leave the relation. A system can become highly effective at generating trust while becoming less trustworthy. Conversely, carefully designed procedures can create conditions in which warranted trust becomes easier to form because relevant commitments, evidence, and avenues of contestation are made more legible.

The next stage of the paper therefore turns from formation to mechanism. It will examine how trust-supporting relations reproduce themselves through cooperation, disclosure, interpretation, evidence production, institutional assurance, and other feedback processes. Poetic composition supplies one route through which events become meaningful; self-reinforcement explains how those meanings can become part of the conditions that shape what happens next.

5 Disclosure, Reciprocity, Reinforcement

Sections 3 and 4 established two pieces of the argument: relational trajectories can exhibit feedback and history dependence, and heterogeneous events can acquire trust-relevant significance through composition. The present section asks the causal question that connects them: through which mechanisms can already formed trust alter the conditions under which later trust is generated?

The phrase *self-reinforcing trust* is otherwise too easy to use circularly. Temporal persistence can arise from stable external conditions, selection into durable relationships, repeated measurement of a latent disposition, institutional constraint, or the absence of perturbation. A claim of relational self-reinforcement

therefore requires an intervening pathway: present trust must change action, exposure, interpretation, institutional arrangement, or the field of possible interaction, and that change must affect subsequent trust-relevant conditions.

Existing work already supplies much of the substantive basis for this architecture. Mayer, Davis, and Schoorman distinguish trust from risk taking; Serva, Fuller, and Mayer identify reciprocal longitudinal relations among trust, perceived trustworthiness, and risk taking; Ferrin, Bligh, and Kohles examine spirals involving cooperation; Ostrom links trust, reciprocity, and reputation in collective action; Huang and Wilkinson place feedback inside a complex-adaptive account of business relationships; and Möllering and Sydow show how self-reinforcement can lock organizations into a trust-based relation [50, 68, 21, 60, 32, 57]. The purpose of this section is therefore taxonomic and diagnostic rather than priority-claiming. It separates distinct feedback channels, identifies the circumstances in which they oppose one another, and distinguishes self-reinforcement that remains evidence-sensitive from reinforcement that narrows correction, alternatives, or observation.

5.1 Mechanisms, Vulnerability, Disclosure

Let $\mathbf{x}_t$ denote the relational configuration introduced in Section 3. Suppose that one component or projection of $\mathbf{x}_t$ represents a trust-supporting orientation, denoted here by τ_t . The notation does not reduce trust to a single universal scalar. It is used only to make a local causal structure visible.

A minimal recursive pathway can be written schematically as

$$m_t = M_k(\tau_t, \mathbf{x}_t, H_t, \Gamma_t), \quad (23)$$

$$\tau_{t+1} = F_k(\tau_t, m_t, \mathbf{x}_t, H_t, \Gamma_t), \quad (24)$$

where M_k denotes a mechanism-specific mediator and k indexes a channel such as disclosure, cooperation, interpretation, affect, monitoring, or institutional assurance. A local self-reinforcing effect exists only when a change in present trust systematically changes the mediator and the induced change in the mediator tends, under the relevant conditions, to support later trust.

For differentiable toy models, a local channel gain can be represented by

$$g_{k,t} = \frac{\partial \tau_{t+1}}{\partial m_t} \frac{\partial m_t}{\partial \tau_t}. \quad (25)$$

A positive $g_{k,t}$ indicates local reinforcement through channel k under that specification. A negative value indicates a locally counteracting channel. This expression is illustrative rather than a claim that empirical trust relations are smooth, one-dimensional, or easily differentiable. In many cases, the relevant mechanism will be discontinuous, threshold-dependent, historically indexed, or better represented qualitatively.

The important conceptual move is separation. Trust at time t and the mechanism through which it affects time $t + 1$ are distinct objects. A person can trust more and therefore disclose more; disclosure can then reveal either trustworthiness or betrayal. A team can trust more and therefore monitor less; reduced monitoring can support autonomy and performance, or it can permit hidden opportunism. A citizen can trust an institution and therefore interpret an ambiguous failure more charitably; that interpretive generosity can protect a warranted relationship from noise, or it can delay recognition of systemic abuse. Self-reinforcement is therefore a structure of recursion, not a claim that the recursion is always beneficial or monotonic.

Definition 5.1 (Relational reinforcement channel). A *relational reinforcement channel* is a mediated pathway through which an existing trust-supporting relational condition changes subsequent action, exposure, interpretation, affect, institution, or interaction opportunity in a way that modifies the production or evaluation of later trust-relevant information.

Remark 5.2 (Persistence is insufficient). Observed persistence of trust does not by itself establish a reinforcement channel. Evidence for reinforcement requires some indication that trust changes an intervening process and that the intervening process contributes to later relational updating.

This distinction also limits the paper’s use of causal language. The following subsections identify candidate mechanisms and empirically supported patterns, but the existence, sign, and strength of a channel must be established within a specified population, relation, and time scale.

Trust changes what can become knowable. This is one of the most important routes by which a relation can begin to reproduce its own conditions. When a participant accepts vulnerability, shares information, delegates a task, asks for help, reveals uncertainty, or permits another person to act without complete control, the relation generates observations that could not have been produced under total withdrawal.

This point is already implicit in influential definitions of trust that place vulnerability and risk at the center [50, 64]. The present framework emphasizes the temporal consequence. Vulnerability is not only something that trust permits after judgment. It is also an action that changes the future evidential environment.

Consider a relation in which actor A is uncertain whether actor B will use sensitive information opportunistically. If A never discloses anything that could be misused, evidence concerning B ’s restraint cannot arise through that channel. Once some trust is sufficient for limited disclosure, B gains an opportunity to protect, exploit, ignore, or responsibly use the information. The observed response then enters later trust formation. Trust has helped produce the test through which its own future can be revised.

This mechanism can be represented by distinguishing exposure v_t from the resulting observation o_{t+1} :

$$v_t = V(\mathbf{x}_t, \Gamma_t), \quad (26)$$

$$o_{t+1} = O(v_t, \mathbf{b}_t, \Gamma_t, \varepsilon_t), \quad (27)$$

where $\mathbf{b}_t$ denotes the counterpart’s behavior and ε_t collects unobserved disturbances. The next relational state depends partly on o_{t+1} . A more trustful state can therefore increase the range of exposures, which increases the range of possible evidence.

Claim 5.3 (Endogenous evidence production). In many trust relations, the evidence available for later trust assessment is endogenous to earlier trust-related action. Consequently, two actors with identical initial information but different willingness to accept limited vulnerability can generate different future information sets.

This claim has two consequences. First, trust can become self-confirming when exposure elicits trustworthy conduct that would not otherwise have been observable. The trusted party can experience recognition, autonomy, or an opportunity for reciprocity and respond in ways that strengthen the relation. Second, trust can become self-correcting when exposure reveals unreliability. A recursive system can therefore contain reinforcement without becoming closed. Indeed, the possibility of disconfirmation is one feature that can distinguish resilient trust from mere insulation against evidence.

The quantity and severity of vulnerability matter. A relation that requires a large irreversible exposure before any informative response becomes visible is normatively different from one that permits gradual, reversible, and proportionate testing. This difference will become important in later sections on governance. Good trust environments need not eliminate vulnerability, but they can shape how evidence is generated through it.

5.2 Cooperation, Reciprocity, Interpretation

A second channel links trust to cooperation and cooperation to later trust. The link should be stated carefully because trust and cooperation are not synonyms. Cooperation can be sustained through contracts, deterrence, reputation, repeated-game incentives, institutional constraint, or aligned interests even when participants do not trust one another. Cook, Hardin, and Levi explicitly emphasize mechanisms through which cooperation can occur without trust [13]. The distinction is analytically valuable precisely because cooperation can then occupy an intervening position in the model rather than being folded into the definition of trust.

When trust does increase willingness to cooperate, however, cooperation creates new relational events. Participants coordinate, depend on one another, incur costs, share credit, encounter conflict, and observe

whether contributions are reciprocated. These events can alter perceived trustworthiness and expectations concerning future interaction. Serva, Fuller, and Mayer’s longitudinal study provides a particularly direct example of reciprocal trust formation across interacting teams [68]. Ferrin, Bligh, and Kohles find support for spiral models linking perceived trustworthiness and cooperation over time [21]. Ostrom’s behavioral account of collective action likewise places trust, reciprocity, and reputation in interaction rather than isolating each into a separate disposition [60].

A minimal dyadic representation can distinguish the two actors:

$$c_t^A = C^A(\mathbf{x}_t^A, \mathbf{x}_t^B, \Gamma_t), \quad (28)$$

$$c_t^B = C^B(\mathbf{x}_t^B, c_t^A, \Gamma_t), \quad (29)$$

$$\mathbf{x}_{t+1}^A = \Phi^A(\mathbf{x}_t^A, c_t^B, H_t, \Gamma_t), \quad (30)$$

with an analogous update for actor B . Mutual reinforcement becomes possible when a trust-supported cooperative move by one participant changes the other’s behavior and the resulting response enters the first participant’s later updating. The relation itself becomes a site of repeated production rather than an exchange of fixed traits.

This framework also clarifies why initial moves can matter disproportionately. A low-trust relation may contain few opportunities for costly cooperation and therefore few opportunities to generate counterevidence to pessimistic expectations. A small credible cooperative act can open a new interaction sequence. Conversely, an early exploitative response can close the sequence before richer evidence develops. These effects resemble path dependence but remain mechanism-specific: the initial event matters because it changes later opportunities, not merely because it happened first.

Proposition 5.4 (Reciprocal amplification under conditional cooperation). *Suppose each participant’s willingness to undertake a trust-relevant cooperative action increases with a sufficiently favorable relational state, and each participant updates favorably after observing costly, voluntary, and appropriately attributed cooperation from the other. Under these conditions, small favorable differences in an initial relational state can be amplified through repeated interaction. The proposition does not imply convergence to high trust; noise, asymmetry, exploitation, saturation, and adverse shocks can interrupt or reverse the process.*

The qualification concerning attribution is essential. A helpful action that is believed to have been coerced may provide little evidence of goodwill. A costly action performed under a transparent external incentive can support reliability expectations without supporting benevolence. The same behavior can therefore feed different dimensions of the relational state.

Trust influences not only what actors do but also how later events are read. This creates an interpretive feedback channel. Every durable relation contains ambiguous observations: delayed responses, incomplete explanations, accidental failures, conflicting testimony, changes of mood, missing information, and situations in which motives are not directly observable. A prior relational history helps determine which interpretations remain plausible and which receive attention.

The possibility of such interpretation is not new. Møllering’s account of trust emphasizes expectation, interpretation, and suspension [55, 56]. The present concern is the recursive consequence. Once a trust-supporting configuration exists, it can alter the interpretation of later ambiguity. A benign interpretation may preserve cooperation long enough for clarifying evidence to arrive. That continued interaction can then generate additional evidence supportive of the relation. In this case, interpretive generosity buffers the relation against noise.

The same mechanism can create epistemic closure. If every counterexample is reinterpreted until it becomes compatible with the existing trust narrative, then trust no longer merely stabilizes interpretation; it suppresses the conditions of meaningful revision. The distinction cannot be drawn from positive affect alone. It depends on whether the interpretive system retains sensitivity to diagnostic evidence.

Let $q_t(h)$ denote the weight assigned at time t to a hypothesis h concerning the counterpart or institution. Let o_t be an ambiguous observation. A trust-supporting history can influence the interpretive mapping

$$q_{t+1}(h) = \mathcal{I}(q_t(h), o_t, H_t, \mathbf{x}_t). \quad (31)$$

The analytical problem is not that prior history enters Equation (31); any coherent learning process uses prior information. The problem arises when the relational state changes the effective weight of evidence in a way that becomes systematically insensitive to potential falsification.

Definition 5.5 (Interpretive shielding). *Interpretive shielding* occurs when an established relational configuration repeatedly transforms potentially adverse evidence into interpretations compatible with continued trust, thereby reducing the probability that counterevidence produces substantial relational revision.

Interpretive shielding is not always irrational. Reliable relations should not collapse after every anomaly. A single delayed message should carry little weight after years of dependable conduct. Robustness to noise is part of what stability means. The normative and epistemic concern begins when shielding extends beyond reasonable noise tolerance and protects the relation from independent checking, alternative accounts, or accumulating contradictory patterns.

This creates a useful distinction between *robust trust* and *closed trust*. Robust trust survives weak or noisy perturbations while remaining revisable under sufficiently diagnostic evidence. Closed trust preserves itself partly by disabling the routes through which diagnostic evidence could acquire force.

5.3 Affect, Monitoring, Autonomy

Trust relations also contain affective and embodied components. McAllister's work distinguishes cognition-based and affect-based foundations of interpersonal trust in organizations [52], while Jones emphasizes the affective character of trust and its connection to optimism about another's goodwill [34]. These traditions support a mechanism that cannot be reduced to explicit evidence aggregation.

A sufficiently trustful relation can lower defensive arousal, reduce the need to continuously inspect for threat, and make open communication easier. Participants may speak with less strategic filtering, admit ignorance, ask for help earlier, or tolerate minor coordination failures without escalation. The resulting interaction can itself become more cooperative and informative. The felt quality of the relation therefore changes the behavior from which later relational evidence is generated.

The reverse process is equally important. Distrust can produce guarded speech, withholding, redundant verification, hostile attribution, and avoidance. These behaviors may then make interaction colder or less efficient, eliciting reciprocal defensiveness. A relation can therefore deteriorate even when the initial suspicion was only weakly grounded. This possibility will be developed more fully in the discussion of distrust and rupture.

Affective regulation also helps connect the present section to poetic composition. Atmosphere, rhythm, bodily familiarity, tone, and spatial setting can contribute to felt safety. Once felt safety changes disclosure or interaction quality, a compositional event becomes part of a causal feedback loop. The poetic structure described in Section 4 therefore need not remain at the level of interpretation. It can alter later behavior and evidence production through affective mediation.

Remark 5.6 (Felt safety and objective safety). Felt safety is a relationally consequential state, but it is not an independent measure of objective safety. Environments can make participants feel secure while preserving substantial exploitation risk. The causal efficacy of an affective state must therefore be separated from the warrant for that state.

This separation is central to the paper's normative argument. A trust mechanism can be psychologically effective and socially stabilizing while remaining ethically problematic.

Trust changes the allocation of attention. Participants who expect constant opportunism must devote time and cognitive resources to checking, documenting, hedging, verifying, and controlling. Trust can reduce some of these costs and allow attention to move toward joint work, exploration, or long-horizon coordination. McEvily, Perrone, and Zaheer examine trust within organizational coordination and discuss its mobilizing and structuring effects [53]. Dirks and Ferrin review evidence concerning the role of trust in organizational settings [14], and later meta-analytic work on trust in leadership reports systematic relationships with organizational outcomes [15]. Mayer and Gavin find that trust in management is related to employees' capacity to focus attention on value-producing activities [51].

Within the present framework, reduced monitoring can create another recursive channel. When monitoring decreases, the trusted party may receive greater autonomy. Autonomy can permit competence, initiative, or goodwill to become visible in ways that heavy surveillance would suppress. Positive performance then supports later trust. The relation has changed the conditions under which trustworthiness can be expressed and observed.

Yet monitoring reduction is one of the clearest examples of a double-edged mechanism. Less surveillance can also reduce detection of opportunism. A self-reinforcing trust relation may therefore accumulate hidden vulnerability. Successful interactions lower monitoring, lower monitoring decreases detection probability, and the absence of detected problems is then mistakenly taken to provide additional evidence of trustworthiness. Under those conditions, the loop can generate confidence faster than it generates knowledge.

Definition 5.7 (Detection-thinning loop). A *detection-thinning loop* occurs when increased trust reduces monitoring or verification, the reduction lowers the probability of observing adverse conduct, and the resulting scarcity of adverse observations contributes to further trust.

The definition identifies a mechanism rather than condemning all reductions in monitoring. Constant surveillance can itself destroy cooperation and autonomy. The governance question concerns calibration: which safeguards can preserve meaningful detectability without converting every relation into continuous suspicion?

One possibility is to distinguish intrusive monitoring from latent accountability. A relation can permit broad autonomy while preserving audit trails, independent complaint channels, periodic review, or triggered verification. Such arrangements reduce day-to-day surveillance without making counterevidence structurally inaccessible. This distinction later becomes important for institutional design.

5.4 Assurance, Shared Experience, Sedimentation

Trust rarely develops in a dyadic vacuum. Legal rules, professional norms, organizational roles, accreditation systems, reputations, warranties, insurance, escrow, third-party oversight, and community sanctions can alter the consequences of vulnerability. Zucker's account of institutional trust production and later work on institution-based trust demonstrate the importance of such arrangements [81, 4].

These arrangements can support trust through at least three distinct routes. First, they can constrain harmful behavior, making reliance safer even when the counterpart's motives remain uncertain. Second, they can provide information about competence, identity, or prior conduct. Third, they can provide recourse if a harmful event occurs. These routes should not be collapsed. A guarantee can make an interaction acceptable without increasing belief in benevolence; a reputation system can increase perceived competence without providing compensation; an enforceable contract can lower downside risk even when interpersonal trust remains weak.

The relational consequence can nevertheless become recursive. Institutional assurance makes an initial interaction possible. The interaction generates shared experience. Shared experience can then support a more direct relation that becomes less dependent on the original guarantee. In this sense, external security can scaffold endogenous relational development.

A schematic sequence is

$$G_t \ ; \ E_t \ ; \ H_{t+1} \ ; \ \mathbf{x}_{t+1}, \quad (32)$$

where G_t denotes a guarantee or assurance, E_t an interaction made possible under that assurance, and H_{t+1} the enlarged shared history. The semicolon deliberately indicates temporal and causal ordering without claiming logical equivalence among the terms.

This mechanism resembles collateral in a broad functional sense: a background arrangement can reduce the perceived cost of testing a relation. The analogy must remain limited because interpersonal, organizational, and political trust cannot generally be reduced to financial security. What matters is the wider principle that an external structure can change the feasible set of vulnerabilities and thereby change which relational evidence can be produced.

Institutional assurance can also generate misplaced transfer. Participants may infer trustworthiness from symbols of certification, official status, legal form, professional dress, institutional affiliation, or technological polish without examining what the relevant assurance actually covers. The poetic and institutional channels then couple: a symbol that originally indexed a real safeguard can acquire a broader aura of reliability than the safeguard warrants.

Research Obligation 5.8 (Specify the object of assurance). Analyses of institutionally supported trust should identify what a guarantee actually secures: behavior, compensation, identity, competence, procedural review, reputation, or some other property. The scope of experienced trust should not be inferred to match the scope of the institutional protection that helped produce it.

Repeated interaction creates more than a larger sample of observations. It can produce a shared history with its own internal references, expectations, routines, remembered crises, jokes, rituals, repaired failures, and jointly recognized turning points. Such history changes both what participants know and how later events become intelligible.

This process can be called *historical sedimentation*. Earlier events do not remain equally active at every moment, but they can become incorporated into routines, expectations, role definitions, and background assumptions. Later trust therefore depends not only on recalled evidence but also on a relational environment partly constructed from previous interactions.

Definition 5.9 (Historical sedimentation). *Historical sedimentation* is the process through which repeated interaction leaves relatively durable traces in expectations, routines, interpretive frames, institutional arrangements, and shared memory, thereby changing the conditions under which subsequent relational events are produced and understood.

Historical sedimentation connects first-order state dynamics with the second-order landscape deformation introduced in Section 3. If a repaired conflict establishes a new norm for disclosure, the next conflict occurs under a different update rule. If repeated successful delegation establishes a routine of autonomy, the relation no longer requires the same evidential threshold for every task. If a history of concealed harm teaches participants that formal apologies are poor signals, later apologies acquire less relational leverage. The trajectory has modified the landscape.

Shared experience can thereby deepen trust, but it can also create closure. Long history can increase mutual intelligibility while making exit costly. Participants may share dependencies, reputations, social networks, property, professional identities, or narratives that make reinterpretation difficult. A relation can become stable partly because too many surrounding relations now depend on its continuation. Such stability should not be mistaken for freely renewed trust.

This point also clarifies why history cannot be represented by a simple count of successful interactions. Two relations with the same number of cooperative outcomes can differ radically if one contains repair after serious conflict, another contains untested routine, and a third contains repeated silence about known harms. The generative structure of the history matters.

5.5 Coupled Channels, Saturation, Reversal

The mechanisms described so far rarely operate alone. Trust can increase disclosure; disclosure can improve mutual understanding; improved understanding can support cooperation; cooperation can create shared success; shared success can change affect; affect can reduce defensive monitoring; reduced monitoring can increase autonomy; autonomy can make competence more visible; successful autonomy can strengthen institutional willingness to delegate. A relation can therefore contain a network of reinforcing channels rather than a single loop.

Let $\mathbf{m}_t = (m_{1t}, \dots, m_{Kt})$ denote a vector of mechanism states. A schematic coupled system is

$$\mathbf{m}_{t+1} = \mathcal{M}(\mathbf{m}_t, \mathbf{x}_t, H_t, \Gamma_t), \quad (33)$$

$$\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t, \mathbf{m}_{t+1}, H_t, \Gamma_t). \quad (34)$$

The matrix of local cross-effects among the components of $\mathbf{m}_t$ can, in a simplified differentiable model, reveal whether mechanisms amplify or counteract one another. The more important conceptual point is that relational reinforcement has an architecture. A highly stable trust configuration may owe its persistence to several partially redundant channels. Removing one channel may therefore have little effect, while simultaneous disruption of several channels can produce abrupt change.

This architecture helps explain resilience. Suppose institutional assurance is weakened but a long shared history, strong reciprocal cooperation, and independent verification remain. Trust may persist because other channels continue to support the relation. Conversely, a relation apparently sustained by many cues may turn out to depend on a single hidden condition, such as one broker, one source of income, one reputational gatekeeper, or one institutional guarantee. The apparent density of the relation can conceal structural fragility.

Definition 5.10 (Reinforcement architecture). A *reinforcement architecture* is the configuration of mutually coupled mechanisms through which present relational conditions alter the production, interpretation, and consequences of future trust-relevant events.

The architecture can be decentralized or highly concentrated. In a decentralized configuration, different sources of evidence and support can partially substitute for one another. In a concentrated configuration, several visible trust cues may ultimately depend on the same actor or resource. This difference is crucial for both robustness and governance.

A useful empirical question is therefore not merely whether trust is high, but how many independent pathways sustain it. Apparent multiplicity can be tested by perturbation: do different channels fail independently, or do they collapse together when one underlying condition changes?

Self-reinforcement does not imply indefinite growth. Most relational mechanisms should be expected to saturate. Once disclosure is already extensive, further trust may produce little additional information. Once cooperation is routine, another successful interaction may add little. Once an institution is granted high reliability, additional symbols of assurance may have diminishing marginal effect.

Nonlinearity also creates reversal. Greater trust can increase exposure until vulnerability becomes too concentrated. Reduced monitoring can improve performance up to a point and then lower detection too far. Interpretive generosity can preserve a relation through ordinary noise and later become a barrier to recognizing a persistent pattern. Strong identification can support cooperation and then make dissent socially costly. Mechanisms with positive local gain can therefore become negative beyond a threshold.

A simple saturating toy relation might take the form

$$\tau_{t+1} = \tau_t + \alpha r_t(1 - \tau_t) - \beta d_t \tau_t, \quad (35)$$

where r_t collects reinforcing events and d_t disruptive events after normalizing τ_t to $[0, 1]$. The factors $(1 - \tau_t)$ and τ_t produce state-dependent sensitivity. Equation (35) is not a proposed general model of trust. It merely demonstrates why reinforcement need not imply explosive growth and why the same event can have different effects at different relational states.

High trust can also increase the impact of betrayal. A harmful act performed by a distrusted party may confirm an existing expectation. The same act performed by a deeply trusted party can reorganize memories, reinterpret prior events, and cast doubt on the mechanisms through which trust had been produced. The perturbation can therefore affect both the current state and the inferred history. In the language of Section 3, the event may deform the landscape rather than simply move the system within an unchanged one.

Remark 5.11 (Positive feedback is local). Descriptions of trust reinforcement should specify the region, time scale, and mechanism for which positive feedback is expected. A channel can be reinforcing at low or moderate trust, neutral near saturation, and destabilizing under extreme dependency or concentrated vulnerability.

This qualification prevents the concept of a trust attractor from becoming a claim that trust inevitably accumulates once it starts.

5.6 Relational Reality, Trustworthiness, Identification

The preceding mechanisms allow a stronger philosophical claim, but it must be formulated carefully. Trust can participate in producing a relational reality that later appears to confirm the trust from which it partly emerged. If I trust, I may disclose. Disclosure may elicit care. Experienced care can deepen trust. Deeper trust can increase cooperation, shared history, and mutual interpretability. After enough iterations, the relation contains real features that did not exist at the beginning: routines of assistance, shared knowledge, mutual expectations, institutional commitments, memories, and dependencies. Trust has not merely represented a pre-existing social object. It has helped participate in the generation of the relation being evaluated.

This does not mean that trust creates truth by belief alone. The counterpart retains agency; material constraints remain; harmful conduct can occur; external facts do not disappear. The claim concerns relational generativity. Some properties of a relationship are partly endogenous to how participants enter and sustain it.

Claim 5.12 (Generative reinforcement). Where trust changes interaction in ways that create durable cooperative, interpretive, affective, or institutional structures, later evidence about the relation can partly reflect structures generated through earlier trust-related action rather than properties that existed independently before the relation.

This is the point at which relational reality becomes epistemically delicate. A self-reinforcing process can generate genuine cooperation and genuine mutual knowledge. It can also generate a closed evidential environment. If trust selects interaction partners, suppresses monitoring, filters testimony, privileges reassuring interpretations, and makes exit costly, the relation may produce a coherent stream of trust-compatible experience while excluding the conditions under which contrary information would arise.

Accordingly, internal coherence is insufficient evidence of trustworthiness. A mature trust relation can be internally rich, emotionally compelling, historically dense, and dynamically stable while resting on selective access to evidence. This is not a rare logical possibility confined to interpersonal life. Organizations, professions, political institutions, religious communities, platforms, markets, and expert systems can all contain mechanisms that shape who interacts, what becomes visible, whose testimony is credited, and which failures are assigned to an exceptional category.

The normative danger is not self-reinforcement itself. Without some recursive stability, cooperation would remain fragile and every anomaly would reset a relation to zero. The danger lies in reinforcement architectures that make confidence easier to reproduce than warranted revision.

Definition 5.13 (Epistemically open reinforcement). A reinforcement architecture is *epistemically open* when its stabilizing mechanisms preserve meaningful routes through which adverse evidence, independent testimony, verification, contestation, and revision can influence the relational state.

Definition 5.14 (Epistemically closed reinforcement). A reinforcement architecture is *epistemically closed* when the same mechanisms that stabilize trust substantially disable, discredit, or prevent the production of evidence capable of revising that trust.

These definitions are deliberately relational. Openness cannot be measured only by the formal existence of information. Evidence may be technically available but inaccessible to a vulnerable participant, socially costly to raise, institutionally ignored, or rendered unintelligible by the dominant interpretive frame.

Three separations govern the interpretation of these mechanisms. Trust remains distinct from trustworthiness; cooperation remains a possible mediator rather than a synonym for trust [13]; and dynamical stability remains distinct from both epistemic warrant and justice. A reinforcement architecture can therefore be effective at reproducing confidence while still being poorly calibrated, unfairly organized, or dependent on constrained exit.

Proposition 5.15 (No normative inference from reinforcement alone). *The existence of a positive relational reinforcement channel does not by itself support an inference to trustworthiness, epistemic warrant, welfare, legitimacy, or justice. Each requires additional criteria and evidence.*

The empirical task is correspondingly mechanism-specific. Longitudinal, experimental, process-tracing, and network designs can each contribute, but the central requirement is to identify the mediator through which present trust changes later trust-relevant conditions. Section 9 develops that research programme in detail.

Research Obligation 5.16 (Mechanism identification). A strong empirical claim of relational self-reinforcement should identify: (i) the initial trust-relevant condition; (ii) the mediator changed by that condition; (iii) the later trust-relevant outcome; (iv) the temporal ordering; (v) plausible alternative explanations; and, where possible, (vi) conditions under which the channel weakens, reverses, or fails.

The final item matters because a theory that predicts reinforcement everywhere is not informative. Boundary conditions are part of the mechanism.

6 Distrust, Rupture, Repair

Recursive dynamics do not belong only to trust. Suspicion, withdrawal, defensive monitoring, restricted vulnerability, and hostile interpretation can also reproduce the conditions under which they later appear reasonable. Such sequences need not be described simply as the disappearance of trust; distrust can coexist with trust across domains and can develop its own reinforcement, historical sedimentation, and basin-like persistence.

This section therefore separates distrust from low trust, examines candidate distrust-reinforcement channels, and then turns to rupture and repair. A serious violation can alter both the current relation and the rules through which later evidence is interpreted. Repair consequently cannot be understood as simple reversal. Work on causal attribution, apology, trust repair, and systemic organizational recovery already demonstrates the importance of differentiated mechanisms and levels of analysis [36, 75, 16, 23]. The relational question developed here is how acknowledgment, explanation, accountability, restitution, structural change, and renewed testing reorganize the history through which future trust and distrust become possible.

6.1 Distrust, Low Trust, Reinforcement

A one-dimensional representation of trust is tempting because it offers a simple picture: high trust at one end, high distrust at the other, and uncertainty in between. That representation is sometimes useful for a local measurement problem, but it is too restrictive for the present framework. A person may trust another’s technical competence while distrusting the person’s motives. A citizen may expect an institution to execute routine procedures reliably while suspecting that it conceals politically inconvenient failures. A research team may trust the accuracy of a collaborator’s calculations while remaining cautious about authorship, credit, or confidentiality. These are not necessarily unstable contradictions. They can be durable multidimensional relations.

Accordingly, let a local relational representation contain at least two partially independent coordinates,

$$\mathbf{r}_t = (\tau_t, \delta_t, \mathbf{z}_t), \quad (36)$$

where τ_t denotes a trust-supporting orientation, δ_t denotes a distrust-supporting orientation, and $\mathbf{z}_t$ collects other relevant relational dimensions. The notation explicitly avoids the restriction

$$\delta_t = 1 - \tau_t. \quad (37)$$

The two coordinates may be negatively related in a particular empirical setting, but that relation should be estimated or argued rather than imposed by definition.

Definition 6.1 (Relational distrust). *Relational distrust* denotes a historically situated orientation in which an actor treats another actor, institution, or relational arrangement as a meaningful source of possible unreliability, opportunism, concealment, harm, or norm violation and organizes interpretation or conduct partly around that possibility.

This definition is deliberately broader than an explicit belief that another party is malicious. Distrust may concern competence, integrity, benevolence, predictability, procedural fairness, information quality, or the design of the relation itself. It can also be selective. One may distrust a channel without distrusting the entire institution, distrust a domain without distrusting all domains, or distrust a present arrangement while remaining open to future repair.

The distinction between distrust and low trust has an important dynamical consequence. Low trust may produce hesitation without generating strong self-reinforcing action. Active distrust can instead reorganize behavior. It may increase verification, reduce disclosure, narrow delegation, change the interpretation of ambiguity, motivate coalition formation, trigger defensive documentation, or encourage exit. Those changes alter the subsequent relational environment. Distrust can therefore have causal productivity of its own.

Remark 6.2 (Coexistence). Trust and distrust can coexist within the same relation when they attach to different dimensions, domains, time horizons, actors, or mechanisms. A model that forces one to eliminate the other may erase precisely the structure that explains relational stability or change.

This coexistence also helps explain why repair can be partial. A repaired competence judgment need not repair an integrity judgment. Renewed procedural confidence need not restore affective safety. A relation can therefore improve without returning to an undifferentiated state of “high trust.”

Distrust becomes dynamically important when it changes conditions that later produce or confirm further distrust. Several mechanisms mirror those discussed for trust, but the symmetry is incomplete.

First, distrust can reduce vulnerability and disclosure. That can be protective, but it also reduces opportunities for the distrusted party to demonstrate restraint, benevolence, or reliability under meaningful exposure. The evidential environment becomes thinner. A person who expects betrayal may therefore avoid the very interactions through which a contrary history could be generated.

Second, distrust can increase defensive behavior. Monitoring, documentation, contractualization, restricted access, duplicated work, anticipatory self-protection, and strategic withholding can all be reasonable responses to risk. Yet the counterpart may experience rejection or hostility in those actions and respond with reduced cooperation. The resulting poorer interaction can then be taken to support the original distrust.

Third, distrust can alter the treatment of ambiguity. Delayed communication may suggest concealment to a distrusting actor; a mistake may suggest sabotage; silence may suggest strategic avoidance. Such interpretations can be correct, incorrect, or underdetermined. The relevant dynamical point is that an established history of distrust changes the prior relational context within which later events acquire significance.

Fourth, distrust can transform institutional arrangements. Repeated failures may generate new audit systems, reporting requirements, access controls, separation of duties, sanctions, and external oversight. These changes can reduce genuine risk and support later repair. They can also entrench a relation of permanent suspicion. Institutionalized distrust therefore has a double possibility: it can create safer conditions for renewed trust, or it can become a self-maintaining governance architecture whose continued existence produces additional signals of suspicion.

A simplified recursive pathway can be written

$$d_t = D(\delta_t, \mathbf{x}_t, H_t, \Gamma_t), \quad (38)$$

$$o_{t+1} = O(d_t, \mathbf{b}_t, \Gamma_t, \varepsilon_t), \quad (39)$$

$$\delta_{t+1} = U_\delta(\delta_t, o_{t+1}, H_t, \Gamma_t), \quad (40)$$

where d_t denotes a distrust-conditioned response such as monitoring, withdrawal, verification, or restricted exposure. The sign of the total loop is not fixed. Verification may reveal misconduct and strengthen warranted distrust. It may reveal reliability and weaken distrust. Withdrawal may avoid harm while also suppressing corrective evidence. A distrust loop therefore requires mechanism-level analysis rather than an assumption that every feedback is self-confirming.

Definition 6.3 (Distrust reinforcement channel). A *distrust reinforcement channel* is a mediated pathway through which a distrust-supporting relational condition changes later action, exposure, interpretation, institutional arrangement, or interaction opportunity in a way that tends, under specified conditions, to preserve or increase later distrust.

The concept allows the paper to distinguish warranted persistence from pathological closure. If repeated verification continuously uncovers new misconduct, sustained distrust may track a persistently unsafe relation. If counterevidence is systematically excluded, reinterpreted, or prevented from arising, the same apparent stability can instead reflect epistemic closure. The external appearance of persistence does not identify which mechanism is operating.

Claim 6.4 (Distrust-attractor candidate). A persistent distrust configuration becomes a plausible *distrust-attractor candidate* when multiple trajectories within a specified neighborhood tend to return toward a similar distrust-supporting configuration after bounded perturbations, and when identifiable relational mechanisms contribute to that return. Persistence alone remains insufficient.

This claim parallels the methodological restraint introduced for trust attractors in Section 3. The word *attractor* should be retained only when the observed or modeled dynamics justify it.

6.2 Rupture, Landscape, Hysteresis

Not every negative event produces rupture. Durable relations routinely absorb mistakes, misunderstandings, delays, conflicts of interest, and temporary failures. A rupture occurs when an event or sequence changes the relational configuration deeply enough that the previous updating regime no longer operates in the same way. The event can be betrayal, deception, abandonment, serious incompetence, abuse of delegated authority, institutional failure, or another violation whose importance is relation-specific.

The distinction between a state change and a rule change is especially useful here. Let e_t denote a rupture-relevant event. A first-order model changes the current relational state,

$$\mathbf{x}_{t+} = J(\mathbf{x}_{t-}, e_t, H_t, \Gamma_t), \quad (41)$$

where t^- and t^+ denote the moments immediately before and after the event for analytical convenience. A more consequential rupture also changes the parameters or rules governing later updates,

$$\theta_{t+} = K(\theta_{t-}, e_t, \mathbf{x}_{t-}, H_t, \Gamma_t). \quad (42)$$

The same later gesture may then be interpreted differently because the landscape through which it is processed has changed.

This formulation captures a familiar relational phenomenon. Before a serious violation, an unexplained delay may be interpreted generously. After a violation involving concealment, the same delay may activate suspicion. Before a data-manipulation scandal, a corrected spreadsheet may appear routine. After the scandal, a similar correction may trigger independent verification. The later event has not changed physically, but its position in relational history has changed.

Definition 6.5 (Relational rupture). A *relational rupture* is a historically consequential event or sequence that produces a substantial change in the relational state, the subsequent updating regime, or both, such that later interactions are evaluated through a materially altered relational history.

The definition avoids tying rupture to a universal threshold. The same act can be minor in one relation and transformative in another. Severity depends on prior commitments, vulnerability, norm structure, attribution, repetition, power asymmetry, the availability of repair, and the meaning the event acquires within the relation.

Rupture also helps connect the present section to the paper’s account of poetic composition. A seemingly small event can become rupture-relevant when it condenses an already distributed history. A short

phrase, missing gesture, repeated silence, or discovered inconsistency may activate many earlier experiences simultaneously. Its causal importance then cannot be read from the isolated event alone. The event operates within a prepared relational landscape.

If rupture changes the relational landscape, recovery should not be expected to retrace the path of decline in reverse. This is the core reason to introduce hysteresis-like reasoning. In a strict physical model, hysteresis has a specific technical meaning. The present paper uses the analogy cautiously to mark a more general historical asymmetry: the state of a relation can depend not only on present conditions but also on the path through which those conditions were reached.

Suppose a relation moved from configuration A to configuration B through a serious violation. Restoring the observable conditions that previously supported A may fail to reproduce A because the violation has become part of H_t . The actors now know something, remember something, or have been changed by something that was absent before. A promise that once generated confidence may now be discounted. A safeguard that once seemed unnecessary may become a condition of continued interaction. A relation may therefore require a different path, different evidence, and different institutional conditions to reach a stable post-rupture configuration.

Let $\mathcal{P}_{A \rightarrow B}$ denote the historical path from A to B and $\mathcal{P}_{B \rightarrow C}^R$ a repair path from B toward some later configuration C . In general,

$$\mathcal{P}_{B \rightarrow C}^R \neq \mathcal{P}_{A \rightarrow B}^{-1}, \quad (43)$$

and there is no requirement that $C = A$.

Proposition 6.6 (Historical asymmetry of repair). *When a rupture changes the relational history or the updating rule, restoration of pre-rupture external conditions is generally insufficient to guarantee restoration of the pre-rupture relational configuration. Repair may require new evidence, new safeguards, new meanings, or new forms of interaction, and its stable endpoint may differ from the earlier relation.*

This asymmetry is particularly important for institutional failure. A policy reversal or leadership apology may remove one immediate cause of harm while leaving organizational routines, incentive structures, reporting channels, or patterns of concealment unchanged. Gillespie and Dietz’s systemic account of organization-level repair is relevant precisely because trust repair after institutional failure cannot be reduced to a dyadic verbal response [23].

Historical asymmetry also explains why demands to “return to normal” can be misleading. The previous normal may no longer be available, desirable, or safe. A repaired relation can instead be a newly constituted relation that contains memory of failure and institutionalizes some of the lessons learned from it.

6.3 Repair, Attribution, Mechanisms

Repair requires more than a positive message. It involves interventions into the mechanisms that produced rupture and the mechanisms that now sustain distrust. Different failures therefore require different combinations of repair processes. The following mechanisms are analytically separable even when they occur together.

Recognition and acknowledgment. A repair process often begins by establishing that a rupture-relevant event will not simply be erased from the shared account of the relation. Acknowledgment can reduce uncertainty about whether the injured party’s experience is even recognized. It can also create a common object around which subsequent explanation, responsibility, and institutional change can occur. Acknowledgment alone, however, does not establish sincerity, adequate remedy, or future reliability.

Explanation and causal clarification. Actors need to know what happened, why it happened, and which mechanisms are likely to persist. Tomlinson and Mayer emphasize causal attributions involving locus, controllability, and stability in trust repair [75]. Within the present framework, causal clarification matters because it changes expectations about recurrence. A failure attributed to an unstable and corrected process creates a different future landscape from one attributed to a stable incentive or enduring integrity problem.

Responsibility and accountability. Where wrongdoing occurred, repair may require acceptance of responsibility and credible consequences. The relevant issue is not punishment in the abstract but whether the post-rupture relation now contains mechanisms that make future conduct more observable, constrained, or answerable. Accountability can alter both perceived trustworthiness and the institutional structure through which future evidence will be generated.

Restitution and material correction. Some ruptures redistribute resources, opportunities, risks, or harms. Symbolic communication without material correction can leave the generative conditions of distrust intact. Restitution, compensation, restoration of access, correction of records, reversal of an unjust decision, or other forms of material repair may therefore be necessary depending on the relation.

Structural change. If the failure was produced by routines, incentives, information asymmetries, concentrated authority, weak oversight, or unsafe reporting channels, then interpersonal reassurance cannot substitute for institutional redesign. Structural repair changes the field in which future interaction occurs. Gillespie and Dietz’s multilevel framework is especially relevant here because organization-level failure can require coordinated changes across multiple system components [23].

Renewed testing and graded vulnerability. Even after explanation and reform, a damaged relation may require new experience. Limited delegation, staged access, reversible commitments, independent verification, and bounded forms of vulnerability can produce new evidence without requiring the injured party to resume the earlier level of exposure immediately. This is the reparative counterpart of endogenous evidence production discussed in Section 5.

These mechanisms can be summarized through a repair intervention vector

$$\mathbf{q}_t = (a_t, c_t, k_t, m_t, s_t, v_t), \quad (44)$$

where the entries represent, schematically, acknowledgment, causal clarification, accountability, material correction, structural change, and renewed vulnerability or testing. The relational update becomes

$$\mathbf{x}_{t+1} = \Phi_R(\mathbf{x}_t, \mathbf{q}_t, H_t, \Gamma_t, \varepsilon_t). \quad (45)$$

No assumption is made that every component must be positive or that larger values always improve repair. Excessive surveillance, performative apology, inappropriate compensation, or forced vulnerability can worsen the relation. The representation simply makes visible that repair is multidimensional.

Definition 6.7 (Relational repair). *Relational repair* is a process through which actors or institutions modify the conditions sustaining post-rupture distrust, generate new evidence or safeguards, and create a viable relational configuration in which future interaction can proceed under revised expectations and boundaries.

This definition does not require restoration of the previous level of trust. Successful repair may instead produce a lower-trust but safer relation, a highly monitored working relation, a restructured institutional relation, or a mutually accepted exit.

Repair mechanisms are not interchangeable because violations communicate different things about future risk. Research by Kim and colleagues shows that responses to alleged violations can differ depending on whether the violation concerns competence or integrity [36]. Their findings do not provide a universal rule for every social relation, but they strongly caution against assuming context-free effects for apology, denial, explanation, or compensation.

The present framework extends that caution. The effect of a repair act depends on at least five relational questions.

First, *what dimension was damaged?* A competence failure, integrity violation, benevolence failure, confidentiality breach, abandonment, and procedural injustice can require different evidence of change.

Second, *how was the event attributed?* If the cause is believed to be stable, controllable, and internal to the responsible actor, expectations of recurrence will differ from those attached to an unstable or externally constrained cause [75].

Third, *what history preceded the event?* A first failure after a long history of reliability may occupy a different position from a nominally identical failure that completes a pattern of earlier warnings.

Fourth, *what power asymmetries structure the repair?* An apology demanded from a subordinate, a settlement imposed through unequal bargaining power, or “reconciliation” conditioned on silence may stabilize interaction while leaving the underlying injustice intact.

Fifth, *which future evidence becomes possible after the repair act?* The most consequential intervention may be the one that changes observability, creates safe reporting, permits independent audit, or makes future commitments reversible enough to test.

These questions suggest a contextual response function,

$$\Delta \mathbf{x}_{t+1} = R(\mathbf{q}_t, \mathbf{v}_t, \mathbf{a}_t, H_t, \Pi_t, \Gamma_t), \quad (46)$$

where $\mathbf{v}_t$ denotes the violation profile, $\mathbf{a}_t$ the causal attribution profile, and Π_t relevant power relations. The same nominal repair action can therefore produce different outcomes across relations.

Remark 6.8 (No universal repair token). No single communicative act universally restores trust. Apology, denial, explanation, compensation, transparency, monitoring, and structural reform acquire reparative significance only through the violation type, attribution, history, power structure, and subsequent evidence.

6.4 History, Protective Distrust, Reopening

Repair becomes theoretically more interesting once analysis moves beyond movement toward a previous state and examines the production of a new history. After a rupture, the event remains part of the relation even when confidence returns. So do the repair attempts, refusals, reforms, memories, and newly created safeguards. The relevant history can be represented schematically by

$$H_{t+1} = H_t \oplus (e_t, \mathbf{q}_t, o_{t+1}, \mathbf{x}_{t+1}), \quad (47)$$

where $\oplus$ denotes historical incorporation rather than simple numeric addition.

The repaired relation can therefore acquire capacities that the pre-rupture relation lacked. It may contain clearer boundaries, better reporting channels, more explicit expectations, greater tolerance for verification, or a richer shared vocabulary for conflict. Conversely, an apparent repair may leave participants more fragile, more dependent, or more reluctant to raise future concerns. The content of the new history matters.

This point also changes how success should be evaluated. A relation that returns quickly to high reported trust may be less robust than one that recovers more slowly while creating better mechanisms for detecting future failure. Short-term affective reassurance and long-term relational resilience are different outcomes.

Claim 6.9 (Repair can increase relational reflexivity). A rupture followed by credible diagnosis, structural correction, and renewed interaction can produce a relation with greater capacity to identify, discuss, and respond to future failure than the pre-rupture relation possessed. This possibility depends on the repair process and should not be presumed from the mere continuation of the relation.

The claim is intentionally conditional. Some ruptures permanently narrow the range of viable relations. Some reveal that exit is preferable to repair. Some cannot be repaired within a morally acceptable distribution of risk. The framework therefore places repair among several possible trajectories rather than making it a universal normative requirement.

A theory concerned with trust generation can easily acquire an implicit bias toward more trust. That bias must be resisted. Distrust can be protective, epistemically responsible, and institutionally necessary. Whistleblower protections, independent audit, peer review, conflict-of-interest rules, segregation of duties, adversarial legal procedures, and scientific replication all institutionalize forms of verification that do not assume unconditional confidence.

From the present perspective, the normative objective is therefore not the elimination of distrust. It is the creation of relational conditions in which trust and distrust can remain revisable in light of evidence,

power, history, and changing circumstances. A repaired relation may retain monitoring that was introduced after failure. Such monitoring does not automatically signify failed repair. It may be part of the new equilibrium that makes renewed cooperation safe.

This suggests the concept of *calibrated reopening*. After rupture, the relation can reopen gradually through bounded exposures whose possible harms remain proportionate and reversible. The process can include independent verification, staged delegation, explicit review points, transparent data, and protected channels for contestation.

Definition 6.10 (Calibrated reopening). *Calibrated reopening* is a post-rupture process in which interaction and vulnerability are reintroduced in bounded, reviewable, and proportionate forms so that new evidence can emerge without requiring premature restoration of the pre-rupture exposure level.

Calibrated reopening links repair to epistemic openness. It creates opportunities for favorable evidence while preserving routes through which adverse evidence can still alter the trajectory. In this respect, resilient repair differs from manufactured reconciliation, which often seeks rapid visible harmony while suppressing the mechanisms of future correction.

6.5 Pseudo-Repair, Manufactured Recovery

The mechanisms of repair can themselves be manipulated. Public apology, ritual, compensation, symbolic reform, transparency dashboards, independent committees, listening sessions, or new codes of conduct can change experienced trust even when the underlying relation remains materially unchanged. The possibility follows directly from the paper’s earlier analysis of poetic composition and relational salience. Repair signals can be composed into a powerful atmosphere of recovery.

The problem is not that symbolic acts are necessarily empty. Symbols, language, ceremony, and acknowledgment can be genuine components of repair. The problem arises when visible repair signals substitute for changes in the mechanisms that generated the original failure or when they make those mechanisms harder to inspect.

Let s_t^R denote the observable signal of repair and m_t^R the material or structural change relevant to future reliability. A divergence can occur when

$$\frac{\partial \tau_{t+1}}{\partial s_t^R} > 0 \quad \text{while} \quad \frac{\partial w_{t+1}}{\partial s_t^R} \approx 0, \quad (48)$$

where w_{t+1} represents a context-specific measure of underlying trustworthiness or safety. The equation is illustrative. Its point is the possibility that a repair signal can improve experienced trust without materially improving the underlying conditions that warranted concern.

Definition 6.11 (Pseudo-repair). *Pseudo-repair* denotes a process that improves the appearance or experience of relational recovery while leaving central mechanisms of failure, power asymmetry, unaccountability, or future vulnerability substantially unchanged.

Pseudo-repair can generate a new self-reinforcing loop. Improved affect lowers scrutiny; lower scrutiny reduces detection; reduced detection produces fewer visible failures; the absence of visible failures is then used to argue that repair succeeded. This is structurally related to the detection-thinning loop introduced in Section 5. The difference is historical: here the loop occurs after an acknowledged or suspected rupture and can transform repair itself into a mechanism of renewed opacity.

The safeguard is not permanent suspicion. It is correspondence among repair signals, material changes, independent observability, and future correctability. A repair process should remain open to evidence that its own interventions are insufficient.

7 Governance, Design, Manipulation

Once trust and distrust are understood through recursive relational processes, governance enters before any explicit instruction to “build trust.” Procedures, guarantees, interfaces, spatial settings, rituals, verification

systems, and opportunities for interaction already shape which relational trajectories are easier to generate. Institutional research likewise shows that trust formation can depend on arrangements extending beyond the immediate dyad [81, 4].

The resulting problem is not trust maximization. Higher experienced trust can coexist with weak trustworthiness, reduced scrutiny, unequal vulnerability, or restricted exit; protective distrust can be warranted under uncertainty or after violation. The relevant question is which conditions should be created, removed, constrained, or kept revisable so that reliance can emerge when warranted, remain corrigible when uncertain, and weaken when contrary evidence appears.

The section develops a taxonomy of trust-generating conditions and governance levers, then examines guarantees, shared experience, atmospheres, interfaces, manufactured relational fluency, and manufactured vulnerability. The guiding Generative Relational principle is to govern the conditions under which trust-worthy relations can form and evolve without making experienced trust the required endpoint.

7.1 Conditions, Levers, Non-Maximization

Suppose an institution can increase a reported trust score through persuasive communication, reassuring architecture, reduced visibility of conflict, or friction that makes questioning costly. If the sole objective is a larger trust measure, each intervention can appear successful. Yet none necessarily improves reliability, integrity, benevolence, procedural fairness, or the ability of affected actors to detect failure. A governance framework that makes observed trust its terminal objective can therefore reward the production of confidence independently of the quality of the relation that confidence concerns.

The present framework begins elsewhere. Let $\mathbf{x}_t$ denote a relational configuration and let Γ_t denote the surrounding relational environment, including institutional, material, informational, spatial, technological, and symbolic conditions. A governance intervention $\mathbf{g}_t$ modifies some part of that environment,

$$\Gamma_{t+1} = \mathcal{G}(\Gamma_t, \mathbf{g}_t, \mathbf{x}_t, H_t), \quad (49)$$

while the subsequent relational state evolves according to

$$\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t; \Gamma_{t+1}, \boldsymbol{\theta}_t, \boldsymbol{\varepsilon}_t). \quad (50)$$

The intervention may therefore affect trust indirectly by changing what actions are available, what evidence can be produced, how costly verification is, how vulnerability is distributed, or what meanings become salient.

Definition 7.1 (Trust-generating condition). A *trust-generating condition* is a feature of the relational environment that changes the probability, direction, stability, or revisability of a trust-relevant trajectory by altering interaction, exposure, evidence, interpretation, affect, institutional assurance, or historical accumulation.

The definition does not imply that every such condition increases trust. A transparent audit mechanism may initially reduce trust by revealing failure. A carefully designed complaint procedure may make distrust more expressible. A reversible pilot may prevent premature commitment. These can still improve the conditions under which trust becomes better calibrated.

Definition 7.2 (Governance of trust-generating conditions). *Governance of trust-generating conditions* denotes deliberate or institutionally stabilized intervention in the environments through which trust, distrust, verification, vulnerability, repair, and exit become possible and evolve over time.

This formulation avoids two shortcuts. First, governance need not manufacture a particular subjective state. Second, a successful intervention need not produce immediate trust. The relevant achievement may instead be improved evidence, safer experimentation, lower asymmetry, more credible accountability, or a better possibility of revision.

Proposition 7.3 (No trust-maximization principle). *Within the present framework, an increase in experienced trust is insufficient for concluding that a governance intervention improved the governed relation. Assessment additionally requires attention to trustworthiness, epistemic openness, vulnerability, reversibility, contestability, and exit.*

The proposition follows from distinctions already established in Sections 5 and 6. Trust can become self-reinforcing even when the generating loop filters contrary evidence, while distrust may remain protective when the underlying relation is unsafe. A governance objective centered on trustworthy relational development must therefore permit both trust formation and trust withdrawal.

Trust-generating conditions are heterogeneous. Grouping them by mechanism helps prevent a vague appeal to “trust building” from hiding the actual intervention. At least seven families are analytically useful.

Institutional assurance. Rules, professional standards, certification, liability, monitoring, independent oversight, enforceable remedies, insurance, escrow, and other third-party arrangements can reduce uncertainty or change the expected cost of violation. They may support initial cooperation where interpersonal history is thin. Zucker’s institutional account and later work on institution-based trust provide direct precedents for this category [81, 4].

Interactional opportunity. Governance can alter whether actors have repeated, meaningful, and appropriately bounded opportunities to cooperate. Joint tasks, staged pilots, shared problem solving, reciprocal commitments, and repeated contact may produce histories that cannot be created through declarations alone. Their value depends on the structure of the interaction: coerced proximity or asymmetric exposure can intensify vulnerability without producing justified trust.

Epistemic access. Transparency, explanation, inspectability, access to records, independent verification, traceability, and channels for disagreement affect what can be known about a relation. These mechanisms matter especially when a trust loop would otherwise become epistemically closed. Their function is not simply to produce positive information. They keep disconfirming information reachable.

Spatial and atmospheric organization. Architecture, sound, lighting, material arrangement, privacy, visibility, seating, thresholds, and other environmental properties can alter felt safety, exposure, attention, and the ease of interaction. The discussion of atmosphere in Section 4 drew on Böhme’s account of atmospheres and their designed production [7]. In the present context, the governance relevance lies in the capacity of environments to modulate relational orientation before explicit propositional judgment is complete.

Symbolic and ritual composition. Ceremonies, professional dress, public commitments, seals, interfaces of authority, repeated forms of greeting, shared narratives, and institutional symbols can condense histories or expectations. Some support coordination by making commitments legible. Others can merely borrow the appearance of reliability. Their effect therefore has to be evaluated separately from the substantive quality of the relation.

Interface and choice architecture. Digital systems determine what information is prominent, which actions are easy, when consent is requested, how warnings appear, and how difficult exit or verification becomes. Human–computer interaction research on dark patterns shows that interface design can steer or deceive users toward decisions that benefit a service while potentially harming users [49]. The relevance to trust governance is direct: an interface can increase felt fluency or apparent legitimacy while simultaneously weakening informed agency.

Temporal organization. Waiting periods, cooling-off intervals, probation, renewal requirements, periodic review, staged commitment, and opportunities to revisit earlier choices govern the tempo of relational formation. Time itself does not create trust automatically, but temporal design affects whether evidence can accumulate before vulnerability deepens and whether commitments can be revised when the environment changes.

These families can be combined. A clinical institution, for example, may use professional certification, confidential spatial design, repeated appointments, explanatory records, ritualized consent, and digital portals simultaneously. The important analytical task is to disaggregate the mechanisms before judging the resulting trust trajectory.

Definition 7.4 (Governance bundle). A *governance bundle* is a coordinated set of interventions across two or more trust-generating condition families whose combined effects alter the relational environment in which trust and distrust evolve.

Governance bundles make interaction effects likely. A transparent record system may support trust when accountability is credible, yet the same transparency can become coercive under severe power asymmetry. A welcoming environment may support difficult disclosure when confidentiality is reliable, yet become manipulative when reassurance is used to induce exposure without adequate protection. Evaluation therefore cannot be reduced to isolated design features.

7.2 Guarantees, Shared Experience, History

A recurring problem in trust formation concerns beginnings. Two actors may lack sufficient shared history to justify substantial vulnerability, while the history needed for trust cannot arise without some initial interaction. Third-party guarantees help explain how this apparent circularity can sometimes be partially resolved.

A guarantee can take many forms: legal liability, a professional licence, escrow, collateral, insurance, a reputable intermediary, an institutional complaints process, a bond, a certification regime, or a community whose sanctions extend beyond the immediate transaction. These mechanisms do not eliminate uncertainty. They change the distribution of risk and the expected consequences of violation. In doing so, they can create a bounded zone in which actors are willing to test a relation that they would not otherwise enter.

Let v_t denote potential vulnerability and let ℓ_t denote expected loss conditional on failure. A guarantee q_t may reduce the actor's effective exposure without changing the counterpart's underlying trustworthiness ω_t ,

$$\ell_t^{\text{eff}} = \ell_t - Q(q_t), \quad \omega_t^{\text{after}} = \omega_t^{\text{before}} \quad \text{unless the guarantee changes conduct or capability.} \quad (51)$$

This distinction matters. A guarantee can make interaction rationally or emotionally possible even when evidence about the counterpart remains limited. The resulting interaction may then generate new evidence about conduct. The guarantee therefore can contribute to the genesis of a relational history without serving as proof of the other party's intrinsic reliability.

Definition 7.5 (Initial safety field). An *initial safety field* is a bounded relational environment in which external protections, reversibility, or limited exposure reduce the cost of exploratory interaction sufficiently for new trust-relevant history to be generated.

The phrase “field” is used locally here and does not claim a complete field theory of governance. Its purpose is to emphasize that the relevant protection can be distributed across several conditions rather than located in one actor. An initial safety field may combine law, institutional reputation, limited stakes, observation, reversible commitment, and the availability of exit.

Guarantees can nevertheless fail in at least three ways. They can be non-credible, creating only apparent safety. They can be inaccessible in practice, particularly for weaker parties. Or they can become substitutes for substantive trustworthiness, allowing harmful relations to acquire legitimacy through borrowed institutional prestige. Governance must therefore examine who stands behind the guarantee, how claims can be enforced, what costs accompany its use, and whose interests the guarantor represents.

A second family of interventions concerns common experience. Repeated cooperation can create evidence, mutual expectations, memories, routines, and forms of affective familiarity. Section 5 described historical sedimentation through repeated interaction. Governance can influence that process by structuring opportunities for actors to encounter one another under conditions where meaningful reciprocal conduct becomes possible.

This possibility is often more subtle than arranging contact. Shared experience has trust-generating potential only when the experience carries relationally relevant information. Standing in the same room need not reveal whether another actor honors commitments. Performing a low-stakes joint task may reveal some forms of reliability while saying little about conduct under severe conflict. A designed experience therefore has a domain and an evidential range.

Definition 7.6 (Relationally informative shared experience). A *relationally informative shared experience* is a jointly encountered sequence whose structure permits participants to observe conduct relevant to a specified dimension of future vulnerability, cooperation, reliability, or repair.

This definition helps distinguish genuine history generation from staged familiarity. Team-building exercises, ceremonies, retreats, simulations, or community events may create warmth and common memory while yielding little information about the dimensions on which future trust will actually depend. They are not therefore valueless. Affective coordination can matter. But the kind of trust they support should not be generalized beyond the experience’s informational scope without further evidence.

Shared experience also has distributive properties. One participant may bear more risk, perform more emotional labor, disclose more personal information, or possess fewer opportunities to leave. A governance arrangement that generates “common experience” through unequal sacrifice can produce apparent cohesion while concealing asymmetry. The ethical unit of analysis must therefore include who contributes what to the history from which trust later emerges.

7.3 Atmospheres, Interfaces, Manufactured Fluency

Section 4 argued that trust-relevant meaning can emerge from heterogeneous and indirect composition. Governance consequently extends beyond explicit rules and statements. Spatial arrangement, rhythm, repetition, sound, material texture, symbols, timing, and remembered events can be intentionally composed. Such interventions deserve analysis precisely because they may alter a relational trajectory without operating through an explicit argument.

A hospital waiting room can make vulnerability feel either exposed or protected. A mediation space can amplify confrontation or permit controlled attention. A public institution can communicate distance through barriers and opaque procedures or accessibility through legible processes and approachable settings. A digital service can borrow familiar visual conventions associated with security or public authority. None of these effects determines trust, and none proves trustworthiness. Yet they can modify salience, affect, willingness to disclose, and the interpretation of subsequent interaction.

We can represent such an intervention by a compositional vector $\mathbf{c}_t$ whose components affect several relational dimensions at once,

$$\Delta \mathbf{x}_{t+1} = \mathbf{J}_c(\mathbf{x}_t, H_t) \Delta \mathbf{c}_t + \boldsymbol{\eta}_t, \quad (52)$$

where $\mathbf{J}_c$ is a context-dependent sensitivity operator. The same change in atmosphere or symbolic form can therefore have different effects for different histories. A security checkpoint may reassure one actor and activate fear in another. Familiar ritual may signal dependable continuity to some participants while signaling exclusion to others.

Remark 7.7 (No universal atmosphere of trust). A designed environment cannot be assumed to possess a universal trust effect. Atmospheric and symbolic interventions are historically, culturally, and positionally mediated. Their evaluation requires attention to heterogeneous participants and prior relational histories.

The governance implication is important. Poetic composition can sometimes support humane interaction where explicit persuasion would be inadequate. Yet its indirectness also creates a special risk of

manipulation. When an institution intentionally produces felt safety while withholding material facts that would change a reasonable actor’s judgment, compositional skill can become an instrument for bypassing scrutiny rather than supporting relation.

Digital environments make the previous problem especially visible. Interfaces can reduce cognitive burden, clarify procedures, expose provenance, and make consent more meaningful. They can also manufacture fluency: a smooth, professional, familiar, or frictionless experience may be interpreted as a signal of competence or legitimacy even when the interface supplies little substantive evidence about the underlying actor.

The literature on dark patterns provides a concrete warning that design can steer behavior through coercion, misdirection, obstruction, urgency, social proof, and other mechanisms [49]. The present paper’s claim is not that every persuasive interface is deceptive. Rather, interface features can enter trust-generating loops, and their normative status depends partly on whether they preserve the user’s ability to inspect, contest, and reverse the resulting commitment.

Definition 7.8 (Manufactured relational fluency). *Manufactured relational fluency* denotes intentionally produced ease, familiarity, coherence, or affective reassurance that lowers friction in a trust-relevant interaction without necessarily increasing access to evidence about the underlying relation.

Fluency can be beneficial. A badly designed interface can make even a trustworthy institution inaccessible. The concern arises when lowered friction is paired with reduced epistemic access or increased asymmetry. Examples include making acceptance effortless while making withdrawal difficult, presenting reassuring signals prominently while burying material conditions, or using social proof to accelerate commitment before verification is possible.

The appropriate response is not to maximize friction. Excessive friction can itself exclude users and make safe institutions unusable. The design problem is instead to align friction with relational significance. High-vulnerability commitments may warrant pauses, verification, explicit explanation, or reversibility even when those features reduce immediate conversion or reported confidence.

7.4 Manufactured Vulnerability, Constraints

Trust matters because it can enable vulnerability under uncertainty. That same feature creates a serious governance problem: actors who control trust-generating conditions may be able to increase another party’s willingness to become vulnerable without proportionately improving the safety or trustworthiness of the relation.

Definition 7.9 (Manufactured vulnerability). *Manufactured vulnerability* occurs when an actor or institution intentionally or systematically alters trust-generating conditions in ways that increase another party’s exposure, dependence, disclosure, commitment, or reduced vigilance without a commensurate improvement in the protections, trustworthiness, or informed agency relevant to that exposure.

The concept is relational rather than psychological. It does not require the exposed actor to be irrational, naive, or deceived about every relevant fact. The surrounding conditions can make increased exposure reasonable from the actor’s local perspective while the designer possesses superior knowledge of risks, alternatives, or downstream lock-in. Manufactured vulnerability can therefore arise through asymmetry even without an explicit false statement.

Several pathways are possible. Institutional prestige can be transferred to an actor who is weakly monitored. A reassuring environment can encourage personal disclosure before confidentiality protections are clear. Repeated low-risk successes can be used to induce a later high-risk commitment outside the domain previously tested. Social proof can imply widespread confidence while hiding selection effects. Exit can become costly only after trust has deepened. Monitoring can be reduced precisely when the relation becomes more consequential.

A simple diagnostic compares the change in induced exposure ΔV with changes in protection ΔP , relevant trustworthiness ΔW , and informed agency ΔA ,

$$M_t = \Delta V_t - \alpha \Delta P_t - \beta \Delta W_t - \gamma \Delta A_t, \quad \alpha, \beta, \gamma > 0. \quad (53)$$

The quantity M_t is not proposed as a universal metric. It records the structural question: did governance increase exposure faster than it improved protection, trustworthiness, and agency? Empirical applications would require domain-specific definitions and weights.

Remark 7.10 (Trust can be genuine while vulnerability is manufactured). Manufactured vulnerability does not require the resulting trust to be subjectively fake. A person may sincerely experience trust because the relational environment successfully generated it. The normative problem lies in the conditions and asymmetries through which exposure was produced, not in whether the experience was psychologically authentic.

This distinction is central to a Generative Relational analysis. A relational reality can be real in its effects and still be unjustly produced. The fact that trust subsequently becomes self-reinforcing does not cleanse the history of its generation.

A complete normative theory lies beyond the scope of this paper. The preceding analysis nevertheless supports several constraints that any governance of trust-generating conditions should at least confront.

Legibility. Material features of the relation should be sufficiently visible for affected actors to understand the kinds of vulnerability, guarantees, monitoring, and institutional backing involved. Perfect transparency is neither possible nor always desirable, but obscuring decision-relevant structure should not be the method by which trust is obtained.

Epistemic openness. Trust-generating systems should preserve channels through which contrary evidence can appear and matter. Audit, complaint, appeal, external review, record access, and independent verification are possible mechanisms. A system that only generates confirmatory information may create stable trust while simultaneously degrading calibration.

Reversibility. Where uncertainty is substantial, actors should often be able to test a relation through bounded commitments before deeper vulnerability becomes irreversible. This principle connects trust governance to the calibrated reopening discussed in Section 6.

Contestability. Affected actors should be able to challenge not only outcomes but also the conditions through which trust is being solicited. This includes questioning a guarantee, disputing a representation of safety, resisting a ritualized expectation of confidence, or asking for an alternative interaction channel.

Proportional vulnerability. The demanded level of exposure should bear some defensible relation to the evidence, protections, and reversibility available at that stage. A system that requires maximal disclosure before demonstrating minimal reliability reverses the order through which trust can be ethically tested.

Exit and relation modification. Trust should not become a mechanism through which later refusal is made impossible. Exit, renegotiation, delegation changes, reduced exposure, and partial withdrawal may be necessary for relational learning. Where complete exit is impossible, meaningful possibilities for relation modification become especially important.

Accountability of the condition designer. Governance analysis must include whoever controls the trust-generating environment. Designers, institutions, intermediaries, platforms, professional bodies, and guarantors can shape relational trajectories while remaining outside the visible dyad. Their incentives and responsibilities therefore belong within the unit of analysis.

Claim 7.11 (Governance criterion). A trust-generating intervention is normatively incomplete when its evaluation reports only whether trust increased. A minimally adequate evaluation must also ask what exposure changed, what evidence remained accessible, who controlled the intervention, what forms of revision or exit remained possible, and whether underlying trustworthiness changed in the relevant domain.

These constraints should not be read as a fixed checklist sufficient for every context. They identify dimensions on which trust governance can fail even when participants report confidence and cooperation.

7.5 Governance, Research, Corrigibility

The Generative Relational perspective developed in this paper shifts the object of governance from a desired terminal attitude to the evolving relations from which attitudes, conduct, evidence, and vulnerabilities emerge. This shift has several consequences.

First, governance concerns processes of relation formation. Who can encounter whom, under what protections, through which mediators, with what information, and at what level of initial exposure all influence the trajectories that can begin.

Second, governance concerns relation updating. Trustworthy systems require ways for new evidence to modify prior expectations. A self-reinforcing relation that cannot learn from disconfirmation can become stable for the wrong reason.

Third, governance concerns relation removal and exit. Some relations should not be repaired. Some guarantees should be withdrawn. Some trust attractors should be escaped rather than strengthened. The possibility of dissolution therefore belongs inside the governance problem from the beginning.

Fourth, governance concerns the evolution of governance mechanisms themselves. If the relational landscape changes historically, a fixed procedure may cease to protect the same interests. Audit systems, guarantees, disclosure rules, interface conventions, and repair processes need their own histories of revision. This echoes the second-order dynamics introduced in Section 3: not only states but update rules can change.

The resulting orientation can be summarized in four verbs: *enable*, *expose*, *revise*, and *release*. Governance should enable bounded interaction where worthwhile relations could emerge; expose actors to relevant evidence rather than only reassurance; preserve revision when history changes; and make release from harmful or exhausted relations materially possible. None of these verbs guarantees trust. That is precisely the point.

Proposition 7.12 (Trust is not the governance endpoint). *For a Generative Relational account, the governance objective is not the production of maximal or permanent trust. It is the maintenance of relational conditions under which trust and distrust can become better calibrated through history while vulnerability, agency, correction, and exit remain normatively visible.*

This proposition also limits the meaning of an attractive trust basin. A deep basin can support valuable cooperation, but depth alone does not justify its preservation. If exit becomes impossible, contrary evidence becomes inadmissible, or vulnerability is systematically displaced onto weaker actors, then attractor stability can become a governance problem rather than an achievement.

The conceptual framework creates several empirical obligations. The first is mechanism identification. Studies should distinguish whether an intervention changes trust through risk reduction, evidence production, affective reassurance, institutional transfer, social proof, interface fluency, repeated interaction, or another pathway.

The second is temporal separation. Immediate increases in reported trust may not predict later calibration. Longitudinal designs are needed to determine whether an intervention creates resilient cooperation, suppresses detection of failure, or merely shifts the timing of distrust.

The third is heterogeneity. Atmospheric, symbolic, and institutional cues may have different effects across histories, social positions, cultures, and prior experiences. Average effects can obscure groups for whom the same intervention signals threat rather than safety.

The fourth is distributive analysis. Researchers should measure who becomes more vulnerable, who obtains information, who bears monitoring costs, who can invoke guarantees, and who can exit. A trust intervention can improve aggregate cooperation while worsening the position of a subgroup.

The fifth is counterfactual design. Where possible, studies should compare trust-generating interventions with alternatives that improve actual trustworthiness or protection without directly targeting confidence. Such comparisons can reveal whether experienced trust is responding to substantive institutional improvement or primarily to presentation.

The sixth is closure detection. Researchers should test whether mature trust systems remain sensitive to negative evidence. A relation that remains stable because members continuously encounter reliability differs from one that remains stable because evidence channels have narrowed.

Research Obligation 7.13 (Dual-outcome reporting). Empirical studies of trust governance should, where feasible, report changes in experienced trust separately from changes in observable trustworthiness, protection, vulnerability, verification access, and exit conditions.

Research Obligation 7.14 (Designer-side analysis). Studies of designed trust environments should identify the actors who select and control the relevant conditions, their incentives, the information they possess, and the asymmetries between designers and participants.

These obligations are particularly important for the later normative analysis. Without them, a governance system may appear successful precisely because its most consequential effects occur outside the variable labeled “trust.”

8 Vulnerability, Power, Justice

Governance of trust-generating conditions raises a further question: who bears the vulnerability, receives the benefits, controls the relevant evidence, and retains the capacity to contest or leave? A trust-supporting relation can be dynamically stable and behaviorally productive while remaining normatively defective. Stability, trustworthiness, and justice must therefore be assessed separately even when they interact.

The ethical importance of entrusted vulnerability is longstanding in trust research [5]; work on autonomy and trust likewise directs attention toward trustworthiness, communication, consent, and institutional conditions rather than a generalized demand for more trust [59]. The present framework adds a temporal and relational problem: trust-producing interventions can change later action, evidence, dependence, and interpretation, so the justice of the evolving process cannot be inferred from the initial intervention or from mutually beneficial outcomes among the most visible parties.

No single metric of just trust is proposed. The section instead develops a normative profile covering vulnerability and benefit, power over the trust-generating field, epistemic standing, third-party externalities, refusal and exit, asymmetric trust demands, historical burdens, and the justice of relational change itself.

8.1 Stability, Vulnerability, Distribution

The attractor vocabulary can invite an understandable but mistaken normative shortcut. If a relation repeatedly returns to a stable region after ordinary perturbations, observers may be tempted to judge the relation healthy, well-functioning, or socially desirable. Yet dynamical persistence supplies no such conclusion by itself. Many undesirable arrangements can be resilient. Some persist because actors repeatedly receive good reasons to rely on one another; others persist because alternatives are blocked, dissent is costly, contrary evidence is filtered, or weaker actors bear most of the cost of maintaining the relation.

Let $\mathcal{A}$ denote a trust-attractor candidate in the sense developed in Section 3. Let $\text{Stab}(\mathcal{A})$ denote its dynamical stability, $W(\mathcal{A})$ an assessment of relevant trustworthiness, and $J(\mathcal{A})$ a normative assessment of the relational arrangement. The framework requires the following separation:

$$\text{Stab}(\mathcal{A}) \not\Rightarrow W(\mathcal{A}), \quad \text{Stab}(\mathcal{A}) \not\Rightarrow J(\mathcal{A}). \quad (54)$$

Even a favorable assessment of trustworthiness in one domain does not settle the wider question of justice. A highly reliable organization can still impose unequal risks, silence affected groups, or externalize burdens onto third parties.

Proposition 8.1 (No stability-to-justice inference). *The persistence, depth, or resilience of a trust-attractor candidate provides no sufficient reason to judge the underlying relational arrangement just. Normative assessment requires additional information concerning vulnerability, power, epistemic access, contestability, distribution, and exit.*

This proposition matters because self-reinforcement can hide normative defects. When a relation generates its own confirming evidence, recruits loyal participants, and lowers the frequency of challenge, observed stability may increase precisely while opportunities for discovering injustice decrease. The normative task is therefore not to weaken every stable relation. It is to avoid making stability itself the criterion by which the relation is justified.

Trust typically permits action under conditions in which complete verification would be costly or impossible. The resulting vulnerability is not necessarily a problem. Many valuable relations require some willingness to rely on others. The justice question begins when vulnerability and benefit are distributed unevenly, especially when the distribution is hidden by aggregate descriptions such as “high trust” or “successful cooperation.”

Consider a relational system with actors $i = 1, \dots, n$. For each actor, let

$$\mathbf{q}_{i,t} = (v_{i,t}, b_{i,t}, k_{i,t}, c_{i,t}, x_{i,t}), \quad (55)$$

where $v_{i,t}$ denotes relevant vulnerability, $b_{i,t}$ relational benefit, $k_{i,t}$ epistemic access to information needed for evaluation, $c_{i,t}$ the cost of contestation, and $x_{i,t}$ realistic exit capacity. These quantities are not assumed to share a common scale. The vector is a diagnostic device, not an invitation to collapse heterogeneous normative dimensions into one utility score.

An intervention can increase aggregate cooperation while changing the profiles very differently. For example, a guarantee may reduce uncertainty for a powerful institution while requiring a weaker participant to disclose more information. A platform may create frictionless interaction while making appeal or withdrawal difficult. A workplace may increase team cohesion while shifting care, emotional regulation, or informal monitoring labor onto a subset of members. The resulting relation may become more stable even though the burden of stabilization becomes more unequal.

Definition 8.2 (Vulnerability asymmetry). A relational configuration exhibits *vulnerability asymmetry* when actors whose participation contributes to the maintenance or benefit of the relation bear materially different exposure to loss, retaliation, dependency, or irreversibility, and those differences are normatively relevant to the relation under examination.

Vulnerability asymmetry is not automatically unjust. Differences in role, responsibility, expertise, or voluntarily accepted risk can matter. The point is that the asymmetry must remain visible. A trust framework that records only reported confidence can erase the very distribution through which confidence becomes possible.

The same caution applies to benefits. Trust can generate lower transaction costs, easier coordination, emotional security, information sharing, status, market access, or institutional legitimacy. These gains may accrue to different actors than those who bear the corresponding exposure. A relational analysis should therefore ask not only “who trusts whom?” but also “who gains from this trust, and who becomes vulnerable in order for those gains to be produced?”

Claim 8.3 (Distributional visibility). Any normative assessment of trust production is incomplete when it reports the level or stability of trust without also identifying the distribution of relevant benefits, vulnerabilities, verification burdens, and opportunities for correction.

8.2 Power, Fields, Epistemic Justice

Section 7 described trust-generating conditions such as guarantees, procedures, interfaces, atmospheres, repeated interaction, and symbolic composition. These conditions are rarely selected from nowhere. Some actors possess greater authority to set them, greater resources to modify them, and greater capacity to define normal behavior within them. Justice therefore concerns not only outcomes within a relational field but also power over the construction and revision of the field itself.

Let γ_t denote a vector of governable conditions affecting relational evolution. We can augment the state update from earlier sections by writing

$$\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t, \gamma_t; \theta_t), \quad (56)$$

where θ_t contains historically evolving update parameters. A further object is then required:

$$\mathbf{p}_t = (p_{1,t}, \dots, p_{n,t}), \quad (57)$$

where $p_{i,t}$ schematically represents the actor's effective influence over selection, interpretation, or revision of γ_t . Again, the notation is diagnostic. It directs attention toward a question that a purely behavioral model can miss: who has the capacity to shape the conditions from which later trust emerges?

Power over the trust-generating field can operate through obvious authority, but it can also operate through design. The actor who selects default settings, controls verification channels, chooses which histories are displayed, defines acceptable evidence, or determines the cost of appeal can alter relational trajectories without issuing an explicit command. The ethical problem becomes especially sharp when such design increases another actor's willingness to accept vulnerability while leaving the designer's own exposure largely unchanged.

Definition 8.4 (Field-shaping asymmetry). *Field-shaping asymmetry* exists when actors possess substantially unequal capacity to select or revise trust-generating conditions while the resulting conditions materially affect the vulnerabilities, interpretations, or options of other participants.

Field-shaping asymmetry is not identical to domination, manipulation, or injustice. Institutions often need delegated authority to design procedures. The concept instead identifies where further normative scrutiny is required. Legitimacy, accountability, transparency, expertise, necessity, and the availability of meaningful challenge can all alter the evaluation.

Trust governance has an epistemic dimension because relations depend on whose reports are believed, which events count as evidence, and which interpretations remain available. Fricker's account of epistemic injustice analyzes harms that arise when persons are wronged in their epistemic capacities, including credibility distortions linked to social power [22]. Dotson further examines forms of silencing in which hearers fail to meet the vulnerabilities of speakers in testimonial exchange [17]. Medina develops a broader account of epistemic resistance and the social conditions that shape epistemic interaction [54]. These works are directly relevant to trust production because a trustful environment can become epistemically closed when some warnings are structurally difficult to voice or receive.

Suppose a community has developed a strong trust attractor around an institution. A participant reports a violation. If mature trust makes members interpret the report through a presumption that the institution "would never do that," the existing attractor can lower the report's effective credibility. The problem is not merely that the group might reach an incorrect factual conclusion. The trust relation itself can participate in determining which speakers are heard and which evidence is admitted into subsequent updating.

This produces a difficult asymmetry. Trust is often praised because it reduces the need for constant suspicion. Yet the same reduction in suspicion can become normatively dangerous when it lowers the receptivity of the system to well-grounded dissent. Conversely, distrust originating from actors who have experienced repeated failures may contain information that a highly trustful majority does not possess.

Definition 8.5 (Epistemic closure of trust). A trust relation exhibits *epistemic closure* to the extent that the mechanisms maintaining confidence systematically reduce access to, uptake of, or appropriate responsiveness toward evidence capable of revising that confidence.

The definition is deliberately mechanism-centered. Stable trust is not closed merely because negative evidence is rare. Closure concerns the relation between stability and the suppression, discounting, or non-production of potentially revising information.

Proposition 8.6 (Distrust can carry epistemic value). *The presence of distrust does not by itself indicate relational failure. Under conditions of asymmetric experience, prior violation, weak accountability, or unequal access to evidence, distrust may preserve information and motivate verification that a trustful majority would otherwise lose.*

This proposition should not be reversed into a presumption that distrust is always epistemically superior. Suspicion can also become self-reinforcing and closed. Section 6 already argued this point. The normative requirement is symmetry of corrigibility: both trust and distrust should remain capable of revision when relevant evidence changes.

8.3 Externalities, Third Parties, Exit

Many theories and empirical studies of trust focus on a dyad, team, or bounded organization. Yet a relation can generate effects beyond the parties whose trust is being measured. The ethical assessment must therefore widen the unit of analysis when trust production displaces costs onto outsiders.

Consider actors *A* and *B* whose mutual trust lowers monitoring and accelerates decision-making. Their relation may be beneficial to both. Suppose, however, that reduced monitoring increases the probability that errors are transferred to clients, subcontractors, residents, future users, or another group *C* that has little voice in the relation. Alternatively, *A* and *B* may build trust by sharing information supplied by *C* without meaningful authorization, by excluding *C* from deliberation, or by stabilizing norms that impose costs on *C*. Nothing in the mutuality of trust between *A* and *B* settles the justice of these wider effects.

Definition 8.7 (Trust externality). A *trust externality* is a normatively relevant benefit, burden, risk, or constraint produced or amplified by a trust relation and borne by actors who are not adequately represented in the formation or governance of that relation.

The concept extends the earlier distinction between trust and trustworthiness. Even when *A* and *B* are genuinely trustworthy toward one another, their stable relation can remain harmful to *C*. This possibility is especially important for organizations and institutions because trust can increase coordination capacity. Greater capacity can magnify both beneficial and harmful effects.

Proposition 8.8 (No mutuality-to-justice inference). *Mutual trust and reciprocal benefit among the direct parties to a relation do not provide sufficient grounds for judging the wider relational arrangement just when relevant third-party effects exist.*

A Generative Relational analysis therefore resists closing the system boundary too early. The practical question is not that every relation must include every possible outsider. Rather, the boundary of normative analysis should expand when credible pathways connect trust production to externalized vulnerability, resource use, evidential exclusion, environmental effects, or transferred risk. Young's work on structural responsibility offers a useful neighboring resource for thinking about responsibility within processes whose outcomes cannot be reduced to one isolated actor's intention [80]. The present argument is narrower: relational benefits can coexist with wider burden transfer, so justice requires attention to the network through which those burdens travel.

A system that values trust can easily construe refusal to be a defect. A worker who asks for verification may be labeled disloyal; a patient who seeks a second opinion may appear uncooperative; a platform user who rejects data sharing may be burdened with additional friction; a citizen who questions an institution may be told that confidence is necessary for collective functioning. These responses reveal why the capacity not to trust deserves explicit normative protection.

Hirschman's classic distinction among exit, voice, and loyalty provides a useful starting point for analyzing responses to organizational decline [30]. For the present framework, the relevant insight is not that exit always solves relational injustice. Exit can be unavailable, costly, or capable of weakening voice. Instead, both voice and exit are relational capacities whose practical availability affects whether trust remains voluntary and corrigible.

Definition 8.9 (Realistic exit). An actor possesses *realistic exit* from a trust relation when withdrawal or substantial reduction of reliance is not merely formally permitted but can be exercised without costs so prohibitive that the permission becomes practically empty relative to the actor's circumstances.

This definition is context-sensitive. Leaving a friendship, employment relation, digital platform, medical system, financial institution, or political community involves very different alternatives and costs. The framework does not demand costless exit. It demands that analysis distinguish legal or formal permission from effective capacity.

Voice is equally important. Where exit is difficult, participants need channels through which they can question evidence, report violations, contest the design of trust-generating conditions, and seek correction

without automatically being classified untrustworthy. A relation that requires visible demonstrations of confidence in order to retain membership can make honest epistemic updating socially dangerous.

Claim 8.10 (Protected refusal). The ethical governance of trust requires some protected space for withholding, reducing, or revising trust without interpreting every act of refusal to be proof of bad faith, disloyalty, or relational defect.

Protected refusal does not prohibit relational consequences. Repeated refusal can reasonably affect cooperation in some contexts. The claim instead rejects a closed structure in which the only acceptable evidence of belonging is the continued performance of trust.

8.4 Trust Demands, Extraction, Unequal Gains

Trust can be requested, encouraged, expected, or institutionally presupposed. The ethical significance of such demands depends partly on who makes them and what evidence or vulnerability accompanies them. A powerful actor may ask a weaker actor to “trust the process” while controlling the process, the data, the appeal mechanism, and the consequences of refusal. In such a case, the language of trust can obscure an asymmetry in which one party is asked to bear uncertainty that the other party has the capacity to reduce.

The problem is not solved simply by providing more reassurance. Reassurance can itself become another trust-generating intervention. The stronger response is to ask whether the demanding party can improve the relation’s objective trustworthiness, distribute vulnerability more fairly, disclose relevant information, accept reciprocal accountability, or reduce the cost of challenge.

Definition 8.11 (Asymmetric trust demand). An *asymmetric trust demand* occurs when one actor or institution seeks another actor’s reliance, lowered verification, or increased vulnerability while retaining substantially greater control over relevant information, conditions of interaction, remedies, or exit costs.

Not every asymmetric trust demand is wrongful. Emergency settings, specialized expertise, fiduciary arrangements, and complex infrastructures can create unavoidable informational asymmetries. The definition identifies the structure that requires justification. Relevant questions include whether the asymmetry is necessary, whether it is reduced when possible, whether oversight exists, whether errors can be corrected, and whether the less powerful actor has meaningful remedies.

The distinction also helps clarify manufactured vulnerability from Section 7. An institution need not deceive participants to create unjustified exposure. It may accurately describe its procedures while using atmosphere, social norms, friction, or dependency to increase reliance beyond what participants would accept under more symmetric conditions of information and choice.

The recursive framework permits a further possibility. Once trust changes behavior, the resulting relation can generate resources for one party: data, attention, unpaid coordination, loyalty, reputational legitimacy, emotional labor, reduced compliance costs, or access to another actor’s network. These resources can then strengthen the conditions under which further trust is requested. Where the benefits of this cycle are persistently captured by one side while vulnerability and maintenance work are concentrated on the other, the process deserves a distinct diagnostic label.

Definition 8.12 (Trust extraction). *Trust extraction* denotes a relational process in which an actor or institution systematically captures substantial benefits made possible by another party’s trust while shifting a disproportionate share of the associated vulnerability, maintenance work, uncertainty, or correction costs onto that party or onto third parties.

The term is intentionally narrower than exploitation in general. It does not claim that every unequal gain from trust is exploitative, nor does it provide a complete theory of exploitation. Its purpose is to identify a mechanism through which trust itself can become a productive relational resource whose returns and risks are distributed asymmetrically.

Trust extraction can coexist with genuine affection, sincere cooperation, or institutional usefulness. This is precisely why the concept matters. Normative defects need not depend on cynical intention. A

self-reinforcing relation can produce increasingly unequal gains through ordinary routines even when no actor initially designed that outcome.

Remark 8.13 (Intent is not the only normative variable). The absence of manipulative intent does not settle the justice of a trust relation. Recursive processes can generate asymmetric burdens through institutional routines, historical inheritance, network position, or cumulative feedback that no single actor fully controls.

This observation connects the present analysis to wider work on structural responsibility [80]. It also reinforces the need for periodic relational review. A relation that began under acceptable conditions may evolve into a different distribution of dependency and benefit.

8.5 Temporal Justice, Relational Change

Trust attractors are historical objects in the minimal sense that their present behavior depends on prior interaction. This path dependence produces a temporal problem of justice. Current actors often enter relational landscapes they did not create. Some inherit reputational advantages, institutional legitimacy, procedural defaults, or networks of confidence. Others inherit histories of violation, exclusion, surveillance, or broken promises. Equal treatment at time t need not therefore create equal relational conditions.

Let H_t denote the relevant relational history available at time t . The state update can be written schematically as

$$\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t, H_t, \gamma_t; \theta_t). \quad (58)$$

If two groups receive the same present intervention γ_t but possess different histories $H_t^{(1)}$ and $H_t^{(2)}$, the intervention may generate different meanings, risks, and trajectories. A request for trust that appears modest against one history can be unreasonable against another.

This is one reason universal trust-building scripts often fail. Reassurance, ritual, transparency, apology, or shared experience do not enter an empty relational field. Their meaning is filtered through prior expectations and memory. The ethical response cannot be to demand that history disappear for the sake of present cooperation. Nor must inherited distrust remain permanent. The point is that repair and trust production should address the conditions under which historical reasons for distrust can be examined, answered, and revised.

Definition 8.14 (Historical burden of trust production). A *historical burden of trust production* is a present cost, skepticism, or vulnerability requirement whose normative significance depends materially on prior relational events rather than on the current interaction considered in isolation.

The definition allows both directions. A history of reliability can lower the reasonable burden of repeated proof. A history of violation can increase the reasonable burden of demonstration and repair. Justice therefore requires attention to temporal asymmetry rather than a permanent rule that all actors must begin from identical confidence.

The preceding analyses suggest a broader object of normative inquiry. Standard assessment often begins after a relation already exists: is the current exchange fair, is the institution trustworthy, or is the present decision legitimate? A Generative Relational approach also asks how the relation itself was formed, how its update rules are maintained, who can revise it, and under what conditions it can be dissolved.

Definition 8.15 (Relational-change justice). *Relational-change justice* concerns the normative conditions governing how consequential relations are initiated, reinforced, revised, weakened, and ended, including the distribution of power, vulnerability, epistemic standing, benefit, burden, and temporal opportunity across those processes.

This definition does not replace distributive, procedural, corrective, recognitional, or epistemic theories of justice. It identifies a process-level question that cuts across them. The same distribution at one time can arise through very different histories. Likewise, a relation that appears acceptable in a static snapshot may be maintained through update rules that make future correction increasingly difficult.

For trust, relational-change justice directs attention toward at least five moments:

1. **Formation:** Were actors able to understand and meaningfully influence the conditions under which reliance began?
2. **Reinforcement:** Did positive feedback distribute gains and vulnerability acceptably, and did it preserve access to disconfirming evidence?
3. **Revision:** Could actors challenge interpretations, procedures, and field-shaping conditions when circumstances changed?
4. **Repair:** After violation, were acknowledgment, remedy, and renewed vulnerability distributed without forcing the injured party to bear the primary burden of restoration?
5. **Dissolution:** Could the relation be weakened or ended without retaliation, impossible loss, or unjust transfer of costs to others?

These moments need not be governed by identical norms. A medical relationship, family relation, public institution, voluntary association, and digital platform create different rights, duties, dependencies, and exit conditions. The value of the framework lies in retaining the temporal sequence and the possibility that justice changes with the relation.

Proposition 8.16 (Justice is trajectory-sensitive). *Where relational history materially changes vulnerability, evidence, dependency, or the rules of future updating, a normatively adequate assessment cannot rely solely on the current relational state. It must also examine the trajectory and the mechanisms through which that state became stable.*

8.6 Normative Profile, Research, Limits

The analysis so far resists a single objective such as maximizing trust, cooperation, stability, or even exit. Each can be valuable under some conditions and harmful under others. The appropriate next step is therefore a profile of constraints and questions rather than an optimization target.

For a trust-generating intervention g , let

$$\mathcal{N}(g, t) = \left(\Delta \mathbf{V}, \Delta \mathbf{B}, \Delta \mathbf{K}, \Delta \mathbf{C}, \Delta \mathbf{X}, \Delta \mathbf{E}, \Delta \mathbf{R} \right)_t, \quad (59)$$

where the components schematically record changes in vulnerability $\mathbf{V}$, benefit $\mathbf{B}$, epistemic access $\mathbf{K}$, contestation costs $\mathbf{C}$, exit capacity $\mathbf{X}$, third-party externalities $\mathbf{E}$, and revisability $\mathbf{R}$. No weights are fixed here. The profile exists to prevent success on one dimension from making other changes invisible.

A preliminary normative evaluation can then ask whether the intervention:

1. improves or preserves relevant trustworthiness rather than merely confidence;
2. avoids concentrating vulnerability on actors with the least power over the trust-generating field;
3. preserves access to relevant evidence and protects the intelligibility of dissent;
4. keeps correction, voice, and realistic exit available in forms proportionate to the stakes;
5. identifies and limits serious third-party externalities;
6. avoids converting historical disadvantage into a demand for greater present vulnerability; and
7. remains open to revision when the relation's distribution of benefit, dependency, or risk changes.

Claim 8.17 (Normative non-equivalence). An intervention that increases experienced trust, cooperation, or relational stability has not thereby demonstrated an improvement in justice. Normative evaluation must remain multidimensional and trajectory-sensitive.

This non-maximizing approach is intentionally modest. A complete theory would need principles for resolving conflicts among the dimensions. For example, stronger verification can improve epistemic access while increasing privacy costs; easier exit can protect autonomy while weakening incentives for internal voice; and deep relational commitment can create valuable continuity while increasing dependency. The current paper does not resolve these conflicts. It makes them explicit so that trust is not permitted to substitute for all of them.

The normative account also imposes an empirical requirement: justice-relevant variables cannot be reconstructed from an average trust score. Studies should disaggregate who becomes vulnerable and who benefits; examine whether critical speakers face credibility deficits, silencing, or retaliation [22, 17]; distinguish formal permission to exit from realistic exit; trace serious third-party effects; compare present interventions with relevant relational histories; and identify who controls the defaults, guarantees, information channels, interfaces, and revision procedures through which trust is produced.

Research Obligation 8.18 (Justice-sensitive observation). Empirical assessment of trust-generating conditions should record, where relevant, the distribution of vulnerability and benefit, the credibility of dissent, realistic exit capacity, third-party externalities, historical asymmetries, and control over the trust-generating field.

These obligations also constrain formal modeling. A model with a single trust variable can illuminate some dynamics while making distributive and epistemic asymmetries impossible to represent. The appropriate level of complexity therefore depends on the research question. Parsimony is valuable, but a model cannot answer a justice question after removing every variable through which injustice could appear.

The paper can now state a limited normative orientation without claiming a complete theory of just trust. Trust-generating conditions are ethically more defensible when they improve the grounds for warranted reliance, distribute vulnerability and benefit in ways that can be justified to affected actors, preserve epistemic openness, protect meaningful challenge, and keep relations revisable when conditions change. They are ethically more concerning when they produce confidence by narrowing evidence, shifting exposure to weaker actors, penalizing refusal, hiding third-party costs, or making an already asymmetric relation increasingly difficult to contest.

Proposition 8.19 (Preliminary justice orientation). *Governance of trust-generating conditions should not seek relational stability independently of the processes that create and maintain it. Normative assessment should examine whether affected actors can know, question, revise, and where appropriate leave consequential relations without bearing unjustifiably concentrated vulnerability or transferring serious unacknowledged costs to others.*

The proposition does not imply that every relation must be continuously negotiated. Durable expectations and routines are among the benefits of trust. Nor does it require equal vulnerability in every interaction. The proposal is instead directed against a narrower error: taking the successful generation of confidence to be sufficient evidence that the relational process producing it is legitimate.

The broader implication is that trust has both generative power and normative opacity. Because it changes behavior, evidence, interpretation, and future possibility, trust can help create the reality within which its continuation appears reasonable. That recursive achievement can support cooperation and care. It can also make inequity harder to perceive. Justice therefore enters not only at the moment of deciding whether to trust, but throughout the history in which the relation becomes capable of sustaining itself.

9 Measurement, Identification, Falsification

The framework becomes useful only if its central claims can be separated empirically. Persistence is not by itself evidence of an attractor; correlation across time is not by itself evidence of self-reinforcement; a memorable event is not by itself evidence of poetic composition. Research design must therefore identify the relational trajectory, the proposed feedback channel, the relevant perturbations or contrasts, and the outcomes through which the claim could fail.

This task is demanding because experienced trust, perceived trustworthiness, actual reliability, vulnerability, monitoring, institutional assurance, interpretation, affect, and exit conditions can move together while remaining analytically distinct. Trust can also change the interactions from which later evidence is produced. The section therefore develops a plural empirical programme: multidimensional measurement, temporally adequate observation, feedback-channel identification, perturbation tests for attractor claims, sequence and context tests for poetic composition, and governance outcomes that include protection, evidence access, contestability, and exit rather than trust scores alone.

9.1 Trajectories, Measurement, Separation

A natural unit of observation for the present framework is not an isolated trust judgment but a relational trajectory. Let

$$\mathcal{T}_{0:T} = (\mathbf{x}_0, \mathbf{a}_0, \mathbf{y}_0; \mathbf{x}_1, \mathbf{a}_1, \mathbf{y}_1; \dots; \mathbf{x}_T) \quad (60)$$

represent an observed history over a specified domain. Here $\mathbf{x}_t$ contains the trust-relevant relational configuration, $\mathbf{a}_t$ actions, and $\mathbf{y}_t$ observations generated through interaction. The empirical meaning of each component must be specified for the setting under study. A workplace team, a patient–clinician relation, a humanitarian organization, an online platform, and an interorganizational partnership need not instantiate the same coordinates.

A trajectory perspective changes the research question. Instead of asking only whether trust is high or low, one can ask which histories lead toward similar regions, which disturbances produce departure, which processes support return, and whether the rules of updating themselves change after major events. This orientation is consistent with the longitudinal evidence already present in organizational trust research. Serva, Fuller, and Mayer, for example, studied reciprocal relations among trust, perceived trustworthiness, and risk-taking over time rather than reducing trust to one cross-sectional observation [68]. Ferrin, Bligh, and Kohles likewise examined spiraling relations between perceived trustworthiness and cooperation [21]. The present framework asks for an additional step: identify how those reciprocal processes alter later relational possibilities and whether their consequences produce recoverable regions, path dependence, or closure.

Research Obligation 9.1 (Trajectory-level observation). When a claim concerns self-reinforcement, path dependence, repair, or attractor formation, empirical observation should include sufficient temporal structure to distinguish a trajectory from a repeated measurement of one static state.

The requirement does not imply that every study must be long in calendar time. Some relations update rapidly. Repeated strategic interaction can generate a meaningful history within minutes, while institutional trust may require years or decades. What matters is that the temporal resolution corresponds to the mechanisms under examination. Nor should duration itself substitute for reinforcement: a meta-analysis of exchange relationships finds only a small average association between relationship duration and trust, with substantial variation across studies and several potentially countervailing mechanisms [77].

The theory repeatedly distinguishes experienced trust, trustworthiness, justification, vulnerability, institutional assurance, and justice. Empirical measurement should preserve those distinctions where possible. Let a study record an observation vector

$$\mathbf{m}_{i,t} = (\tau_{i,t}, \omega_{i,t}, \nu_{i,t}, \mu_{i,t}, \kappa_{i,t}, \chi_{i,t}), \quad (61)$$

where τ denotes experienced trust, ω observable indicators of relevant trustworthiness, ν vulnerability or exposure, μ monitoring or verification intensity, κ epistemic access, and χ realistic exit capacity. These symbols are placeholders for domain-specific measures rather than a claim that the dimensions admit a common scale.

The distinction is necessary because several theoretically important patterns are invisible in a trust-only outcome. Experienced trust may rise while verification access falls. Cooperation may increase while one subgroup becomes more exposed. Institutional assurance may raise confidence without changing the behavior of the underlying actor. A repair intervention may improve reported trust while leaving the violated procedure unchanged. Conversely, actual trustworthiness may improve before experienced trust recovers.

Claim 9.2 (Non-equivalence of empirical outcomes). An observed increase in experienced trust should not be taken to provide sufficient evidence of increased trustworthiness, improved protection, reduced vulnerability, or greater justice.

A useful design therefore combines self-report with behavioral, institutional, and relational indicators. Depending on the setting, these may include willingness to disclose information, delegation, acceptance of risk, frequency of verification, error rates, fulfillment of commitments, complaint access, appeal outcomes, concentration of dependence, response to negative evidence, and the material cost of exit. Qualitative material can add dimensions that standardized instruments may miss, especially the changing meaning of events across different histories.

The objective is not methodological maximalism. More variables do not automatically produce better explanation. The purpose of dimensional separation is to prevent a single trust score from carrying theoretical burdens that belong to different objects.

9.2 Identification, Self-Reinforcement, Attractors

The central empirical challenge is to distinguish self-reinforcement from ordinary temporal persistence. A high correlation between τ_t and τ_{t+1} is insufficient. Persistence can arise from stable personality, unchanging external conditions, repeated exposure to an independently trustworthy actor, institutional constraint, measurement inertia, or many other processes. The present framework requires a mediated loop through which the current relation changes conditions relevant to its own later continuation.

Let R_t denote a candidate reinforcement channel, such as disclosure, cooperation, reciprocal vulnerability, interpretive leniency, reduced monitoring, institutional assurance, or accumulated shared history. A minimal causal sequence can be represented schematically by

$$\mathbf{x}_t \longrightarrow R_t \longrightarrow \mathbf{z}_{t+1} \longrightarrow \mathbf{x}_{t+1}, \quad (62)$$

where $\mathbf{z}_{t+1}$ contains conditions or evidence affected by the channel. The arrows in Equation (62) represent a formal causal hypothesis, not a substitute for verbal explanation. The substantive claim must identify what each link means in the empirical setting.

For example, a disclosure mechanism might predict that higher trust increases voluntary disclosure; disclosure creates opportunities for protective or exploitative response; the response becomes evidence; and that evidence changes later trust. A cooperation mechanism might predict that trust permits coordination that would otherwise not occur; successful coordination changes perceived reliability; and the changed perception alters willingness to cooperate again. A monitoring mechanism might predict that trust lowers verification; reduced verification changes which failures become observable; and the resulting evidential environment affects later trust.

Three types of contrast are especially useful.

Channel interruption. Intervene on the proposed mediator while holding other conditions as comparable as possible. If reduced monitoring is the proposed reinforcing mechanism, introducing independent audit without changing the partner's behavior can test whether persistence depends partly on evidential thinning. If reciprocal vulnerability is proposed, designs can vary whether one party has an opportunity to reciprocate.

Alternative-path comparison. Compare relations that reach similar present trust levels through different histories. If the framework is correct, equal current trust need not imply equal responses to future perturbation. A relation stabilized through repeated verified reliability may react differently from one stabilized primarily through symbolic assurance or dependence.

Mechanism substitution. Replace one trust-supporting pathway with another. For example, increase actual monitoring quality while reducing reassuring presentation, or strengthen institutional protection while leaving interpersonal familiarity unchanged. Such designs help distinguish confidence generated by substantive protection from confidence generated by fluency or atmosphere.

Research Obligation 9.3 (Feedback-channel identification). A claim of relational self-reinforcement should identify at least one trust-dependent mediator whose alteration changes the probability, resilience, or recoverability of later trust-relevant configurations.

This obligation also creates a route to falsification. If proposed mediators do not change later trajectories, the self-reinforcement claim should be weakened or abandoned for that setting. The framework should not redescribe every instance of stability after the fact.

The term *attractor* imposes a stronger evidential burden than *feedback*. A feedback loop can exist without producing an attractor, and an observed cluster of states can occur without attraction. Standard dynamical systems terminology associates attraction with the behavior of trajectories from a neighborhood of initial conditions [71]. In complex empirical systems, exact reconstruction of a phase portrait may be impossible. The weaker notion of a trust-attractor candidate therefore remains useful, but it should be tied to observable signatures.

At least five signatures deserve attention.

Local return after perturbation. After a disturbance that displaces the relation, trajectories tend to return toward a similar region without requiring complete external reset. The speed, path, and degree of return may vary.

Neighborhood dependence. Relations beginning from sufficiently similar configurations display some convergence in later relational behavior. This does not require identical outcomes, but attraction should imply more than persistence of a single case.

Basin sensitivity. The effect of a perturbation depends on the region from which the trajectory starts. The same ambiguous event may be absorbed in one history and trigger a large transition in another.

Multistability or hysteresis. Different stable relational regimes can persist under similar external conditions, and reversing an external change may not immediately restore the prior relation. Historical lock-in, increasing returns, and path dependence provide neighboring theoretical resources for interpreting such patterns [3].

Changing recovery near transition. When a relation approaches a qualitative transition, recovery from small perturbations may change. Work on critical transitions in complex systems shows why altered recovery behavior can sometimes precede regime shifts, although those results cannot simply be imported into trust without independent validation [67]. In trust research, the appropriate question is whether empirically observed recovery dynamics carry information about increasing fragility in the specific relational system.

These signatures should be considered jointly rather than turned into a checklist that automatically proves attraction. A study can also encounter cyclic, drifting, metastable, or externally driven dynamics. Those outcomes are not failures of research. They indicate that the attractor vocabulary is too strong or incomplete for the observed case.

Definition 9.4 (Empirical trust-attractor candidate). An *empirical trust-attractor candidate* is a recurrent relational region for which observed trajectories provide evidence of local return, path-sensitive resilience, or neighborhood-level convergence through identifiable endogenous mechanisms, while alternative explanations remain explicitly examined.

Research Obligation 9.5 (Attractor alternatives). Studies invoking trust-attractor language should compare the attractor hypothesis with plausible alternatives such as stable exogenous conditions, measurement persistence, strategic equilibrium, institutional constraint, seasonal recurrence, or unobserved common causes.

9.3 Perturbation, Rupture, Temporal Order

Sections 3 and 6 placed perturbation and repair near the center of the dynamical account. Empirically, perturbations are valuable because they reveal properties that are difficult to observe under equilibrium-like conditions. A relation can appear stable while its resilience remains unknown.

Perturbations need not be artificially harmful. Ethical research can use low-stakes ambiguity, controlled information delays, mild coordination failure, changes in verification access, or naturally occurring disruptions. Historical and organizational studies can analyze actual crises without producing them. The critical design question is whether the disturbance is theoretically linked to the domain of trust under study.

Several quantities may be informative. Let d_t denote distance from a specified trust-supporting relational region $\mathcal{S}_T$. Following a perturbation at time t_0 , one may examine a recovery profile

$$\rho(k) = d_{t_0+k}, \quad k = 0, 1, \dots, K. \quad (63)$$

The function ρ need not be monotonic. Oscillation, overshoot, delayed repair, and partial recovery may all carry substantive meaning. Comparative work can ask whether histories, institutional guarantees, or repair practices change the shape of ρ .

Repair requires additional observations because a return in reported trust does not show that the relation has returned to its previous configuration. The post-repair relation may contain new monitoring practices, altered memory, changed vulnerability, institutional safeguards, or domain-specific distrust. A successful design should therefore compare configuration profiles rather than one scalar before and after repair.

Claim 9.6 (Repair leaves historical residue). Where rupture and repair alter memory, verification practices, institutional conditions, or vulnerability, equal pre-rupture and post-repair trust scores do not imply restoration of the same relational state.

This claim creates several empirical questions. Do repaired relations become more robust to similar future violations? Do they become more cautious while remaining cooperative? Does apology work differently after competence-related and integrity-related breaches, consistent with established trust-repair research [36, 75]? Do organizational reforms change the conditions of later evaluation, not merely immediate confidence [23]? These questions connect the framework to an existing empirical literature while extending attention toward longer relational trajectories.

A major claim of the paper is that history can matter through more than the accumulation of positive and negative evidence. The order, spacing, recurrence, and framing of events may change their relational significance. This claim is empirically testable.

Suppose two histories contain the same multiset of events

$$\mathcal{E} = \{e_1, e_2, \dots, e_m\}, \quad (64)$$

but present those events in different orders π and π' . Let $U(\pi(\mathcal{E}))$ denote a later trust-relevant update. A temporal-order hypothesis predicts that, for some histories,

$$U(\pi(\mathcal{E})) \neq U(\pi'(\mathcal{E})). \quad (65)$$

This is the empirical counterpart of temporal non-commutativity introduced in Section 4.

Experimental designs can vary whether a supportive act occurs before or after an ambiguous failure, whether a guarantee precedes or follows disclosure, whether repair begins immediately or after denial, or whether repeated small acts precede one salient symbolic event. Longitudinal observational designs can use naturally varying sequences while accounting for selection processes as carefully as possible.

History should also be studied through memory and reinterpretation. Participants may later revise what earlier events meant. Interviews, diaries, archival records, message histories, and repeated narrative elicitation can reveal this process. The purpose is not to privilege retrospective accounts over behavioral data. It is to observe a component of relational state that may itself change through time.

Research Obligation 9.7 (History-preserving design). When the hypothesis concerns sequence, memory, or reinterpretation, data collection should preserve event order and relevant contextual links rather than reducing the history prematurely to cumulative counts or averages.

A null result is informative. If permutations of theoretically relevant events produce no detectable difference under adequate power and measurement, then claims about temporal composition should be narrowed for that domain.

9.4 Poetic Composition, Design, Manipulation

The concept of poetic relational composition is especially vulnerable to post hoc overinterpretation. Any memorable event can be given the label “poetic” after it changes a relation. Empirical work therefore needs prospective criteria.

Section 4 defined the concept through heterogeneous, partial, indirect, temporally dispersed, and context-sensitive elements whose composition can acquire relational significance. At least four empirical properties follow.

Heterogeneity. The relevant configuration contains cues from more than one modality or relational register, such as language, timing, gesture, space, memory, institutional symbolism, or embodied response.

Context dependence. The effect of an element changes with relational history or surrounding cues. A gesture that raises trust in one trajectory may be neutral or threatening in another.

Compositional interaction. The joint effect of elements cannot be adequately described by simply adding their isolated average effects. Interaction does not by itself prove poetic structure, but it provides one measurable implication of composition.

Retrospective integration. Later events can change the significance assigned to earlier events. What was initially ordinary may become evidence of care, manipulation, warning, or continuity after subsequent developments.

A factorial experiment can test some of these properties by varying cue combinations and histories. Sequence experiments can test order dependence. Experience-sampling methods can record when participants notice particular cues. Qualitative process tracing can examine how disparate events become linked in later interpretation. Physiological or behavioral measures may be useful in some domains, but they should not be privileged direct readouts of trust.

Let $e_1, \dots, e_m$ denote components of a candidate composition and $\Delta\tau(e_j)$ their isolated average trust effects. A simple interaction contrast is

$$\mathcal{C} = \Delta\tau(e_1, \dots, e_m) - \sum_{j=1}^m \Delta\tau(e_j). \quad (66)$$

A nonzero $\mathcal{C}$ can indicate interaction, but it is not sufficient for the theoretical claim. The study must still show why the elements are relationally heterogeneous, how context changes their meaning, and whether the composition participates in later trajectory change.

Research Obligation 9.8 (No post hoc poetic labeling). A study invoking poetic relational composition should specify in advance which features make the configuration heterogeneous, context-sensitive, sequential, or symbolically integrated, and should report null or contrary results without reclassifying them after observation.

This requirement protects the concept from becoming a decorative synonym for complexity. The claim is narrower: some trust-relevant changes may depend on relational composition whose significance is not recoverable from isolated cues alone.

Section 7 argued that institutions can arrange conditions under which trust becomes easier or harder to generate. Empirical work should therefore study not only participants but also designers, incentive structures, and the architecture of the environment.

One useful design separates at least three layers:

1. changes in experienced trust or felt safety;
2. changes in actual trustworthiness, protection, or accountability; and
3. changes in the participant's capacity to inspect, contest, refuse, or exit.

The layers can move in different directions. An interface can become smoother and more reassuring while cancellation becomes harder. An institution can add symbolic transparency while reducing access to underlying records. A guarantee can lower perceived risk while transferring actual risk to a third party.

Studies of deceptive or manipulative interface design already demonstrate the importance of examining design choices rather than user attitudes alone. Work on dark patterns, for example, documents design practices that steer user behavior through interface structure [49]. The present framework extends the question toward trust: which design features raise confidence, which improve actual protection, and when the two diverge.

A particularly informative design compares reassurance-only interventions with substantive institutional improvements. Let G denote a presentational or atmospheric intervention and W a change in underlying trustworthiness or protection. A factorial design can compare

$$(G, W) \in \{0, 1\} \times \{0, 1\}. \quad (67)$$

If $G = 1, W = 0$ substantially increases experienced trust while verification, protection, and behavior remain unchanged, the result provides evidence of a trust-production mechanism distinct from substantive improvement. If $G = 0, W = 1$ improves protection before confidence changes, the design reveals a lag between trustworthiness and experienced trust. The combined condition can test interaction between presentation and substance.

The ethical design of such studies requires care. Researchers should not expose participants to significant unprotected vulnerability merely to observe trust. Low-risk simulations, archival analysis, natural experiments, and field studies with existing institutional variation will often be preferable.

9.5 Distribution, Networks, History

The justice claims in Section 8 create additional data requirements. Dyadic measures are insufficient when trust relations generate externalities or when one party controls the field in which trust is produced. Empirical work should therefore ask who receives benefits, who bears exposure, who performs verification work, who controls information, who can challenge the arrangement, and who can leave without disproportionate loss.

For actors $i = 1, \dots, n$, let

$$\mathbf{j}_{i,t} = (b_{i,t}, v_{i,t}, r_{i,t}, c_{i,t}, x_{i,t}, p_{i,t}), \quad (68)$$

where b denotes benefit, v vulnerability, r verification or repair burden, c contestation cost, x realistic exit capacity, and p influence over trust-generating conditions. The vector should not be collapsed into a single justice score unless an independently defended normative theory justifies such aggregation.

Third-party analysis is especially important. Suppose trust between A and B reduces their transaction costs but increases risk for C . A study that samples only A and B can report a successful trust-building intervention while missing the principal normative burden. Network, supply-chain, organizational, community, or policy-level data may be necessary to identify such externalities.

Research Obligation 9.9 (Boundary justification). Researchers should justify the relational boundary within which benefits, burdens, trust, and vulnerability are measured. When plausible externalities extend beyond that boundary, the limitation should be stated rather than silently interpreted to mean absence of effect.

Qualitative work has a special role here because burdens can be difficult to anticipate from formal role descriptions. Informal emotional labor, adaptation to opaque procedures, reputational risk, fear of retaliation, and dependence on unofficial intermediaries may sustain a trustful arrangement without appearing in administrative data.

Trust relations are nested. Interpersonal trust can depend on organizational rules; organizational trust can depend on professional communities, regulators, infrastructures, or political institutions. Conversely, repeated local interactions can alter organizational routines and collective expectations. A relational account should therefore avoid assuming that the appropriate scale is fixed in advance.

Let $\mathbf{x}_t^{(s)}$ denote a relational configuration at scale s . Cross-scale coupling can be represented schematically by

$$\mathbf{x}_{t+1}^{(s)} = \Phi^{(s)}\left(\mathbf{x}_t^{(s)}, \mathbf{x}_t^{(s-1)}, \mathbf{x}_t^{(s+1)}, \mathbf{u}_t^{(s)}\right). \quad (69)$$

The notation does not assume a literal hierarchy in every case. It records the possibility that neighboring scales constrain one another.

Empirically, this means that apparent interpersonal distrust may reflect institutional history rather than a dyadic failure, while apparent institutional trust may depend on a dense network of local intermediaries. Network designs can examine whether trust-supporting relations cluster, whether bridging actors transmit assurance or suspicion, and whether a perturbation propagates across relational boundaries. Multilevel qualitative work can trace how the same event receives different meaning at different scales.

Scale also matters for justice. A highly trustful team can be embedded in an organization that transfers costs outward. A trusted institution can rely on precarious intermediaries. The relational unit that best explains local trust may therefore differ from the unit required for normative evaluation.

Some of the strongest tests of the framework may come from comparative historical cases. Institutions often leave records of scandals, reforms, public inquiries, complaints, procedural changes, communication campaigns, changes in monitoring, and subsequent shifts in participation. Such cases can reveal whether trust repair changes the relational landscape over periods that are impractical for experiments.

A useful comparative design selects cases that differ in mechanism rather than only outcome. Two institutions may both recover public confidence, but one may do so through demonstrable procedural reform while another relies primarily on leadership change, symbolic recommitment, or favorable external events. Two communities may both sustain low trust, but one through repeated direct violations and another through inherited historical memory. Comparing these paths can test whether similar present states conceal different future susceptibilities.

Historical reconstruction also permits study of relational memory. Archives do not provide transparent access to the past; what was recorded, by whom, and under what institutional conditions matters. A trust history should therefore include the production and preservation of evidence itself when those processes shape later interpretation.

Process tracing can be especially useful when the hypothesis concerns a sequence of mechanisms: an initial guarantee permits interaction; interaction produces shared experience; shared experience lowers monitoring; lower monitoring conceals a later failure; scandal produces rupture; reform changes verification; and later trust develops under the revised system. The objective is not to force the case into a predetermined cycle. It is to determine which links are supported and where the proposed mechanism fails.

9.6 Falsification, Research Programme, Restraint

A generative relational framework is useful only if observations can count against it. Several results would require revision of the present claims.

First, if trust persists but interruption of proposed endogenous channels does not alter later persistence, the self-reinforcement hypothesis is weakened. Stable exogenous conditions may provide a better account.

Second, if relations displaced by comparable perturbations show no tendency for return, basin-like language may be inappropriate even when trust is persistent.

Third, if different relational histories with similar present states respond similarly to theoretically relevant perturbations, claims about path dependence should be narrowed.

Fourth, if sequence manipulations produce no meaningful differences, temporal non-commutativity should not be asserted for that domain.

Fifth, if heterogeneous symbolic or atmospheric compositions have effects that are fully explained by isolated cues and standard additive models, the stronger claim of poetic relational composition is unnecessary.

Sixth, if trust-building interventions increase experienced trust only when underlying protection or trustworthiness improves, concerns about manufactured trust may be less important in that setting than the framework anticipates.

Seventh, if a relation remains highly stable while continuously admitting counterevidence, protecting dissent, and supporting low-cost exit, stability should not be taken to indicate epistemic closure or domination.

Proposition 9.10 (Framework corrigibility). *The concepts developed in this paper should be retained only where they improve explanation relative to simpler alternatives. Relational complexity does not justify theoretical complexity by itself.*

Boundary conditions are equally important. Trust dynamics may differ when interactions are anonymous rather than repeated, when exit is impossible rather than easy, when stakes are existential rather than modest, when legal duties replace discretion, or when actors possess radically asymmetric information. The same vocabulary can organize questions across these domains without assuming identical mechanisms.

Null results should be reported substantively. Failure to observe attraction, sequence dependence, or poetic composition can identify domains where simpler models suffice. Such findings help prevent the framework from becoming immune to evidence.

The paper's claims can be developed through a staged programme rather than one omnibus study.

Stage I: construct valid relational measurements. Identify domain-specific dimensions of trust, trustworthiness, vulnerability, monitoring, evidence access, and exit. Establish which dimensions can be observed repeatedly without distorting the relation excessively.

Stage II: identify feedback mechanisms. Use longitudinal observation, experiments, or quasi-experiments to test whether present trust changes disclosure, cooperation, interpretation, monitoring, or other mediators that alter later relational conditions.

Stage III: study perturbation and return. Observe how relations respond to controlled, natural, or historical perturbations. Compare recovery profiles across different prior histories.

Stage IV: test path dependence and composition. Vary event order, contextual cues, symbolic configurations, and relational histories. Determine where temporal sequence and heterogeneous composition add explanatory power.

Stage V: examine designed trust environments. Compare changes in experienced trust with independent indicators of trustworthiness, protection, verification access, contestability, and exit. Identify designer-side incentives and asymmetries.

Stage VI: widen the normative boundary. Measure distributional effects, third-party burdens, inherited histories, and power over relational conditions. Determine whether successful trust production creates or conceals relational injustice.

Stage VII: develop formal models only where evidence warrants. Use the accumulated empirical structure to decide which variables, feedbacks, and state spaces deserve formal representation. Formalization should compress supported mechanisms rather than precede them with unnecessary precision.

This order is not mandatory. Historical cases may motivate formal models before experiments are available; simulations may reveal which observations are most informative; qualitative work may overturn an assumed measurement model. The staging instead expresses a methodological preference: increasingly strong formal claims should be accompanied by increasingly specific empirical commitments.

The concepts developed throughout the paper are intentionally evocative. *Attractor*, *landscape*, *poetic composition*, and *self-reinforcement* can make relational phenomena newly visible. Their analytical value, however, depends on discipline in use. An attractor should organize questions about return and basin structure. Poetic composition should organize questions about heterogeneous contextual meaning. Self-reinforcement should identify a feedback mechanism. Governance should specify which conditions are changed and by whom. Justice should identify distributions, powers, and affected parties rather than merely attach a favorable or unfavorable label.

The empirical programme therefore returns the paper to its initial claim in a more demanding form. Trust can participate in generating the conditions of its own continuation, but whether it does so, through which channels, for how long, and with what consequences must be shown rather than assumed. Some relations may exhibit strong endogenous reinforcement. Others may persist because the external environment remains stable. Some may contain multiple locally stable regimes. Others may drift without any attractor-like structure. Some poetic compositions may reorganize relational meaning; others may be theoretically interesting but empirically inert.

The same openness applies normatively. A trust attractor can support valuable coordination, mutual vulnerability, and repair. It can also stabilize manipulation, suppress correction, or distribute costs unfairly. Empirical research must therefore observe not only whether trust forms, but what kind of relation is being generated around it.

Research Obligation 9.11 (Four-part reporting). Research drawing on the present framework should, where feasible, report four separable layers: the observed relational trajectory, the proposed generative mechanism, the evidence for dynamical stability or transition, and the normative distribution of vulnerability, agency, and consequences.

This obligation prepares the concluding stage of the paper. The theoretical contribution is not a claim that trust always forms an attractor or that all trust possesses poetic organization. It is a proposal to study trust through its generative relations, recursive histories, compositional conditions, and powers of self-transformation, while preserving empirical and normative routes for saying when those proposals do not fit.

10 Time, Possibility, Commitment, Autonomy

The previous sections examined trust through relational states, recursive feedback, rupture, governance, justice, and empirical identification. A further consequence follows once history is allowed to alter not only the present relation but also the range of futures that the relation can reach. Trust can open possibilities by enabling disclosure, coordination, specialization, and long horizons. It can also close possibilities by concentrating dependence, reducing alternatives, and making revision costly. The relevant object is therefore not only the current level of trust but the future relational capacity generated by a trajectory.

10.1 Possibility, Horizons, Evidence

Let $\mathbf{x}_t$ denote the relational state and θ_t the historically formed update rules introduced earlier. We add a third object, the practically reachable possibility set Ω_t :

$$\Omega_t = \Omega(\mathbf{x}_t, \theta_t, H_t, E_t), \quad (70)$$

where H_t is relational history and E_t collects material, institutional, and social conditions. Two relations can therefore have similar present trust states while supporting different futures. One may preserve alternative partners, independent evidence, and feasible exit; another may have become organized around one increasingly indispensable relation.

This motivates a three-level description:

$$\mathbf{x}_{t+1} = \Phi(\mathbf{x}_t; \theta_t), \quad (71)$$

$$\theta_{t+1} = \Psi(\theta_t, \mathbf{x}_t, H_t), \quad (72)$$

$$\Omega_{t+1} = \Xi(\Omega_t, \mathbf{x}_t, \theta_t, a_t, E_t). \quad (73)$$

The first equation concerns movement within a relational landscape, the second concerns historical change in the update rules, and the third concerns changes in what future relations and actions are practically reachable. The third level is especially important for a Generative Relational account: a relation can generate opportunities that did not previously exist, but it can also eliminate options whose disappearance becomes visible only later.

Trust often expands Ω_t . A collaborator may gain access to information, roles, communities, or projects that would otherwise remain unavailable. Institutional trust can lower transaction costs and make complex forms of coordination possible [46]. Yet distrust can also preserve possibilities when it motivates redundancy, independent verification, or refusal of premature commitment. The normative question is therefore not whether trust enlarges the number of options. It concerns the quality, diversity, reversibility, and distribution of future relational capacity.

Evidence production is part of the same dynamics. When trust changes whom an actor speaks with, what is disclosed, which institutions are consulted, or which alternatives are abandoned, it changes the future evidence available for revising that trust. Hence:

$$\mathcal{E}_{t+1} = \mathcal{E}(\mathcal{E}_t, \Omega_t, a_t, H_t), \quad (74)$$

where $\mathcal{E}_t$ is the accessible evidence set. A narrow relational horizon may therefore produce a narrow evidential horizon. This is one reason self-reinforcing trust can become informationally closed without any deliberate deception.

The distinction between available and perceived possibilities is also necessary. Let Ω_t^A denote materially or institutionally available futures and Ω_t^P futures that actors can presently recognize. Then $\Omega_t^P \subseteq \Omega_t^A$ may hold because of limited information, learned dependence, fear, or institutional invisibility. Governance that protects relational possibility must therefore address both actual alternatives and the capacity to perceive them.

Poetic composition enters here in a specific way. A small event can alter the salience of a future that was already technically available but previously unrecognized. A remembered gesture, a shared difficulty, or a new symbolic frame can redirect attention toward a different relational trajectory. This does not create every material possibility, but it can change which possibility becomes actionable. The compositional account in Section 4 thus extends from present meaning to future relational orientation.

10.2 Commitment, Irreversibility, Hysteresis

Possibility preservation cannot be the only norm. Some valuable relations require durable commitment, sunk investment, specialization, and the deliberate closing of alternatives. Classic work on commitment emphasizes the role of side bets, investments, and alternatives [6, 65, 66]. The difficulty is that productive

commitment and harmful lock-in can have similar observable features at one moment. Both may involve persistence, high exit costs, and narrowed alternatives.

We therefore distinguish trust from commitment. Trust concerns a relation to vulnerability and expectation; commitment concerns the willingness to sustain or prioritize a relation through time. They can reinforce each other without being equivalent. A compact state representation is

$$\mathbf{z}_t = (\mathbf{x}_t, c_t, e_t, r_t), \quad (75)$$

where c_t represents commitment, e_t exit cost, and r_t reconfiguration capacity.

Commitment can transform the relational horizon through an operator $\mathcal{C}$:

$$\Omega_{t+1} = \mathcal{C}(\Omega_t, c_t, I_t), \quad (76)$$

where I_t includes accumulated investments and obligations. Some possibilities disappear because resources have been allocated elsewhere; others appear only because the commitment exists. Long-horizon collaboration, intimate care, and institutional specialization commonly have this structure.

Irreversibility is therefore multidimensional. Financial, informational, legal, social, emotional, and temporal investments need not move together. A relation can be easy to leave legally but hard to leave epistemically because alternative information channels have disappeared. Another can be socially costly to exit while remaining materially reversible. Rather than one scalar, an irreversibility profile is more informative:

$$\mathbf{q}_t = (q_t^m, q_t^i, q_t^s, q_t^l, q_t^e, q_t^r). \quad (77)$$

Path dependence appears when entry and exit thresholds differ. A relation may require strong initial evidence before commitment, yet once investments accumulate, much stronger negative evidence may be needed to reverse it. Such hysteresis can be rational when commitment creates real value, but it can also shelter deteriorating relations. The empirical task is to determine whether persistence reflects newly generated value, increasing exit costs, interpretive shielding, or a mixture of these mechanisms.

The relevant normative concept is relational option value: not the maximization of alternatives, but preservation of enough high-quality options to keep later revision meaningful. A useful profile includes diversity of alternatives, recoverability, evidence access, and the distribution of reconfiguration costs. This yields a governance principle of staged commitment: larger, harder-to-reverse commitments should normally be preceded by bounded trials, review points, independent information, and protected rollback where feasible. Such design does not oppose trust. It protects the conditions under which trust can remain warranted after commitment deepens.

10.3 Dependence, Autonomy, Reconfiguration

Dependence does not by itself negate autonomy. Relational-autonomy scholarship emphasizes that agency develops within social relations rather than outside them [47, 11, 58]. The relevant question is whether dependence leaves actors with meaningful capacities to interpret their situation, raise objections, renegotiate terms, obtain independent evidence, and reshape the relation.

Power-dependence theory provides a complementary insight: power can emerge from asymmetries in the availability and value of alternatives [19]. For trust dynamics, a dependence profile can be represented by

$$\mathbf{d}_{i,t} = (d_{i,t}^m, d_{i,t}^e, d_{i,t}^s, d_{i,t}^i, d_{i,t}^l), \quad (78)$$

covering material, epistemic, social, institutional, and legal dependence. The profile matters more than a single total because different forms of dependence generate different vulnerabilities and repair possibilities.

Relational autonomy can likewise be represented through capacities rather than independence:

$$\mathbf{a}_{i,t} = (v_{i,t}, k_{i,t}, n_{i,t}, g_{i,t}, r_{i,t}), \quad (79)$$

where v denotes effective voice, k independent knowledge access, n meaningful alternatives, g negotiated-governance capacity, and r reconfiguration capacity. O'Neill's account of autonomy and trust is relevant

here because trustworthy institutions can support agency precisely by making dependence more intelligible and answerable [59].

Trust can improve this profile. Reliable relations can provide resources, capabilities, confidence, and access that expand what an actor can do. Yet successful trust can also produce adaptive narrowing: independent channels are gradually dropped, alternatives atrophy, and the actor becomes less able to evaluate the relation except through the relation itself. Epistemic dependence is especially consequential because it changes not only the costs of exit but the capacity to know whether exit should be considered. Fricker’s account of epistemic injustice helps illuminate the unequal credibility conditions that can intensify this problem [22].

The resulting concept is *reconfiguration capacity*: the practical ability to revise the structure of a relation without requiring complete rupture. We can write

$$\mathcal{R}_{i,t} = R(v_{i,t}, k_{i,t}, n_{i,t}, g_{i,t}, b_{i,t}), \quad (80)$$

where $b_{i,t}$ represents buffers such as savings, legal protection, institutional support, or redundant social ties. Exit remains important [30], but voice and internal reconfiguration often become relevant earlier than full withdrawal.

This distinction also refines the attractor framework. A stable trust configuration can support autonomy when repeated return preserves evidence, voice, and revision. Another equally stable configuration can erode autonomy when return depends on narrowing alternatives or suppressing independent interpretation. Stability alone therefore cannot distinguish a generative relation from a dependency trap.

Proposition 10.1 (Possibility-sensitive evaluation). *Evaluation of durable trust should examine not only present cooperation or experienced confidence, but also the future possibility set, the distribution of irreversibility, and the parties’ reconfiguration capacities.*

This proposition links the temporal extension back to Sections 7–9. Governance should not attempt to keep every future open, nor should it celebrate commitment merely because it is durable. It should protect the relational capacities through which commitment can remain revisable, evidence can still enter, and dependence does not silently eliminate agency. The unresolved question of when trust should remain deliberately unsettled returns in Section 14.3.1.

11 Collectives, Networks, Scales, Institutions

Trust often operates beyond a dyad. Groups coordinate through partially overlapping interpersonal relations, roles, procedures, reputations, and institutional guarantees. Scaling the framework therefore requires more than averaging individual trust. The collective problem concerns how local relations combine, how network structure changes propagation, and how signals move across levels without becoming equivalent across those levels.

11.1 Collectives, Networks, Distributed Memory

Let a trust network be $G_t = (V_t, E_t, W_t)$, where nodes represent relevant actors or roles, edges represent relations, and W_t contains multidimensional relational weights. A collective state is not generally recoverable from the mean edge value:

$$\mathbf{x}_t^C \neq \frac{1}{|E_t|} \sum_{e \in E_t} \mathbf{x}_{e,t}. \quad (81)$$

The same average can coexist with radically different topologies: centralized reliance on one broker, distributed redundancy, polarized clusters, or loosely connected communities. Granovetter’s work on embeddedness and Tilly’s work on trust networks both support attention to structure rather than isolated attitudes [26, 74].

Local interaction can nevertheless generate macro regularities. A compact representation is

$$\mathbf{X}_{t+1}^C = F(G_t, \{\mathbf{x}_{e,t}\}, I_t, M_t), \quad (82)$$

where I_t denotes institutional arrangements and M_t distributed memory. Macro stability may therefore persist even when individual members disagree, leave, or revise their beliefs. This is important because collective trust need not imply a collective mind.

Institutions can carry relational memory across personnel turnover through records, procedures, precedents, reputational systems, and role expectations. Zucker's account of institutional trust production and Kramer's work on organizational trust help motivate this extension [81, 38]. Such memory can stabilize cooperation, but it can also transmit outdated assumptions. Distributed memory is therefore both a resource and a possible channel of lock-in.

Network topology also affects vulnerability. Dense mutual monitoring can improve detection but can produce correlated interpretation. Centralization can make coordination fast while creating a single point of epistemic or operational failure. Redundant ties can preserve correction channels, but they may be expensive. The relevant collective norm is therefore distributed corrigibility: enough plurality and redundancy must remain for errors in one part of the network to be detected and challenged elsewhere.

A collective trust-attractor candidate should accordingly require more than stable aggregate confidence. It should exhibit recurrent macro coordination together with mechanisms that reproduce the supporting network or institutional configuration. The stronger claim also requires attention to whether the system can absorb local disagreement without converting it into exclusion or silence. Collective resilience may depend on bounded distrust, independent monitoring, and subgroup plurality rather than on uniform confidence.

11.2 Scales, Coupling, Translation

Local, organizational, professional, and institutional trust should remain separate states even when they interact. Let $\mathbf{x}_t^{(\ell)}$ denote the state at level ℓ . Cross-level dynamics can be written

$$\mathbf{x}_{t+1}^{(\ell)} = F_\ell\left(\mathbf{x}_t^{(\ell)}, \{C_{\ell m,t}\mathbf{x}_t^{(m)}\}_{m \neq \ell}, E_t\right), \quad (83)$$

where $C_{\ell m,t}$ is a translation or coupling operator. The operator is not merely a statistical coefficient. It can represent accreditation, media reporting, legal recognition, professional standards, managerial hierarchy, audit, or other mechanisms through which one level becomes informative at another.

This formulation prevents a common mistake: confidence in a local professional need not imply confidence in the organization, and institutional legitimacy need not imply interpersonal trust. Coupling can transmit influence without establishing equivalence. Bachmann and Inkpen's discussion of institution-based trust is especially relevant because institutional arrangements can support relations while remaining analytically distinct from the actors they stabilize [4].

Three cross-scale dynamics deserve particular attention. First, amplification occurs when a local signal is repeatedly strengthened across levels; damping occurs when translation procedures absorb or contextualize it. Second, desynchronization occurs when trust changes at different speeds across levels. Third, the coupling operators themselves can evolve. A scandal, successful repair, or new regulatory regime can alter which local signals future institutions treat as credible.

Cross-scale rupture can therefore redistribute trust rather than simply reduce it. Confidence may move from a central institution toward local experts, from a profession toward peer networks, or in the opposite direction. This redistribution matters for governance because a repair strategy aimed only at restoring the former center can miss the newly generated relational ecology.

11.3 Ecology, Repair, Corrigibility

The combined system can be represented by a multilevel ecological state

$$\mathcal{X}_t = \left(\{\mathbf{x}_t^{(\ell)}\}_{\ell=1}^L, \{C_{\ell m,t}\}, G_t, I_t, M_t\right). \quad (84)$$

A cross-scale trust-attractor candidate would require recurrent organization of this larger state, not simply return of one indicator. Such a claim is correspondingly demanding and may often be unnecessary. In many empirical settings, documenting persistent coupling or recurrent redistribution will be the more defensible result.

Power enters through control over translation. Actors who determine which local reports reach organizational decision-makers, which communities receive institutional recognition, or which metrics become official evidence can shape trust across scales without directly changing every underlying relation. This is a form of possibility-shaping power because translation rules influence which futures become institutionally reachable.

Collective repair should therefore be multi-level. Interpersonal apology cannot repair a damaged reporting system; procedural reform cannot by itself restore a betrayed personal relation. Repair may require differentiated interventions at different levels and explicit attention to the interfaces between them. Braithwaite and Levi's work on trust and governance and Ostrom's work on reciprocity and collective action provide useful neighboring traditions for this problem [9, 60].

Proposition 11.1 (Cross-scale corrigibility). *A trust ecology is more corrigible when no single translation interface monopolizes the movement of relevant evidence across levels and when independent pathways remain available for challenge, verification, and repair.*

This proposition connects network structure to the earlier justice analysis. Distributed trust can generate collective capacity, but it can also distribute vulnerability and invisibility. The next section therefore turns from scale to institutional recognition, legitimacy, and authority: the mechanisms through which some relations gain official standing while others remain weakly visible.

12 Recognition, Legitimacy, Authority, Contestation

Institutions do more than stabilize expectations. They classify actors, authorize roles, recognize claims, and allocate standing. These operations can feed back into trust by changing visibility, credibility, access, and the consequences of challenge. This section separates four objects that frequently travel together but should remain analytically distinct: trust, trustworthiness, legitimacy, and recognition.

12.1 Recognition, Visibility, Possibility

Recognition is a relational operation through which an actor, claim, role, or status becomes institutionally legible. We can represent a recognition operation by

$$\rho_t : (a, c, H_t, I_t) \mapsto (s_t, v_t, p_t), \quad (85)$$

where s_t is standing, v_t visibility, and p_t access to procedures or resources. Recognition therefore changes future relational possibilities rather than merely recording a pre-existing fact. It can amplify a speaker's credibility, open channels of appeal, or make certain harms visible to institutions.

Non-recognition has the converse effect. A trustworthy actor can remain institutionally weak if the relevant procedures do not recognize its evidence or standing. Recognition can also shelter weak trustworthiness when status itself becomes a substitute for continuing evaluation. Suchman's work on organizational legitimacy is useful here because legitimacy depends on socially constructed judgments about appropriateness, while remaining distinct from the empirical reliability of every recognized actor [72].

Recognition chains can become recursive. Prior recognition increases visibility; increased visibility generates more recorded evidence; recorded evidence then supports later recognition. This mechanism can be benign, but it also creates recognition advantage and bottlenecks. The justice question concerns who controls these procedures, which forms of evidence they accept, and whether unrecognized actors retain routes for contestation.

12.2 Legitimacy, Procedure, Misalignment

Legitimacy concerns the justified or accepted right of an institution or authority to exercise a role. Trust may support legitimacy, and legitimate procedures may support trust, but neither relation is guaranteed. A simple coupled system can be written

$$T_{t+1} = F(T_t, L_t, W_t, E_t), \quad (86)$$

$$L_{t+1} = G(L_t, T_t, P_t, H_t), \quad (87)$$

where W_t denotes perceived trustworthiness and P_t procedural conditions. Tyler's work on procedural justice is relevant because fair procedures can influence acceptance independently of favorable outcomes [76]. Levi and Stoker likewise emphasize the importance of separating political trust from judgments of trustworthiness [39].

The distinction produces several possible regimes: high trust with contested legitimacy; legitimate but distrusted institutions; recognized actors with weak trustworthiness; trustworthy actors with low recognition; and mutually reinforcing trust-legitimacy closure. These regimes matter because the same observed compliance can emerge from very different relational structures.

Legitimacy can acquire attractor-like dynamics when procedures, expectations, and recognition reproduce one another. Yet the same caution used for trust attractors applies here. Recurrent compliance is not enough. A stronger claim requires evidence that the institutional arrangement actively regenerates the conditions of legitimacy and returns after perturbation.

12.3 Authority, Challenge, Contestation

Authority introduces another distinction. Compliance with authority can reflect trust, legitimacy, coercion, habit, dependency, or strategic accommodation. Hence authority should not be inferred from trust and trust should not be inferred from obedience. The central governance problem is how a relation can coordinate under authority while preserving credible challenge.

We call the practical capacity to question, appeal, request reasons, introduce counterevidence, or seek review *contestational capacity*. Pettit's work on contestability provides an important neighboring account of freedom from uncontrolled power [61]; Estlund's work on democratic authority likewise highlights the need to distinguish authority from simple correctness [20].

Contestational capacity need not weaken trust. When challenge is protected and consequential, actors can rely on a relation without assuming that every current decision is correct. This produces *contestational elasticity*: the ability of a relation to absorb disagreement, review, and correction without either immediate collapse or forced consensus. By contrast, authority shielding occurs when accumulated trust or legitimacy makes challenge increasingly costly or institutionally inaudible.

A stable authority relation can therefore have two very different forms. A corrigible configuration returns to coordination after challenge while preserving the channels that made challenge possible. A closed configuration returns to obedience partly because challenge channels have narrowed. Both may appear stable in aggregate data. Only the former demonstrates resilience compatible with ongoing epistemic correction.

Voice, exit, and protected suspension should be treated jointly. Hirschman's distinction between voice and exit remains useful [30], but some contexts require an intermediate option: actors may need to suspend reliance temporarily while evidence is reviewed. This protected suspension prevents a false choice between immediate compliance and complete rupture.

12.4 Governance, Correction, Repair

Recognition, legitimacy, and authority all shape future evidence. Governance should therefore preserve independent correction channels across these domains. At minimum, this means separating recognition from immunity, legitimacy from unquestionability, and authority from permanent procedural compression.

Three design principles follow. First, recognition procedures should permit new evidence and appeals from actors who lack prior status. Second, legitimacy should be renewable through procedure and performance rather than inherited indefinitely from past confidence. Third, authority should retain challenge channels whose use does not itself mark the challenger as untrustworthy.

Repair must also be differentiated. A legitimate institution can acknowledge procedural failure without immediately restoring interpersonal trust; a trusted professional can retain interpersonal confidence while the institution loses legitimacy. Cross-scale repair therefore requires clarity about which relational object was damaged.

Proposition 12.1 (Institutional non-equivalence). *Trust, trustworthiness, legitimacy, recognition, and authority may mutually influence one another, but evidence concerning one does not by itself establish the others.*

This proposition limits the tendency of successful institutions to compress multiple forms of standing into one generalized confidence judgment. It also prepares the next section: institutions that learn, anticipate, and respond to crisis must revise these relations over time rather than simply preserve them.

13 Learning, Anticipation, Crisis, Adaptation

Trust changes how organizations learn from the past and prepare for the future. High trust can encourage reporting, experimentation, and coordination; it can also suppress dissent, reduce search, and make successful routines difficult to question. Under crisis, these tensions intensify because the time available for verification and contestation contracts. The common problem across learning, anticipation, and emergency is therefore the preservation of corrigibility under changing temporal conditions.

13.1 Learning, Memory, Exploration

Institutional learning should be understood through more than stored information. It includes changes in relational state, decision rules, organizational memory, and the future possibility set:

$$\mathbf{x}_{t+1} = F_x(\mathbf{x}_t, o_t, a_t), \quad (88)$$

$$\boldsymbol{\theta}_{t+1} = F_\theta(\boldsymbol{\theta}_t, o_t, a_t), \quad (89)$$

$$M_{t+1} = F_M(M_t, o_t, a_t), \quad (90)$$

$$\Omega_{t+1} = F_\Omega(\Omega_t, \mathbf{x}_{t+1}, \boldsymbol{\theta}_{t+1}). \quad (91)$$

This formulation distinguishes adaptation within existing rules from learning that revises the rules themselves. Argyris's distinction between single- and double-loop learning is a direct neighboring tradition [2]; Levitt and March emphasize the role of routines, histories, and organizational memory [41].

Trust can enlarge the learning aperture by making people willing to report uncertainty, mistakes, and dissent. Edmondson's work on psychological safety provides an important empirical precedent [18]. Yet trust can also narrow the aperture when established success reduces search or when disagreement is interpreted as relational disloyalty. March's exploration- exploitation framework helps clarify the resulting tension [48]. A system that exploits trusted routines too successfully may stop generating the evidence required to know when those routines have become obsolete.

Failure is therefore not uniformly valuable. Small, bounded failures can reveal limits before large commitments become irreversible [69]. But experimental risk is itself distributive: organizations should not learn by repeatedly exposing the same low-power actors to loss. Learning governance must therefore connect experimentation to the justice analysis of Section 8.

A learning-attractor candidate concerns recurrent adaptation patterns rather than a permanently fixed rule. A corrigible system repeatedly returns to a state where anomalies can be reported, investigated, and retained in memory. A defensive system repeatedly returns to established routines by reclassifying anomalies or forgetting them. The second pattern can be highly stable while being poorly adaptive.

13.2 Anticipation, Preparedness, Uncertainty

Anticipation differs from prediction. Under deep uncertainty, the task is not necessarily to select one future and assign it maximal confidence. It can instead involve maintaining several plausible futures, monitoring weak signals, and preparing reversible responses. Work on anticipatory governance and decision-making under uncertainty provides relevant neighboring concepts [28, 62, 70].

Let $\mathcal{S}_t = \{s_1, \dots, s_k\}$ denote an active scenario ensemble. Trust can attach to a particular forecast, but it can also attach to the anticipatory process: the procedures through which signals are collected, disagreements retained, and responses revised. Process trust is especially important when no single forecast can be strongly justified.

Preparedness can be represented by a profile

$$\mathbf{p}_t = (d_t, m_t, r_t, c_t, e_t), \quad (92)$$

covering detection, monitoring, response capacity, coordination, and exit or rollback. Holling's work on resilience is relevant because recovery need not mean return to an identical state [31]. Preparedness may instead protect the capacity to reorganize after surprise.

This creates the preparedness paradox: successful prevention can remove the visible event that would have justified the preparation. Later actors may therefore interpret absence of catastrophe as evidence that the preparation was unnecessary. Trust in preparedness institutions must consequently be supported by transparent counterfactual reasoning, audit, and periodic evidence renewal rather than by fear or permanent deference.

Anticipatory lock-in occurs when one scenario progressively captures resources, authority, and attention, making revision difficult even when evidence changes. Staged commitment and protected plurality provide a countermeasure: invest enough to act, but preserve monitoring and credible alternatives until uncertainty has genuinely narrowed.

13.3 Crisis, Compression, Emergency

Crisis changes the temporal geometry of governance. Let τ_a denote available decision time and τ_v , τ_d , and τ_c the time normally required for verification, deliberation, and contestation. A compression condition appears when

$$\tau_a < \max\{\tau_v, \tau_d, \tau_c\}. \quad (93)$$

Under such conditions, ordinary procedures cannot all operate at their usual depth. Emergency reliance therefore becomes distinct from ordinary trust: actors may reasonably follow role-based authority because there is insufficient time to reproduce the entire justificatory process.

Research on crisis leadership, naturalistic decision-making, high-reliability organizing, and sensemaking helps clarify these conditions [8, 37, 79, 78]. Pre-crisis institutional design matters because emergency discretion is safer when roles, boundaries, and review procedures were established before compression occurred.

The key variable is challenge bandwidth: how much questioning, alternative evidence, and review can still pass through the system under time pressure. Too little bandwidth can create authority shielding; too much unstructured challenge can make coordination impossible. Distributed command and role clarity can help by allowing local action without routing every decision through one central node [12, 1].

Emergency measures also leave residue. Temporary surveillance, exceptional authority, shortened appeal procedures, and concentrated communication channels can persist after the original temporal constraint disappears. The governance obligation is therefore procedural re-expansion: as time pressure declines, verification, voice, contestation, and ordinary accountability should regain space. Gross and Ní Aoláin's work on emergency powers highlights the importance of this transition [27].

13.4 Adaptation, Residue, Recovery

The three domains can now be connected. Learning concerns revision after experience; anticipation concerns preparation before outcomes are known; crisis concerns action when ordinary temporal procedures cannot fully operate. Trust can support all three, but the same recursive mechanisms can also produce closure.

A resilient institution therefore requires temporal pluralism. It needs slow processes for investigation and memory, intermediate processes for monitoring and preparedness, and fast processes for emergency response. No single trust relation should permanently dominate all three temporal regimes. Fast emergency authority should not determine long-horizon truth; slow deliberation should not prevent urgent action when delay is itself harmful.

Proposition 13.1 (Temporal corrigibility). *Trustworthy adaptation requires mechanisms that permit the governing tempo of a relation to change with the situation and mechanisms that restore slower verification, contestation, and learning once temporal compression recedes.*

This proposition also limits attractor language. A healthy adaptive system may not return to one fixed state. Its stability may lie in recurrent restoration of learning capacity, scenario plurality, and review after disturbance. The final section therefore turns to the boundary conditions of the framework itself and to the conditions under which these concepts should be revised or abandoned.

14 Boundaries, Corrigibility, Conclusions

The framework developed in this paper is intentionally broader than a single-variable model of trust, but breadth creates its own risks. A relational vocabulary can become unfalsifiable if every new complication is redescribed as another relation. Attractor language can become decorative if persistence is treated as sufficient evidence. Poetic composition can become vague if any meaningful context is allowed to count. The final task is therefore to specify where the framework remains uncertain and what would justify narrowing, revising, or rejecting its stronger claims.

14.1 Units, Openness, Hidden Causes

The unit of analysis remains context-dependent. A dyad, team, institution, network, or cross-scale ecology may each be appropriate, but conclusions at one level should not be transferred automatically to another. Collective states need not correspond to a collective subject, and system-level regularities can coexist with substantial micro-level heterogeneity.

Open systems create a second limitation. Apparent return toward a trust configuration may be driven by external forcing rather than endogenous attraction. Repeated subsidies, legal compulsion, selection effects, or common environmental shocks can imitate self-reinforcement. Empirical work must therefore compare internal feedback explanations with simpler common-cause and external-input models before invoking attractor dynamics.

Temporal resolution also matters. A relation may appear stable at a monthly scale and highly volatile at a daily scale; a crisis response may look transient over one year and institutionally persistent over ten. Claims about basins, hysteresis, or return should therefore specify the observation interval and test whether the result survives reasonable changes in temporal granularity.

14.2 Measurement, History, Reflexivity

Relational history is not a neutral archive. What gets recorded depends on power, institutional procedures, memory, and later reconstruction. This creates a selection problem for any historical account of trust. Researchers should distinguish between events that occurred, events that were recorded, and events that later became retrievable or institutionally consequential.

Measurement can also alter the relation being measured. Repeated surveys about trust may make participants attend to doubts they had not previously articulated; monitoring can reduce the very vulnerability whose consequences are being studied. Reflexive effects should therefore be treated as data rather than mere contamination when they reveal how observation enters the relational system.

Poetic composition needs especially sharp exclusion criteria. The concept earns analytical value only when heterogeneous cues interact in ways not well captured by additive evidence models, when sequence or context materially changes meaning, or when small events have leverage because of their position in a relational history. If ordinary Bayesian updating or direct incentive change explains the result adequately, poetic language should not be retained merely for elegance.

14.3 Normative Conflict, Adversarial Cases, Suspension

The normative dimensions developed in the paper do not reduce to one ranking. Trustworthiness, autonomy, justice, collective capacity, reversibility, and security can conflict. Preserving exit can reduce the depth of commitment; distributed redundancy can lower efficiency; rapid crisis action can temporarily reduce contestation. The framework therefore supplies dimensions for judgment rather than a universal objective function.

Adversarial cases are crucial. Manipulators can exploit disclosure, atmosphere, institutional recognition, emergency urgency, or procedural legitimacy to manufacture confidence. A strong theory should be able to describe these cases without calling every successful trust-production mechanism desirable. It should also explain why distrust, refusal, or suspension can sometimes be the more protective relational response.

14.3.1 Suspension, Uncertainty, Reversibility

Some relations should remain unresolved for a time. When evidence is incomplete and the costs of a false commitment are high, governance can protect a middle condition in which cooperation is bounded, evidence gathering continues, and neither trust nor distrust is forced prematurely. This protected suspension is not indecision without structure. It is an institutionally supported capacity to remain revisable.

14.4 Synthesis, Corrigibility, Conclusions

The paper's first contribution is a mechanism-explicit account of recursive relational formation. It does not claim that reciprocal trust, trust spirals, path dependence, or self-reinforcing relations were previously absent from the literature. Instead, it separates present trust from the channels through which trust changes later action, evidence, interpretation, or relational structure. *Relational self-reinforcement* names this broader mechanism class and requires a specified comparison, mediator, and temporal pathway rather than being inferred from persistence alone.

The second contribution concerns stricter criteria for the existing *trust-attractor* vocabulary. The paper does not claim the term or the dynamical-systems orientation as novel. It argues instead that attractor language is stronger than feedback, duration, or path dependence and should be used only when a specified system displays relevant dynamical signatures under perturbation. Where return, basin dependence, threshold behavior, multistability, hysteresis, or comparable signatures are absent, the weaker language should be preferred. This restriction is central rather than a caveat: it makes the attractor concept capable of failing.

The third contribution concerns composition. The paper does not claim that framing, affect, ritual, embodiment, context, or symbolism are newly discovered causes of trust. *Poetic relational composition* names the narrower hypothesis that heterogeneous and temporally distributed elements can acquire joint significance through interaction, sequence, and relational history in ways not adequately reconstructed by fixed context-independent cue weights. The concept explains how a small event can acquire large relational leverage without becoming sufficient evidence of trustworthiness. If additive evidence or direct incentives explain a case adequately, poetic vocabulary adds no explanatory gain.

These three contributions meet in the governance problem. Conditions that produce trust can be designed, and successful trust production can diverge from the properties that warrant reliance. The paper therefore rejects trust maximization as a general objective. Governance should instead preserve the conditions under which reliance can become informative, revisable, contestable, repairable, and realistically terminable. Justice widens the same question by asking who bears vulnerability, who controls the trust-generating field, whose evidence receives standing, which third parties absorb costs, and who retains capacity to reconfigure or leave the relation.

Three separations carry most of the normative burden of the argument: experienced trust is not trustworthiness; dynamical stability is not justice; and relational generation is not epistemic warrant. The later analyses of commitment, autonomy, collective organization, legitimacy, learning, anticipation, and crisis elaborate consequences of these separations rather than adding independent foundations to the theory. Their common question is whether recursive trust remains open to evidence and capable of changing the rules and possibilities through which it reproduces itself.

The framework is intentionally corrigible. Future research may show that some proposed mechanisms are rare, that simpler models explain most apparent attractor behavior, or that poetic composition adds little beyond established accounts in particular domains. Such results should narrow the theory. Formalization should follow identified mechanisms rather than substitute for them, and relational complexity should be retained only when it explains observable patterns that simpler accounts leave unresolved.

Proposition 14.1 (Corrigible relational theory). *A relational theory of trust should be retained only to the extent that its additional mechanisms, historical variables, or cross-scale structures explain observable patterns that simpler accounts do not explain with equal adequacy.*

The resulting perspective is neither a celebration of trust nor a generalized suspicion of it. Trust can generate cooperation, capability, repair, and new relational futures; it can also generate closure, dependence, and manufactured vulnerability. The central task is to identify the conditions under which that recursive generativity remains epistemically open, normatively defensible, and capable of revision.

References

- [1] Chris Ansell, Arjen Boin, and Ann Keller. Managing transboundary crises: Identifying the building blocks of an effective response system. *Journal of Contingencies and Crisis Management*, 18(4):195–207, 2010.
- [2] Chris Argyris. Double loop learning in organizations. *Harvard Business Review*, 55(5):115–125, 1977.
- [3] W. Brian Arthur. *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press, Ann Arbor, 1994.
- [4] Reinhard Bachmann and Andrew C. Inkpen. Understanding institutional-based trust building processes in inter-organizational relationships. *Organization Studies*, 32(2):281–301, 2011.
- [5] Annette C. Baier. Trust and antitrust. *Ethics*, 96(2):231–260, 1986.
- [6] Howard S. Becker. Notes on the concept of commitment. *American Journal of Sociology*, 66(1):32–40, 1960.
- [7] Gernot Böhme. *The Aesthetics of Atmospheres*. Routledge, London, 2017.
- [8] Arjen Boin, Paul 't Hart, Eric Stern, and Bengt Sundelius. *The Politics of Crisis Management: Public Leadership under Pressure*. Cambridge University Press, Cambridge, 2 edition, 2016.
- [9] Valerie Braithwaite and Margaret Levi, editors. *Trust and Governance*. Russell Sage Foundation, New York, 1998.

- [10] Patrick R. Brown, Andy Alaszewski, Trish Swift, and Andy Nordin. Actions speak louder than words: The embodiment of trust by healthcare professionals in gynae-oncology. *Sociology of Health & Illness*, 33(2):280–295, 2011.
- [11] John Christman. Relational autonomy, liberal individualism, and the social constitution of selves. *Philosophical Studies*, 117(1–2):143–164, 2004.
- [12] Louise K. Comfort. Crisis management in hindsight: Cognition, communication, coordination, and control. *Public Administration Review*, 67(s1):189–197, 2007.
- [13] Karen S. Cook, Russell Hardin, and Margaret Levi. *Cooperation Without Trust?* Russell Sage Foundation, New York, 2005.
- [14] Kurt T. Dirks and Donald L. Ferrin. The role of trust in organizational settings. *Organization Science*, 12(4):450–467, 2001.
- [15] Kurt T. Dirks and Donald L. Ferrin. Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology*, 87(4):611–628, 2002.
- [16] Kurt T. Dirks, Roy J. Lewicki, and Akbar Zaheer. Repairing relationships within and between organizations: Building a conceptual foundation. *Academy of Management Review*, 34(1):68–84, 2009.
- [17] Kristie Dotson. Tracking epistemic violence, tracking practices of silencing. *Hypatia*, 26(2):236–257, 2011.
- [18] Amy Edmondson. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2):350–383, 1999.
- [19] Richard M. Emerson. Power-dependence relations. *American Sociological Review*, 27(1):31–41, 1962.
- [20] David M. Estlund. *Democratic Authority: A Philosophical Framework*. Princeton University Press, Princeton, NJ, 2008.
- [21] Donald L. Ferrin, Michelle C. Bligh, and Jeffrey C. Kohles. It takes two to tango: An interdependence analysis of the spiraling of perceived trustworthiness and cooperation in interpersonal and intergroup relationships. *Organizational Behavior and Human Decision Processes*, 107(2):161–178, 2008.
- [22] Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, Oxford, 2007.
- [23] Nicole Gillespie and Graham Dietz. Trust repair after an organization-level failure. *Academy of Management Review*, 34(1):127–145, 2009.
- [24] Erving Goffman. *Frame Analysis: An Essay on the Organization of Experience*. Harper & Row, New York, 1974.
- [25] Nelson Goodman. *Languages of Art: An Approach to a Theory of Symbols*. Hackett Publishing Company, Indianapolis, 2 edition, 1976.
- [26] Mark Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510, 1985.
- [27] Oren Gross and Fionnuala Ní Aoláin. *Law in Times of Crisis: Emergency Powers in Theory and Practice*. Cambridge University Press, Cambridge, 2006.
- [28] David H. Guston. Understanding ‘anticipatory governance’. *Social Studies of Science*, 44(2):218–242, 2014.

- [29] Russell Hardin. *Trust and Trustworthiness*. Russell Sage Foundation, New York, 2002.
- [30] Albert O. Hirschman. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Harvard University Press, Cambridge, MA, 1970.
- [31] C. S. Holling. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4:1–23, 1973.
- [32] Yimin Huang and Ian F. Wilkinson. The dynamics and evolution of trust in business relationships. *Industrial Marketing Management*, 42(3):455–465, 2013.
- [33] Gareth R. Jones and Jennifer M. George. The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, 23(3):531–546, 1998.
- [34] Karen Jones. Trust as an affective attitude. *Ethics*, 107(1):4–25, 1996.
- [35] Gideon B. Keren. Framing, intentions, and trust–choice incompatibility. *Organizational Behavior and Human Decision Processes*, 103(2):238–255, 2007.
- [36] Peter H. Kim, Donald L. Ferrin, Cecily D. Cooper, and Kurt T. Dirks. Removing the shadow of suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. *Journal of Applied Psychology*, 89(1):104–118, 2004.
- [37] Gary A. Klein. *Sources of Power: How People Make Decisions*. MIT Press, Cambridge, MA, 1998.
- [38] Roderick M. Kramer. Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, 50:569–598, 1999.
- [39] Margaret Levi and Laura Stoker. Political trust and trustworthiness. *Annual Review of Political Science*, 3:475–507, 2000.
- [40] David Levick, Robert Woog, and Kel Knox. Trust and goodwill as attractors: Reflecting on a complexity-informed inquiry. *World Futures*, 63(3–4):250–264, 2007.
- [41] Barbara Levitt and James G. March. Organizational learning. *Annual Review of Sociology*, 14:319–338, 1988.
- [42] Roy J. Lewicki and Barbara Benedict Bunker. Developing and maintaining trust in work relationships. In Roderick M. Kramer and Tom R. Tyler, editors, *Trust in Organizations: Frontiers of Theory and Research*, pages 114–139. SAGE Publications, 1996.
- [43] Roy J. Lewicki, Daniel J. McAllister, and Robert J. Bies. Trust and distrust: New relationships and realities. *Academy of Management Review*, 23(3):438–458, 1998.
- [44] J. David Lewis and Andrew Weigert. Trust as a social reality. *Social Forces*, 63(4):967–985, 1985.
- [45] Chris P. Long and Sim B. Sitkin. Trust attractors: A dynamical systems approach to trust research. In Roger C. Mayer and Barbara M. Mayer, editors, *A Research Agenda for Trust: Interdisciplinary Perspectives*, pages 63–75. Edward Elgar Publishing, 2024.
- [46] Niklas Luhmann. *Trust and Power*. John Wiley & Sons, Chichester, 1979.
- [47] Catriona Mackenzie and Natalie Stoljar, editors. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford University Press, New York, 2000.
- [48] James G. March. Exploration and exploitation in organizational learning. *Organization Science*, 2(1):71–87, 1991.

- [49] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark patterns at scale: Findings from a crawl of 11k shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–32, 2019.
- [50] Roger C. Mayer, James H. Davis, and F. David Schoorman. An integrative model of organizational trust. *Academy of Management Review*, 20(3):709–734, 1995.
- [51] Roger C. Mayer and Mark B. Gavin. Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal*, 48(5):874–888, 2005.
- [52] Daniel J. McAllister. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1):24–59, 1995.
- [53] Bill McEvily, Vincenzo Perrone, and Akbar Zaheer. Trust as an organizing principle. *Organization Science*, 14(1):91–103, 2003.
- [54] José Medina. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. Oxford University Press, Oxford, 2013.
- [55] Guido Möllering. The nature of trust: From georg simmel to a theory of expectation, interpretation and suspension. *Sociology*, 35(2):403–420, 2001.
- [56] Guido Möllering. *Trust: Reason, Routine, Reflexivity*. Elsevier, Amsterdam, 2006.
- [57] Guido Möllering and Jörg Sydow. Trust trap? self-reinforcing processes in the constitution of inter-organizational trust. In Masamichi Sasaki, editor, *Trust in Contemporary Society*, pages 141–160. Brill, Leiden, 2019.
- [58] Jennifer Nedelsky. *Law’s Relations: A Relational Theory of Self, Autonomy, and Law*. Oxford University Press, New York, 2012.
- [59] Onora O’Neill. *Autonomy and Trust in Bioethics*. Cambridge University Press, Cambridge, 2002.
- [60] Elinor Ostrom. A behavioral approach to the rational choice theory of collective action: Presidential address, american political science association, 1997. *American Political Science Review*, 92(1):1–22, 1998.
- [61] Philip Pettit. *Republicanism: A Theory of Freedom and Government*. Oxford University Press, Oxford, 1997.
- [62] Ray Quay. Anticipatory governance: A tool for climate change adaptation. *Journal of the American Planning Association*, 76(4):496–511, 2010.
- [63] Paul Ricœur. *The Rule of Metaphor: Multi-Disciplinary Studies of the Creation of Meaning in Language*. University of Toronto Press, Toronto, 1977.
- [64] Denise M. Rousseau, Sim B. Sitkin, Ronald S. Burt, and Colin Camerer. Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3):393–404, 1998.
- [65] Caryl E. Rusbult. Commitment and satisfaction in romantic associations: A test of the investment model. *Journal of Experimental Social Psychology*, 16(2):172–186, 1980.
- [66] Caryl E. Rusbult, John M. Martz, and Christopher R. Agnew. The investment model scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal Relationships*, 5(4):357–387, 1998.
- [67] Marten Scheffer, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461:53–59, 2009.

- [68] Mark A. Serva, Mark A. Fuller, and Roger C. Mayer. The reciprocal nature of trust: A longitudinal study of interacting teams. *Journal of Organizational Behavior*, 26(6):625–648, 2005.
- [69] Sim B. Sitkin. Learning through failure: The strategy of small losses. *Research in Organizational Behavior*, 14:231–266, 1992.
- [70] Andy Stirling. Keep it complex. *Nature*, 468:1029–1031, 2010.
- [71] Steven H. Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Chapman & Hall/CRC, 3 edition, 2024.
- [72] Mark C. Suchman. Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3):571–610, 1995.
- [73] Jonathan H. W. Tan and Dinithi N. Jayasekara. Perfect conformity to observable minimal rituals engenders trust: An experimental test of the signaling hypothesis. *Applied Psychology*, 74(1):e12555, 2025.
- [74] Charles Tilly. *Trust and Rule*. Cambridge University Press, Cambridge, 2005.
- [75] Edward C. Tomlinson and Roger C. Mayer. The role of causal attribution dimensions in trust repair. *Academy of Management Review*, 34(1):85–104, 2009.
- [76] Tom R. Tyler. *Why People Obey the Law*. Princeton University Press, Princeton, NJ, 2006.
- [77] Bart S. Vanneste, Phanish Puranam, and Tobias Kretschmer. Trust over time in exchange relationships: Meta-analysis and theory. *Strategic Management Journal*, 35(12):1891–1902, 2014.
- [78] Karl E. Weick. The collapse of sensemaking in organizations: The mann gulch disaster. *Administrative Science Quarterly*, 38(4):628–652, 1993.
- [79] Karl E. Weick and Kathleen M. Sutcliffe. *Managing the Unexpected: Sustained Performance in a Complex World*. Jossey-Bass, San Francisco, 3 edition, 2015.
- [80] Iris Marion Young. *Responsibility for Justice*. Oxford University Press, New York, 2011.
- [81] Lynne G. Zucker. Production of trust: Institutional sources of economic structure, 1840–1920. In Barry M. Staw and L. L. Cummings, editors, *Research in Organizational Behavior*, volume 8, pages 53–111. JAI Press, 1986.