

Norm Formation among Heterogeneous Generative-Relational Subjects: An Evolutionary Inquiry

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Norm formation among heterogeneous generative-relational systems raises a problem of interaction when participants possess different capacities, histories, practical expectations, and means of interpreting one another. An initial act of contact already relies upon a local practice, while subsequent encounters may change the very conditions under which behavior is perceived as appropriate, obligatory, or meaningful. This preliminary inquiry examines that problem through an evolutionary approach to normative interaction. Drawing on established research on conventions, social expectations, animal normativity, and multi-agent systems, it distinguishes recurrent behavior, provisional practical commitments, shared rules, normative endorsement, and legitimacy. The proposed research program considers normative initialization, provisional anchoring, partial translation, reciprocal revision, asymmetric power, and stable pluralism alongside convergence and conflict. Its generative-relational orientation permits interaction rules and interpretive capacities to change without presuming a common fitness measure or a fundamental temporal order. The manuscript develops a conceptual framework and outlines conditional formal and empirical investigations. It claims neither a general law of cross-system norm formation nor a common moral standard among human, nonhuman, artificial, and hypothetical extraterrestrial participants.

Keywords: norm formation; heterogeneous generative-relational systems; normative initialization; evolutionary interaction; coevolution; social conventions; cross-system interpretation

Discussion Paper Note

This manuscript begins an independent preliminary inquiry into the evolution of normative relations among heterogeneous generative-relational subjects and systems. It uses *subject* provisionally for interacting participants under investigation, without making an attribution of phenomenal consciousness or moral personhood. Its comparative scope includes human societies, nonhuman ecological relations, artificial agents, and explicitly hypothetical extraterrestrial encounters, with separate evidential standards for each.

The current source contains complete Sections 1–14, covering introductory foundations, cross-system identification, normative initialization and translation, evolutionary modeling, relational co-evolution, normative stability, power-conditioned norm stabilization, a preliminary coupled formal framework, cross-system comparison, and an empirical research-design program. A source-verified bibliography and the full 14-section planning outline accompany the manuscript. Further comparative validation, model testing, experimental programs, and ethical assessments remain research tasks. Concepts such as normative initialization, provisional anchoring, and coevolutionary translation are working proposals, not established laws or claims about ultimate ontology.

Except where otherwise indicated, copyright © 2026 Wanhong Huang. The source follows the reference template's Creative Commons Attribution–NonCommercial 4.0 International (CC BY-NC 4.0) notice; applicable statutory exceptions and third-party rights remain unaffected. Licence: <https://creativecommons.org/licenses/by-nc/4.0/>.

1 Introduction

The formation of norms across heterogeneous generative-relational systems presents a problem of interaction under uncertain conditions of interpretation. Human communities, nonhuman organisms, artificial agents, and hypothetical extraterrestrial interlocutors can differ in their sensory capacities, material dependencies, histories of learning, communicative resources, and the relations through which their practices become intelligible. Where such systems encounter one another, an initial act of observation, approach, withdrawal, signal transmission, or deliberate silence may carry consequences before the parties have established any mutually recognizable procedure for interpreting it. The inquiry developed in this paper examines how practical expectations, locally binding conventions, and potentially shared norms may form and change under those conditions. It brings research on social-norm evolution into conversation with a generative-relational approach to heterogeneous interaction, while reserving separate questions concerning normative authority and moral justification.

The problem begins with the conditions of initial engagement. In familiar coordination problems, participants ordinarily possess an identifiable set of available actions, a sufficiently common description of the encounter, and an established interpretation of relevant outcomes. An encounter with a structurally unfamiliar system may lack some or all of these premises. A signal interpreted by one party as an invitation could be treated by another as an intervention into a protected relation; an apparent refusal could reflect sensory inaccessibility, resource constraints, or a form of communication the observer has not learned to recognize. These possibilities provide conceptual counterexamples to the assumption that an unambiguously neutral opening move is always available. They do not establish any empirical claim about extraterrestrial life or the internal normative experiences of other species. Their analytical function is to expose how an initial interaction protocol already relies upon contingent practical judgments, even when those judgments are tentative and the counterpart's expectations remain largely unknown.

Existing research offers substantial resources for analyzing coordination and the emergence of social regularities. Young's work on the evolution of conventions demonstrates how recurrent coordination under incomplete information and occasional perturbations can select stable patterns of behavior in specified games [51]. His later account connects such patterns to expectations, signaling, sanctions, and the stochastic dynamics of social norms, while emphasizing the difficulties of applying formal models to particular historical cases [52]. Shoham and Tennenholtz develop an influential computational treatment of social-convention emergence in multi-agent settings, including the effects of memory and updating on convergence [44]. The normative multi-agent systems literature has also treated rule adoption, monitoring, and norm emergence as substantive computational research problems, rather than as consequences newly introduced by the present GR terminology [9, 43]. Institutional grammar research independently distinguishes rules, norms, and shared strategies as recurrently constituted features of human institutions [17]. These contributions establish substantial prior work on the evolution of conventions, computational norm emergence, and the institutional organization of rules. The present inquiry proposes a comparative extension for evaluating established explanations when participants differ in available actions, interpretive procedures, memory structures, or criteria of appropriate conduct, and when interaction changes the conditions of subsequent encounters. This extension is proposed as a set of diagnostic requirements and research designs; Its incremental explanatory value requires comparison with existing agent-based and institutional models.

A corresponding problem concerns what qualifies as a norm. Observed regularity, strategic coordination, habitual practice, inferred expectation, normative endorsement, and justified authority designate different objects of inquiry. Bicchieri's account distinguishes social norms from other recurrent practices by attending to empirical and normative expectations and to conditional preferences [8]. That diagnostic distinction is especially important in cross-system settings, where the analyst may possess behavioral observations while lacking reliable access to the participant's expectations. Comparative research on possible animal normativity likewise proposes operational constructs such as socially maintained normative regularities, while leaving contested questions about the underlying psychological prerequisites open [50]. An animal's distinctive ecological affordances, for example, provide evidence concerning its

possible activities and dependencies; they do not independently show that the animal endorses an explicit moral proposition. Equally, a stable convention among software agents need not imply shared subjective understanding or moral personhood.

Recent experiments with interacting large language model agents illustrate both the opportunities and the interpretive constraints of studying artificial conventions. Ashery, Aiello, and Baronchelli report the formation of population-wide naming conventions and collective biases in controlled decentralized LLM populations, together with changes induced by committed minority agents [5]. Such evidence bears on observable coordination under specified prompts and interaction structures. The stronger attribution of practical understanding, internally endorsed obligation, or autonomous normative authority requires additional evidence and independent conceptual criteria. The research program proposed here therefore uses the term *subject* provisionally for a relational participant whose activities, responses, or representations enter the interaction under investigation. The term by itself leaves open questions of phenomenal consciousness, legal status, moral considerability, and degrees of normative agency. Those questions require domain-sensitive treatment across human, animal, and artificial cases.

Within the present generative-relational approach, the principal object is the changing relationship among heterogeneous systems and the practical conditions through which an interaction becomes possible, interpretable, repeatable, or contestable. A participant’s modes of recognizing signals, adopting rules, and revising expectations may be historically constituted through relations that themselves change as the encounter develops. Heterogeneity can therefore involve a difference in more than a preference value within a common game: it may involve differences in the construction of the game, in which moves can be perceived, in whether a repeated behavior is treated as a rule, and in what counts as a violation. The approach regards such differences as research hypotheses requiring operational specification. Generative-relational terminology supplies an analytical orientation; it does not confer a privileged account of ultimate ontology or a universally valid theory of norm formation. Its additional explanatory value must be evaluated against descriptions in which rules, interpretive resources, and relational conditions remain fixed: the proposed extension matters empirically where measured changes in those conditions improve a causal account or a held-out prediction of interactional practices.

The first proposed mechanism is *provisional normative initialization*. A participant begins with a contextually chosen practice or protocol even though the counterpart’s requirements are incompletely known. A tentative restriction on intrusive contact, a request for a recognizable response, or a rule permitting disengagement could each serve as an initial practice under certain conditions. The choice has an epistemic basis and a normative cost: even observation and waiting may affect the counterpart, while measures considered protective by one party may be interpreted differently by another. Accordingly, provisional initialization means taking responsibility for a practical starting rule under uncertainty. It does not establish that this rule is universally harmless, mutually acceptable, or morally justified. A particular model might compare initial protocols by specified risks or reversibility conditions; the choice of such criteria would remain part of the analysis rather than a neutral measurement of all possible interests.

A second candidate mechanism is *provisional normative anchoring*. Once interaction is underway, an agent may adopt a revisable local rule or a relation to a community in order to sustain further inquiry and practice. Anchoring can reduce the need to reconsider every coordination choice and may support continuity across encounters. Its effects depend on the agent’s memory, alternatives, dependence on other participants, and opportunities for dissent and revision. Nor is anchoring the sole route to continued practice. Distinct norms can coexist through negotiated boundaries, context-sensitive switching, division of practical responsibilities, translation procedures, or institutional arrangements that preserve disagreement. In some circumstances participants may continue interacting without developing a shared norm. The analysis therefore treats anchoring as a conditional mechanism to be compared with these alternatives, rather than as an inevitable consequence of extensive heterogeneity or a diagnosis of physical criticality.

Normative interpretation itself may evolve within the encounter. One participant may discover a signal distinction previously unavailable in its conceptual repertoire; another may acquire a new practice of withholding action while the first completes a response. Their interaction can establish a locally mean-

ingful procedure even when neither party initially possesses a description of the eventual convention. Such processes may involve changes to communication interfaces, sensorimotor opportunities, social relations, or institutional expectations alongside changes in recurrent choices. This is a broader target than selecting among fixed strategies in a preassigned payoff matrix. It also requires methodological restraint: observable success in coordination does not establish that the parties attach identical meanings to the convention. Interpretive mappings may be partial, asymmetrical, or dependent on a mediator, and successful encounters may coexist with persistent disagreement.

The term *evolutionary* is adopted in a deliberately plural sense. It identifies the historical transformation of practices and relations through processes that may include cultural transmission, reinforcement or other forms of learning, selection among recurrent strategies, negotiation, and institutional change. Established models of stochastic evolutionary games, repeated interaction, and computational norm emergence can each illuminate specific parts of this process, although their assumptions and explanatory targets differ. The present framework allows interacting systems to possess heterogeneous revision mechanisms and, where justified, different representations of action and value. It does not posit a common fitness function, a necessary movement toward agreement, or a scale on which every normative arrangement can be ranked. A computational update index expresses a modeling choice and carries no independent proof that metaphysical generativity has a primitive temporal order.

This qualification becomes especially consequential when studying power. An observed convention may stabilize because participants recognize a useful coordination practice; it may also stabilize because one party controls the communication interface, the resources needed for participation, or the penalties attached to departure. Institutional analysis has long shown the importance of rules, biophysical conditions, and the situations in which participants interact [40]. In heterogeneous encounters, asymmetric control may shape which normative alternatives are visible, which expressions are understood, and which revisions can be enacted. Such asymmetries must enter an explanation of norm formation, although a complete political economy of material dependency, infrastructure, ownership, and heterogeneous value circulation constitutes a distinct inquiry. For the present analysis, the primary question is how power conditions the trajectory by which a practice becomes stable, disputed, revised, or abandoned. Stable compliance is analytically separable from mutual endorsement and from legitimacy.

A related risk arises when explanatory concepts developed in human social life are projected without qualification onto other systems. Human normative expectations can sometimes be investigated through language, reported commitments, and deliberative practices. These evidential routes are partially absent or significantly altered in the study of nonhuman animals and software agents. Cross-species ecological coordination may involve learned regularities whose normative status remains disputed. An AI system may reproduce language associated with approval, refusal, or commitment without providing sufficient evidence for the same psychological mechanisms observed in people. Hypothetical extraterrestrial encounters extend the limits of possible interpretation still further, but function here as thought experiments concerning initial uncertainty rather than as evidence about actual civilizations. The inquiry consequently treats the admissibility of normative descriptions as an explicit empirical and philosophical issue in each setting.

Several outcomes must remain available to the theory. Interactions may produce a shared convention, stable plural practices, reciprocal interpretive capacities, asymmetric accommodation, persistent conflict, or a breakdown of contact. Coordination can improve while one party's practical alternatives contract. A jointly formed protocol can be maintained by sanctions even when some participants contest it. Conversely, continued cooperation may be achieved through local translation without convergence toward an identical internal representation of the rules. These possibilities motivate comparative models that distinguish observable behavior, expectations when measurable, material dependencies, opportunities for exit, and the participants' capacity to propose and realize changes to their interaction rules. They also supply potential counterexamples to claims that heterogeneity must generate belonging, that adaptation always yields convergence, or that more stable norms are ethically preferable.

The contribution of this paper is an initial conceptual and methodological architecture for studying the evolution of normative relations across structurally heterogeneous participants. It identifies the ini-

tialization problem, distinguishes provisional anchoring from normative endorsement, and specifies the interpretive and institutional conditions that require treatment in cross-system models. It proposes to examine reciprocal change in practical repertoires and relational arrangements while retaining comparisons with established norm-evolution research. The theoretical development is preliminary: no common generative potential metric, cross-species norm detector, empirically validated interaction model, or general theorem of normative convergence is claimed. The mathematical framework and illustrative examples developed later in the manuscript will require explicit state descriptions, observation assumptions, and separate tests of behavioral coordination, normative understanding, and justified authority.

The argument proceeds by first clarifying the relevant dimensions of heterogeneity and the evidential conditions for attributing norms. It then examines initial-contact protocols and the potential stabilizing role of provisional normative commitments, followed by the difficulties of translation between partially incompatible interpretive systems. The next stages compare evolutionary and learning models, investigate reciprocal changes in relational organization, and examine stability, conflict, and the effects of enforcement. Human–AI interaction, nonhuman ecological practices, and hypothetical extraterrestrial encounter serve as differently grounded comparative cases. A conditional formal framework subsequently specifies candidate norm-updating processes and their empirical limits; a concluding discussion addresses moral status, the design of discriminating tests, and open problems in the interpretation of evolving normative relations. Throughout, the explanatory development of a norm remains distinct from any claim that the norm ought to command acceptance.

2 Generative-Relational Heterogeneity in Norm Formation

This section establishes the analytical meaning of heterogeneity for an evolutionary inquiry into norm formation among generative-relational (GR) subjects and systems. Its purpose is to distinguish differences in feasible practice, perceptual access, interpretation, historical learning, and institutional conditions from the comparatively narrow differences in preferences or strategies commonly represented within a shared model of coordination. The discussion proceeds by examining the conditions under which participants act, the forms of evidence through which their actions become intelligible, the distinct mechanisms that can sustain regular conduct, and the consequences of relational asymmetry for the interpretation of apparently similar behavior. The method is conceptual and comparative, supported by selected work in ecological psychology, theories of agency, animal normativity, and artificial-agent interaction. The resulting distinctions identify the explanatory requirements for a cross-system model; they do not provide empirical evidence that these systems possess equivalent mental states or moral standing.

2.1 Material Conditions and Feasible Practices

A generative-relational description begins by specifying the conditions under which a participant can persist, respond, and enter into further relations. These conditions can include metabolic dependence, sensorimotor organization, technological infrastructure, institutional permissions, and access to materials or information. An organism’s ability to move through water, an airborne animal’s ability to navigate through the atmosphere, and a software agent’s ability to invoke a remote tool are different practical possibilities. Their distinctions arise from bodily or implemented organization together with features of an environment. The common vocabulary of “action” provides a useful abstraction, while the mechanisms through which an action becomes feasible remain system-specific.

Gibson’s ecological approach treats perception in relation to the practical possibilities afforded by an environment to an organism [24]. This perspective supplies a starting point for cross-system comparison: environmental opportunities depend upon the capacities of the participant, and the significance of a feature can differ across organisms. For example, the same stretch of water may be a medium of locomotion for one organism, an obstacle for another, and an object of remote measurement for an artificial platform. Such differences need not imply that the respective participants hold opposing moral propositions concerning the proper use of water. A judgment about what a being *ought* to do requires evidence

concerning expectations, evaluation, social regulation, or some other defensible basis for a normative attribution.

For present purposes, let an agent-relative feasible-practice repertoire denote the actions, transitions, and relational engagements available to a particular system under stated material and institutional conditions. Feasibility is thus conditional: a device may possess a software routine for a particular operation while lacking the permission, energy, connectivity, or actuator access necessary to perform it. A human participant may understand a practice yet be prevented from performing it by resource constraints or institutional exclusion. The possibilities available to a living organism may change as its ecological relations, health, development, or social environment change. In each case, a practical repertoire must be assessed with reference to the relevant conditions of enactment.

This conditionality matters for norm formation. A coordination proposal can be meaningful to one participant and infeasible for another; a prohibition may be trivial for a participant that lacks the corresponding action, and consequential for one whose livelihood depends on it. Apparent disagreement over rules may therefore originate partly in differences in achievable action, costs of compliance, or dependence on particular environments. A model that assumes a common action set risks assigning an artificial symmetry to the negotiation. It may portray a refusal as a divergent preference when the requested action lies outside a participant’s feasible repertoire. Conversely, observable compliance may result from a severely restricted range of alternatives. These are empirically and ethically different situations.

2.2 Perceptual Access and Interpretive Conditions

Heterogeneity also concerns what participants can detect, how they associate observations with the activity of others, and how uncertainty enters their interpretation. Human researchers may observe a bird’s migratory route through instruments, reconstruct environmental conditions through remote sensing, and infer patterns of response from repeated measurements. None of these operations directly provides the organism’s complete experience of the route. The observations support claims about measurable conduct and environmental association within the limits of the method. Claims about the animal’s evaluation of a practice, its interpretation of another individual’s behavior, or its experience of a violated expectation require additional evidence.

In human–AI interaction, a common linguistic interface can obscure a different evidential problem. A person and a language-model-based agent may both produce an apology, a promise, a refusal, or a statement describing a rule. The human utterance may be embedded in autobiographical continuity, interpersonal accountability, social recognition, and consequences for subsistence or reputation. An AI utterance may depend on training history, current instruction priorities, prompt context, memory architecture, tool access, and deployment oversight. These factors vary considerably across artificial systems. Observable linguistic similarity consequently warrants a comparison of outputs, while equivalence of underlying interpretation, experience, or normative commitment requires separate argument.

Controlled experiments with populations of large language model agents have reported conventions and collective biases arising from local interactions [5]. Their evidential contribution is specific: prescribed architectures and interaction settings can produce stable population-level patterns. The existence of such patterns leaves open questions about the modeled agents’ internal interpretation of rules, the effects of persistent memory and independent action, and the extent to which an observed convention generalizes beyond the experimental environment. In a GR inquiry, the empirical study of an artificial population therefore specifies its deployment setting and interaction history along with its measured behavior. Otherwise, stable output distributions can be mistaken for a complete account of socially situated normative practice.

Interpretation may itself be reciprocal and historically revisable. A participant can learn that an initially threatening gesture functions as a greeting, or that silence has a significance different from the one initially expected. Such learning need not be shared, simultaneous, or symmetric. One participant may acquire a reliable prediction of the other’s reactions while the other remains unable to identify its counterpart’s intentions. In still other cases, both participants may coordinate through a stable interface

without converging on identical meanings. The present approach treats interpretive correspondence as an empirical achievement to be examined, with partial understanding and systematic misinterpretation retained as genuine possibilities.

2.3 Normative Organization and the Evidence of Rule-Governed Conduct

A third dimension of heterogeneity concerns the organization of practices that observers may describe as normative. A regular pattern of action can arise through biomechanical constraint, environmental reinforcement, habituation, coordination incentives, learned expectations, explicit rule following, or coercive enforcement. Distinguishing these mechanisms is necessary because a similar surface regularity may be produced by different forms of organization. An annual migration, a repeated social greeting, and a software system's consistent refusal to execute a prohibited request can each exhibit regularity. Their regularities alone do not establish the same kind of rule, expectation, or normative authority.

Work on human social norms gives one point of conceptual comparison. Bicchieri's account emphasizes the relevance of empirical and normative expectations in identifying social norms, alongside their conditional influence on behavior [8]. Young's analyses of conventions describe how coordination patterns may become established through repeated local interaction and population-level dynamics [51, 52]. Such theories provide different tools for studying the formation and maintenance of practices. Their empirical applicability to heterogeneous systems depends on whether the relevant expectations, strategies, population structure, and response mechanisms can be identified in each case. Importing a human norm category into a nonhuman or computational domain without evaluating these prerequisites risks treating a convenient analogy as an established result.

The study of nonhuman animals illustrates why the evidential question must remain open. Westra and colleagues survey competing positions on animal social norms and develop a framework for examining the psychological capacities and behavioral evidence relevant to such claims [50]. Their work provides grounds for inquiry while documenting disagreement over the prerequisites for normative cognition. A learned practice maintained through social influence might support a hypothesis about animal normativity; the same observation may also admit reinforcement-based or ecological explanations. Evidence for a group's response to deviations may strengthen some interpretations, yet the interpretation of the response itself can remain disputed. Interspecies analysis must preserve this distinction between evidence of socially organized practice and evidence of an explicit evaluative proposition.

The vocabulary of agency introduces a related philosophical issue. Barandiaran, Di Paolo, and Rohde propose individuality, interactional asymmetry, and normativity as central properties in a theoretical characterization of agency [7]. Their proposal is informative because it directs attention toward the organization that makes action meaningful relative to the maintenance or regulation of a system. It is a philosophical and theoretical account, however, and its adoption would require further argument in the diverse empirical domains considered here. GR terminology does not settle whether a language model, a robot, a living organism, or a collective institution meets a particular conception of agency. It provides a comparative research vocabulary while leaving those classifications open to evidence and debate.

For clarity, the manuscript uses the term *participant* for an identifiable system whose conduct can enter the interaction under examination. *Subject* remains a provisional research designation within the paper's title and theoretical problem. Neither term automatically attributes consciousness, personhood, capacity for explicit moral deliberation, or equal moral standing. The evidential standards for identifying norms are developed separately because the threshold appropriate to observable coordination need not suffice for ascribing felt obligation, reflective endorsement, or legitimate authority.

2.4 Historical Learning and Relational Organization

Generative-relational heterogeneity further involves the formation histories through which capacities and practices become available. In a human community, a practice may depend on long-term socialization, linguistic tradition, shared infrastructures, and institutional memories. An animal population may transmit behavior through social learning under particular ecological conditions. An artificial agent may

inherit statistical regularities from pretraining, behavior-shaping objectives from subsequent training, session-specific instructions, or stateful memories acquired through interaction. These descriptions refer to different mechanisms of persistence and change. They should not be converted without argument into a common developmental sequence or a universal scale of advancement.

The distinctions also affect how normative change can occur. A participant might change its action while retaining an interpretation of the governing rule. Another might reinterpret the same action under a newly learned convention. A third might help alter the interaction protocol, the communication medium, or the institution that determines which actions are feasible. These processes differ in the structures affected, the evidence required to detect them, and their consequences for other participants. Studies of institutional diversity have emphasized that rules interact with the features of the action situation and the conditions in which participants operate [40]. A GR extension asks whether interaction can change both the local practices and the relational conditions through which those practices are generated. Such reciprocal change is a proposed research object, with causal mechanisms to be specified for each setting.

Historical constitution is especially relevant when two systems produce comparable outputs. A software agent may generate a research proposal whose surface structure resembles one written by a human researcher. The resulting documents can be compared as texts, and each may contribute to subsequent work. Their conditions of production can nevertheless differ in labor arrangements, resources, institutional accountability, access to information, and the ability to revise or challenge governing rules. Similar artifacts may therefore arise from different relational histories. An explanation of norm formation that relies exclusively on output equivalence would miss some of the conditions under which either participant can sustain its commitments and participate in later revision. This is a substantive modeling concern, without constituting proof of an unobservable inner difference.

Temporal organization adds another axis of heterogeneity while leaving the ontology of time open. Organisms, human institutions, and computational agents may have different response rates, memory persistence, renewal cycles, and horizons of dependence. A delay can be harmless for one participant and destructive of an interaction opportunity for another. Persistence of a convention across many rapid model calls has a different observational meaning from transmission across biological generations or institutional transformations. Researchers therefore need to state the clocks, sampling intervals, memory assumptions, and observation windows used in each empirical model. Such operational timing is a feature of the modeling procedure; it does not establish a primitive temporal order for generativity as a whole.

2.5 Relational Asymmetry and Comparative Identification

The dimensions developed in this section intersect within concrete relationships. One participant may possess greater interpretive reach while depending materially on another; a system with broad technological capability may have limited knowledge of a partner’s ecological needs; a socially influential actor may control sanctions without determining the content of a practice’s underlying significance. The presence of multiple asymmetries therefore challenges attempts to order participants along a single scale of sophistication, intelligence, independence, or generative potential. Within a specified domain, some of these properties may be measurable and comparable. Their combination does not automatically define a global hierarchy.

Table 1 organizes the principal analytical dimensions and the associated questions of evidence. The table is an inventory for model construction, not a taxonomy of morally superior or inferior kinds of beings. Several dimensions can overlap, and a single participant may change category-relevant characteristics across contexts or historical intervals.

For a preliminary formal representation, the heterogeneity of a participant i may be indexed by a context-dependent descriptive profile $H_i(c) = (A_i, O_i, I_i, U_i, C_i)$, where the components refer, respectively, to feasible practices, accessible observations, interpretive processes, update mechanisms, and relevant enabling or limiting conditions in context c . This is an analyst’s organizational device; its components need not be mutually independent or measurable by the same instruments. Their types can differ

Table 1: Analytical dimensions of heterogeneity relevant to cross-system norm formation.

Dimension	Comparative object	Evidence or identification issue
Material feasibility	Conditions for executing or sustaining a practice	Capability, permissions, resources, ecological dependence
Perceptual access	Detectable cues and accessible evidence	Sensory/technical interface; limits of observation
Interpretive organization	Attribution of significance to signals and actions	Response to perturbation; ambiguity and translation error
Normative organization	Expectations, evaluations, constraints, and sanctions	Distinguishing convention, obligation, and mere regularity
Historical learning	Persistent memory, transmission, and revision	Longitudinal records; separation of training and interaction
Relational power	Control over alternatives, interfaces, and enforcement	Exit options, asymmetry, sanctioning and imposed compliance

across participants, and a correspondence between two profiles may be partial. The expression makes no claim that these components exhaust a participant’s constitution or that the profile is a metaphysically privileged decomposition. Detailed operators and compatibility assumptions are reserved for a later formal investigation.

Two consequences follow for comparative research. First, a model of interaction must identify which capacities and relationships actually constitute the domain of comparison. A shared string output, action label, or experimental reward may provide a limited basis for comparing specific outcomes while leaving crucial generative conditions unmeasured. Second, a theory of norm formation should preserve several kinds of explanation for a stable convention: reciprocal understanding, ecological adaptation, strategic accommodation, coercive stabilization, and coincidence under environmental constraints may be observationally similar in restricted data. Their differentiation requires interventions, longitudinal evidence, and independent observation of feasible alternatives whenever such evidence is available.

The proposed GR extension concerns feedback between practices and the relational conditions that make those practices intelligible and feasible. It permits the possibility that participants learn each other’s rules, develop distinct but compatible conventions, modify their interaction interface, or enter stable relations whose asymmetries remain contested. The theory does not require convergence on a common inner representation, a common welfare metric, or an ontologically fundamental precedence between participant systems. Whether a particular interaction exhibits norm formation, practical coordination, asymmetric enforcement, or a different kind of regularity remains an empirical and interpretive question. This evidential distinction supplies the starting point for the subsequent analysis of criteria for identifying norms across heterogeneous systems.

3 Criteria for Identifying Norms across Heterogeneous Systems

The identification of a norm is a distinct analytical task from the description of behavioral similarity or the reconstruction of generative-relational (GR) heterogeneity. Institutionally oriented research has already distinguished rule-governed prescriptions, socially maintained norms, and shared strategies; this separation motivates a more demanding account of the evidence available in cross-system encounters [17]. This section develops criteria for evaluating when an observed practice warrants claims about a regularity, a convention, a socially maintained expectation, a practical commitment, or a more demanding form of normative orientation. The criteria are constructed through comparison of social-norm theory, evolutionary accounts of conventions, and recent approaches to nonhuman normativity, followed by an examination of artificial agents and asymmetric encounters. The aim is to make evidential requirements explicit while permitting different mechanisms to realize similar practices. This inquiry remains diagnostic: it does not offer a universal psychological definition of a norm, establish normative authority

through behavioral success, or resolve the moral status of its participants.

3.1 Behavioral Regularities and Their Generating Conditions

A repeated action is initially an observation to be explained. Its recurrence may arise from the structure of a physical environment, an individual's acquired habit, a shared technical protocol, population-level coordination, or an expectation whose violation attracts criticism. These mechanisms can produce records that look similar at the level of their outcome frequencies. A diagnostic account therefore begins by specifying the observation window, the feasible alternatives, and the conditions under which the practice recurs. The observed fact that several participants select the same response supplies evidence of a regularity within those conditions. Its classification as a socially normative practice requires further inquiry into the forces that maintain it.

Consider three illustrative patterns. A school of fish exhibits coordinated changes of direction; passengers at a station repeatedly form a queue; artificial language agents converge on one label during a coordination game. Each pattern is observable and potentially stable. The first can be investigated through sensory coupling and collective movement, the second through learned expectations and possible social correction, and the third through the architecture of interactions, incentives, and memory. The presence of shared conduct is a legitimate starting observation in each case. The additional inference that participants understand a rule as appropriate or binding has a different evidential burden. In the fish example, no propositional claim about what one *ought* to do follows from synchronized swimming alone.

The distinction between descriptive and injunctive norms is pertinent to such an inference. In the focus theory developed by Cialdini, Kallgren, and Reno, descriptive norms concern what people perceive as commonly done, while injunctive norms concern what they perceive as approved or disapproved; personal norms are addressed separately [15]. These are constructs developed to investigate human social behavior. Their value for the present inquiry lies partly in separating the prevalence of an action from the participant's interpretation of its appropriateness. Extending them to heterogeneous beings requires a locally warranted account of perception and interpretation. When those capacities cannot be established, the analyst can retain a description of regularity without substituting a conjecture about evaluative experience.

A second distinction concerns the difference between an available alternative and an effective alternative. A human respondent may be capable of making a different choice in principle, while facing a credible threat of expulsion for deviation. A deployed AI agent may possess many possible text continuations, yet be restricted by higher-priority instructions or output filters. A nonhuman animal may depart from a repeated practice only at a significant energetic or survival cost. Studies of conformity should therefore investigate feasible departure under the conditions that participants actually face. The scarcity of alternatives can make an action highly predictable while offering limited evidence that the participant has adopted a normative commitment to it.

3.2 Social Maintenance, Expectations, and Conventions

The presence of other participants becomes more informative when their behavior changes the likelihood or form of a practice. In an account of social maintenance, the behavior of others may influence learning, continued participation, tolerance of departures, reinforcement, or correction. A socially maintained practice can thus be distinguished from one generated independently by several agents encountering similar material constraints. Distinguishing social maintenance from a norm understood as a binding obligation remains a further task.

Westra and colleagues propose the concept of a *normative regularity* for comparative research on nonhuman animals: a socially maintained pattern of behavioral conformity within a community [50]. Andrews, Fitzpatrick, and Westra further develop a dimensional approach that allows comparative investigation of different psychological and social mechanisms without presupposing a uniquely human

package of normative capacities [2]. These proposals make a useful diagnostic move. Evidence of selective tolerance, learned group-typical practice, or active social feedback may support claims about social maintenance, while the inference to reflective obligation or explicit moral judgment remains a separate question. The term *normative regularity* is retained here with its authors' proposed meaning rather than treated as a synonym for any recurrent animal behavior.

For human social norms, the diagnostic requirements can be more demanding because a researcher can often investigate beliefs, reference groups, and conditional preferences directly. Bicchieri's framework distinguishes empirical expectations concerning others' behavior from normative expectations concerning others' views of appropriate conduct, and examines the conditional influence of those expectations on choices [8]. This approach also reveals why observed agreement can conceal divergent reasons for action. Two participants may queue together, for example, although one expects reciprocal cooperation and the other anticipates punishment for refusal. Their shared action provides limited evidence for shared endorsement.

The evolutionary account of conventions introduces another important analytical possibility. A stable coordination practice may arise because participants expect the practice to be followed and obtain benefits when their choices match. Young's analyses describe the emergence and persistence of conventions through repeated interaction and stochastic population dynamics [51, 52]. The resulting self-reinforcement is explanatorily significant even where participants lack an explicit account of the practice's moral worth. A coordination convention can support reliable cooperation, generate exclusionary effects, or coexist with critical disagreement. Its evolutionary stability provides evidence for a sustaining mechanism within a particular model, while leaving open the practice's legitimacy.

For a heterogeneous interaction, identifying the relevant community or reference population may itself be difficult. A human participant may orient toward professional colleagues; an AI agent may process the instructions of its developer alongside messages from a user and feedback from other agents; an animal may respond differently to familiar group members and unfamiliar individuals. These reference relations need not coincide with the visible set of participants in an experiment. A cross-system analysis should therefore identify who can influence whose actions, whose response counts as feedback, and which communication or observational channels allow those influences to operate. The existence of a population does not by itself establish a shared normative reference group.

3.3 Violation Responses and Practical Commitments

An informative diagnostic strategy studies what happens when a recurrent practice is interrupted. The response to deviation can help distinguish an environmentally induced regularity from a socially maintained convention and can sometimes reveal the existence of rule-directed expectations. Evidence may include corrective feedback, selective affiliation, withdrawal of cooperation, attempts to renegotiate a convention, or explicit criticism. The same response nevertheless admits competing interpretations. Aggression following a deviation may arise from threat perception, resource competition, frustration, enforcement of a practice, or other mechanisms. A norm-oriented interpretation gains support when the predicted response varies systematically with the social meaning of the deviation, under controlled alternatives.

Intervention and comparison are especially useful. Investigators may compare behavior when the same physical outcome is caused accidentally or intentionally, when departures are witnessed or unobserved, when group membership changes, or when an established coordination rule is publicly amended. These contrasts can reveal whether the response is sensitive to an expected practice, the actor's perceived responsibility, or a contextual signal. Their applicability depends on the organism or system: experiments appropriate to adult human participants should not be presumed valid for animals or artificial agents. Naturalistic observation, gentle perturbation, and analysis of interaction records may provide different, often complementary forms of evidence. Ethical limits on intervention are particularly salient when a proposed test creates distress, ecological damage, or an asymmetry in a participant's capacity to withdraw.

A practical commitment introduces a further distinction. A participant may continue to follow a rule across changing incentives, carry it into new contexts, defend it in response to challenge, or participate in its revision. Such patterns can provide evidence of commitment with a specified functional interpretation. For a human participant, testimony and reasons may help establish the kind of commitment involved, although verbal reports also have limitations. For an artificial agent, a generated statement such as “I commit to this rule” is evidence of textual behavior. Stronger evidence for persistent, action-guiding commitment would require a defined agent architecture, stable memory or state, identifiable options for deviation, and observed consistency and revision across independent encounters. Even then, functional persistence would not settle phenomenal consciousness or moral personhood.

The contrast between practical adoption and reflective endorsement matters for first contact. An unfamiliar agent might adopt a signaling protocol because it enables exchange, while remaining uncertain about the partner’s reasons for using it. Another might respond to apparent sanctions without regarding the protocol as appropriate. A third might propose an alternative that it considers more acceptable. These possibilities justify distinguishing *practical accommodation*, *commitment to a practice*, and *endorsement of a normative claim*. In a radically heterogeneous encounter, the analyst may be able to establish the first and document aspects of the second while leaving the third indeterminate. Such indeterminacy is an evidential result, and it should be preserved in subsequent modeling.

3.4 Normative Identification in Human, Animal, and Artificial Interaction

The interpretation of similar behaviors across heterogeneous systems requires evidence tailored to each form of organization. Human participants may provide explanations of expectations and obligations through interview, deliberation, and observed social practice. Even here, experimenter demand, preference falsification, institutional pressure, and the gap between declared and enacted rules can complicate classification. Responses to departures and independent measures of expectations may be more informative than either testimony or conformity rates considered alone [8]. An investigator should also specify the subgroup whose expectations matter rather than infer a single community-wide norm from a limited sample.

For nonhuman organisms, social transmission and maintenance can be studied through longitudinal records, population comparisons, patterns of social feedback, and interventions that respect welfare constraints. Current research on animal normativity proposes multiple dimensions and observes that the psychological prerequisites for norm attribution remain disputed [50, 2]. For this reason, it can be legitimate to report a socially maintained practice or candidate normative regularity without claiming the presence of explicit normative propositions. The broader GR investigation does not require a bird, fish, mammal, or other organism to possess the same conceptual vocabulary as a human researcher in order for its behavior to be relevant to cross-system relations.

Artificial-agent populations introduce an analogous but technically distinct identification problem. Ashery and colleagues report the formation of conventions and collective bias in controlled populations of large language model agents [5]. Their results warrant claims about the behavior of those systems under the experimental coordination protocol. A stable convention can arise from prompts, model priors, payoff definitions, memory windows, and interaction topology. Its appearance does not establish that the agents independently authored an ethical obligation or recognized a partner as a moral subject. Researchers can strengthen causal claims by modifying one element of the protocol at a time, comparing alternative models or memory conditions, and testing whether a convention persists when the original incentives or instruction regime change.

The same evidential discipline applies to hypothetical extraterrestrial agents, with an additional restriction: no directly observed extraterrestrial normative system is currently available as evidence for such comparison. A first-contact scenario can function as a conceptual stress test. It asks which diagnostic categories would survive if common language, familiar bodies, shared institutions, and presumed agreement on worthwhile outcomes were absent. Such scenarios can expose tacit human assumptions; they should be identified as thought experiments rather than empirical case studies. Their research value

is to specify observations a protocol would need before assigning normative significance to an unfamiliar signal.

3.5 Normative Authority and the Limits of Behavioral Evidence

A successful diagnostic account must separate the existence of a practice from the authority claimed for it. A rule can be maintained by reciprocal expectations, repeated rewards, institutional permissions, or threats of exclusion. These mechanisms can account for stability without supplying a justification for enforcement. Likewise, a community's approval of a practice is a fact about that community's evaluative relations; the ethical defensibility of the practice calls for a further argument. A comparison of heterogeneous GR systems is particularly sensitive to this distinction because the actors controlling resources, interpretive interfaces, and sanctions may also influence which behaviors appear to count as voluntary conformity.

Norm identification and normative legitimacy therefore occupy different analytical tasks. The former investigates observable regularities, mechanisms of maintenance, available alternatives, and evidence of expectation or commitment. The latter asks whether the claims made upon affected participants can be justified in the relevant moral and political context. Consent, agency, welfare, reciprocal recognition, contestability, and structural dependence may matter differently across cases, and some participants may be unable to express their interests through a shared language. The present inquiry does not convert this plurality into a universal test of legitimacy. It preserves the difference between an empirical explanation of compliance and a defensible attribution of authority.

This distinction becomes acute when the rules governing interaction are imposed by a powerful participant. An agent can coordinate reliably because the conditions for refusing or proposing alternatives have been removed. Under such conditions, the resulting practice may be operationally stable while its normative character remains uncertain. The relevant observations include who establishes the protocol, whether participants can alter it, what occurs when they depart from it, and which consequences follow from refusal. These are also political-economic questions about control over generative conditions. The present section uses them to protect diagnostic inference; their detailed treatment as structures of power belongs to the separate political-economic analysis of heterogeneous GR subjects.

3.6 An Evidential Protocol for Cross-System Classification

Table 2 summarizes a proposed protocol for recording claims at the level that available evidence can support. Its categories are analytical distinctions, and a single observed practice may meet several criteria. They form neither an ontological ladder of beings nor a proof that norms develop in a fixed sequence. A researcher can combine observations from multiple sources, state competing explanations explicitly, and revise classifications as an interaction history develops.

A disciplined investigation may proceed through a sequence of methodological decisions without presuming an intrinsic order of normative capacities. It first identifies the particular practice and its enabling conditions, including the available opportunities to depart from it. It then asks whether behavior depends on others' observed conduct, anticipated reactions, explicit communications, or material constraints. When possible, investigators compare multiple candidate mechanisms through ethically admissible changes in context or incentives. The interpretation is reported with a confidence statement matched to the evidence: an observed regularity, a plausible convention, a supported social expectation, a candidate practical commitment, or an explicitly articulated normative justification. These descriptions remain open to subsequent revision.

The protocol also requires attention to *observational equivalence*. A rule-sensitive participant and a strategically compliant participant may behave identically throughout a short experiment. A socially learned animal practice and an independently reinforced habit may also be indistinguishable within an impoverished observation record. Such equivalence motivates richer longitudinal and intervention evidence; it cannot be eliminated through a new label alone. When the necessary discriminating observations are unavailable or ethically inappropriate, a qualified conclusion is methodologically preferable to

Table 2: Proposed evidential distinctions for identifying normative phenomena across heterogeneous GR systems.

Analytical object	Indicative evidence	Key interpretive limitation
Behavioral regularity	Recurrence under specified ecological, institutional, or technical conditions	Common behavior can arise through independent constraints
Social maintenance	Learning, tolerance, feedback, or correction dependent on other participants	Social influence alone need not imply felt obligation
Coordination convention	Reciprocal behavioral expectations and response to partners' choices	Stability can reflect incentives or default protocols
Normative expectation	Evidence that others' approval, disapproval, or rule claims matter	Human reports and behavioral proxies have different warrants
Practical commitment	Persistence, defense, or revision of a rule across controlled contexts	Functional consistency alone cannot establish subjective endorsement
Authority and legitimacy	Explicit claims, affected-party participation, reasons and contestation	Actual enforcement and group approval do not prove justification

a more confident but unsupported attribution. The ability to leave a claim unresolved is an important part of the proposed framework.

The criteria developed here thus provide a basis for studying normative development without converting differences in behavior into presumed differences in moral worth, intelligence, or ontological status. They support an evolutionary investigation in which practices may be maintained through heterogeneous mechanisms, and participants may gradually establish new channels for interpreting and revising those practices. The next analytical step concerns initial encounter: under conditions in which common norms and mutually interpretable evidence remain uncertain, the participants still face consequential choices concerning whether and how to begin interaction. Criteria for norm identification clarify which conclusions can be drawn during that process and which must remain provisional.

4 The Initialization Problem of Heterogeneous Encounters

The initial encounter between heterogeneous generative-relational (GR) systems presents a problem of practical coordination under incomplete normative interpretation. An initiating participant may have to select a signal, observation procedure, distance, intervention, or decision to defer while lacking reliable access to the meanings and expectations of another participant. This section analyzes the structure of that problem, the limits of identifying an uncontroversial initial action, and the effects of information, material asymmetry, and irreversibility on early contact. It develops a bounded account of provisional initialization by combining conceptual counterexamples, elementary compatibility analysis, and comparisons with established practices in decision-making under deep uncertainty and extraterrestrial-contact governance. The objective is to clarify the conditions under which interaction can begin responsibly without claiming a universally valid first-contact rule.

4.1 Initial Action under Uncertain Normative Interpretation

An encounter is initially identified through some observation: a signal, an alteration in the environment, a recurring response, or a directly observable participant. Recognition that another system is present may precede any reliable interpretation of what its actions signify. The observing system nonetheless faces practical alternatives. It may approach, transmit, monitor, interrupt an existing activity, withdraw, or allocate resources to further observation. Each option involves assumptions about worthwhile inquiry, permissible disturbance, acceptable delay, and the status of the other participant. Their normative significance cannot be settled simply by describing the options as information acquisition.

The relevant asymmetry lies partly between the familiar conditions of the initiating participant and the unknown conditions of its counterpart. A human institution may treat a greeting as a cooperative opening, an unattended boundary as freely accessible, or an unanswered signal as evidence of indifference.

Another form of life or artificial agent may associate these events with different effects or expectations. In addition, the counterpart may be unable to perceive the attempted greeting, may be organized collectively rather than individually, or may attribute significance to aspects of the encounter that the observer cannot yet detect. A scientifically careful account therefore distinguishes the physical occurrence of a contact event from its interpretation as a communicative or normative act.

The initial normative problem extends to cases in which the initiator chooses to remain silent. Silence may be a deliberate gesture, an absence of communicative capability, a response to danger, or behavior that the other participant cannot recognize. Delayed action may preserve some possibilities while allowing others to disappear. The claim that a first encounter requires a choice is consequently best understood conditionally: when an agent has a live decision among feasible alternatives that affect the interaction, its selection may depend on some practical criterion. An encounter can also begin without any identifiable agent exercising such control, as when ecological contact occurs through unplanned environmental change. In those cases, normative responsibilities arise through later interpretations and responses rather than through a deliberate first move.

The present analysis calls *normative initialization* the selection or institution of provisional conduct under conditions in which the participants lack an established, mutually interpretable rule for their encounter. The term designates an analytical problem rather than a discrete stage that every relationship must pass through. Some interactions begin within inherited ecological regularities, others within institutional protocols, and others through contingencies in which each party initially mistakes the nature of the other. Initialization may therefore be distributed across participants and events. It may also be revised retrospectively as subsequent interaction transforms the interpretation of what counted as the encounter's beginning.

4.2 Compatibility Limits among Candidate Initial Practices

The difficulty of choosing an initial practice can be stated without attributing a common utility function to the participants. Consider a finite set $\mathcal{A}$ of initial actions available to a specified initiating agent in a specified circumstance. Let $\mathcal{H}$ be a set of candidate interpretations of the counterpart's relevant conditions and practices. For each $h \in \mathcal{H}$, define $\mathcal{A}_h \subseteq \mathcal{A}$ as the actions regarded as acceptable *under that candidate interpretation*. These sets are part of the analyst's provisional representation. They need not represent the counterpart's own articulated judgments, and they may be wrong or incomplete.

A common-acceptability requirement can be expressed by the following condition, referenced here as Equation 1.

$$\mathcal{A}_{\text{common}}(\mathcal{H}) := \bigcap_{h \in \mathcal{H}} \mathcal{A}_h. \quad (1)$$

If $\mathcal{A}_{\text{common}}(\mathcal{H})$ is empty, there is no action in the specified feasible set that is acceptable under every included candidate interpretation. This is a conditional consequence of the definition in Equation 1. It neither demonstrates that an actual counterpart would reject all possible acts nor proves that the considered models exhaust its norms. Conversely, a nonempty intersection offers common acceptability only relative to the analyst's interpretation set and acceptability criteria. It does not independently establish the legitimacy or practical safety of the selected action.

Consider a minimal thought experiment involving two feasible actions: transmitting a signal and remaining silent during a fixed observation interval. Under interpretation h_1 , transmitting is understood as respectful recognition and silence as avoidance; under interpretation h_2 , unsolicited transmission is treated as intrusion and silence as restraint. The two candidate acceptable-action sets are disjoint. The analyst may add a third option, such as passive observation, but this addition changes neither the conceptual issue nor the need to assess its effects: observation may itself produce detectable emissions, intrusion, environmental disturbance, or missed opportunities. This example does not assert that unfamiliar civilizations actually possess either interpretation. It shows how apparent good intentions can fail to select a mutually acceptable procedure when normative criteria are underdetermined.

Three forms of uncertainty should be kept separate. *Interpretive uncertainty* concerns the meanings attributed to signals or conduct. *Practical uncertainty* concerns what effects an action may have on material and relational conditions. *Normative uncertainty* concerns which effects are acceptable, who is entitled to evaluate them, and which claims deserve authority. More data about the first two may improve understanding of the third, yet empirical information alone may leave competing claims to authority unresolved. A robust interaction protocol must therefore distinguish what can be investigated from what remains a matter of justification.

4.3 Contact Modalities and the Transformation of Evidence

The conditions of initialization differ across contact modalities. A remote signal can be recorded and examined before any transmission in reply. Physical proximity may alter ecological conditions through movement, contamination, noise, resource competition, or disturbance even before a symbolic exchange is established. Artificial-agent encounters may be shaped by shared platforms, system instructions, tool privileges, logging arrangements, and the possibility of changes to persistent memory. For terrestrial animals, a human observer's arrival, apparatus, or feeding intervention may change the very behavior from which a putative social norm is inferred. These differences call for tailored methods of observation and restraint.

Contemporary extraterrestrial-search governance supplies a narrowly defined institutional comparison. The International Academy of Astronautics' June 2026 *Declaration of Principles Concerning the Conduct of the Search for Extraterrestrial Intelligence* recommends independent verification and open reporting of confirmed evidence. For a reply to a confirmed detection, it calls for international consultation through broadly representative institutions before transmission [29]. These provisions concern astronomy-based technosignature detection and potential responses. They represent a human institutional proposal for responsible scientific conduct, with a specified scope; they are neither a binding constitution for interaction with an encountered extraterrestrial civilization nor evidence that another intelligence would accept the same consultation rules.

A different concern arises when the encounter includes material exchange. Article IX of the 1967 Outer Space Treaty addresses the avoidance of harmful contamination of celestial bodies and adverse environmental change on Earth resulting from extraterrestrial matter; it also provides for consultations among state parties in specified circumstances of potentially harmful interference [49]. The treaty governs states and their activities rather than defining reciprocal obligations with extraterrestrial organisms. Nevertheless, the distinction between signal-based detection and material intervention illuminates an important analytical issue: the form of contact determines which classes of irreversible or difficult-to-observe effects must be considered. A protocol appropriate for radio astronomy may be insufficient for ecological contact, physical arrival, or the transfer of potentially reproducing material.

Observation cannot always be treated as epistemically passive. Measurement technologies have footprints, interventions change local incentives, and feedback may modify a participant's expectations. The relation between observer and observed may consequently alter the conditions under which further evidence is produced. In an evolving GR investigation, observation protocols should be treated as part of the interactional configuration whenever their effects are materially consequential. This requirement does not imply a universal observer effect, nor does it attribute subjective experience to systems for which such attribution remains unsettled. It directs inquiry toward identifying which relevant changes could be induced by the research practice itself.

4.4 Irreversibility, Learning Opportunities, and Decision Criteria

Early contact can have path-dependent consequences. An action may introduce a replicating organism, reveal a vulnerable location, establish a costly precedent, damage a local environment, or create a durable record that shapes later interpretations. Some changes can be reversed through withdrawal or repair; others have uncertain or effectively irreversible consequences at the decision horizon under consideration.

The distinction is contextual: reversibility depends on what variables matter, to whom, within what time frame, and with which available restorative capacities.

Research on decisions under deep uncertainty provides useful tools for analyzing such cases. Robust decision making investigates strategies across plausible models and seeks options whose performance remains tolerable over diverse futures, often allowing adaptive revision [33]. Work on uncertainty and irreversibility in environmental decision making also emphasizes the value of retaining flexibility when later learning could improve a consequential choice [35]. These approaches offer technical precedents for exploring uncertainty and adaptation. Their application to heterogeneous first contact would still require an explicit account of who specifies the models, what counts as satisfactory performance, and how otherwise incommensurable interests enter the evaluation.

A precautionary orientation may recommend small and interpretable probes, observation before intrusive intervention, and opportunities to pause or discontinue an exchange. These choices can support later revision, provided the counterpart can perceive them and relevant harms remain controllable. Yet a universal rule of minimum immediate disturbance would be too strong. Observation can forgo an urgent chance for rescue or communication; withdrawal can interrupt a fragile relation; and apparently reversible probing may elicit permanent changes in another system. The value of waiting depends on the situation and cannot be inferred from uncertainty alone.

The proposed principle of *revision opportunity* should therefore be treated as a provisional evaluative commitment. It asks whether an initial practice preserves meaningful opportunities for the participants, or those affected by their actions, to question and alter emerging interaction rules. Such opportunities concern more than the initiator's ability to revise its own plan. They also depend on whether another participant can affect the exchange, whether its responses will be recognized as informative, and whether adjustments remain possible after dependencies have emerged. Revision opportunity is one candidate criterion for responsible initialization. It does not confer automatic moral priority over every competing concern and cannot substitute for an independent argument about protection, consent, harm, and justice.

4.5 Asymmetric Power and the Authority to Initiate

The capacity to initiate contact is often distributed unequally. A technologically advantaged agent may control the observation instruments, translation interface, physical access, informational records, or capacity to interrupt the encounter. The other participant may have few means to represent its position within the interaction's emerging conventions. Asymmetry affects the information obtained from a first move and the practical consequences attached to acceptance, refusal, or silence. A compliant response under a strongly controlled interface may provide evidence of adaptation to constraints while leaving endorsement or legitimate authority undetermined.

In human institutions, decisions to communicate on behalf of a population raise questions of representation and accountability. In a hypothetical extraterrestrial encounter, the entity initially contacted may represent only a local group or organization, while the terrestrial initiator may be an individual scientist, corporation, state, or informal coalition. Neither side's presumed internal unity is warranted without evidence. A protocol that treats a single voice as the authorized representative of an entire heterogeneous population risks converting an early interpretation into a binding political arrangement. The IAA consultation proposal illustrates an effort to widen human participation in one narrowly defined decision [29]; the choice of representative institutions and the treatment of nonhuman interests remain open political and normative questions.

The same problem arises in human–AI encounters, although the relevant architecture is different. A user, developer, model provider, organization, and tool operator can possess different powers over an AI agent's output, memory, objectives, and access to external systems. The model's apparent conversational agreement may reflect those governance structures rather than autonomous endorsement. Conversely, a provider's control over an interaction environment may also limit a human user's opportunities to contest or modify the terms of use. In either direction, the analysis asks who has practical power to set the initial protocol and who can revise it. This comparison concerns observable infrastructure and institutional

relationships, without presupposing AI consciousness or an equivalent moral standing between human and artificial participants.

Power therefore enters the initialization problem through the distribution of control over possible actions and over their interpretation. A rule may acquire practical force because an initiator can deny access to energy, communication, shelter, data, or institutional recognition. Such force can establish a stable convention, but it provides limited evidence that the convention was jointly authored or regarded as legitimate. The detailed political economy of these asymmetries belongs to the companion GR inquiry into heterogeneous subjects. For the evolutionary study of norm formation, their significance is more specific: initial asymmetries can condition which practices are available for imitation, negotiation, refusal, and subsequent revision.

4.6 Provisional Protocols for the Early Encounter

A provisional initial-contact protocol can be conceived as a *revisable arrangement for acting while interpretive and normative uncertainty remain substantial*. Its design may include explicit statements of uncertainty, a record of competing interpretations, bounded channels of experimentation, opportunities for withdrawal, and procedures for reviewing observations before expanding an exchange. Each element has an identifiable methodological purpose. Maintaining alternatives reduces premature commitment to one explanatory model. Limiting intervention can constrain some foreseeable forms of harm. Recording decision grounds permits later accountability. Providing means of response can help test whether an emerging interaction pattern is reciprocal or merely imposed.

Table 3 sets out a candidate set of design considerations. The items are neither necessary and sufficient conditions for ethical contact nor a universal protocol suitable for all forms of life. They are potential safeguards to be considered according to a particular system’s material and communicative capacities, risks, and relevant governance arrangements.

Table 3: Candidate design considerations for provisional interaction under heterogeneous normative uncertainty.

Design concern	Possible provisional practice	Limitation requiring review
Interpretive uncertainty	Record alternative meanings and discriminating observations	Relevant meanings may lie outside the initial model set
Material disturbance	Bound intervention and assess possible cascading effects	Apparently passive monitoring may itself change conditions
Revision capacity	Permit pauses, modifications, and observable responses where feasible	The counterpart may lack the means to signal a preference
Decision authority	Document initiators, affected parties, and contestation channels	Representation may be asymmetric or internally disputed
Learning and accountability	Preserve records of actions, interpretations, and outcomes	Records and interfaces can themselves be controlled or selective

A protocol should also specify what might prompt revision. Repeated divergence between expected and observed responses, newly detected ecological impacts, evidence that an apparent representative lacks authority, or the discovery of previously excluded participants may all require reconsideration. Continued interaction may alter the repertoire of possible signals and responses, which means that later choices need not belong to the initially enumerated action set. This is an important difference from models that assume a fixed strategy space. Where the material conditions of communication change, the space of intelligible practices may itself change.

The ethical standing of these practices remains provisional. Minimizing worst-case damage, preserving options, seeking consent, promoting communication, and protecting continuity are distinct objectives. Their relative claims may conflict, and another GR system might assign them unfamiliar significance. The present account accordingly treats protocol selection as a documented normative decision subject to later criticism. It offers a research program for improving initial conduct through learning and institu-

tional accountability, while leaving the justification of any particular intervention sensitive to its context and affected participants.

4.7 Initialization as a Condition of Normative Evolution

The initialization problem exposes a feature of evolutionary norm formation that can be obscured by models beginning with an agreed game, fixed action repertoire, or preassigned reward function. An initial exchange can determine which signals become visible, which participants can respond, and which departures attract attention. Later adaptation takes place within conditions that earlier actions may already have transformed. Research should therefore compare not only the conventions that eventually appear, but also the histories and constraints under which their alternatives became available or inaccessible.

Several empirically distinguishable outcomes remain possible. Participants may establish a mutually understood convention, develop stable context-specific procedures, continue to interact amid unresolved interpretation, or cease contact. A rule may be maintained by coercive dependence or instrumental adjustment without a shared normative commitment. A failure of coordination may reflect incompatible environments, limited communication channels, or an initiation protocol that obstructed recognition of the other's responses. Discriminating among these possibilities requires longitudinal observation, controlled changes where ethically permissible, and explicit separation of behavioral agreement from normative endorsement.

Normative initialization links the identification of possible rules with the analysis of how practices develop through interaction. It establishes neither a necessary first stage in an ontological sequence nor a universal prescription for encounter. Its contribution is to make explicit the interpretive, material, and political conditions under which provisional practices begin to shape a shared history. The following section examines how such practices may become sufficiently stable to support continued inquiry and action while retaining possibilities for translation, contestation, and revision.

5 Provisional Normative Anchoring during Early Interactions

Provisional normative anchoring concerns the practical continuity of interaction when participants encounter heterogeneous rules, capacities, and interpretations without a settled shared normative framework. Its analytical role is to connect the initialization of an encounter with the processes through which participants may learn, negotiate, preserve, or revise their practices. This section develops a conditional account of anchoring, differentiates practical adoption from endorsement and belonging, examines the consequences of asymmetric relationships, and identifies alternative means of sustaining interaction. Its method combines conceptual distinctions, a limited formal representation, and proposals for observational discrimination. Anchoring is treated as a possible mechanism of coordination whose occurrence and effects require investigation, rather than as a necessary stage of normative development.

Initialization, anchoring, and translation identify different analytical questions rather than obligatory consecutive stages. Initialization concerns the selection of a practice for an encounter under uncertain interpretation; anchoring concerns the contingent persistence or revisable adoption of a working practice; translation concerns correspondences through which participants come to interpret one another. A system may sustain coordination through translation or parallel procedures without a durable anchor, and an anchor may operate before any mutually reliable translation is available.

5.1 Practical Stabilization under Incomplete Normative Knowledge

Interactions can continue while important aspects of another participant's conduct remain uninterpreted. An unfamiliar response may be understood provisionally as an invitation, warning, refusal, or event without communicative intent. Repeated reconsideration of every interpretation can be costly where decisions have immediate material consequences, yet rapid fixation on one explanation can produce avoidable misunderstandings. Practical stabilization arises when an agent or institution adopts a locally operative way

of responding across some range of encounters. The practice may guide whether to approach, pause, exchange signals, restrict an intervention, or seek a different form of evidence. Its use indicates that the participant has established a workable decision condition for the present context; the evidential warrant for that condition remains a separate matter.

Existing accounts of social norms help delimit this proposal. Bicchieri’s analysis distinguishes social norms from customs and other collective behaviors by attending to empirical and normative expectations and to conditional preferences [8]. Young’s evolutionary account investigates conventions and norms sustained through coordination, sanctioning, identity signaling, and related incentives [52]. These theories offer substantive explanations within their stated domains. Provisional anchoring has a narrower, more preliminary analytical role: it identifies a locally operative commitment that can support recurring decisions while the evidence for a shared norm is incomplete. Anchoring may subsequently develop into a convention or socially maintained normative expectation, but such development depends upon identifiable social mechanisms. Mere repetition of a chosen response remains compatible with instrumental adaptation, learned habit, externally imposed compliance, and limited communication.

The stabilizing practice is also distinct from a metaphysical commitment to a privileged present or an ontologically primitive succession of generative levels. A study may record interactions chronologically and investigate changing decisions under a time-indexed model. That modeling choice specifies empirical observations and update conditions; it does not establish that relational generativity itself possesses a universal temporal order. Similarly, a persistent practice need not correspond to a physical attractor. Terms such as basin, critical point, or dissipation acquire technical significance only after an appropriate state space, dynamics, observables, and evidence have been supplied. Practical stabilization can be characterized in ordinary institutional, cognitive, or behavioral terms without importing those additional physical claims.

5.2 Working Commitments and Normative Endorsement

The phrase *provisional normative anchoring* denotes an analytically proposed arrangement in which an agent treats one or more practices as locally operative guides to action while preserving, to a specified extent, the possibility of reevaluation. Its normative character initially concerns a rule’s practical use within deliberation or coordination; evidence of psychological commitment, explicit moral judgment, or legitimate authority must be established separately. In human practical philosophy, Bratman’s planning account describes intentions as components of partial plans that organize further reasoning and coordination while permitting reconsideration [12]. This provides a useful comparison for the organization of human action. The analogy does not license the attribution of human intentions to an artificial agent or a nonhuman animal on the strength of stable outputs alone.

Four distinctions are essential for comparative inquiry. *Epistemic recognition* concerns the agent’s working interpretation of what a counterpart does or expects. *Practical adoption* concerns a rule currently used to select conduct. *Normative endorsement* concerns an evaluative judgment that the rule deserves acceptance within some domain. *Legitimate authority* concerns the defensible standing of a person, institution, or procedure to require adherence from others. These properties may coexist, yet evidence for one provides no automatic warrant for another. A participant can use an unfamiliar greeting to avoid disruption while remaining uncertain whether the greeting is obligatory. Another may recognize a rule and decline to endorse it. A collective may follow a convention because infrastructure excludes alternatives, although its members would prefer another arrangement.

Belonging adds a further dimension. A participant may identify with a community that sustains a rule, accept practical commitments without affiliating with that community, or retain multiple affiliations that authorize different practices in different circumstances. The evidence appropriate to each possibility differs. A stated human identification can support an inference of perceived membership, subject to contextual interpretation. Repeated behavioral matching by a language model may demonstrate context-sensitive compliance or convention learning while leaving questions of subjective identification open. In nonhuman communities, conformity and social maintenance require careful behavioral inves-

tigation; comparative research explicitly warns against assuming uniform psychological mechanisms across species [2]. Belonging may stabilize a provisional commitment in some settings; other forms of anchoring require no shared group identity.

5.3 Contextual Commitments and the Capacity for Revision

Provisional commitments can be specified by the contexts in which they apply, the repertoire of permitted practices, and the circumstances that warrant reconsideration. A working rule might prescribe a particular distance during observation, a sequence of low-impact signals, or a procedure for recording disagreements. Its scope should be explicit where possible. The question of revision is equally important: changes in material risk, newly observed responses, credible objections, or evidence of misidentified representatives may alter the grounds for retaining a practice. Revision may involve modifying a rule, suspending its application, adopting a parallel procedure, or relinquishing the earlier arrangement.

A minimal formal representation can make this scope visible without claiming a universal theory of normative behavior. Let i identify an analytically specified participant; let h denote the interaction history available to that participant, and c its currently relevant context. Let $\mathcal{A}_i(h, c)$ be the feasible action set under that participant's material and institutional constraints. A candidate anchoring rule α_i selects a possibly plural working subset, as represented by Equation 2.

$$\alpha_i(h, c) \subseteq \mathcal{A}_i(h, c). \quad (2)$$

The selection can be empty when an agent has no locally admissible action, and may contain several alternatives when a procedure permits discretion or controlled exploration. Distinct participants can possess different action spaces and ways of representing history and context. Equation 2 describes an analyst's choice of representation, which becomes empirically informative only when its terms and observable indicators are specified. It does not supply the normative criterion by which the subset should be chosen or an assertion about the internal phenomenology of its user.

Revisability can then be investigated through the conditions under which α_i changes in response to additional encounters. An agent may preserve a stable general commitment while revising local enactments, or preserve surface conduct while changing its interpretation of the practice. These possibilities generate different empirical predictions. A human interlocutor may continue an established greeting despite learning that the counterpart attributes a different meaning to it, now using the greeting as a jointly agreed convention. An AI agent with persistent memory may revise its procedure after repeated corrections, whereas a stateless model may merely adjust to an immediate prompt and reproduce its earlier pattern under a reset. Institutional participants may publicly revise a rule while retaining incentives that reward the original conduct. Longitudinal analysis consequently needs records of observed practice, available alternatives, interpretation, and the conditions of revision.

5.4 Belonging, Multiple Affiliations, and Practical Continuity

The role of belonging becomes especially salient when heterogeneous relationships impose incompatible requirements. Participation in a scientific collaboration, professional institution, local community, or interspecies care arrangement may involve expectations that cannot be met simultaneously in every context. Under such conditions, alignment with a recognizable group or practice may reduce the need to reopen foundational decisions during each encounter. Shared routines can support predictability, coordinated attention, division of labor, and mutual accountability. These are plausible practical mechanisms whose magnitude and desirability vary with participants and circumstances. In particular, the presence of divergent expectations does not by itself imply that an agent experiences an identity crisis or a dynamical critical state.

The choice of an anchor need not be exclusive. A participant can follow one procedure during scientific measurement, another during community deliberation, and a further rule during emergency response. The practical continuity of this arrangement may depend on stable *transition procedures*: rules that indicate which context is relevant, how conflicts are recognized, and how a participant moves between

practices. The same pattern may be seen when a bilingual or cross-cultural participant shifts forms of address without abandoning either community. In an artificial-agent setting, different task policies may likewise be activated under specified permissions and histories, although this behavioral organization alone provides no evidence of personal identity or moral self-legislation.

A further possibility is affiliation with a procedure for managing disagreements. Participants may share a commitment to documenting objections, checking interpretations, or reviewing adverse effects while maintaining diverse substantive values. Such procedural affiliation can sustain practical interaction without requiring a unified account of the good. Yet a procedural framework is itself a normative choice: someone selects its admissibility standards, allocates speaking opportunities, and decides when revision is authorized. Its fairness and representativeness therefore require separate inquiry. This is particularly important where some affected beings cannot express objections through the communication channels recognized by the coordinating institution.

The relation between truth-seeking and belonging is accordingly contingent rather than adversarial by definition. Belonging can support epistemic correction when a group rewards criticism, replicable evidence, and the revision of shared understandings. It can also sustain misleading beliefs when conformity is strongly rewarded or contrary evidence is excluded. Young's account notes several distinct mechanisms through which norms remain self-enforcing [52]. Which mechanisms characterize an observed form of belonging is an empirical question. For the present inquiry, the central hypothesis is that provisional affiliation sometimes supports continued engagement with heterogeneous others by supplying predictable practical arrangements; exclusionary and evidence-resistant forms remain competing possibilities.

5.5 Alternative Arrangements for Sustained Interaction

Anchoring supplies one possible means of reducing recurrent practical uncertainty. Other arrangements may fulfill similar functions without converging on a dominant local norm. Participants may maintain parallel practices, negotiate case-specific accommodations, translate only those distinctions needed for a shared task, or defer a consequential action while conducting further observation. A stable rule for switching between incompatible practices may be more appropriate than permanently selecting one. Institutional innovation can also change the available options, creating new channels for participation and mechanisms for resolving disagreement.

These alternatives have different informational and material requirements. Negotiation presupposes some observable channel through which proposals and refusals can be distinguished. Translation requires partial interpretive correspondence, which may emerge unevenly. Deferment may preserve some future options but also carries opportunity costs. Parallel practices can lower direct conflict while introducing unequal access to resources and unequal recognition. Ostrom's institutional analysis emphasizes how rules in use, biophysical conditions, and community characteristics together shape situations of interaction [40]. Extending such inquiry across radically heterogeneous GR systems calls for attention to the possibility that even the categories of actor, resource, compliance, and recognized objection are differently constituted. This extension remains a proposal for further research.

The contrast among strategies should not be framed as a binary choice between norm fixation and disorder. In controlled populations of large-language-model agents, Ashery and colleagues observed the emergence and modification of collective conventions, including effects of committed minorities [5]. Those experiments establish particular properties of artificial interaction under specified tasks and model conditions. They do not establish that analogous patterns in human or ecological communities share the same psychological sources or political implications. More generally, an observed stable convention may have resulted from symmetrical coordination, adaptive response to asymmetry, explicit agreement, or restrictions that made departure impracticable.

Table 4 summarizes candidate diagnostic contrasts. Its entries are research prompts, not a universal classification algorithm. They help identify what an inquiry would need to observe in order to distinguish anchoring from other ways of maintaining interaction, while preserving the possibility of mixed strategies

Table 4: Candidate mechanisms sustaining interaction under heterogeneous normative uncertainty and their diagnostic limits.

Mechanism	Candidate observable indicators	Competing interpretation or limit
Provisional anchoring	Context-bounded practice retained across encounters, with recognizable revision triggers	Habit, external compliance, or limited available alternatives may yield similar behavior
Plural procedures	Distinct practices maintained in parallel across participant groups	Apparent pluralism may reflect separation or unequal ability to participate
Norm translation	Altered interpretations or jointly negotiated signals improving task coordination	Behavioral success may occur without shared internal meanings
Context switching	Reliable changes of operative practice when a recognizable context changes	Switches may reflect mechanical routing rather than reflective commitments
Negotiated revision	Recorded proposals, objections, and modified practice after interaction	Dominant participants may control which objections are heard
Deferred intervention	Additional observation or delayed action in response to uncertainty	Delay can create irreversible losses or favor the status quo

and changing forms over a common history.

5.6 Asymmetric Power, Contestability, and Normative Standing

Provisional anchoring can be organized through an asymmetrical relationship. A participant with greater control over communication interfaces, physical territory, sanctions, computational resources, or evidence may make one practice effectively unavoidable for another. Observed stability can then reflect the distribution of alternatives more than agreement about what is appropriate. This possibility becomes particularly consequential when an institution claims that a stable pattern indicates consent, endorsement, or consensus. Each inference requires evidence independent of repeated compliance.

The inquiry should therefore document who established the operative rule, which participants were affected, who could contest it, and what resources were required to pursue alternatives. A rule may protect a vulnerable participant's capacity to continue interaction; it may also institutionalize control by the party able to define which responses count as intelligible. In cross-species relations, a human observer's categories may fail to register an animal's aversive response. In human–AI relations, the developer's instruction hierarchy, memory architecture, and available tools influence which forms of rule revision are technically possible; those facts alone settle neither an AI system's moral standing nor the legitimacy of its governance. The political economy of such asymmetries is treated more fully in the companion GR program on heterogeneous subjects, while the present section isolates their consequences for how conventions are formed and maintained.

Contestability is a useful candidate institutional property because it creates a channel through which an accepted practice can become questionable. It requires more than an abstract permission to disagree. Participants need some effective way to express or manifest departures, to have those departures registered, and to obtain consideration of the rule's continuation or revision. Where mutual communication is limited, ecological, behavioral, or institutional evidence can serve as partial grounds for reconsideration, subject to the familiar risks of proxy representation. The availability of contestation does not prove the procedure just, and the inability to contest does not alone establish that an observed practice is illegitimate. Both claims require a fuller account of affected interests, authority, and relevant institutional context.

The proposed framework consequently leaves normative legitimacy open. The evolutionary persistence of an anchor supplies evidence concerning how a practice is maintained. Its justification depends on additional ethical and political arguments that may remain contested across the interacting systems. This analytical separation is especially important for first-contact scenarios, in which the initiating actor may initially possess the principal means of describing the other's conduct. Transparent documentation of interpretations, alternatives, disturbances, and revision decisions can improve accountability within a

human research institution; its adequacy for the other participant remains a matter for ongoing investigation.

5.7 Empirical Discrimination and the Transition to Normative Translation

A research program on provisional anchoring can distinguish competing mechanisms through controlled and longitudinal comparisons. One design could vary the number of independently responding counterparts while holding interaction count and task difficulty fixed. A second could manipulate whether norms conflict, whether actors share a reward function, and whether alternative practices are genuinely available. Another could compare contexts with explicit revision permissions, opaque enforcement, or shared authority over rule modification. Relevant observations include persistence following a prompt reset, responses to credible correction, context-sensitive switching, attempts to negotiate conflicting expectations, distribution of opportunities to contest, and changes in the underlying interaction interface. Each indicator requires interpretation against the capacities and constraints of the systems under study.

Evidence inconsistent with an anchoring mechanism is equally informative. A purported anchor may disappear as soon as a salient cue is removed, indicating transient accommodation rather than sustained commitment. A stable response may persist despite the absence of any opportunity to choose another action, suggesting constrained compliance. Apparent shared membership may fail to predict any behavior beyond the wording of the immediate conversation. A participant may coordinate effectively through task-specific translation with little evidence of durable affiliation. These findings would narrow the scope of the proposed anchoring account rather than undermine the broader question of heterogeneous norm formation.

For interactions involving animals, the inference from visible behavior to cognitive or moral processes must remain particularly limited, in keeping with comparative proposals that distinguish multiple dimensions of socially maintained regularities [2]. For artificial agents, differences in memory, deployment permissions, training, and interaction protocols should be measured before inferring continuity of normative commitments. Hypothetical extraterrestrial examples can clarify logical possibilities and reveal hidden assumptions in current protocols, but provide no empirical distribution of likely forms of extraterrestrial normative life.

Provisional normative anchoring is thus best understood as a revisable analytical hypothesis about one possible condition of continued interaction. Its explanatory value depends upon distinguishing practical stabilization from mere repetition, endorsement from observed compliance, and belonging from monopolization of an operative norm. The next stage of inquiry concerns the processes through which interacting systems develop interpretable correspondences among their expectations and practices. Normative translation may reinforce an existing anchor, reshape its meaning, or support a different arrangement altogether; its analysis requires attention to partial understanding and to the changing relations through which evidence acquires significance.

6 Normative Translation across Heterogeneous Interpretive Systems

Normative translation concerns the formation of usable correspondences among participants whose expressions, practices, and grounds for evaluating conduct differ. Its role in the present inquiry is to explain how continued interaction may alter what participants can recognize as a request, permission, refusal, expectation, or proposed rule. The word *normative* identifies the practical role attributed to a correspondence within an encounter; reliable matching of signals alone establishes communicative coordination, while attribution of shared commitments requires further evidence. This section develops a context-sensitive account of translation; assesses the evidential contribution of shared environments and negotiated signs; distinguishes coordination from agreement about normative authority; and examines the effects of unequal control over interpretive infrastructures. The method combines philosophical analysis of translation, research on communicative grounding and emergent communication, and a limited

relational representation. The resulting framework is a proposal for comparative investigation, with its empirical content requiring specification for each class of participants.

6.1 Interpretive Correspondences and Their Domains of Application

Translation within heterogeneous encounters may concern at least three objects. The first consists of *signals*: utterances, gestures, inscriptions, movements, interface events, or other distinguishable occurrences. The second consists of *practical orientations*: patterns of response through which a participant acts on an expression or event. The third consists of *normative roles*: the conditions under which participants treat a performance as requested, permitted, prohibited, accountable, or open to challenge. A correspondence between two signals can be useful for prediction while leaving their normative roles indeterminate. Conversely, participants may come to coordinate a prohibition through different signal systems without adopting identical descriptions of the protected activity. The levels therefore require distinct evidence.

The philosophical literature on radical translation supplies a relevant caution. Quine’s analysis of translation and reference emphasizes the underdetermination that can arise when interpretations are constrained by public responses and observable circumstances [41]. Yet this caution has a substantial counterpoint. Davidson contests the coherence of radically untranslatable conceptual schemes and argues against treating total interpretive failure as evidence for self-contained, incomparable languages [18]. The present approach takes neither a general indeterminacy thesis nor a universal guarantee of mutual interpretation as an empirical premise. Instead, the question is operational: which correspondences can be supported by the observations, interventions, and practices available in a specified encounter? A failure to construct a translation may reflect limited access, insufficient exposure, incompatible tasks, or an inadequate interpretive model. It does not establish the existence of a metaphysically inaccessible inner scheme.

For a restricted analytical representation, let $\mathcal{P}_i$ and $\mathcal{P}_j$ denote the identifiable practical expressions or practices of participants i and j in a stated research design. These sets are analyst-defined observational domains, and they need not be identical in granularity, content, or cardinality. A provisional translation relation is introduced in Equation 3.

$$\tau_{ij}^{c,h} \subseteq \mathcal{P}_i \times \mathcal{P}_j. \quad (3)$$

Here c indexes an explicitly described interaction context and h records the observational history available to the investigator. The relation is set-valued: multiple expressions may be associated with one practice, and one expression may support several candidate correspondences. Each admitted pair should be accompanied in the research record by the observations or interventions that warrant its provisional inclusion. This representation expresses an analyst’s empirical hypothesis about correspondence; it is neither a map between private experiences nor evidence that participants possess interchangeable concepts. A relation recorded at one context or history does not acquire an unrestricted domain of validity through its notation.

The contents of $\mathcal{P}_i$ also require restraint. A human informant might distinguish an apology from an acknowledgment through reflective explanation, whereas an automated agent might be evaluated through log traces, policy constraints, and behavior under controlled prompts. An animal’s retreat from a human approach may indicate avoidance or disturbance without allowing a reliable inference that the animal has classified the approach as a norm violation. Within Equation 3, such differences affect what can reasonably be entered as an observable practice and what strength of inference a correspondence supports. They are part of the research object rather than measurement noise to be eliminated by assuming a shared psychology.

6.2 Environmental Grounding and Reciprocal Interpretive Evidence

An interacting pair may acquire evidence about one another through changes in a shared environment. A movement toward a resource, a pause before crossing a boundary, a signal followed by an opening of access, or the withdrawal of an offered object can provide publicly observable consequences against which

candidate interpretations are assessed. The epistemic advantage of such grounding is conditional on the participants having access to relevant aspects of the environment and possessing means of detecting the consequences. A shared location alone does not imply comparable sensory access, practical interests, or an agreed segmentation of events.

Research on human communication offers a well-developed account of the interactive maintenance of common ground. Clark and Brennan describe communicative grounding as the collaborative establishment and updating of sufficient mutual understanding for an ongoing activity [16]. Their analysis concerns human participants and their communicative capacities. In a heterogeneous GR study, the corresponding methodological proposal is to examine publicly testable sequences of action and response, while allowing that some participants may lack propositional language, accessible declarations of belief, or stable opportunities for reciprocal correction. The distinction is important: an investigator may establish a reliable predictor of another agent's conduct even when there is little evidence that the other agent recognizes the prediction as part of a shared communicative project.

Consider a schematic encounter between two agents capable of modifying their positions around an obstacle. One agent regularly produces a signal before advancing; the other moves aside. Repetition supports several competing accounts: the signal might be a request, a threat, a warning, a coordination cue, or an incidental correlate of advancing. A research design could vary the signal independently of movement, alter the available paths, reverse which agent controls passage, and observe whether either participant attempts to repair failed coordination. These manipulations may discriminate among interpretations. They cannot, by themselves, establish that both agents evaluate the resulting convention as fair, obligatory, or freely accepted. Such evaluation requires further evidence appropriate to their capacities.

Environmental interventions also change what is being interpreted. A newly introduced communication channel can make distinctions visible that were previously inaccessible. A change in territorial arrangements may alter whether a movement is interpreted as an invitation or intrusion. In the GR vocabulary, the interaction conditions and the participant's interpretive repertoire may be jointly revised. Consequently, translation is potentially a generative process: the parties may acquire new ways of describing and acting upon their relationship, with subsequent encounters taking place under modified practical conditions. This proposal has a clear empirical restriction. A study must observe or otherwise justify the asserted change in interpretation; improvements in task performance alone remain compatible with ordinary stimulus-response learning.

6.3 Emergent Communication and the Formation of Normative Roles

Communication systems can develop through constrained interaction even where an established shared code is unavailable. Galantucci reports laboratory experiments in which pairs of human participants formed graphical communication systems while coordinating in a task that excluded familiar written symbols; the observed systems integrated explicit signs with implicit procedures for coordinating behavior [22]. In artificial-life research, Steels demonstrated a computational mechanism for self-organizing spatial vocabulary, and later investigated the emergence of distinctions, lexical forms, and elementary syntactic patterns in a specified visually grounded robotic language-game architecture [46, 47]. These studies supply concrete evidence that conventions for communicative coordination can arise through particular interaction designs. They neither establish that all heterogeneous beings share an analogous learning mechanism nor demonstrate the independent moral standing of artificial participants.

The transition from an emerging sign convention to a putative *normative* convention requires additional analysis. An agent may learn that signal s predicts a resource transfer. Another possibility is that participants begin treating s as an acceptable means of requesting a transfer, recognize circumstances in which a request may be denied, and modify their behavior when the procedure is challenged. The second pattern contains more than a signal-outcome association. It concerns the role of the performance in structuring what subsequent responses are treated as appropriate and what departures trigger correction or renegotiation. The indicators described in Section 5 and the criteria for norm identification in Section 3 are therefore relevant, although each remains sensitive to the capacities and evidence available

for the system in question.

A hypothetical pair can illustrate the distinction. Suppose two autonomous devices share access to a charging station and discover that a specific light sequence reliably precedes an exchange of priority. Over multiple trials, each device adapts to the sequence. These observations could support a limited claim about emergent coordination. A stronger norm-formation hypothesis would predict systematic adjustment when access conditions change, sensitivity to justified exceptions within an explicit test design, and identifiable mechanisms through which failed exchanges modify the operative rule. The relevant architectural evidence might include update logs, policy representations, constraints imposed by developers, or externally enforced arbitration. None of these observations alone establishes that either device experiences obligation or possesses independent legislative agency.

Normative co-creation, in this preliminary sense, occurs when interaction helps establish or transform a practice that governs subsequent coordination and when there is evidence that the participating systems contribute to the changing conditions. The degree and character of participation may be highly asymmetric. One participant may propose most signals while another contributes through refusal, nonresponse, altered trajectories, or preferences inferred from behavior. For human participants, verbal proposals and explicit objections can provide additional evidence. For some other participants, an intervention might reveal the investigator's capacity to control behavior more clearly than it reveals a shared understanding. Claims of co-creation should therefore specify whose contributions have been identified, which aspects of the practice changed, and what alternative explanations remain.

6.4 Partial Translation, Plural Meaning, and Repair

A testable account of normative translation can assess the local reliability of proposed correspondences. A correspondence may be effective for coordinating access to a space while remaining ineffective for decisions about ownership, reciprocity, or permissible exclusion. The parties may share an operational signal while interpreting its normative scope differently. Conversely, distinct signals may perform sufficiently similar functions in a particular practice without constituting synonymous expressions. An empirical inquiry can record this pattern as *partial translation*: observed correspondences are restricted to stated settings, participants, and measurable outcomes.

Partial translations may be particularly important where the systems' practical repertoires differ. One participant may possess explicit procedures for challenging a rule, while another has only a restricted capacity to withdraw, avoid, or disrupt the interaction. A human observer might be able to document the other's departure yet remain uncertain whether it is a refusal, a response to sensory overload, or a reaction to environmental conditions. A robust research protocol would therefore preserve competing interpretations and seek discriminating evidence rather than collapsing them into a single normative category. The appropriate standard of evidence may differ across domains. In some human institutional contexts, an explicit objection is available; in an interspecies setting, independent welfare or ecological observations may be more relevant; in a software setting, internal traces may support analysis of implemented updating rules without revealing a human-like experience of meaning.

Interactive repair offers a way to investigate the changing scope of a translation. Suppose one participant acts on a signal in a way the other subsequently corrects. An observer can examine whether the correction is registered, which behavior changes, whether the revised correspondence generalizes across contexts, and whether the parties regain coordination when the old signal is deliberately made ambiguous. A successful repair sequence warrants a claim about adaptive communicative behavior. Stronger claims about reciprocal normative understanding require evidence of the parties' ability to sustain, revise, or contest the relevant practical expectations through their available capacities. Such evidence might be verbal in human encounters, procedural in an engineered system, or indirect and cautiously interpreted in an animal study.

Failures and ambiguities are analytically productive. If a supposedly shared signal ceases to work when a familiar environmental cue is removed, the earlier success may have rested upon independent use of that cue. If the agents adapt to different partners while retaining incompatible accounts of the

Table 5: Candidate diagnostics for contextual normative translation and competing interpretations in heterogeneous encounters.

Analytical target	Candidate observations	Principal interpretive limit
Signal coordination	Repeated response to a distinguishable cue across matched trials	Shared stimulus contingencies can generate similar outputs without mutual interpretation
Contextual transfer	Correspondence survives changes of task, partner, or environmental cues	Transfer may reflect general-purpose adaptation rather than an acknowledged rule
Communicative repair	Failed exchanges elicit identifiable correction and subsequent modification	Correction may be externally programmed or directed by one party alone
Normative sensitivity	Deviations elicit context-sensitive objections, allowances, or rule proposals	Reward optimization and fixed sanctions may explain the same regularity
Participatory revision	Multiple parties measurably influence the operative correspondence or procedure	Unequal options and interface control can conceal unilateral direction
Authority and justification	Affected participants can explain, contest, or revise the grounds of a rule, where capacities permit	Behavioral compliance and social persistence provide insufficient grounds for legitimacy

practice, stable plural meaning may coexist with local cooperation. If a dominant participant repeatedly overrides another’s attempts to modify a signal, the record may document imposed conformity rather than reciprocal translation. These outcomes distinguish the empirical presence of coordination from the theoretical claims of convergence, equivalence, and agreement.

Table 5 anticipates the cross-system observational comparisons developed below. The diagnostic contrasts show how reliable coordination, contextual transfer, repair, and participatory revision can provide progressively different kinds of evidence. Each contrast also admits competing explanations that should be examined before attributing shared understanding or legitimate authority.

6.5 Interpretive Infrastructure and the Distribution of Authority

Translation interfaces are also sites of institutional and material power. The party that defines the permissible signal vocabulary, manages the data record, decides which actions count as relevant observations, or controls access to communication channels can influence which interpretations become credible. In human–AI interactions, an interface may translate a user’s nuanced expression into a limited set of machine-readable labels, while a model’s response is presented as a natural-language explanation of an internal process that the user cannot directly inspect. Those two directions of mediation are structurally different. Their practical effects depend on model design, logging, deployment rules, user rights, and the availability of alternative channels. The visible fluency of a system is insufficient evidence for symmetrical authority over interpretation.

Human social epistemology provides a bounded comparison. Fricker characterizes hermeneutical injustice in terms of structurally unequal access to shared interpretive resources that obstruct an affected person’s ability to make sense of an important experience [19]. This account concerns human knowers and socially constituted forms of disadvantage. Its relevance here is analogical: it directs attention to how control over interpretive resources can constrain which grievances, expectations, and objections can enter an institutional record. It does not automatically confer human epistemic rights upon AI systems or nonhuman organisms, and the legitimacy of representation in those settings requires separate argument.

An example illustrates the stakes without presupposing a contested moral conclusion. A conservation agency might translate observed avoidance behavior into a species-management category, then construct an intervention around that category. If the classification overlooks a relevant seasonal or ecological dependency, the resulting policy may harm the organisms it purports to protect. A better translation requires more than semantic accuracy in a vocabulary: it may require changing the observational practices, involving local ecological expertise, testing rival interpretations, and revising the conditions under which evidence is collected. Similarly, an AI service can constrain a human user through limited available options even where its language appears inclusive and accommodating. The analytical issue is the

distribution of capacities to propose, challenge, and alter interpretive categories.

This power-sensitive analysis should be distinguished from an exhaustive political economy of heterogeneous subjects. The companion GR political-economy inquiry concerns ownership and control of material and institutional conditions, the creation and circulation of heterogeneous values, and the authority to set normative terms. The present section asks a more specific question: how does power affect the evidence and interaction mechanisms through which normative correspondences form? Both cooperative translation and imposed standardization can yield stable patterns of conduct, making an account of interpretive authority necessary for credible causal explanation.

6.6 Comparative Diagnostics for Translation and Normative Learning

A comparative program requires measurements that distinguish proposed translation mechanisms from observationally similar alternatives. In a human interaction study, a researcher might combine interpreted dialogue, independent reports of intent, controlled changes of context, and records of negotiated correction. In a multi-agent simulation, instrumented states, memory resets, partner substitution, and changes to the permissible signaling channel can test how coordination depends on architecture and history. In an animal interaction study, controlled environmental variation and longitudinal observation may help identify learned expectations, subject to limitations on access to subjective or explicitly moral states. An extraterrestrial first-contact scenario, by contrast, remains a conceptual stress test for the assumptions of the methodology rather than an available empirical population.

The proposed research should distinguish at least three outcomes. First, a practice may become more predictable while interpretive uncertainty remains largely intact. Second, the participants may establish reliable local correspondences, supported by repair and transfer, while maintaining differences in normative evaluation. Third, their interaction may alter the available categories, rule-making procedures, or practical opportunities through which subsequent correspondences are constructed. These outcomes support different strengths of claim and can occur together. Their comparison requires records of interventions, alternatives, and changes to the relationship, as well as observations of behavior.

The distinctively generative-relational proposal is that interpretive correspondence may develop jointly with the repertoires, relations, and material conditions that make interpretation possible. This is a research hypothesis, not a conclusion established by the translation literature alone. Its examination can proceed through longitudinal designs and carefully specified computational models in which communicative practices, available actions, and capacities for revision are allowed to vary. Such inquiry could clarify when translation enables cooperation, when it sustains plural normative arrangements, and when an apparently shared convention reflects asymmetric constraint. The resulting distinctions prepare the analysis of evolutionary norm formation without assuming a universal semantic space, a common reward metric, or convergence toward a single moral order.

7 Evolutionary Modeling of Norm Formation in Heterogeneous Interactions

An evolutionary analysis of normative interaction must specify how recurrent practices are selected, transmitted, revised, and maintained when the participants possess different action repertoires, histories of observation, learning procedures, and material dependencies. The present section compares established modeling traditions, identifies the assumptions required for their cross-system application, and proposes a limited transition representation for heterogeneous encounters. Its method is comparative and formal: established results are discussed within their original domains, while the proposed generative-relational extension is presented as a modeling hypothesis. The section establishes methodological requirements for the subsequent examination of jointly changing norms and relational organization. It does not offer an empirical test or a general law of normative evolution.

7.1 Evolutionary Games and the Selection of Conventions

Evolutionary games provide a disciplined account of how population patterns may change when a strategy's persistence depends upon interaction with other strategies. In a stylized coordination environment, multiple equilibria can be available while repeated encounters and adjustment processes favor particular conventions. Young's study of convention evolution examines agents with limited information who respond to sampled histories, sometimes making mistakes; stochastic stability supplies a criterion for comparing long-run outcomes within the specified perturbed processes [51]. Kandori, Mailath, and Rob likewise investigate how learning and mutation can select long-run equilibria in games under particular population and adjustment assumptions [30]. These contributions show that historically accumulated encounters may shape collective regularities even where agents lack comprehensive strategic knowledge.

Three methodological qualifications matter for heterogeneous generative-relational (GR) subjects. First, the available actions in a game are defined by the modeler. A pair of social agents may be assigned the same strategies in a coordination game even though their actual capacities and environmental affordances differ. Second, a payoff function records a specified evaluative mapping. It may represent reproductive success, economic return, predictive error, programmed reward, or another measurable consequence; its definition establishes neither the participant's complete values nor the moral authority of the selected convention. Third, stochastic stability concerns the stationary behavior of a particular perturbed stochastic model. Its application requires information about the possible states, adjustment rule, perturbation mechanism, and relevant limiting regime. Treating a stochastically stable outcome as the normatively preferred outcome would add an independent ethical assertion.

The distinction between convention selection and normative identification is especially important. A recurrent coordination rule can minimize disruptions while remaining unexamined, contested, or imposed. Shoham and Tennenholtz explicitly model the emergence of social conventions among computational agents and investigate the influence of memory and update frequency in their simulation framework [44]. The wider normative multi-agent systems literature has examined norm creation, recognition, diffusion, and enforcement under varied simulation assumptions [9, 43]. These antecedents already furnish norm-sensitive computational frameworks. Shoham and Tennenholtz's results concern identifiable computational processes. They establish mechanisms for conventional coordination within the stipulated game; the normative expectations, self-understandings, and legitimacy of participants require different evidence. Accordingly, evolutionary-game models are useful for explaining the establishment of recurrent rules, while claims about normative commitment remain conditional on an additional identification protocol.

7.2 Cultural Transmission and the Historical Organization of Learning

Cultural evolution extends the analysis beyond individual strategic revision by investigating how practices, representations, and learned dispositions pass among individuals and groups. Cavalli-Sforza and Feldman's quantitative approach distinguishes modes of cultural transmission and examines their potential evolutionary consequences [13]. Boyd and Richerson analyze cultural inheritance, guided variation, and biased transmission as interacting elements of a dual-inheritance research program [10]. Their modeling vocabularies clarify why a locally persistent practice may depend upon who learns from whom, which evidence is available, and which forms of transmission receive institutional or social support.

The transfer of these models to heterogeneous systems requires careful individuation of the transmitted object. An instruction, ecological routine, artifact, protocol, gesture, and institutional obligation may be reproduced by different mechanisms. For example, a human organization can teach a safety rule through explicit training, while a deployed AI agent may implement a similar behavioral restriction through a system instruction, fine-tuning, or a tool permission. Both arrangements can generate a stable pattern. Their transmission histories, revision pathways, and available opportunities for dissent are different. The term *transmission* therefore identifies a research question about a replicable practice or influence relation; it does not guarantee the reproduction of a shared mental state.

The choice of population units is consequential. Models may track persons, organizations, families

of learned procedures, AI deployments, or context-specific interaction roles. They may also track coupled ecological and institutional processes in which an organism's practices depend on changes to an environment partly produced by other organisms. Aggregating unlike units into a single population risks erasing the distinct causal mechanisms responsible for their stability. A comparison can retain separate populations with explicitly defined contact links, or study particular cross-population events while leaving some internal dynamics unmodeled. Such selection should be tied to available observations and the intended explanation.

Cultural adaptation also need not converge toward one model-wide ranking of practices. Participants may transmit different norms within overlapping contexts, maintain specialized roles, or preserve multiple routes for correcting errors. A model of plurality can record which conventions function in which relationships and whether participants can move among them. Its evaluative criteria may include persistence, predictability, adaptability, harm, or access to revision; these measures remain analytically distinct. No weighted aggregation acquires normative authority merely because it is computationally convenient.

7.3 Repeated Interaction and Adaptive Learning Mechanisms

Learning in repeated games provides a further route for explaining conventional patterns. Fudenberg and Levine examine how adaptive processes can account for equilibrium-related behavior when agents learn through experience rather than relying upon exhaustive common knowledge [21]. Different update mechanisms respond to different evidence. An agent may adjust in response to sampled action frequencies, accumulated returns, inferred partner types, or observed sanctions. A repeated-game account may model reputational expectations and threats of exclusion; a reinforcement-learning account may update action values or policies in response to a specified reward stream [48]. These approaches identify causal candidates to be tested, not interchangeable descriptions of a single evolutionary process.

An appropriate heterogeneous model should assign the learning rule to the relevant participant or participant class. A human participant may reason verbally about what a rule means, revise an explicit commitment, and respond to institutional sanction. A software agent may update a value function within a training episode, retain an external memory, or remain behaviorally fixed until redeployment. Some artificial systems combine these mechanisms; others lack persistent memory. A nonhuman organism may learn through reinforcement or social interaction while remaining inaccessible to verbal reports. Specific architectures and observed capacities must determine which proposed transitions are meaningful in each case.

Reward deserves separate treatment because it is frequently used to conceal assumptions. A developer-specified scalar reward can guide an artificial learning process and permit comparisons *within* that process. A biological organism's survival-related reinforcement, a human's livelihood, and an organization's legal obligations may each enter different local mechanisms. Their numerical values are not automatically commensurable. Furthermore, a reward-maximizing policy can imitate a norm-sensitive response without demonstrating an expectation that the action is obligatory. Conversely, norm maintenance sometimes involves accepting locally costly actions or revising the evaluative frame of the interaction. Researchers should therefore record the provenance of each reward or feedback variable, the participant who can modify it, and competing explanations for any observed convergence.

The observational role of memory is equally important. An identical pair of current actions can be generated under short-window response rules, long historical dependence, externally imposed defaults, or negotiated commitments. Longitudinal interventions can separate these possibilities by changing past interaction sequences while holding current cues as constant as the design permits. In computational agents, memory ablations and policy resets offer a direct test of certain mechanisms. Human and animal studies require ethically acceptable proxies and cannot assume experimental access to every relevant state. These differences limit the ease with which results can be transported across domains.

7.4 Interaction Topology and Unequal Opportunities for Adjustment

Evolutionary outcomes often depend on the architecture of encounters. Interaction partners may be sampled globally, restricted by a network, assigned by an organization, or selected through a platform interface. Under a particular graph-based evolutionary game, Ohtsuki and collaborators derive an approximate condition relating the benefits and costs of cooperation to the average number of neighbors [38]. The scope of their result depends upon the specified games and update rules; it does not furnish a general measure of cooperation for heterogeneous normative systems. Its relevance here is methodological: the organization of contact can alter the consequences of locally similar behavior.

Cross-system relations add an additional difficulty. The participant who chooses partners may differ from the one who observes outcomes, controls the communication interface, defines eligible actions, or maintains the material conditions of continued participation. In a human–AI deployment, for instance, an organization may set access permissions and platform policies while the human user and artificial agent directly exchange messages. The resulting interaction structure includes at least three functional positions. Treating the dyad alone as the entire social environment can obscure the institutionally controlled opportunities for learning and rule revision. In ecological relations, habitat modification may have analogous effects upon opportunities for encounter without reproducing human forms of institutional authority.

This distinction matters when interpreting convergence. A population may settle into a single recurrent response because heterogeneous participants have mutually adjusted. The same aggregate pattern may arise when one participant faces a narrow set of feasible responses or when a controlling institution filters all alternative messages. A credible evolutionary account should report the distribution of accessible alternatives, exit conditions, and opportunities to propose revisions. Power asymmetry is thereby an observable condition affecting transition dynamics; the complete political economy of ownership, resource extraction, and generative value belongs to a separate inquiry.

7.5 A Minimal Heterogeneous Transition Representation

A comparative modeling program benefits from a common notation that preserves locally different state descriptions. Consider a finite index set of modeled participants I , each with a local state space Z_i , an available-action correspondence $A_i(z_{i,t}, r_t, e_t)$, and a set of observations O_i determined by its perceptual or technical interface. The modeling index t enumerates recorded interaction events. It is a methodological coordinate and makes no claim about an ontologically primitive ordering of generativity. Local states $z_{i,t}$ may include learned dispositions, memory, and a representation of operative rules where such components are identifiable. The relational configuration r_t and relevant environmental record e_t describe constraints visible at the chosen observational resolution.

For each participant, a local stochastic transition kernel can represent a hypothesis about how its state changes following an encounter. The minimal conditional representation is stated in Equation 4.

$$z_{i,t+1} \sim K_i(\cdot \mid z_{i,t}, o_{i,t}, a_{i,t}, r_t, e_t), \quad a_{i,t} \in A_i(z_{i,t}, r_t, e_t), \quad i \in I. \quad (4)$$

The observation $o_{i,t} \in O_i$ may be generated by a participant-specific observation map, while K_i is defined over that participant’s own state space Z_i . Consequently, two participants need not possess identical action labels, learning variables, or update distributions. A deterministic transition can be represented by a kernel concentrated at one state. The construction assumes that a suitable stochastic description exists at the resolution of interest, with measurable state and observation spaces if probability-theoretic operations are required. In applications, identifying such a kernel may remain infeasible.

Equation 4 is deliberately incomplete as a coupled dynamical system. It conditions each local adjustment on the relational and environmental state at the recorded event, while the production of observations and the subsequent updating of r_t and e_t require separate mechanisms. A later formal framework will specify possible couplings and their evidential limits. The present representation is useful because it makes an often-hidden assumption explicit: a shared state space or payoff scale is an optional modeling

Table 6: Analytical roles and transfer conditions of candidate evolutionary models for heterogeneous norm formation.

Model family	Analytical question and typical observables	Key assumptions and cross-system caution
Evolutionary and stochastic games	Convention frequencies, equilibrium selection, mutation and adjustment histories	Explicit strategies, interaction payoffs, perturbations and population/update structure
Cultural transmission models	Spread, transformation and retention of practices through learning or imitation	Identifiable transmitted traits, transmission channels and units of reproduction
Repeated games and adaptive learning	Partner-conditioned action adjustment, memory and expectation updating	Observable feedback, bounded information, specified rules of adaptation and continuation
Reinforcement and multi-agent learning	Policy changes, reward-conditioned behavior, intervention and memory-ablation response	Agent-specific reward provenance, architectures and access to alternative actions
Graph and network dynamics	Effects of contact topology, partner choice and concentrated influence	Specified networks, encounter mechanisms and permission structures
Probabilistic generative-model approaches	Perceptual inference and action selection under modeled uncertainty	Defined model classes and observation channels; normative endorsement requires separate evidence

simplification and requires justification. It also prevents the assertion that the same externally observed response implies equality of the participant-specific kernels.

Consider an illustrative encounter in which one participant can send a request or wait, whereas another can approach or withdraw. The first participant’s request may be interpreted by the second as an attempt at coordination, a territorial challenge, or an irrelevant signal. If repeated events yield a predictable pair of responses, an analyst can characterize the stability of that pair. Establishing a shared convention requires evidence about contingent expectations and the participants’ available adjustments. Establishing mutual normative understanding requires further evidence of interpretation, correction, or acknowledged obligation according to the capacities of the subjects involved. The same observed sequence is compatible with several local transition rules and several relational mechanisms.

7.6 Model Comparison and Cross-System Evidential Requirements

The explanatory value of an evolutionary model depends on the question it answers, the states it distinguishes, and the observations able to discriminate it from alternatives. Table 6 summarizes major modeling families and identifies the additional assumptions relevant to heterogeneous normative encounters. The table identifies the assumptions and transfer conditions relevant to model selection. The methods can be complementary, but their compatibility has to be established in each application.

Active-inference and related probabilistic generative-model approaches supply an additional, conditional comparison. Friston’s free-energy formulation provides a model family connecting perception, action, and learning through probabilistic prediction and inference [20]. Such a framework can be relevant where evidence supports a participant’s internal generative model or a justified computational approximation. It does not provide, on its own, an empirical account of intersubjective moral obligations or their legitimacy. The inferential and normative uses of *expectation* must remain distinguished: anticipating an event is different from expecting another participant to comply with a rule. One may measure the former while the latter remains an open interpretive question.

A cross-system research design can compare candidate mechanisms by perturbing the information available to an agent, changing partner identity, controlling the visibility of past encounters, modifying the cost of an action, or allowing participants to propose rule revisions. Evidence favoring one mechanism might consist of a predictable difference in adjustment under these interventions. Persistent behavior after a simulated memory reset, for example, could challenge a model that assigns the behavior exclusively to accessible episodic memory, while remaining compatible with other forms of stored policy or external constraint. The result depends on what the reset actually removes. Similarly, reliable adaptation to new partners may support a generalization mechanism without demonstrating ethical consent or mutual understanding.

Observational equivalence places a major limit on interpretation. Two learning mechanisms can produce indistinguishable action records over the observed contexts, and a single reward-based policy

may support distinct descriptions of its apparent norm sensitivity. Identification requires access to additional variables or interventions that generate different predictions. Where comparable observations are unavailable, theoretical pluralism may be the appropriate outcome of the research rather than a methodological failure. Model comparison should document which causal distinctions can actually be tested and which remain speculative.

7.7 Scope of Evolutionary Explanation for Normative Relations

The preceding comparison supports a focused research proposal. Existing evolutionary and learning models can explain aspects of the spread, persistence, adaptation, and contestation of regularities under specified conditions. A generative-relational extension would additionally track cases where participants' available actions, interpretive categories, interaction topology, and rule-revision possibilities change through the encounter itself. This possibility motivates models with evolving descriptions of the interaction, although it introduces difficult questions of model identifiability and of how changing state spaces can be compared responsibly. A fixed-state approximation remains valuable when its scope is explicit and empirically adequate.

Three distinctions should remain operative throughout such work. A selected or persistent convention is an empirical or formal outcome relative to specified dynamics. Normative endorsement concerns what kinds of commitment and justificatory participation can be attributed to particular participants on defensible evidence. Legitimate authority concerns whether rules can be justified to affected parties under an articulated ethical or institutional framework. The three objects may be related, but they demand different explanatory and evaluative methods. In particular, a single convergence score is insufficient to establish the character of a cross-system normative order.

The framework also leaves room for stable pluralism, context-specific coexistence, asymmetrically imposed conduct, cyclic adjustment, and persistent disagreement. Which patterns arise depends on the specified coupling of local transitions, material conditions, and institutional powers. Claims about inevitable convergence, universal evolutionary progress, or thermodynamic criticality would exceed the results surveyed here. The systematic treatment of these possibilities requires an account of how norm repertoires and relational organization develop together. That account is the task of the subsequent analysis of normative coevolution.

8 Coevolution of Norms and Relational Organization

Normative interaction takes place within arrangements that determine who can encounter whom, which conduct is feasible, how actions are interpreted, and which participants can alter the rules governing subsequent encounters. The present section examines the possible joint development of these arrangements and the practices treated as regular, expected, or appropriate by the participants. It first distinguishes behavioral adjustment from changes in relational organization, then considers ecological, institutional, and communicative mechanisms through which the conditions of interaction may be modified. The analysis subsequently addresses asymmetric adaptation, historically contingent pathways, and evidence capable of distinguishing concurrent change from causally coupled development. Its method is comparative and mechanism-oriented. Ecological and institutional research provides established accounts of particular feedback processes; the proposed generative-relational (GR) interpretation extends the questions that an inquiry into norm formation might ask. The section develops analytical criteria for coevolution, while the specification of a complete coupled formal model remains a separate task.

8.1 Relational Conditions as Variables of Normative Development

A model of norm formation often treats an interaction setting as a background condition: participants repeatedly select available behaviors, obtain feedback, and gradually develop a pattern of coordination.

Such an approximation can be effective when roles, permissible actions, channels of observation, and enforcement mechanisms remain reasonably stable during the period studied. In other encounters, repeated practice may transform precisely those conditions. A jointly constructed communication interface can make additional requests possible; an institutional agreement can alter who is entitled to issue a request; and control over an infrastructure can restrict the practical alternatives available to one participant. A participant can therefore change its behavior under established relations while also participating in the alteration of the relations under which behavior becomes possible and meaningful.

For the present inquiry, *relational organization* denotes an analytically specified configuration of access, roles, dependencies, channels of interaction, and procedures for adjusting these conditions. The configuration may include physical proximity, ecological affordances, legal or institutional permissions, technical interfaces, historical records, and enforceable expectations. These aspects are not presumed to have the same causal character. A migration corridor, an application-programming interface, and an international negotiating procedure operate through different material and institutional mechanisms. They can be compared insofar as each helps determine which interactions occur and what forms of adjustment can follow, while their differences remain part of the explanation.

Similarly, a *normative repertoire* refers to candidate practices, associated expectations, and available responses to deviation at an identified observational scale. Its contents need not consist of explicitly articulated propositions. In a human organization, an operative repertoire may contain instructions, rights of appeal, and acknowledged obligations. In a computational system, an analyst may observe rule-conditioned responses and accessible revision operations while leaving questions about subjective endorsement open. In nonhuman animal interactions, observations may support claims about learned social expectations under particular conditions without warranting attribution of the full conceptual or institutional apparatus of human normativity [2, 50]. Such differentiation is necessary for the term *coevolution* to identify a possible relation among heterogeneous processes without asserting that all participants undergo an identical normative transition.

The proposed analytical distinction is between *within-organization adaptation* and *organization-changing adaptation*. In the first case, a participant changes a response while the modeled opportunities and institutional channels remain fixed. In the second, interactions also modify at least one relevant relation: access to an activity, practical authority to initiate it, the interpretive resources through which it is understood, or the procedure for contesting an established rule. Both mechanisms may operate together. Three levels of claim require further separation. *Concurrent change* records changes in behavior and relational conditions within a common observation window. *Interaction-conditioned reconfiguration* requires evidence that encounters contributed to altering those conditions. A stronger claim of *normative-relational coevolution* additionally requires a supported feedback pathway: changes in practices or candidate expectations alter relational conditions, and those modified conditions subsequently influence the development or maintenance of practices. Such feedback may be asymmetric in magnitude, control, and benefit; it need not represent reciprocal endorsement. Their distinction requires a declared observational resolution. A newly learned action may count as an expansion of an individual's repertoire in one model while remaining a strategy adjustment within a richer, fixed action space in another. Claims about evolving possibility spaces therefore depend partly on the chosen representation and should be supported by evidence about the mechanism that renders a practice newly accessible or inaccessible.

8.2 Ecological Construction and the Reconfiguration of Feasible Practices

Biological research on niche construction offers an instructive account of feedback between activities and conditions of later activity. Odling-Smee, Laland, and Feldman analyze cases in which organisms modify biotic or abiotic aspects of environments and thereby change selection pressures and ecological inheritance [37]. The relevance to the present inquiry is a structural one: actions can help produce environmental conditions that alter subsequent opportunities for action. This does not establish that ecological feedback is itself norm formation. The empirical meaning of a social norm, an adaptive ecological routine, and an environmentally induced behavioral regularity remains distinct.

Consider a hypothetical coastal setting in which human land-use rules, fishing practices, and marine animal movements interact. A conservation measure may change human access to particular zones, affect the condition of habitats, and modify the frequency and location of encounters with animals. Changes in animal movement can in turn prompt revision of monitoring methods, management boundaries, or locally accepted practices. A study might find feedback between ecological change and human institutional adjustment, yet further evidence would be required to characterize an animal's response as conforming to a norm rather than tracking a modified habitat or learned expectation. The illustration demonstrates the need to represent environmental and normative mechanisms simultaneously while permitting a different evidential standard for each.

Such cases complicate the assumption that the feasible action repertoire of a participant can be treated as permanently given. Material conditions partly determine which practices can be attempted and maintained. Changes in physical resources, infrastructural access, energy availability, or the survival of partner populations may therefore modify the circumstances under which a practice can be sustained. Conversely, the adoption of a rule can influence these very conditions through prohibitions, incentives, conservation efforts, or collective production. An analysis confined to frequencies of observed actions may miss the altered opportunity structure responsible for their persistence.

Ecological feedback should nevertheless be described at a resolution suitable to the evidence. Observing that a habitat changes while a community's rules change does not, by itself, identify a reciprocal causal mechanism. Joint changes may reflect seasonality, independent interventions, demographic variation, or an unobserved external influence. Where possible, longitudinal measures of environmental conditions, access restrictions, learned practices, and encounters should be combined with comparison sites or carefully bounded interventions. The role of niche construction here is to supply a tested family of ecological questions and modeling strategies, while the identification of social and normative mechanisms remains a separate empirical undertaking.

8.3 Institutional Revision and the Organization of Reciprocal Expectations

Institutions can make relations themselves a subject of collective practice. Rules may specify eligible participants, permitted actions, decision procedures, monitoring obligations, sanctions, and processes of amendment. Crawford and Ostrom's institutional grammar distinguishes rules, norms, and shared strategies, and treats their reconstitution through repeated interactions as an established object of inquiry [17]. Ostrom's institutional analysis examines how rules, communities, and biophysical conditions interact in diverse action arenas; it also documents the significance of locally variable arrangements and opportunities for collective rule adjustment [40]. Aoki's comparative institutional analysis addresses institutional complementarities and change, including the co-development of organizational conventions and features of the participants who operate within them [3]. Both approaches provide substantively developed precedents for analyzing relationships between conduct and the arrangements that help stabilize it. The GR proposal draws upon these lines of inquiry without claiming that institutional feedback originates in the present framework.

Suppose a collaborative research network initially assigns access to computing resources through an administrator's discretionary approval. Participants develop practical expectations about which projects will be accepted and which requests are likely to be refused. Disputes over allocation may subsequently lead to a published eligibility rule, an appeal process, or a shared governance body. The revised arrangement can change behavior in at least two ways: participants adjust their requests under the new rule, and participants acquire new means to challenge or modify the rule itself. A norm concerning acceptable requests may thus co-develop with a norm concerning legitimate revision procedures. The second-level procedure changes the possibilities for further normative development even when the immediate allocation pattern changes little.

This distinction matters especially when participants possess unequal capacities to document, communicate, or enforce expectations. An institution that admits new members may formally extend an opportunity while leaving some participants without the practical means to exercise it. An accessible ap-

peals process may similarly depend on translation services, legal assistance, compute access, or freedom from retaliation. Accordingly, an analysis of rule change must consider effective participation together with formal authorization. Observed acceptance of a rule may reflect agreement, resignation, strategic accommodation, limited alternatives, or uncertain interpretation. Discriminating among these accounts requires attention to participation records, available revision routes, and responses to counterfactual opportunities for objection.

Institutional change can also preserve multiple, partially incompatible norms. A negotiated procedure may allocate distinct areas of authority, authorize context-specific exemptions, or establish a mechanism for determining which local rule applies to an encounter. These arrangements can be durable without producing a shared ethical theory. Their durability may rest on mutual recognition, enforceable compromise, asymmetrical dependence, or a mixture of these mechanisms. Descriptive coevolution concerns how procedures and expectations develop together; assessing their legitimacy additionally requires independent inquiry into representation, justification, and the affected parties' opportunities for contestation.

8.4 Co-development of Communication and Interpretive Capacities

Communication presents a further form of relational development because a sign system can evolve during the interactions through which it is used. Galantucci's experimental research shows that human participants can develop novel graphical communication systems in constrained coordination tasks, using both explicit signs and implicit behavioral procedures [22]. Computational language-game studies similarly investigate the emergence of shared reference and simple linguistic distinctions within specified artificial-agent architectures [46, 47]. These results motivate a careful question: how can a change in communication resources alter which normative expectations participants are capable of expressing, recognizing, or contesting?

One plausible mechanism is *interpretive scaffolding*. Participants first use a signal to coordinate a limited action, then establish a more stable mapping between the signal and the relevant situation, and subsequently use that mapping to express a request, correction, prohibition, or proposed revision. Each step may increase the range of possible exchanges. This sequence is a candidate developmental route, rather than a required historical order. Other systems might begin with explicit institutional rules, rely on bodily or ecological cues, or maintain separate modes of communication for different relationships. A previously agreed symbol may also become obsolete when conditions of participation change.

For example, imagine two differently embodied agents that initially distinguish only the occurrence of contact. Repeated interaction could establish a cue associated with withdrawing from a shared workspace; a second cue might later indicate a request to resume work. If an agent can subsequently contest whether the withdrawal cue was appropriate, the communication medium may support a richer practice than immediate behavioral coordination. An observer would need to distinguish successful signal discrimination from interpretation of another participant's claim or reason. Behavioral performance alone can be compatible with distinct internal mechanisms, and some species or artificial systems may lack the relevant communicative capacities. The analytical advantage of the GR approach lies in recording how the opportunities for interpretation and correction may change, subject to those evidential limits.

A more difficult question arises when a communication interface is controlled by a third party. A platform may determine which utterances pass through a translation layer, which categories are offered for expressing objections, and which records remain available for later evaluation. A channel that increases short-term coordination can narrow the possibilities of dissent or norm revision. Accordingly, the evolution of mutual understanding should be examined together with the governance and material conditions of the medium. This issue connects the present inquiry to the political economy of heterogeneous GR subjects, although the detailed analysis of ownership, value circulation, and infrastructural dependence belongs to that separate project.

8.5 Asymmetric Adaptation and the Political Conditions of Coevolution

Reciprocal causal influence and equal power are different properties of an interaction. Two systems may affect one another while only one retains authority to set operational thresholds, modify interface permissions, or discontinue the relationship. A subordinate actor can adapt to the dominant actor's requirements, while the dominant actor adjusts a small number of operational settings in response. Under an appropriately broad description, both change; the distribution of available revision rights nevertheless remains unequal. A study that reports only coupled behavioral trajectories can obscure this asymmetry.

Human–AI interaction offers an especially salient comparison. A human user may modify their work practices to accommodate a model's interface, output conventions, or access conditions. The deployed AI system may also produce different responses as a conversation develops, while the provider or deploying organization controls persistent memory, training updates, system-level instructions, compute allocation, and tool permissions. The relevant unit of analysis therefore matters: the deployed model, a conversational session, the provider, a user community, and an organization using the model have different capacities for changing the relationship. Altered outputs in a session do not automatically demonstrate autonomous revision of the model's governing norms. Conversely, human participation in an interaction does not guarantee meaningful control over the infrastructure on which participation depends.

Distinguishing *reciprocal co-development* from *one-sided accommodation* requires examining the distribution of effective options. Relevant questions include who can propose new rules, what happens following refusal, which actor can leave without losing essential resources, whether a dissenting interpretation reaches the other party, and whether an affected participant can cause a governing procedure to be reconsidered. Reciprocal co-development may exist along some dimensions and be absent along others. A system can support collaborative communication while remaining centrally governed, and a formal right of revision can coexist with material obstacles to exercising it.

These observations do not settle the ethical standing of artificial systems or the moral legitimacy of particular institutional arrangements. They identify explanatory conditions that should be assessed before describing an interaction as jointly negotiated. The political significance of evolving norms is that rule formation can alter subsequent distributions of agency and dependence. Normative improvement, by contrast, requires articulated and defensible evaluative criteria. A transition from unilateral instruction to consultation may widen participation in a measurable institutional sense, but whether consultation adequately respects each affected party depends upon its capacities, interests, and the relevant account of justification.

8.6 Path Dependence, Local Pluralism, and Discontinuity

The development of normative relations may depend on the order and particular circumstances of encounters in the *modeled history*. An initially chosen communication convention can privilege a group familiar with the medium; an early allocation rule can determine which participants later accumulate resources; and a negotiated exception can create a precedent for future amendments. These mechanisms make historical trajectories analytically relevant without treating the modeler's sequence of recorded events as an ontologically fundamental direction of time. Path dependence is a property to investigate under specific update rules, institutional memories, and opportunity structures.

A path-dependent outcome need not be uniquely stable or globally shared. Separate communities may maintain compatible local conventions while using different justificatory vocabularies. A collaborative practice can support translation between some roles and preserve uncertainty between others. Some participants may hold an explicit agreement while others exhibit learned compliance, or a negotiated arrangement may alternate between distinct procedures depending on the setting. Such outcomes suggest the usefulness of modeling *stable pluralism* as a maintained configuration of differentiated rules and interfaces. Stability here refers to observable persistence relative to a specified interval and disturbance class; it supplies no independent warrant for the justice of the arrangement.

Relational change may also be discontinuous. A new technological interface, a legal amendment, the loss of a habitat, or a breakdown in trust can abruptly remove practices that had appeared well estab-

Table 7: Candidate dimensions, evidence, and rival explanations in the study of normative–relational coevolution.

Dimension	Potential evidence of reciprocal change	Rival account or identification limit
Feasible practices and ecological conditions	Changes in available actions associated with participant-produced environmental or infrastructural modifications	Independent ecological trends, unobserved external intervention, shifts in measurement scope
Institutional permissions and revision channels	Amendments following participant proposals; subsequent changes in practical access or challenge procedures	Symbolic consultation without effective access, centrally imposed modifications
Interpretive resources and communication media	New, reliably used repair signals or contestation categories emerging through interaction	Fixed translation templates, imitation, superficial coordination without shared understanding
Interaction topology and dependence	Altered access to partners, exit options, mutual monitoring, or concentrated control	Background demographic variation, platform-wide changes unrelated to local negotiation
Normative expectations and responses	Changing expectations and reactions to deviation linked to changes in relational conditions	Habit, reward conditioning, coercive compliance, selection of participants

lished. In particular models, the discontinuity may take the form of a threshold transition, a shift in an attractor basin, or a change in the represented state space. The appropriate description depends on the available data and specified mechanism. Physical criticality, thermodynamic dissipation, and universal phase-transition language require stronger independent evidence than a visible discontinuity in conduct. The subsequent analysis of normative stability and transition will distinguish these possibilities more systematically.

A further complication concerns the identity of the evolving norm. If participants revise both the words expressing a commitment and the practical conditions of its application, apparently continuous use of a label may conceal a change in content. Conversely, a revised label can leave a practical expectation largely intact. Longitudinal research should therefore track contexts, obligations, permissions, response patterns, and routes of revision alongside linguistic descriptions. The identity of a practice across encounters is an interpretive and operational question, and claims of normative preservation should specify what has been preserved.

8.7 Analytical Criteria for Identifying Relational Coevolution

The preceding comparisons suggest that an inquiry into relational coevolution should record multiple dimensions of change and the evidence connecting them. Table 7 separates the object of possible change from indicative observations and competing explanations. It is intended as a preliminary diagnostic protocol. The rows are neither presumed to occur together nor assigned a common numerical scale.

A minimally adequate inquiry would specify the participants and relational units, establish a baseline for each relevant dimension, document interactions and external interventions, and identify a plausible causal pathway linking changes in conduct to changes in the organization of subsequent conduct. For the stronger coevolutionary interpretation, the inquiry should additionally trace how the modified organization feeds back into later conduct or candidate normative expectations; correlated trajectories alone support the weaker description of co-change. Where experiments are feasible, researchers can compare conditions with fixed versus revisable interaction rules, or with and without participant access to the interface that structures exchange. In field settings, process tracing, comparison of institutional records, observations of ecological conditions, and interviews with capable human participants may offer complementary evidence. Such sources should be interpreted according to the capacities of the relevant subjects; verbal reports cannot be required of every nonhuman system, and simulated agents' textual claims are not straightforward self-reports of moral endorsement.

The candidate mechanisms admit discriminating observations. If an apparent shift in norms can be explained entirely by a stable exogenous reward schedule, the evidence may favor within-organization adaptation. If modifying a communication interface changes behavior but leaves the range of contestable claims and effective permissions unchanged, an asserted growth in normative autonomy would require additional support. If one party's repeated objections have no discernible influence on rules or revision procedures, the record may support asymmetric stabilization over reciprocal institutional development.

Finally, if the available data cannot distinguish these mechanisms, the analysis should retain the uncertainty rather than infer coevolution solely from correlated time series.

In this sense, the proposed GR perspective offers a research question with a definite empirical burden: identify how interaction-dependent modifications to relational conditions affect the practices and expectations subsequently available to heterogeneous participants. It encourages attention to changing possibilities, interpretive mediation, and revision rights while retaining the specificity of biological, institutional, and computational accounts. The next section considers the conditions under which such changing arrangements become relatively stable, undergo transitions, or preserve several incompatible normative repertoires. A formal joint transition architecture, including the representation of potentially changing state and action spaces, will be developed later as a separate provisional model.

9 Dynamics of Normative Stability and Transition

The evolutionary study of norms requires an account of how interactional practices persist, become vulnerable to disturbance, and give way to alternative arrangements. Such an account becomes especially demanding among heterogeneous generative-relational (GR) subjects, whose environments, action repertoires, interpretations, and capacities for revision may differ. This section examines the distinction between observed persistence and model-dependent stability; the selection and replacement of conventions; path-dependent switching barriers; the maintenance of plural normative practices; and the evidential standards required for claims about abrupt or critical change. The discussion combines a limited stochastic illustration with comparative analysis of institutions, networked practices, and interacting systems. Its purpose is to establish diagnostic distinctions for a later coevolutionary formalism, while preserving the separate questions of normative endorsement and legitimacy.

9.1 Criteria for Persistence, Dynamical Stability, and Normative Endorsement

A practice can persist for several reasons. Participants may understand and endorse an expectation, may coordinate through a convenient routine, may lack access to alternatives, or may comply with an externally enforced condition. Persistent patterns can also result from ecological constraints or the regular operation of a computational interface. The empirical observation that a pattern recurs within an observation window therefore establishes a fact about conduct under specified measurement conditions. Further interpretation requires evidence about the mechanism maintaining the pattern. The present analysis treats *persistence* as the observed recurrence or maintenance of an identified practice over a stated interval, reserving *stability* for a claim that depends upon a defined disturbance class and an account of how the system responds to it.

The disturbance class matters. A human–AI exchange may exhibit stable turn-taking while the interface, model weights, moderation instructions, and user permissions are held constant. A different version of the system, a revised permission regime, or a user who can challenge the system’s rules may produce a different result. Similarly, a recurring ecological interaction can survive seasonal variation yet become impossible when a habitat is altered. Stability relative to small variations in a reward parameter is distinct from stability under a change in the available action set. A model that fixes the environment is therefore informative about a particular conditional mechanism; it is incomplete as an account of changing relational conditions.

Four notions should be distinguished in comparative work. *Observational persistence* describes recorded behavior; *dynamical stability* describes the response of a specified model state or invariant distribution to designated perturbations; *institutional reproduction* concerns the continued operation of procedures, permissions, enforcement, and supporting infrastructure; and *normative endorsement* concerns an agent’s practically or explicitly supported acceptance of a rule. These notions can overlap while remaining empirically and conceptually separable. An imposed rule may be institutionally durable while privately rejected by capable human participants; an endorsed norm may lose its enabling infrastructure;

and an artificial population can converge upon a token convention through reward-directed updates without thereby demonstrating moral acceptance.

In many contexts an additional distinction is needed between *local* and *global* stability. A practical arrangement may remain durable within one network, habitat, institution, or task domain while other arrangements remain viable elsewhere. Heterogeneous participants need not occupy a single shared state space, and a numerical measure of aggregate agreement can conceal durable differences among groups. The operative unit of analysis must therefore be specified before a claim of stability is generalized across agents, populations, or institutional levels.

9.2 Stochastic Residence and the Selection of Conventions

Evolutionary game theory provides rigorous examples of long-run convention selection under particular game forms and behavioral assumptions. Young’s analysis of repeated coordination with incomplete information and occasional mistakes characterizes stochastically stable conventions; Kandori, Mailath, and Rob similarly investigate long-run equilibria under learning and mutation [51, 30]. These results establish how specified mechanisms can select among equilibria, and they supply an instructive contrast with the question of whether the chosen practice constitutes a warranted normative commitment. The equilibrium concept derives its meaning from the modeled strategy set, payoff and information assumptions, and perturbation process.

A small probabilistic illustration makes the distinction between a long-run distribution and practical residence especially visible. Suppose an analyst represents two observable coordination regimes by states A and B and uses a time-homogeneous, two-state Markov chain during a fixed institutional configuration θ . Let α_θ denote the transition probability from A to B , and β_θ the probability of the reverse transition. Under $0 < \alpha_\theta, \beta_\theta < 1$, the transition matrix and stationary probabilities are given in Equation 5.

$$P_\theta = \begin{pmatrix} 1 - \alpha_\theta & \alpha_\theta \\ \beta_\theta & 1 - \beta_\theta \end{pmatrix}, \quad \pi_\theta(A) = \frac{\beta_\theta}{\alpha_\theta + \beta_\theta}, \quad \pi_\theta(B) = \frac{\alpha_\theta}{\alpha_\theta + \beta_\theta}. \quad (5)$$

The probabilities refer to observed regime changes in an explicitly chosen discretization. They are neither utilities nor degrees of moral acceptance. The representation also presumes that the analyst can identify A and B consistently over the observation interval. If the meaning of a practice, the available acts, or the communication interface changes sufficiently, maintaining a fixed two-state model may cease to be adequate.

Even where the stationary proportions are identical, the temporal experience of a practice may differ markedly. For example, equal transition rates $\alpha_\theta = \beta_\theta = 0.01$ and $\alpha_\theta = \beta_\theta = 0.30$ both yield stationary shares of one-half, yet they imply very different lengths of uninterrupted residence in the same regime. For an initial state A , the probability of remaining continuously in A over T updates, under the stated chain assumptions, is given in Equation 6.

$$\Pr_\theta(X_1 = A, \dots, X_T = A \mid X_0 = A) = (1 - \alpha_\theta)^T. \quad (6)$$

Consequently, similar long-run frequencies are compatible with distinct switching rates and experiences of continuity. This elementary example is a formal illustration, with no empirical calibration and no claim that an actual norm has only two possible states. Its analytical value lies in identifying a measurement problem: frequency of a convention, duration of a convention, and responsiveness to interventions answer different questions.

The distinction also clarifies the status of *attraction basins*. In a specified deterministic model, a basin consists of modeled initial conditions whose trajectories approach a given attractor. In a stochastic model with persistent noise, the analogous empirical concern may instead involve probabilities of residence, escape, or return. Neither usage automatically supplies a unique measure of normative strength. Where heterogeneous systems possess different local state descriptions or changing conditions of interaction, a purported common attractor requires a justified cross-system representation. Descriptions of basin structure should remain attached to the explicit model that generates them.

9.3 Historical Dependence and Barriers to Convention Switching

The replacement of an established convention can depend on what participants have already learned, which partners they can access, and the costs incurred when a practice is abandoned. Granovetter's threshold models demonstrate, for a specified class of collective behaviors, how the distribution of individual thresholds influences aggregate outcomes and how similar aggregate dispositions can coexist with different collective patterns [25]. Threshold mechanisms are useful here as candidate descriptions of interdependent adoption decisions. They do not, by themselves, establish that an adoption pattern is a moral norm or that participants agree on the grounds of its legitimacy.

Path-dependent change has a further analogue in Arthur's account of technologies adopted under increasing returns and historical contingencies [4]. In that model family, early contingencies can influence which technological path becomes established. An inquiry into norm formation may find similar mechanisms where learning costs, installed communication protocols, institutional investment, or network compatibility reward continued use of an existing practice. This is an analogy concerning a possible mechanism. Its adequacy must be evaluated through evidence about the feedbacks maintaining a particular normative arrangement.

For heterogeneous GR systems, a switching barrier can have several distinguishable components. An agent may need to acquire a new interpretation of another agent's signal, construct a new communication interface, secure access to alternative partners, or obtain permission to modify an institution. A human user can understand why a platform's default rules are problematic yet remain dependent on the service for work or essential communication. An artificial agent may be capable of generating an alternative rule description while lacking the technical permission to modify its operating constraints. In both cases, observable continuity can result from restrictions on effective alternatives; their specific causal architectures and implications for normative agency differ.

Such barriers also expose a limitation of single-cost models. The consequences of switching may be described in distinct and only partially comparable terms: material expenditure, ecological damage, interpretive uncertainty, vulnerability to retaliation, loss of public standing, or decreased access to future alternatives. Assigning these costs to a shared scalar objective requires additional assumptions about comparability and whose evaluative perspective is represented. A policy analyst may adopt an explicitly situated metric for a limited purpose, while the participants may assess the same transition through incompatible practical or normative terms. The model should make that disparity visible rather than conceal it within one payoff parameter.

9.4 Networked Persistence and the Maintenance of Normative Pluralism

Interaction topology can support a mosaic of conventions. Separate communities may reproduce differentiated practices through local learning, shared partners, and repeated encounters, while interfaces allow a participant to move between practices when context changes. A professional practice, an ecological coordination pattern, and a computational communication protocol may each exhibit a degree of stability in its own domain. Plural persistence then need not be regarded as an incomplete stage on the way toward a single system-wide norm.

A heterogeneous network also makes the identification of convergence sensitive to aggregation. A population-level average can increase even as differences between subgroups persist, and the same average may be produced by distinct mixtures of local practices. Analysts should report which agents interact, which edges carry communication or enforcement, and whether apparently shared conventions are used for comparable reasons. Where a switching rule depends on context, the contextual map is part of the evolving organization, even if the visible rule repertoire remains unchanged.

A useful comparative distinction is that between *stable pluralism* and *unresolved fragmentation*. Stable pluralism, as an operational proposal, involves persistent differentiated practices accompanied by workable interfaces for coordination, boundary negotiation, or mutual avoidance in the environments examined. Fragmentation describes differentiated practices whose interfaces fail or remain unspecified under the relevant encounters. The distinction is intentionally pragmatic and context-relative. A stable

plural arrangement may still contain power disparities or unacceptable exclusions; stable separation may support coexistence in one setting while entrenching domination in another. The political-economic organization of these asymmetries is developed separately in the companion GR political-economy inquiry.

The question of cross-system identity also becomes acute. Two groups may use the same word for a rule while assigning it different permissions, obligations, exceptions, and contestation procedures. Conversely, different representations may sustain equivalent practical commitments in the circumstances studied. Longitudinal comparison should therefore trace the application conditions and effects of a practice, using translation relations whose scope is explicitly limited. Where interpretive capacities themselves co-develop, apparent norm persistence may disguise substantial revision in what the norm enables or constrains.

9.5 Institutional Revision and Hysteretic Transition Patterns

Some systems react differently to an intervention depending on their current configuration and history. Within a specified model, such dependence may be represented through a *hysteresis* pattern: the condition under which a system adopts one convention differs from the condition under which it subsequently abandons that convention. The mechanism could arise through accumulated commitments, infrastructure investment, partner expectations, or changes in the feasibility of alternative practices. A demonstrated loop under controlled, reversibly varied conditions would support a stronger inference than observing one historical transition.

Consider a cooperative communication arrangement that becomes attractive only after a sufficient number of participants can interpret a new signal. Once the arrangement is established, participants may invest in training or infrastructure that lowers the level of current participation required to sustain it. Conversely, institutional sanctions may make an externally imposed practice hard to leave, even after widespread disapproval becomes observable. Both mechanisms can generate dependence on the preceding configuration, but one involves coordination-supporting investment and the other may involve restricted exit. Their normative evaluations require separate inquiry into consent, effective options, and justification.

The empirical identification of hysteresis must account for changes in background conditions. In social settings, a researcher cannot usually vary a parameter back and forth while holding history, institutions, incentives, and participant composition fixed. An observed delay following policy reversal might reflect organizational inertia, new information, strategic obstruction, or an irreversible loss of alternatives. The label *hysteresis* is therefore best reserved for a specified mathematical model or for evidence that discriminates a history-dependent switching mechanism from these rivals. Ordinary institutional persistence may be a more accurate initial description.

This qualification is particularly important where a normative transition changes the model's feasible acts. An amendment can create new appeal procedures, a platform revision can introduce new authorization categories, and an ecological transformation can change the species and agents able to participate. Under these conditions, the old and revised arrangements may require different local state spaces. A single smooth trajectory connecting them would be an additional modeling construction, not an independently established property of the encounter. Section 11 develops a preliminary formal framework in which such representational changes can be stated more explicitly.

9.6 Abrupt Change and the Evidential Scope of Criticality

Abrupt or threshold-like changes have been documented in controlled social coordination experiments. Centola and colleagues, for example, report convention-switching behavior associated with the size of a committed minority within a specific experimental naming and coordination setting [14]. The study provides evidence of a threshold effect in its experimental domain. It does not determine a universal critical fraction for human societies, and still less does it establish a universal transition point for encounters among humans, artificial agents, other species, or hypothetical extraterrestrial systems.

Table 8: Operational indicators and inferential limits in the comparative study of normative stability and transition.

Analytical claim	Potential evidence	Rival mechanism or scope limit
Persistence of a practice	Longitudinal recurrence under documented contexts and observation intervals	Repeated ecological constraint, habit, selection of encounters, or fixed interface defaults
Dynamical residence or return	Estimated switching and return probabilities under controlled or tracked perturbations	Time-varying transition rules, unidentified hidden states, aggregation across local regimes
History-conditioned switching	Different transition responses associated with recorded prior configurations or investments	Exogenous policy changes, cohort replacement, changing material conditions
Plural normative stability	Durable differentiated practices together with identifiable coordination or boundary procedures	Aggregate averages concealing incompatible local expectations or enforced separation
Institutional stabilization	Effective rule reproduction and contestation channels documented across interventions	Formal rights without practical access, one-sided compliance, infrastructure monopoly
Threshold or critical transition	Reproducible response near specified parameters; diagnostics tied to an explicit model	Discrete intervention, measurement discontinuity, demographic selection, metaphorical phase-transition language

The mathematical theory of critical transitions concerns more specific dynamics. Kuehn's treatment examines bifurcations, fast-slow systems, stochastic dynamics, and conditions under which proposed early-warning indicators have theoretical support [31]. Applying such models to normative practices requires a defensible correspondence between observable variables, the assumed dynamical equations, the parameter changes, and the regime considered. A sudden change in recorded adoption, without this correspondence, supports a description of rapid change but leaves the stronger dynamical interpretation open.

Several explanatory rivals deserve explicit treatment. An abrupt shift in observed behavior may result from a discrete legal decision, a software deployment, a sudden change in what is measured, or the selective disappearance of one participant group. A population may also cross a behavioral threshold because the distribution of local preferences changes while the interaction mechanism remains fixed. Some of these mechanisms can be represented by nonlinear dynamical systems; others are better examined with institutional records, causal process tracing, or comparative qualitative methods. The choice of method should follow the claim under examination.

Thermodynamic vocabulary warrants even greater care. Dissipation, free energy, and physical phase transition are well-defined in physical theories with specific state variables and conservation or statistical assumptions. A conflict among social norms can involve costly coordination, concentrated power, or declining trust without thereby constituting thermodynamic dissipation. Similarly, the presence of an attractor in a fitted social model does not establish a physical phase transition in social reality. The GR vocabulary of generative capacity and relational change remains compatible with mathematical dynamics while maintaining the distinction between representation, empirical mechanism, and ontology.

9.7 Comparative Diagnostics for Normative Stability and Transition

The distinctions developed above suggest an evidential protocol for comparing transition processes across heterogeneous systems. Table 8 links candidate claims to minimally relevant observations and alternative mechanisms. Its purpose is to organize a research design, not to award a common numerical stability or moral-quality score to different kinds of subjects.

A comparative study can begin by identifying the unit of analysis, the participants' available practices, the conditions governing entry and exit, and the interpretable evidence of normative expectations. It can then record variation in interaction topology, communication interfaces, sanctioning capacity, institutional rules, and ecological conditions, distinguishing changes initiated by participants from imposed or incidental changes. Where interventions are possible and ethically appropriate, researchers can compare a fixed-rule arrangement with one in which participants possess effective avenues of procedural revision. Longitudinal observations, experiments, software and institutional change logs, and situated interpretation by capable human participants can complement one another. Evidence requirements should remain appropriate to the agents involved; a nonverbal organism or computational process cannot simply be

assessed through the same interview-based measures used for a human institution.

A transition claim should also specify the possible observations that would weaken it. Evidence of high recurrence without recovery following perturbation weakens a claim of robust dynamical stability. Evidence that all changes follow external platform updates weakens a claim of endogenous normative coevolution. An apparently universal threshold that varies substantially across network arrangements or communication media should be reported as setting-dependent. Stable aggregate agreement combined with incompatible local interpretations weakens the inference of a shared norm. Persistence produced by threat or restricted exit may support a claim about effective enforcement while leaving endorsement and legitimacy unsettled.

Normative stability therefore requires separate analysis of maintenance mechanisms, responses to departures, and revision opportunities. Some arrangements may persist through coordination; others through material dependence, selective interaction, coercion, or ecological constraint. Some may switch rapidly, while others retain differentiated repertoires across changing contexts. The GR contribution proposed here is to include the organization of future interaction among the possible objects of transition. The next section examines in greater detail how asymmetric control of enforcement, evidence, and infrastructure can stabilize norms, including cases where behavioral alignment arises without reciprocal endorsement. The fully coupled representation of changing relational conditions remains an open formal task developed in Section 11.

10 Power-Conditioned Norm Stabilization

The stabilization of an interactional rule can depend upon the distribution of capacities to enforce conduct, determine the subjects of negotiation, circulate evidence, and maintain the material interfaces through which participation occurs. This section examines these mechanisms as conditions of normative evolution among heterogeneous generative-relational (GR) subjects. Its immediate objective is to explain how a practice can become recurrent and institutionally durable while participants continue to differ in their reasons for following it, their interpretations of its meaning, or their effective opportunities to challenge it. The discussion proceeds through sanctions, agenda setting, interpretive authority, infrastructure dependence, and human–artificial intelligence (AI) interactions. It then develops a limited observational argument and proposes comparative criteria for distinguishing coordination, coerced compliance, and justified acceptance. The approach is mechanism-oriented and interpretive, informed by institutional theory and studies of power; it does not treat a particular account of political economy as already established for every class of GR system.

Norm recognition and enforcement already feature in normative multi-agent systems research [43]; institutional-grammar scholarship also separates rules, norms, and shared strategies [17]. The present discussion adds a comparative question about the distribution of enforcement, interpretive, and amendment capacities across participants whose control over their interfaces may differ. The expression *power-conditioned norm stabilization* refers here to a family of processes in which an asymmetry in participants' effective capacities changes the reproduction, visibility, or revision of an interactional rule. The relevant capacities can include the ability to impose consequences, exclude alternatives from deliberation, authorize classifications, limit access to a shared system, or alter the procedures by which a rule is contested. These mechanisms can overlap, and their analytical boundaries depend on the material and institutional setting. Power is therefore treated as a condition to be identified through its operation, rather than as a scalar property inferred from a participant's general intelligence, technological sophistication, or success in securing observed conformity. The expression *norm stabilization* is used provisionally: when evidence establishes only recurrent behavior under enforcement, the observed object should be classified as a stabilized compliance pattern; its status as a norm remains subject to the diagnostic criteria in Section 3. A technological advantage may expand an actor's options for enforcement or assistance; its effects on the formation and legitimacy of norms remain contingent on the relations and practices in which that advantage is exercised.

10.1 Sanctions and Behavioral Alignment under Asymmetric Enforcement

Sanctions may make a rule reproducible by altering the consequences associated with deviation. A participant can anticipate loss of access, material penalties, exclusion from a community, reputational consequences, or diminished opportunities for further interaction. Such consequences can stabilize an observable practice even if some affected participants regard the rule as mistaken or lack a shared interpretation of its rationale. In human settings, theories of social norms have already distinguished conformity supported by empirical and normative expectations from conduct maintained through other incentives [8]. The present analysis uses that distinction to ask which mechanisms sustain a recurring practice when the participants do not necessarily share the same means of forming, expressing, or revising commitments.

Enforcement can be reciprocal, collectively supervised, centrally administered, or embedded in an interface that automatically rejects disallowed actions. These arrangements have different implications for normative evolution. A collectively agreed response to damaging conduct may help participants maintain a negotiated standard and preserve conditions for future cooperation. An opaque sanction imposed by a participant controlling essential infrastructure can produce outwardly similar conformity while leaving the affected participants with little influence over the rule. Ostrom's analysis of collective resource governance provides a substantive human institutional precedent for treating monitoring, sanctions, and local rule arrangements as conditional features of particular governance systems [39]. It does not establish that an identical enforcement mechanism is appropriate for other animals, computational agents, or hypothetical extraterrestrial societies.

The key empirical distinction concerns the path through which a participant's conduct becomes rule-consistent. Evidence that a practice recurs while enforcement is present supports a finding of effective behavioral stabilization. An inference concerning endorsement requires additional evidence, obtained through methods suited to the type of participant. Among humans, private reports, opportunities to disagree without penalty, and observed choices under changed sanction exposure may be relevant. Kuran's analysis of preference falsification explains how publicly expressed human preferences can diverge from privately held ones in response to perceived social pressure [32]. Comparable discrepancies in AI-generated text should first be investigated through instruction hierarchies, decoding, optimization objectives, and deployment constraints; public and private human preferences cannot simply be imputed to a computational system. For nonhuman organisms, behavioral changes under social pressure would require species-appropriate controls before being described as norm-guided conformity.

An enforcement arrangement can also change the direction of subsequent normative development. Participants may learn to avoid a disapproved act, devise workarounds, adopt a shared interpretation of the sanction, or coordinate to revise the underlying rule. In a sufficiently restrictive setting, visible departures from the imposed practice may become rare while the capacity to develop alternatives remains active in less observable interactions. Stability in the measured behavioral trace thus depends partly on the investigator's access to suppressed or displaced practices. The distinction is essential when interpreting apparently consensual outcomes in asymmetrically governed settings.

10.2 Agenda Setting and the Boundaries of Normative Deliberation

Power can influence the formation of a rule by shaping the set of proposals that participants are permitted or able to consider. Bachrach and Baratz's account of the two faces of power draws attention to nondecision-making and the control of the political agenda alongside overt choices among presented alternatives [6]. Within its original human political context, this analysis provides a reason to examine the organization of deliberation as well as the final decisions it produces. Its extension to heterogeneous GR systems is a proposed analytical comparison: who can present a candidate rule, what kinds of objection can enter the interaction, and which matters remain outside a shared interpretive interface?

Consider a hypothetical encounter between an administrative organization and an AI-mediated service used by a community. Members may have opportunities to choose among three settings while being unable to propose changes to the categories represented by those settings. The service can thereby register repeated selections and infer a preferred default, although the range of acceptable options was

determined elsewhere. A similar structure can arise in an international negotiating procedure, a research consortium's access policy, or a human-managed ecological intervention. These cases differ in the forms of participation available to affected beings; their common analytical feature is that observed agreement concerns a bounded agenda whose construction itself requires explanation.

Agenda control may stabilize practices through at least three distinguishable mechanisms. First, selection among available options can reproduce the categories that organize later choice. Second, recurrent exclusion of alternative proposals may limit the evidence through which participants learn that disagreement exists. Third, control over amendment procedures can make an established arrangement more durable than its levels of endorsement would independently warrant. A methodological implication follows: a study of norm formation should document the history of proposed alternatives, refused questions, available amendment procedures, and actual participation in setting these procedures. A count of votes, compliant actions, or model responses provides a partial record when the production of the choice set is itself asymmetrical.

The capacity to set an agenda also admits beneficial and collectively authorized uses. Limited menus, professional standards, and emergency procedures may lower coordination costs, prevent particular harms, or make communication possible under severe uncertainty. The relevant research question concerns the grounds and consequences of their adoption: which actors participated, how exclusion was justified, whether meaningful challenges are possible, and which alternatives remain feasible. A power-sensitive analysis should therefore avoid equating all constraints with illegitimate domination. Equally, practical convenience supplies limited evidence about the authority of the actor deciding which alternatives others may pursue.

10.3 Interpretive Authority and the Circulation of Normative Evidence

The identification of a norm requires interpretive judgments about whether a behavior expresses an expectation, a strategic response, an accidental regularity, or an obligation recognized by participants. Actors may differ in their ability to produce the evidence on which such judgments are made and in the authority accorded to their interpretations. In institutional settings, records, categories, and official descriptions can determine which actions are recognized as compliance and which departures become visible. An interactional rule may thus stabilize partly through a durable practice of classifying the conduct of others.

Fricker's analysis of epistemic injustice provides a human social-philosophical precedent for examining unequal standing in processes of knowing and interpreting [19]. The GR question is related but distinct. It asks how access to communication channels, measurement systems, archives, and classificatory procedures conditions the emergence and apparent durability of rules across agents with different interpretive capacities. Some asymmetries arise through institutional authority, others through sensor limitations, incompatible representational systems, inaccessible technical documentation, or one-sided control over translation services. An AI-mediated interface may recognize particular categories of grievance while lacking an avenue for users to make other forms of concern intelligible to its operators. Describing that interface requires evidence about design decisions and actual use, rather than an attribution of intentional exclusion to the software.

Interpretive authority may operate on both sides of a rule. It helps constitute which behavior is considered appropriate, and it influences whether disagreement is recorded as normative opposition, noise, technical error, or failure of understanding. These classifications can direct future interventions. A community whose objections are repeatedly categorized as errors may be subjected to further system calibration, while a governance process that recognizes objections as reasoned challenges may invite procedural amendment. Identical observed conduct can therefore acquire different institutional histories according to the available interpretive channels. Translation, discussed in Section 6, supplies one potential means of revising these classifications, while Section 3 sets out the evidential caution needed before attributing normative commitments across heterogeneous cases.

Reciprocal interpretation also need not imply symmetric expertise. A specialist may accurately diagnose a technical failure or ecological hazard while affected participants hold indispensable knowledge of

the consequences for their practices. Different interpretive contributions can be complementary. The analytical issue is whether their respective evidential standing is specified, whether uncertainty is recorded, and whether an authoritative interpretation can be contested when new evidence becomes available. Such conditions are especially relevant when norm formation itself changes which forms of expertise and testimony count as acceptable.

10.4 Infrastructural Dependence and the Feasibility of Normative Alternatives

A practice may become stable because the arrangements that make alternatives practicable are controlled asymmetrically. Technical interfaces, access credentials, resource allocation procedures, habitat boundaries, and communication infrastructures are heterogeneous mechanisms through which an actor's available activities may be enabled or restricted. Star and Ruhleder's account of infrastructure examines the relationship between system design, access, communication, and organizational change in a situated scientific collaboration [45]. Its relevance here is methodological: the technical substrate of an interaction can itself be part of the relational organization through which expectations become durable.

An organization may, for example, permit participants to request reconsideration of an imposed standard while requiring the appeal to pass through a platform to which some participants have unstable access. The rule may then appear widely observed because alternative procedures are difficult to initiate. In another setting, a shared infrastructure may lower the cost of participation for previously excluded actors and enable more reciprocal rule making. Dependence therefore has several possible consequences, and the causal pathways should be traced rather than assumed. Relevant measurements include access reliability, the cost of switching to an alternative service, the portability of records, the availability of independent communication, and the ability to continue productive activity during a dispute.

The relation between *formal exit* and *effective exit* deserves particular care. Formal permission to discontinue participation can coexist with conditions under which discontinuation threatens livelihood, access to essential services, social relationships, or the continuation of a research activity. These consequences can influence behavior even without a direct command. Conversely, a costly exit need not imply coercion in every setting; investments in a mutually beneficial infrastructure may generate dependence through voluntary coordination. Identifying domination requires additional information about control, contestability, alternatives, and the affected participants' standing. An inquiry into norm stabilization can record these dependencies while leaving the more comprehensive political economy of production, ownership, and value circulation to the companion research program on heterogeneous GR subjects.

A further distinction concerns the capacity to modify the infrastructure itself. Participants may adapt their behavior to a fixed interface while having no authority to alter its permissions. Others may possess administrator access, the ability to revise the schema of recognized actions, or the technical knowledge needed for maintenance. Such distribution can produce a stable division between those who adjust their conduct and those who adjust the conditions of conduct. Repeated compliance in this setting constitutes evidence of a particular organization of interaction; evidence of reciprocal normative authority would additionally require effective participation in the modification of rules and their enabling interfaces.

10.5 Human–AI Interactions and the Distribution of Rule-Revision Capacity

Human–AI interaction illustrates the importance of distinguishing visible behavior from the institutional and technical arrangements through which that behavior is produced. A human user and an AI service may exchange similar verbal commitments, issue requests, provide corrections, and appear to reach an agreement about future conduct. The service's response is nevertheless conditioned by training, system instructions, interface design, memory architecture, tool permissions, evaluation procedures, and deployment choices. The human participant's conduct is shaped by different histories and material requirements, including professional roles, social obligations, economic dependence, and available avenues of objection. Similar dialogue is therefore compatible with substantially different capacities to form, express, enforce, or revise a norm.

Current AI services can make interactional rules visible through stated policies, automatic moderation, refusal patterns, ranked outputs, and the terms under which a user may access or contest the service. A user may influence responses through feedback or permitted configuration changes, while a provider or institutional deployer retains control over model updates, data handling, authorization policies, and service continuity. These roles vary across deployments. Some systems are open to local modification, some are administered by organizations other than their developers, and some expose extensible governance interfaces. Analytical claims about authority therefore require identification of the relevant human individuals, institutions, and technical components. The apparent AI interlocutor is not necessarily the locus of the power exercised through its outputs.

Research on organizational algorithmic auditing offers an applicable, bounded form of institutional evidence. Raji and colleagues propose an end-to-end internal auditing framework in which decisions and their justifications are documented across an AI development lifecycle [42]. For the present inquiry, development records, deployment logs, version histories, user appeals, and post-deployment evaluations can help distinguish behavior produced by a revised model from behavior produced by a changed interface or institutional rule. Such documents do not settle the subjective or moral status of an AI system. They can, however, illuminate which actors possess powers of revision and which participants can challenge a behavioral standard after it has become established.

The converse direction of influence also merits attention. A widely adopted AI service may condition the communication habits, professional judgments, and practical expectations of its human users. Repeated suggestions or classifications can become familiar reference points even when the service lacks independent authority to endorse them. This influence should be investigated through measured behavioral changes, access to alternatives, and the surrounding organizational decisions. It would be inaccurate to infer an autonomous political intention from an observed system-level influence, just as it would be incomplete to describe the influence solely through the individual intentions of users. The GR perspective proposes a study of the situated network of dependencies in which the interactional norm is reproduced.

These issues extend beyond an exclusively human account of AI governance. Future computational systems may exhibit more persistent learning, organizational initiative, or capacities for rule revision than currently documented services. Assessments of those possibilities would require evidence adequate to the relevant claim. The present section remains agnostic about machine consciousness and moral personhood while maintaining that human–AI interactions already present concrete questions of institutional power, practical dependency, and normative influence.

10.6 Behavioral Compliance, Normative Endorsement, and Observational Equivalence

Power-conditioned stabilization presents an identification problem. The observation that participants repeatedly follow a rule may be consistent with several generating mechanisms: learned coordination, practical convenience, endorsement, avoidance of sanctions, lack of alternatives, or the filtering of non-conforming responses by an interface. These mechanisms can coexist. The observational record may therefore support a finding of compliance while underdetermining the mechanism responsible for it.

A minimal illustrative model makes this limitation explicit. Suppose, only for a restricted population of human decision makers and a specified action, that an observed choice to comply is governed by an unobserved net inclination v_i toward that action, a context-dependent external inducement c_i , and a decision threshold τ_i . The stylized observation rule is defined in Equation 7.

$$y_i = \mathbf{1}\{v_i + c_i \geq \tau_i\}, \quad y_i \in \{0, 1\}. \quad (7)$$

Here $y_i = 1$ denotes observed compliance with the specified practice. The latent term v_i aggregates the participant’s reasons for adopting the action independent of the modeled inducement, while c_i denotes a net effect of the identified inducement; neither term is a measure of moral legitimacy. Equation 7 is a deliberately simplified identification example, with an additive decision rule assumed for exposition. At $\tau_i = 1$, the distinct assignments $(v_i, c_i) = (2, 0)$ and $(v_i, c_i) = (-1, 3)$ generate the same observed

compliance. The first assignment is consistent with an internally favorable action inclination; the second requires a stronger external inducement to offset an unfavorable one. Observing y_i alone cannot determine which assignment generated it.

This illustration is neither a psychological law nor a general theory of normative motivation. Its significance rests on the structure of the observational inference. Identifying the underlying mechanism requires additional measures, assumptions, or interventions. For capable human participants, one might combine protected expressions of disagreement with changes in enforcement conditions and longitudinal observation. For artificial systems, the relevant additional evidence concerns policy constraints, optimization procedures, version histories, and controlled changes to access or instructions; the latent human inclination v_i in Equation 7 should not be transferred to machines as an unexamined mental attribute. Other organisms may require ecological, developmental, or behavioral designs appropriate to their capacities. A model's latent constructs therefore need not be shared across the heterogeneous participants being compared.

A further identification problem arises when the interaction itself determines which responses are recorded. If a platform discards disallowed requests before they reach an observer, the resulting data describe the distribution of visible requests conditional on filtering. If a governance process excludes certain affected groups from consultation, an apparent consensus describes the participating group and its available proposals. The evidence may be reliable within those bounds while insufficient to establish a wider agreement. Claims about compliance should accordingly state the population, the observation interface, the selection mechanism, and the practices that were actually possible.

10.7 Contestability and the Empirical Identification of Power-Conditioned Stability

The analysis suggests that power-conditioned stabilization is most usefully evaluated through a comparison of the mechanisms that make a rule persistent and the opportunities available for its revision. In an institutional setting, contestability can be operationalized through actual access to appeal, transparent reasons for rejection, independent review, and the ability to propose alternative procedures. In technologically mediated interactions, relevant evidence can include user-accessible records, interoperability, documented policy changes, permissions to modify decision rules, and the consequences of discontinuing a service. For interactions involving nonhuman living beings, the ethical and epistemic tasks differ: researchers may investigate welfare-relevant responses and material alternatives while avoiding unsupported assumptions of propositional objection or human-style deliberative consent.

These considerations provide possible conditions for more accountable relations; they are not a universally valid protocol. Reversibility may matter where a mistake risks irreversible harm, while prolonged reversibility may be unattainable in ecological or institutional systems. Participation procedures require choices about representation and expertise, and those choices may themselves reproduce power inequalities. An exit option may protect autonomy in one setting while shifting burdens onto vulnerable participants in another. Local arrangements must therefore be evaluated in terms of their actual operation and the stakes of the affected relations, with space for competing procedural and substantive accounts of legitimacy.

Table 9 organizes candidate observations that could help separate different stabilization mechanisms in concrete research settings. The categories are intended for diagnosis within specified settings. A single empirical indicator can be compatible with multiple mechanisms; the table accordingly pairs each mechanism with a rival explanation or identification limit. The proposed comparison can be implemented through historical documents, participant observation, field interviews when appropriate, institutional and software audit trails, and controlled interventions where ethical and feasible. The research design must also consider whether the act of observation changes the risks of dissent.

Several counterexamples should remain part of this research program. A norm can be strongly sanctioned and widely endorsed, so sanction exposure alone cannot establish the absence of agreement. A default can be selected by a participatory body and provide a useful coordination service. An infrastructure may be jointly governed even where technical permissions remain unequal because specialist roles

Table 9: Candidate diagnostics for power-conditioned norm stabilization in heterogeneous interactions. Each indicator requires setting-specific interpretation.

Mechanism	Candidate observation	Alternative explanation or limit
Enforcement exposure	Practice changes under identifiable variation in sanction conditions	Concomitant information, risk, or material changes
Agenda restriction	Alternatives excluded from a recorded deliberation or interface	Legitimate scope limits, feasibility, unobserved proposals
Interpretive authority	Disputes depend on a controlling actor's classification or translation	Genuine expertise differences, measurement limitations
Infrastructural dependence	Compliance persists when alternative access is costly or unavailable	Mutual investment, coordination benefits, participant selection
AI rule governance	Output standard changes with documented platform or policy revision	Concurrent model, user-population, or task changes
Contestability	Disagreement produces a traceable revision, appeal, or reasoned response	Formal appeals with little effective access or influence

are necessary. A participant may accept a rule for reasons that are intelligible within its local normative framework despite having limited influence over the rule's initial formulation. Conversely, a formally participatory procedure may leave the allocation of effective power almost untouched. These possibilities require close observation of the conditions through which practices and permissions are reproduced.

The central claim of this section is accordingly bounded. Asymmetric control of sanctions, deliberative agendas, interpretation, and infrastructure can provide mechanisms for reproducing interactional regularities under conditions of heterogeneous endorsement and uneven revision capacity. Their presence is relevant to the explanation of norm formation and stability, while the justice or legitimacy of their exercise requires further normative evaluation. The accompanying GR political-economy inquiry may examine the ownership, production, and circulation of the conditions of power in greater depth. In the present manuscript, the result is a methodological requirement for the forthcoming formal model: candidate normative states and update processes should remain analytically distinguishable from the relational capacities that condition which rules are enacted, observed, challenged, and revised.

11 A Preliminary Formal Framework for Normative Coevolution

A formal investigation of normative coevolution among heterogeneous generative-relational (GR) participants requires a representation in which interaction can alter both participants' practices and the conditions under which subsequent practices become feasible or intelligible. This section proposes a minimal family of coupled transition models for that purpose. It specifies participant-dependent state and action spaces, locates normative and interpretive variables within an evolving relational configuration, and separates observable conduct from the latent mechanisms that might generate it. A small illustrative construction then exhibits negotiated coordination, asymmetric fixation, context-dependent pluralism, and continued alternation as distinct outcomes of different specified update rules. The final analysis establishes an elementary observational identification limit and describes the evidence required to give the framework empirical content. The construction is a candidate modeling language informed by research on conventions and multi-agent learning. Its claims concern specified transition mechanisms and observations, with no implied ontology of normative subjectivity.

11.1 Typed Participants and Contextual State Descriptions

Let I be a finite index set of modeled participants. For the probability-theoretic version of the model, equip each local state space and each observation and action space with a specified σ -algebra; standard Borel spaces are a sufficient, though stronger than always necessary, regularity assumption for the constructions considered here. Products of these spaces carry their product σ -algebras. These conditions are

assumptions of the present formal model and are not claims about the mathematical structure of lived experience. The index identifies roles in a particular encounter and does not determine whether a participant is a human person, an artificial system, a collective institution, or an organism. For each $i \in I$, define a local description space X_i , a candidate normative-repertoire space $\mathcal{N}_i$, and an interpretation or memory space M_i . A modeled local configuration is $z_i = (x_i, n_i, m_i) \in Z_i = X_i \times \mathcal{N}_i \times M_i$. The term *candidate* is consequential: n_i may represent an analyst’s classification of practical expectations, a software policy, a publicly expressed commitment, or, where evidence supports the attribution, an agent’s internalized normative orientation. Distinguishing these cases is part of model specification. Their encoding within a common tuple does not make them psychologically or morally equivalent.

Write R for the space of relational configurations and E for environmental or infrastructural conditions. A relation $r \in R$ may record asymmetric permissions, channels of communication, access to an enabling resource, enforcement authority, or an observed interaction topology. The environment $e \in E$ may describe available signals, equipment, habitat conditions, or an institutional setting. These spaces are likewise chosen by the investigator for an explicit research purpose. A history variable h_t retains relevant records of interactions, interventions, and changes in context. The event index t enumerates modeled observations and updates; it introduces no metaphysically privileged temporal ordering of GR phenomena. A fixed measurable space H may include finite histories of every length as a disjoint union of event-sequence spaces, with the corresponding disjoint-union σ -algebra. Using an actual history summary instead requires demonstrating that it retains the information needed for the intended transition specification.

The complete analytical state is defined in Equation 8.

$$w_t = ((z_{i,t})_{i \in I}, r_t, e_t, h_t) \in \mathcal{W} = \left(\prod_{i \in I} Z_i \right) \times R \times E \times H. \quad (8)$$

Here H is a specified space of histories or history summaries. When h_t records complete event sequences, the realized history length changes with t , although H itself remains fixed. With a full history encoded in the augmented state and a time-invariant event mechanism, a Markov representation is possible at this descriptive level; a shorter summary is sufficient only under an explicitly justified conditional-sufficiency assumption. The use of a product space permits distinct types X_i , $\mathcal{N}_i$, and M_i ; it does not posit a shared psychology, a common scalar valuation, or a translation isomorphism between participants. Participants may also be aggregates whose internal heterogeneity is suppressed at the chosen observational resolution. The consequences of aggregation belong in the study’s error analysis.

Actions are represented through participant-specific correspondences $A_i(w_t) \subseteq \bar{A}_i$, where $\bar{A}_i$ is a fixed ambient action alphabet introduced for bookkeeping. For the stochastic construction, each $A_i(w)$ is assumed nonempty, with a measurable graph in $\mathcal{W} \times \bar{A}_i$, and the modeled action-selection law must assign it probability one. These assumptions are part of the admissible model class; they may fail when available actions or participants cannot be reliably enumerated. Membership in $A_i(w_t)$ means that an action is feasible or permitted in the modeled conditions. A change to r_t , e_t , or $z_{i,t}$ can therefore change the feasible action set without changing $\bar{A}_i$. This convention provides a mathematically economical way to encode institutional revision, technological enablement, loss of access, and newly acquired communicative practices. At this descriptive resolution, $A_i(w_t)$ abbreviates a context-sensitive feasible-action correspondence: the local notation $A_i(z_{i,t}, r_t, e_t)$ used in Section 7 is recovered whenever history affects feasibility only through these displayed state components. The ambient alphabet is a modeling device: it need not contain every action an evolving real-world system might invent. An observed action outside the alphabet signals an inadequacy of the current model and calls for revision of its descriptive domain.

11.2 Observation, Interpretation, and Coupled Updates

Each participant may access a different observation space O_i and receive an observation through a context-dependent kernel O_i . Joint observations can be correlated, as when one shared event simultaneously alters several participants’ information. Define a measurable joint observation kernel $O(\mathbf{do} \mid w_t)$

on the product observation space, with participant-specific marginals. Let $\iota_i(w_t)$ denote the participant's modeled, accessible pre-action information, with a specified measurable information map; a local policy may then be written $\pi_i(da_i | \iota_i(w_t), o_i)$. Distinct full states can induce the same accessible information, so the policy must respond identically to those states when their observed signals coincide. Policies are assumed measurable and supported on $A_i(w_t)$ for the relevant states and observations; feasibility must therefore be compatible with the information actually available to the decision mechanism. Where conditional independence of action selections is justified, the joint action law can factor into local policies. Correlated or institutionally mediated choices instead require a specified joint action kernel $\Pi(da | w_t, \mathbf{o})$, supported on the product of feasible-action sets. In either case, the joint observation-and-action law is $B(d\mathbf{o}, d\mathbf{a} | w_t) = O(d\mathbf{o} | w_t)\Pi(d\mathbf{a} | w_t, \mathbf{o})$, with the product interpreted as composition of kernels. No independence of observations or common access to hidden states is assumed.

After observations and actions, the participants' local descriptions, relational configuration, and environment may change jointly. The coupled transition representation is given in Equation 9.

$$\begin{aligned} (\mathbf{o}_t, \mathbf{a}_t) &\sim B(\cdot | w_t), \\ (\mathbf{z}_{t+1}, r_{t+1}, e_{t+1}) &\sim K(\cdot | w_t, \mathbf{o}_t, \mathbf{a}_t), \\ h_{t+1} &= U_H(h_t, \mathbf{o}_t, \mathbf{a}_t, \mathbf{z}_{t+1}, r_{t+1}, e_{t+1}). \end{aligned} \tag{9}$$

The observation-and-action kernel B and coupled update kernel K are candidate empirical specifications, not universal generative laws. An observed change in several state components represents co-change within this formal description; evidence for normative–relational coevolution additionally requires identifying a feedback mechanism linking evolving practices with changes in relational conditions, as distinguished in Section 8. In a particular operational model, that claim calls for two distinguishable causal contrasts: varying a feasible practice or proposal can alter a subsequent relational condition, and varying that condition can alter a later practice or norm-like expectation. Such contrasts require specified interventions, temporal observation, and assumptions that address common causes; a correlated coupled update kernel does not establish them by itself. To be a well-defined stochastic process, B and K must be measurable probability kernels on the declared product spaces; the history update U_H must be measurable as well. These conditions are formal assumptions, and they do not imply that the kernels can be recovered uniquely from observational data. A deterministic dynamical model arises when the kernels assign unit mass to specified outcomes. The history update U_H may retain a bounded, predeclared summary or append an event to a complete record. Its informational fidelity has to be assessed relative to the intended inference. For example, discarding rejected negotiation proposals may produce an apparently stable convention that conceals continuous institutional disagreement.

Equation 9 extends the local update representation in Equation 4. A local kernel can describe one participant's adjustment conditional on its observations and the current context; a coupled kernel additionally specifies how simultaneous adjustments and shared conditions interact. The representation is compatible with the general idea of multi-agent stochastic interaction developed in Markov games [34], yet it deliberately leaves payoff functions optional and makes rule repertoires, feasible practices, and interpretive resources eligible for change. Littman's original analysis concerns a specific class of zero-sum games; the present extension makes a separate proposal about evolving normative descriptions and carries no result about equilibrium existence or convergence from that study.

The internal structure of K should be specified at the resolution justified by evidence. For one application it may be reasonable to posit a sequential mechanism: an observed interaction changes interpretive memory, interpretive memory changes subsequent rule selection, and the chosen practice alters institutional permissions. Another application may require a simultaneous or correlated update, as when platform revisions affect several parties at once. A factorization into independently updating local kernels is a substantive assumption and can be tested only against evidence capable of detecting cross-participant dependencies. In particular, an observed change in one participant's conduct does not by itself show that the other participant's normative orientation has changed.

11.3 Normative Repertoires and Partial Interpretive Relations

A candidate repertoire n_i can be specified as a collection of context-indexed dispositions, rule representations, or institutional prescriptions, provided that its empirical interpretation remains explicit. For a human group, a recorded rule might concern an expectation about turn taking, together with evidence of member recognition or sanction. For a deployed artificial agent, it might be a versioned set of instructions and accessible procedures for selecting a response. For a nonhuman animal population, it might be an analyst's provisional hypothesis concerning socially maintained behavioral expectations. These instantiations require different validation standards. In particular, a model variable called n_i does not establish consciousness, moral judgment, or personal autonomy.

Interpretation can be represented with a time- and context-indexed relation $\mathcal{T}_{ij,t} \subseteq U_i \times U_j$ between selected expression or practice descriptions U_i and U_j . A pair in $\mathcal{T}_{ij,t}$ records an empirically supported correspondence for a specified interactional purpose, such as a reliable response to a warning in a particular environment. The relation may be partial, many-to-many, and asymmetric in evidential support. Its revision can be encoded within $m_{i,t}$ and $m_{j,t}$, or within shared relational resources in r_t . This treatment preserves the possibility that successful interaction depends on a locally useful alignment while divergent meanings and interpretations persist.

A particularly significant distinction arises when the ability to interpret an interaction depends on a resource controlled by a third party. An AI-mediated translation interface, for example, may determine which expressions are transmitted, recorded, or classified as requests for rule revision. A change to the interface can alter $\mathcal{T}_{ij,t}$ through r_t even if the participants' candidate repertoires remain unchanged. Conversely, sustained interaction may produce a shared signal or new category that expands feasible actions in $A_i(w_t)$ and $A_j(w_t)$. These possibilities make the relationship between normative updating and changing interpretive conditions empirically investigable. They also prevent a purely behavioral match from serving as sufficient evidence for equivalent commitments, as analyzed in Section 6 and the preceding diagnostic sections.

11.4 A Two-Participant Illustration with Divergent Normative Outcomes

Consider two hypothetical participants A and B repeatedly encountering a simple coordination interface. For explanatory economy, each has two publicly observable responses, labeled 0 and 1 by the analyst, along with a local expected-response marker $q_{i,t} \in \{0, 1\}$. The labels are external interface codes. They do not presuppose common meanings, comparable internal evaluations, or moral standing. The initial markers differ, with $q_{A,0} = 0$ and $q_{B,0} = 1$. The interface state r_t specifies who may revise the protocol, whether an amendment channel exists, and which context c_t governs the exchange. The following cases are distinct possible specifications of K in Equation 9; they are not predictions about actual populations. Most illustrate coordination or rule-conditioned adjustment without meeting the stronger feedback-supported coevolution criterion. To satisfy that criterion, a case would additionally need a specified pathway whereby the participants' practices change an operative relational rule and the changed rule affects subsequent practices.

In a *negotiated-coordination* regime, the protocol admits an explicit proposal-and-confirmation step. Let a successful mutual acknowledgment at event t set the next pair of public markers to $(0, 0)$, while an unacknowledged proposal preserves the current pair. The update then has a jointly accepted coordination state under the specified acknowledgment rule. Participants could nevertheless assign different private significance to marker 0. Evidence of an agreement would require independent justification of the acknowledgment criterion, the capacity to decline, and the continued operation of the revision channel. The observable fixed point alone establishes only a stable public coding under the chosen update procedure.

In an *asymmetric-fixation* regime, suppose A controls access to an essential interface and permits participation by B only when B emits 0. A protocol may force the observed responses to $(0, 0)$ while leaving B 's locally preferred marker at 1. For a human participant, an investigator might discover this divergence through confidential reports or removal of the access condition. For an artificial agent, the

Table 10: Four conditional pathways in a two-participant interaction illustration. Identical public responses may be generated by different relational mechanisms; feedback-supported coevolution is not established by these traces alone.

Update regime	Public pattern under assumptions	Further evidence required
Negotiated coordination	Stable joint marker after acknowledged proposal	Meaningful refusal, acknowledgment, ability to revise
Asymmetric fixation	Stable joint marker maintained by access constraint	Control of alternatives, sanction exposure, latent disagreement
Contextual pluralism	Distinct recurring markers linked to specified contexts	Reliable context recognition and retention of local alternatives
Alternating revision	Two-period marker sequence under synchronous copying	Timing and symmetry of updates; sensitivity to changes in rules

relevant evidence would concern documented control logic and changes under revised deployment permissions. The outward pattern is identical to the negotiated-coordination outcome, but its generating mechanism differs. Here the asymmetry lies in r_t and in the conditions governing feasible actions, and the inference to endorsement is underdetermined.

In a *context-dependent pluralism* regime, let $c_t \in \{\alpha, \beta\}$ describe two publicly identified interaction settings. The adopted interface rule is 0 in context α and 1 in context β , with each participant retaining a procedure for recognizing the context and selecting the corresponding response. Both local repertoires may preserve distinct expectations about which rule is appropriate outside the shared interface. If contexts recur and their recognition is reliable, the resulting coordination can remain stable while involving multiple active rules. The persistence of a context-conditioned procedure, rather than the dominance of one marker across all encounters, supplies the evidence of coordination. This example also shows why an analyst should not infer universal normative convergence from agreement within a limited domain.

In an *alternating-revision* regime, suppose each participant simply adopts the other participant’s publicly displayed marker from the immediately preceding encounter. The deterministic update is defined in Equation 10.

$$(q_{A,t+1}, q_{B,t+1}) = (q_{B,t}, q_{A,t}), \quad (q_{A,0}, q_{B,0}) = (0, 1). \quad (10)$$

Under exactly this rule the joint marker state alternates between $(0, 1)$ and $(1, 0)$, with period two. The result follows by direct substitution into Equation 10; it is a property of a stipulated discrete dynamical system. The example illustrates that reciprocal adjustment alone does not guarantee convergence. If the observed responses eventually ceased alternating, possible explanations would include changes to memory, asynchronous updates, interface conditions, or the revision rule. The example gives no reason to treat oscillation as a thermodynamic instability or a culturally general developmental stage.

Table 10 compares the four regimes by linking their observable traces to the mechanisms that remain to be established. The rows represent analytically specified cases, not empirical classifications of existing human, animal, or AI communities.

The toy construction serves two methodological purposes. First, it locates outcome diversity in explicitly changed update mechanisms rather than assuming that one universal process yields every form of normative order. Second, it identifies a measurable distinction between the *stabilization of observable conduct* and the *development of capacities to negotiate, interpret, or revise norms*. A research design interested in coevolution should measure both when feasible. Stable compliance can occur without increased interpretive reciprocity, while growing interpretive capacity can coexist with continuing disagreement or alternating practices.

11.5 Observational Equivalence and Limits of Normative Identification

The candidate framework has an important identification limit. Let $Y_{0:T}$ be the observable record over a finite measurement interval and let θ index a specified joint model, including kernels and the observation interface. Two different parameterizations or model structures may produce the same probability law

for all recorded observations even when they differ in their latent normative states. The problem is formalized by the following deliberately restricted proposition.

Proposition 11.1 (Observational equivalence under a nondegenerate compliance process). *For any $0 < \alpha, \beta < 1$ and any common initial distribution of a binary public record Y_0 , there exist two distinct latent mechanisms—one in which an internally maintained practical commitment selects the public response and one in which an externally controlled interface determines it—that generate the same complete law of $Y_{0:T}$ for every finite T . Thus, even a nonconstant binary compliance history does not by itself identify which of these mechanisms operates in the stated model class.*

Construction. Let the public response $Y_t \in \{0, 1\}$ follow a two-state Markov transition law with $\Pr(Y_{t+1} = 1 \mid Y_t = 0) = \alpha$ and $\Pr(Y_{t+1} = 0 \mid Y_t = 1) = \beta$. In model θ_1 , a latent practical-choice state C_t follows precisely this transition law, the public response is $Y_t = C_t$, and the modeled agent faces no access-induced sanction. In model θ_2 , the agent’s stipulated private preference remains fixed at 0, while an externally controlled interface indicator L_t follows the same transition law and determines which response is displayed, $Y_t = L_t$. The distribution of C_0 in the first model and of L_0 in the second equals the stipulated distribution of Y_0 . For each finite sequence $y_0, \dots, y_T$, its probability under either model is the same initial mass multiplied by the same sequence of transition probabilities. The observed joint laws therefore agree for all finite T , although the proposed sources of response selection differ. This establishes the stated existence result. The models are intentionally unrestricted enough to admit these two constructions; the proposition is an identification counterexample within that class, not a universal claim of observational equivalence under additional institutional measurements or interventions.

The construction in Proposition 11.1 is an elementary logical demonstration of underidentification within a chosen observation class. It does not assert that every actual interaction is observationally indistinguishable or that normative commitments can never be investigated empirically. Interview evidence, records of proposals and objections, longitudinal responses to changed incentives, and independently audited institutional permissions may distinguish mechanisms in appropriate settings. Such investigations carry their own measurement error, ethical limits, and potential effects on participants. An investigator should also avoid presenting the numerical equality of output distributions as evidence that the generating representations are GR-equivalent or share the same ontological status.

The full coupled model in Equation 9 introduces further identification problems. A change in observed conduct may arise from z_i , r , e , or the observation interface, with several channels acting together. A change in the estimated interpretation relation $\mathcal{T}_{ij,t}$ may reflect improved shared understanding, altered signal filtering, changes in task distribution, or an analyst’s reclassification. Distinguishing these explanations requires intervention or independent measurement, and sometimes the available evidence will remain incomplete. A useful formalization makes these identification limits explicit and resists untested flexibility in its latent variables.

11.6 Empirical Operationalization and the Boundary of Formal Claims

A practical use of the framework begins by identifying the observation unit and the feasible scope of inference. In a repeated human–AI interaction, possible units include turns, delegated tasks, versioned policy changes, access requests, and documented rule-revision events. An applied study could code r_t using permission structures and platform governance records; e_t using interface and resource configurations; $A_i(w_t)$ through experimentally verified available actions; and $m_{i,t}$ through appropriate human elicitation or documented technical state. Where an internal representation is unavailable, the study should retain a behavioral or institutional proxy and state its insufficiency for psychological attribution.

For interactions involving nonhuman animals, relevant observations may include changed ecological affordances, spatial use, stable inter-individual responses, and sensitivity to manipulated social or environmental cues. These measures could support hypotheses about learned coordination or norm-like social regulation under carefully specified criteria. They would provide different evidence from human expressions of reflective moral endorsement. The use of one abstract state representation across cases is

consequently a *comparative bookkeeping practice*, subject to different validation protocols for different kinds of participant.

The coupled specification also accommodates competing explanations. An apparent expansion of normative repertoire may be a new interface affordance with unchanged expectations. Apparent convergence may result from censoring disallowed responses. Apparent reciprocity may result from both parties following rules provided by a common third party. Apparent structural coevolution may amount to independent participant adaptation under a fixed institutional configuration. A credible empirical investigation should compare these alternatives through preregistered observable distinctions where feasible, controlled perturbations, process tracing, and sensitivity to alternative classifications. Existing work on stochastic convention selection [51, 44], institutional rule development [39], and multi-agent learning [21] supplies methodological precedents, while the proposed GR synthesis requires separate validation.

The proposed model further leaves several theoretical questions open. Normative repertoires may lack a unique measurable encoding; changes to the set of available participants can require an extension of I ; dynamically created forms of action may outgrow the ambient alphabet; and different observational resolutions can produce incompatible descriptions of the same encounter. Conditions for stationary distributions, attractors, equilibrium selection, or convergence would depend on substantial additional assumptions. A participant's authority to revise an interface cannot be inferred from the existence of a revision operator in the analyst's model. Finally, normative legitimacy remains an evaluative question informed by the relevant ethics, law, and practical relations, rather than an outcome entailed by the stochastic update equations.

Accordingly, the present framework should be judged first by its discriminating capacity. It supplies a vocabulary for specifying which agents are being described, which aspects of their normative practices are measured, how interaction histories affect interpretation, which relational conditions enable or restrict revisions, and which rival mechanisms may produce a given observation. Its central research hypothesis is that at least some instances of cross-system norm formation will require jointly evolving interpretive and relational variables to explain the observed processes adequately. The hypothesis could fail in domains where models with fixed interaction structures predict the relevant observations as well, or where purported normative variables cannot be identified beyond ordinary behavioral adaptation. That possibility is a reason to pursue comparative and empirical evaluation, developed in the following sections, while maintaining a clear separation between formal representation, explanatory adequacy, and normative justification.

12 Comparative Patterns of Norm Formation across Heterogeneous Systems

Comparative inquiry can assess the generative-relational (GR) account of norm formation only when the evidence supplied by different kinds of interaction remains distinguishable. Human deliberation, human–AI coordination, interspecies communication, and hypothetical encounters with unfamiliar intelligence differ in the observability of expectations, the means through which commitments are expressed, and the institutional or ecological conditions enabling revision. This section develops a bounded comparison of those settings. It examines the mechanisms that could support a cross-case analogy, identifies domain-specific evidential limits, and proposes discriminating observations for subsequent research. The comparison distinguishes simultaneous changes in practices and conditions, interaction-dependent institutional or ecological modifications, and the stronger reciprocal-feedback claim involved in coevolution. It is interpretive and methodological: published studies support particular claims about studied human and nonhuman populations, whereas extraterrestrial contact is treated entirely as a counterfactual test of the framework. The coupled formal description introduced in Section 11 is used as an aid to organizing the comparison, without claiming that its state variables have a uniquely shared empirical interpretation across cases.

12.1 Comparative Units and the Preservation of Domain-Specific Evidence

The unit of comparison proposed here is a *norm-formation episode*: a bounded set of encounters in which an identifiable practice is proposed, repeated, contested, maintained, or revised. Episodes can involve persons, institutions, trained computational systems, living organisms, or mixtures of these. An episode must be specified with respect to its observation boundary, participants, feasible actions, environmental and infrastructural conditions, and record of rule-relevant changes. A computational agent deployed through a platform cannot be understood merely as a free-standing language-producing individual; the organizational and technical arrangements that select its training, memory, access, and outputs may contribute directly to the reproduction of the observed practice. Likewise, a human community cannot be reduced to one representative cultural disposition.

Four comparative questions orient the analysis. First, which recurrent conduct is observable, and what competing processes can generate it? Second, how are deviations interpreted and managed, through learning, avoidance, correction, sanctions, or changes in material conditions? Third, who can initiate or contest an adjustment to the interaction protocol? Fourth, what supplementary evidence would establish the proposed interpretation of a practice as convention, normative expectation, endorsed commitment, or legitimate authority? The same questions can be asked across settings, while the appropriate measurements and available answers remain specific to each setting. An analyst's use of the word *norm* should therefore accompany an explicit evidential classification of the sort developed in Section 3, rather than silently moving between distinct meanings of that word.

Comparability can arise at the level of relations among observation, adaptation, and control. Two systems may both display a recurring greeting signal; in one case its maintenance may involve human expectations and accountability, while in another it may be generated by a constrained decoding policy. Both cases admit a comparison of signal persistence, available alternative signals, and interface revision. Evidence of mutual moral commitment remains appropriate to neither case merely because their public transcripts resemble one another. The proposed GR contribution lies in making the differences in historical formation, enabling conditions, and revision authority explicit when examining similar interactions.

12.2 Human–AI Coordination under Versioned Rules and Unequal Governance

Human–AI interaction offers a comparatively accessible setting for examining changes in practical rules under materially heterogeneous conditions. Consider a hypothetical research team that repeatedly uses a language-model service to summarize sources and draft reports. Team members gradually adopt a procedure for checking uncertain statements, recording corrections, and negotiating which decisions require human approval. The visible transcript may show a repeated correction-and-confirmation pattern. Its production, however, can involve several jointly relevant mechanisms: human professional expectations, interface prompts, retrieval permissions, model version changes, organizational review policy, and learned user habits. The observation of orderly exchange underdetermines which of those mechanisms explains its persistence.

Human–AI interaction design research provides an empirical starting point without resolving the normative question. Amershi and colleagues proposed human–AI interface guidelines, evaluated with design practitioners and existing AI-infused products, including attention to timing, failures, and correction mechanisms [1]. Such design work supports treating feedback and repair as features of interaction arrangements. It does not establish that a deployed AI service endorses the values represented in a user-facing guideline. The separate experimental literature on language-model populations reports that conventions can arise through repeated decentralized exchanges among model-based agents [5]. These results concern defined laboratory populations, tasks, model architectures, and prompting procedures. Their bearing on reflective endorsement, sustained independence of normative revision, or interactions with human institutions remains a distinct empirical matter.

For a versioned human–AI service, the analyst can ask whether adaptation occurs at the level of an individual user's instructions, a model's available memory, its underlying parameters, a retrieval corpus, or the platform's deployment rules. Each possibility implies a different intervention. A response pattern

may be changed through the user's prompt while remaining constrained by a provider-defined content policy. A work team's customary reporting practice may become dependent on a software integration whose approval requirements are selected elsewhere. Thus reciprocal responsiveness in dialogue can coexist with asymmetric control over durable permissions and infrastructural conditions. End-to-end auditing research emphasizes the relevance of choices made throughout the organizational life cycle of an AI system [42]; a comparison of norm formation should retain those choices in its relational description.

A feasible research strategy would observe matched tasks across configurations that vary memory persistence, access to a rule-editing interface, human review rights, and model or deployment versions. Measures might include task correction, objections to a proposed rule, acceptance or rejection of amendments, the survival of amendments after system changes, and the availability of accountable reasons. These observations could discriminate user habit formation from interface-imposed regularity. Claims about a model's inner commitments or conscious interests would require independent conceptual and empirical support. The current paper assigns neither moral personality nor phenomenal subjectivity to a present-day AI system by virtue of its participation in an interaction episode.

12.3 Human–Animal Coordination and Ecologically Situated Practices

Relationships between people and nonhuman animals offer another kind of heterogeneity: participants may have overlapping needs and learnable social cues while differing substantially in available signals, perceptual access, ecological dependence, and capacities for reflecting upon rules. Experimental work by Hare and colleagues found that domestic dogs in the studied tasks could use human communicative cues to locate hidden food, with comparisons involving apes, wolves, and young puppies [26]. These findings support a claim about particular social-cognitive capacities and their possible domestication history. They leave open the much stronger question of whether following a person's cue involves a represented social obligation or moral endorsement.

A more extended human–animal episode might involve a community adjusting routes, feeding practices, or boundaries in response to recurring animal movement. Humans can revise signs, fences, schedules, and monitoring procedures; animal movement may respond to available resources, threats, learned associations, and species-specific social information. This is a plausible field for studying co-development of feasible practices and changed ecological relations. The object described as a shared *rule*, however, may be a human institutional arrangement coordinated with an animal's learned pattern. Observed stability would not independently justify attributing the same rule representation to both parties. Conservation research has also emphasized the practical importance of socially transmitted animal cultures and traditions [11]. Such work provides reasons to attend to animal histories and learned ways of living, while any inference concerning animal normativity requires a further evidential argument.

Recent comparative approaches to animal norms propose dimensional criteria rather than an all-or-nothing assumption that nonhuman social life either contains human-equivalent moral norms or consists solely of automatic behavior [50, 2]. This orientation is compatible with the present framework's distinction among regularity, social sensitivity, expectations, and normative standing. An investigator could consider developmental learning, reactions to violations, differences between familiar and unfamiliar interaction partners, and the effects of altered environmental affordances. Such observations would have to be interpreted relative to the species and local situation. The absence of an articulated verbal objection cannot be treated as consent; conversely, a learned avoidance response alone need not demonstrate a propositional judgment about legitimacy.

The normative evaluation of a human–animal relationship has an additional representational problem. Human institutions frequently specify the arrangements governing habitats, breeding, movement, and care. Responsibility for these decisions may be grounded in ethical reasons independent of whether an animal can take part in a human-style negotiation. The political-economic asymmetry governing land, resources, and ecological management is therefore analytically pertinent, while the specifically evolutionary question concerns how situated routines and expectations change with those relations. This preserves a boundary between the present G3 inquiry and the broader inquiry into control over heteroge-

neous generative conditions developed for the separate G2 political-economy paper.

12.4 Human Intercultural Encounters and the Variation of Established Normative Repertoires

Human intercultural interaction offers stronger opportunities to study articulated expectations, retrospective accounts, disagreement, negotiated consent, and public reasons. Its distinctive evidential advantage is the availability, in suitable circumstances, of participant testimony and identifiable institutions for challenging a rule. At the same time, translation and representation may be unequal, and an investigator's description of a *culture* may obscure differences among persons, organizations, generations, and circumstances. Comparisons should begin with locally situated practices and identify the actual participants, histories, and institutions within each sampled setting.

Cross-cultural behavioral studies illustrate the importance of institutional and social variation without serving as direct demonstrations of cross-cultural norm co-creation. Henrich and colleagues reported substantial differences in outcomes of economic games across fifteen small-scale societies and analyzed associations with everyday social and economic organization [27]. Gelfand and colleagues investigated differences in the strength of norms and tolerance of deviation across a thirty-three-nation sample [23]. These studies establish researchable diversity within their sampling, instruments, and interpretive frameworks. They do not show that members of each studied population possess identical commitments, nor do they establish how two differently situated communities would negotiate a new jointly accepted rule upon contact. The latter is a longitudinal interaction question that requires direct evidence of translation, revision, and institutional participation.

For an illustrative encounter, consider two organizations cooperating on the governance of a shared research dataset. One organization expects wide reuse following publication; another attaches locally defined obligations to consultation, credit, or the return of benefits. Both may endorse scientific inquiry while holding different expectations about permissions and continuing responsibilities. A provisional agreement could specify reciprocal review, acceptable reuse, mechanisms for objection, and a means of modifying terms after experience accumulates. Such an agreement offers an observable record of norm negotiation. Whether it also qualifies as procedurally fair depends on whose interests were represented, the actual ability to refuse, available remedies, and the consequences of economic or institutional dependence. These conditions may change as collaborative work proceeds.

The relevant comparative observation is consequently more than a change in the frequency of a behavior. Researchers could trace drafts of agreements, the translation of contested terms, alternative proposals rejected or accepted, consultation records, and modifications in access or decision authority. They should also investigate whether terms apparently translated into a common vocabulary retain different local significance. A stable final agreement can coexist with persistent plurality of purposes and values. This is a case in which normative coordination may be empirically documented while internal consensus remains only partial. It also cautions against assigning a single utility function to the parties merely because they have ratified one operational arrangement.

12.5 Hypothetical Extraterrestrial Contact as a Reversal of Epistemic and Material Power

An encounter with an extraterrestrial civilization belongs to the present comparison as a disciplined thought experiment. It supplies no observed alien behavior and supports no inference about actual extraterrestrial cognitive architecture, preferences, language, or political organization. Its analytical role is to expose hidden assumptions in protocols developed among familiar human participants, especially assumptions about common communication channels, the feasibility of inaction, reciprocal interpretability, and the relation between capability and legitimate authority.

Suppose a technically powerful extraterrestrial civilization were able to stabilize Earth's energy supply while retaining exclusive control of the infrastructure that makes such stabilization possible. Its representatives might describe the arrangement as protection or assistance. Human institutions could

Table 11: Comparative evidential diagnostics for normative interaction among heterogeneous GR participants. Empirical access and epistemic limits differ across cases.

Encounter setting	Candidate evidence of practice formation or revision	Interpretive boundary or competing mechanism
Human–AI collaboration	Versioned prompts, permissions, repair records, user objections, documented policy amendments	Similar outputs may be caused by filtering, model updates, or user adaptation; endorsement is not identified
Human–animal relations	Learned cue responses, habitat and movement changes, social transmission, environmental interventions	Adaptive learning and ecological affordances may explain behavior without shared propositional norms
Human intercultural cooperation	Negotiated texts, participant accounts, rejected proposals, appeal procedures, altered institutional permissions	Public agreement may conceal unequal representation, partial translation, or constrained alternatives
Extraterrestrial first contact (hypothetical)	Stipulated message and resource-control scenarios; conditional protocol analysis	No alien observations or validated shared semantics; protocol authority and possible effects remain unknown

appreciate the material benefits while contesting the foreign civilization’s authority to determine which forms of human experimentation, governance, or cultural continuity remain feasible. The same recorded pattern of cooperative resource use could therefore be generated by consensual collaboration, unavoidable dependence, or a mixture of both. A further comparison asks whether humans exercising substantial technological control over other terrestrial organisms employ parallel justificatory structures. The resulting power-position reversal can test the consistency of human evaluative arguments; it does not settle the moral status of every relationship by analogy.

Initial contact is particularly difficult because waiting, signaling, entering a territory, or withholding information may each have different consequences under an unfamiliar normative system. The 2026 International Academy of Astronautics principles concerning SETI detections recommend broad international consultation before a response to a confirmed detection [29]. These principles constitute a human institutional procedure for an astronomy-based detection setting. Their authority and practical content should be assessed within that domain. They supply neither a verified account of alien norms nor a fully specified framework for physical arrival, coexistence, or joint government. In the thought experiment, a more powerful visitor’s ability to define the communication interface, determine Earth’s representatives, or control life-sustaining technology would create distinct questions about who can initiate and revise practical rules.

A useful variant considers an extraterrestrial visitor whose technology exceeds current human capabilities but whose ecological or interpretive understanding of Earth’s living systems is limited. Its superior engineering could coexist with severe mistakes about human historical commitments or the requirements of other organisms. Equating technical capacity with universal epistemic access would then produce unreliable descriptions even before evaluating the visitor’s political claims. The case illustrates a candidate separation among the ability to alter another system’s generative conditions, the ability to understand that system, and the authority to make binding rules for it. The separation is analytically useful under the stipulated scenario and remains open to further normative examination.

12.6 Comparative Diagnostics for Translation, Revision, and Control

The preceding cases support a structured comparison of research questions and limits. Table 11 brings together the principal candidate evidence and the corresponding inferential boundary. The entries intentionally classify types of observation rather than rank kinds of participants by moral importance, cognitive superiority, or developmental level. The extraterrestrial row indicates a theoretical test case, whereas the other rows identify fields in which suitable empirical records can, in principle, be collected.

Several comparison rules follow. Claims about *behavioral adjustment* require observable change against a specified baseline. Claims about *norm formation* require an additional account of the practice’s

regulation, expectations, and relation to possible alternatives. Claims about *mutual understanding* require evidence of interpretation and correction adequate to the participants' communicative capacities. Claims about *legitimate shared authority* require normative and institutional evaluation beyond the fact that a rule became stable. The sequence of these statements expresses increasing evidential demands for stronger claims; it is a methodological ordering of inferences, without presupposing a fundamental hierarchy of beings or histories.

The table also clarifies the conditional role of the framework proposed in Section 11. A change in an individual's behavioral state, a change in its interpretation of signals, and a change in the authority to modify a shared interface may appear in the same observed encounter. Their relations can be modeled through a coupled transition kernel, as in Equation 9, when the relevant state and observation variables can be operationalized. The empirical status of those variables varies substantially among a participant interview, a software policy log, animal spatial data, and a fictional alien exchange. The notation enables disciplined comparison of dependencies; it establishes no uniform measure of normative agency.

12.7 Comparative Limits and the Research Agenda for Normative Coevolution

The proposed inquiry should retain potential counterexamples. Some human–AI interactions may be explained adequately by stable interface constraints and individual habit learning. Some human–animal routines may be understood through ecological incentives and training without introducing social-norm variables. Intercultural institutions may maintain established procedures despite substantial changes in the participants' expressed preferences. In each case, invoking evolving normative representations could add complexity without improving explanatory power. Conversely, institutional rule amendments, mutually interpretable corrections, changes in the feasible action space, or new procedural rights might supply evidence that a fixed-practice model misses important dimensions of change.

A further limit concerns comparisons made by the investigator. Researchers can impose a common observational vocabulary onto heterogeneous interactions through coding schemes, surveys, telemetry, or interface logs. Such procedures make patterns comparable in a restricted analytical sense, while also selecting which differences can appear in the record. Decisions about measurement and translation may privilege well-instrumented or institutionally dominant actors. Consequently, data access and classification should themselves be examined as part of the relational conditions of inquiry. This methodological demand is especially important when a platform controls AI logs, a human institution controls ecological monitoring, or a powerful organization determines which parties count as participants in negotiations.

The comparisons suggest a focused, falsifiable GR research hypothesis: in some heterogeneous encounters, changes in the accessible practices, interpretive channels, and powers of rule revision will improve explanations of developing conventions beyond accounts that hold those conditions fixed. Establishing that claim requires matched rival models, intervention or process evidence, and careful definitions of what is being measured. A positive result would support the relevance of coevolving relational variables in the studied domain; it would not imply general normative convergence, equivalent inner experience across species, or an emergent moral authority. A negative result could show that simpler local learning or fixed-institution models adequately explain the case. Both outcomes would clarify the scope of the proposed framework.

This comparative investigation therefore supplies a bridge from the conceptual and formal analysis of heterogeneous norms to the research design developed in the next section. The central comparative conclusion is methodological: interactional resemblance becomes informative when it is examined together with historical formation, available alternatives, interpretive capacities, environmental dependence, and control over revision. Such a comparison can preserve substantive differences among human, nonhuman, artificial, and hypothetical participants while identifying the specific mechanisms through which a shared practice may be constructed, maintained, contested, or transformed.

13 Research Design for Evaluating Heterogeneous Norm Formation

An evolutionary account of norm formation among heterogeneous generative-relational (GR) participants becomes empirically informative when its proposed mechanisms can be distinguished from more economical explanations of the same encounters. Such explanations include established norm-emergence models with explicit creation, recognition, and enforcement procedures [43], as well as institutional models that represent revisions of rules and shared strategies [17]. This section develops a research design that connects the conceptual diagnostics of Section 3, the coupled representation in Section 11, and the comparative settings of Section 12. Its objective is to specify observable units, meaningful interventions, rival causal models, inferential limits, and possible disconfirming results. The discussion combines a controlled human–AI interaction design with longitudinal institutional and ecological approaches. These designs support domain-specific tests of candidate mechanisms; the availability of an experimental manipulation in one domain supplies no guarantee that the same manipulation has an intelligible analogue in another. The proposed program is a protocol for future inquiry and reports no original data.

13.1 Units of Observation and Graduated Normative Outcomes

A suitable primary unit is a *situated interaction episode*: a delimited sequence of actions among identifiable participants, occurring through specified channels and under documented rules of access, interpretation, and revision. A sequence may be embedded in a team, institution, model deployment, or ecological setting. Researchers should record the version of the procedure in force, the alternatives available to each participant, and the conditions under which the episode enters the observational sample. Episodes involving the same participant, institution, or shared communication interface will usually be dependent; a statistical design should accordingly represent relevant clustering, repeated observations, and cross-episode interference.

Four classes of outcome require separate operationalization. *Behavioral coordination* concerns observed response regularities and successful task completion. *Norm-like regulation* concerns contingent expectations, responses to deviations, correction practices, and the persistence of a rule across changes in immediate incentives. *Interpretive development* concerns the capacity to recognize, explain, revise, or repair the significance of an interaction under conditions adequate to the particular participants. *Institutional rule revision* concerns observable changes in which actions are permitted, who may propose amendments, and whose objections can alter an operative rule. These are distinguishable analytical variables. Their co-occurrence may justify an enriched account of normative interaction, while none is an automatic measure of moral approval, consciousness, or legitimacy.

In research with human participants, interviews, confidential elicitation, behavioral responses to counterfactual vignettes, and experimentally varied expectations can help distinguish public conformity from personal commitments. Bicchieri’s account of diagnosing social norms distinguishes empirical and normative expectations and conditional preferences [8]; its detailed measurement practices pertain to human social research. A validated measure of human endorsement should be assessed for comprehension, anonymity, cultural translation, and response pressure. For artificial systems, versioned output traces, policy specifications, actual revision permissions, documented training updates, and human governance decisions are observable; these records support claims about implemented rules and adaptation without establishing experienced obligation. For nonhuman organisms, ethically collected observations of learning, social transmission, deviations, environmental affordances, and repair-like interaction may motivate a dimensional analysis of normative behavior [50, 2]. The investigator should specify the particular behavioral and ecological evidence supporting each inference.

Attribution of a norm also depends on the available alternatives. Two individuals may utter the same response while one can refuse without penalty and the other must comply to retain access to an essential resource. Recording only the utterance masks this contrast. An episode-level dataset should therefore associate responses with feasible options, sanction exposure, interpretation channels, and revision authority whenever these can be established. Unobserved alternatives should be coded as unknown, with reconstruction assumptions documented separately.

13.2 Causal Questions and Counterfactual Comparisons

The proposed framework permits several distinct causal questions. Does granting reciprocal amendment rights alter how a shared practice develops? Does access to a repair channel improve the mutual interpretation of proposed rules? Does dependence on a single controlling participant sustain outward agreement after the participants' expressed expectations diverge? Does incorporating recorded changes in interaction structure explain outcomes beyond participant-level learning under fixed conditions? Each question requires a separately specified treatment or exposure, observation window, comparison, and estimand. Standard causal-inference methodology emphasizes that longitudinal exposures, time-varying confounders, and treatment–confounder feedback complicate identification [28]; the GR formulation does not remove these problems.

Consider a study in which the unit of randomization is an intact collaborative team, and a well-defined binary treatment D_g changes the team's actual protocol-amendment rights. Let $Y_g(d)$ denote a prespecified, independently assessed measure of *coordination on held-out tasks after a standardized rule-disagreement episode* for team g under protocol condition $d \in \{0, 1\}$. Its scoring rubric, task difficulty, assessment window, and treatment-blind coding must be fixed in advance and applicable in both arms. Actual amendment opportunities and enacted amendments are measured separately as protocol-fidelity and process variables; counting the presence of reciprocal amendment rights as the outcome would make the contrast partly true by design. A design-specific causal estimand is given in Equation 11.

$$\Delta_{\text{coord}} = \mathbb{E}_{g \sim \mathcal{P}}[Y_g(1) - Y_g(0)]. \quad (11)$$

Here $\mathcal{P}$ is the study's explicitly defined population of eligible teams; $d = 1$ assigns the reciprocal amendment procedure, whereas $d = 0$ assigns a one-sided but disclosed amendment procedure within a low-stakes experimental interface. The definition presupposes sufficiently well-specified intervention versions, consistent treatment implementation, meaningful outcome measurement, positive assignment probability for each arm, and a treatment assignment regime that supports the intended potential-outcomes contrast. Random assignment can identify the intention-to-treat effect for the eligible teams when between-team interference and differential missingness are adequately controlled. The quantity is a causal effect of the specific *amendment-rights intervention* on held-out coordination; it is neither a measure of normative endorsement nor direct evidence that reciprocal normative coevolution occurred. Interaction within a team is part of its outcome, while interactions *between* teams can violate the assumed separation of experimental units. Equation 11 is a proposal for a causal question, not an estimate or an ethical justification for restricting rights outside a controlled study.

A second family of tests concerns explanation rather than treatment effects. Compare a baseline model $\mathcal{M}_0$ in which institutional permissions and interpretive channels are fixed with a relational model $\mathcal{M}_1$ whose measured permissions, translation procedures, and access conditions can vary. Match observation channels and assess predictive performance on predeclared held-out interactions or transfer environments; compare complexity, calibration, uncertainty, and the interpretability of the estimated mechanisms. An advantage for $\mathcal{M}_1$ would support the explanatory usefulness of its additional variables for the sampled setting only if improvements survive appropriate capacity controls and sensitivity analyses. It would not establish GR as a privileged ontology. An equivalent or superior $\mathcal{M}_0$ would constrain claims that evolving relational organization is necessary for the particular observations.

13.3 Controlled Human–AI Interaction Protocols

Human–AI collaboration permits a relatively tractable initial study because interface permissions, policy changes, prompts, and the model version can be documented. One candidate protocol uses small human teams collaborating with an artificial assistant on a repeated classification or resource-coordination task in a sandbox. The task should create opportunities for ambiguous classifications, disagreement, and revision of a local working rule while avoiding real-world livelihood, healthcare, or other high-stakes consequences. Participants receive transparent information about the assistant, the experimental rules, data handling, and their ability to withdraw. The research objective concerns the joint process of practice

formation, including human responses to a software-mediated counterpart; it does not require treating the model as a moral person.

A factorial design could manipulate two *interface-level* conditions: whether a rule amendment requires acknowledgment from both sides or is controlled by a designated human administrator, and whether the communication interface permits context-sensitive clarification and repair or is restricted to a fixed coding scheme. The phrase *both sides* here denotes two operational roles in a protocol, with clearly specified human oversight; it conveys no claim of equivalent rights, intentions, or experience. The assistant's ability to submit a proposal, accept a new instruction, or change a displayed policy must be implemented and logged rather than inferred from agreeable language. Design documentation should distinguish model outputs, executable tool permissions, and decisions attributable to human operators. Contemporary human–AI design and audit research provides a practical foundation for recording such interfaces and governance choices [1, 42].

During an initial observation period, record response agreement, errors, repair exchanges, rejected amendments, time to coordination, and whether a proposal actually changes future feasible actions. A scheduled context change can then assess transfer: does a learned public rule generalize when a label's meaning, a user's task, or the interface's permissions change? A separate, ethically acceptable relaxation of a simulated access constraint can test whether a participant's conformity depended on restricted alternatives. Human participants may provide confidential accounts of reasons and perceived pressure; these remain reports subject to response bias. AI behavior should be analyzed through versioned system instructions, logging, replication with controlled prompts, and documented configuration changes. The study should retain a stable-version arm to distinguish evolving governance from changes caused by a model update.

Several controls are essential. Rewards for task performance should be held constant where the proposed comparison concerns amendment rights. Exposure to other teams' outputs should be monitored to reduce contamination across randomized groups. A human-only collaboration arm, a rule-based assistant arm, and a model-based assistant arm may help identify which patterns are explained by ordinary human coordination, scripted filtering, or adaptive interaction. Such arms need not be presumed psychologically comparable. The trial should also record relevant participant experience with AI systems and investigate differential attrition. A positive effect on held-out coordination would estimate the effect of the specified amendment-rights intervention under the study assumptions. Changes in actual rule amendments would be interpreted as separately logged process evidence, subject to the mechanical opportunities permitted in each arm. Any stronger inference to reciprocal normative coevolution or mutual endorsement would require time-resolved mechanism evidence and additional argument.

13.4 Longitudinal and Cross-System Research Designs

Human institutions offer a second route to studying norm and relation coevolution. In a consenting community organization or research team, investigators could prospectively document decisions, negotiated rule amendments, rejected proposals, minutes of deliberation, authority transfers, and subsequent practice. Repeated confidential interviews and structured observations can distinguish publicly invoked principles from participants' explanations and perceived alternatives. Administrative records can locate changes in permissions and the distribution of resources. When institutional reform occurs through a known process, its rollout may provide a comparison across sites or periods. The design must nevertheless address self-selection, concurrent reforms, missing records, and the possibility that procedural change follows earlier shifts in expectations. Qualitative process tracing can examine the causal sequence through which an amendment altered conduct or the feasible range of objections; longitudinal quantitative analyses may complement that account where their assumptions are credible [28, 39].

Cross-cultural inquiry requires additional precautions because translated survey categories and experimentally assigned game outcomes may carry different meanings across settings. Investigators should use participatory instrument development, locally reviewed terms, multiple evidence sources, and checks for measurement invariance when quantitative comparison is attempted. Group-level summaries should

be presented alongside individual variation. A construct that measures normative expectation in one setting may need substantial conceptual adaptation elsewhere; mechanical use of identical item wording cannot guarantee comparable meaning. Disagreement over the relevant categories is itself a potential observation about the conditions of translation, provided that the researchers document how interpretations were negotiated.

Research with other living systems calls for domain-appropriate methods. Noninvasive longitudinal monitoring, naturally occurring habitat changes, ethically justified enrichment, and observations of social learning may establish that an interaction pattern changes with social partners or environmental affordances. Comparisons can include matched ecological conditions, different histories of social exposure, and changes in available signaling opportunities. Animal welfare, ecological disturbance, and requirements of species-specific care take precedence over an experimental desire to force a convention or induce conflict. Investigators can test hypotheses about learned coordination or socially transmitted routines without claiming access to human-like verbal commitments. Explicit norm attributions should be tied to the dimensional criteria and disagreements described in the relevant animal-normativity literature [2, 11].

Potential extraterrestrial encounters have a different epistemic status. Simulations, role-structured deliberation, and scenario analysis can test the *internal consequences of assumed protocols* when interpretive expectations are unknown and capabilities are unequal. They cannot supply observed alien preferences, normative attitudes, or empirically validated first-contact success rates. For this domain, research products should consist of transparent assumptions, failure-mode analysis, uncertainty descriptions, and contestable procedural recommendations. Their value lies in inspecting the human commitments embedded in proposed contact protocols.

13.5 Model Competition and Disconfirming Evidence

Research on GR normative coevolution should make its empirical exposure to failure explicit. A candidate relational mechanism competes with several plausible alternatives: independent learning driven by task rewards; imitation or local habit; stable institutional rules combined with changing participant composition; response filtering imposed by a platform; and compliance induced by sanctions or loss of access. The alternatives may coexist, so the comparison should target the additional explanatory or causal role of particular relational variables, rather than demand that each episode instantiate exactly one mechanism. An outcome measure structurally available in only one randomized arm is unsuitable as the primary comparative estimand; the study should separately describe protocol permissions, actual amendments, and arm-comparable behavioral or interpretive outcomes. A coevolution hypothesis additionally calls for time-resolved evidence of feedback in both directions between practices and interaction conditions; detecting two changing series or a one-way platform intervention is insufficient for that stronger claim. A feasible controlled design might independently randomize an opportunity to propose a revision at one stage and the implementation of a specified revision at a later stage, while recording who actually determines the operative rule. The first manipulation tests whether a participant's intervention can affect a relational condition; the second tests whether that condition affects subsequent practices. Their joint interpretation requires documented treatment fidelity, an explicit sequence of events, and controls for common causes and spillovers. Neither manipulation alone demonstrates that a norm, as distinct from a public response, has coevolved.

Table 12 presents a provisional correspondence between proposed mechanisms, evidence that would discriminate among explanations, and results that would weaken the particular GR extension. Its entries describe potential tests, not performed experiments or guaranteed falsifiers of an unrestricted philosophical vocabulary.

A stronger empirical program should register each hypothesis in terms of an estimand or measurable prediction, an expected effect direction where justified, a meaningful interval of effects, a planned sample or stopping rule, and criteria for interpretable negative evidence. Exploratory discovery of candidate representations can precede confirmatory testing, but the two modes of inquiry should remain distin-

Table 12: Proposed tests for mechanisms of heterogeneous norm formation. Negative results constrain specific operational hypotheses within the studied domain.

Candidate mechanism	Discriminating design or observation	Potential constraining result
Relationally enabled revision	Randomized amendment rights; blinded coordination outcome and separate fidelity checks	Equivalent held-out coordination under comparable exposure and adequately measured uncertainty; process checks confirm the contrast
Interpretive co-development	Clarification-channel intervention; repair and transfer across changed contexts	A fixed-code or task-learning account predicts the same repair and transfer patterns
Power-conditioned compliance	Ethically reversible access constraints; independent assessment of perceived alternatives	Behavior and expressed reasons remain unchanged under verified removal of the simulated constraint
Coevolving relational organization	Longitudinal measurements of rules, topology, and access; held-out model comparison	Fixed-relation models match or exceed relational models with comparable complexity and observation access
Norm-like regulation	Expectations, deviations, correction and conditional-preference measurements where appropriate	Stable behavior is fully explained by affordances, immediate rewards, imitation, or filtering

guishable in reporting. Preregistration helps make that distinction visible [36]; it does not eliminate measurement error, protocol deviations, model misspecification, or interpretive disagreement. Simulation studies additionally require independent parameter-recovery tests, checks of identifiability, and ablations that remove relational variables one at a time. Claims of predictive advantage should be evaluated on withheld sequences or environments, with explicit treatment of dependence and leakage.

Measurement limitations may prevent a decisive result. If investigators cannot reliably observe operative permissions, separate model updates from user learning, or measure the alternatives available to a participant, several mechanisms may remain observationally equivalent. Within such a study, the responsible finding may be *non-identification*. If improved instrumentation then distinguishes the mechanisms, the research program gains a more precise empirical scope. Conversely, repeated well-powered findings that fixed-structure learning models predict the relevant changes as well as richer relational models would reduce support for the additional GR variables in those domains. The framework remains subject to revision, including the possibility that its proposed distinctions add little explanatory value in some settings.

13.6 Research Ethics, Governance, and Participation

The conditions of research participation are part of the phenomenon under investigation. A study of norm-making authority can inadvertently reproduce the very asymmetries it examines if an institution determines all categories, withholds amendment channels, or treats silence as consent. Human-participant protocols should therefore obtain appropriate ethics approval, provide accessible withdrawal and appeal routes, minimize compelled public disclosure, protect confidential accounts of disagreement, and make risks from power imbalance reviewable. When experimental conditions assign unequal amendment rights, the asymmetry should be disclosed, limited to a low-stakes sandbox, and reversible without consequences for the participants' real-world entitlements. Investigators should distinguish participants' assent to participate in an experiment from endorsement of any simulated rule.

AI-system studies require a traceable separation among model behavior, deployment permissions, platform policy, and human operational authority. Changes in hosted model versions and hidden moderation configurations can otherwise be mistaken for negotiated norm revision. The project should disclose relevant access restrictions and retain audit records consistent with privacy, contractual and security obligations. Studies involving ecological systems should integrate independent welfare or conservation assessment and accept that some informative interventions would impose unacceptable burdens. These limits are substantive constraints on research design, not merely external inconveniences.

Governance also concerns the categories used to report results. Participants and affected communities

may reasonably contest the investigator's terms for rule compliance, autonomy, understanding, and repair. A procedural channel for reviewing measurement categories can be studied as one possible instance of relational revision. Such participation cannot, by itself, certify the legitimacy of a rule or settle disputes about moral standing; it can render the evidential basis and the distribution of interpretive authority more transparent. The study should document who could influence its design, whose concerns were excluded, and how disagreements about interpretation were handled.

13.7 A Staged Empirical Programme and Its Inferential Boundaries

The investigation can proceed through a limited sequence of increasingly demanding designs. An initial descriptive phase maps actual actors, interface roles, feasible practices, local expectations, and available records. A controlled phase tests selected mechanisms in a human–AI sandbox with reproducible configurations and registered comparisons. A longitudinal phase follows institutional revisions and interpretive changes under independently documented governance conditions. Domain-specific comparative work then evaluates which constructs retain meaning across human institutions and particular nonhuman ecological settings. Simulation and first-contact scenarios remain explicit sensitivity analyses of stipulated assumptions. The ordering is pragmatic and concerns the organization of research, rather than a claim that generative systems are intrinsically arranged in such a sequence.

Across these phases, a useful result would specify the domain and evidence for which changes in relational access, interpretive resources, or rule-revision authority materially improve the explanation of normative practices. Equally useful would be an established boundary: cases in which ecological adaptation, individual learning, fixed conventions, or simple reward-sensitive behavior adequately explain the available observations. Either kind of result would sharpen the meaning and usefulness of heterogeneous normative coevolution as a proposed research object. The normative legitimacy of emergent practices would remain an additional subject for ethical and political assessment, informed by but irreducible to the empirical history of their formation.

14 Conclusion

Norm formation among heterogeneous generative-relational (GR) subjects concerns the development of workable, contested, or mutually intelligible practices among participants whose capacities and conditions of participation may differ substantially. The preceding investigation has examined this research object through conceptual distinctions, evolutionary models, a provisional coupled representation, comparative cases, and prospective empirical designs. This concluding section specifies the resulting analytical contribution, distinguishes established arguments from conditional mechanisms, identifies the institutional significance of normative asymmetries, and sets out the evidence needed to advance the inquiry. The central proposal is that interaction can transform both the practical rules participants follow and the relationships through which those rules are interpreted, negotiated, and sustained. Its scope remains that of an exploratory theory rather than a demonstrated general law of normative development.

14.1 The Analytical Scope of Heterogeneous Norm Formation

The starting distinction concerns observable similarity and relational constitution. Participants may produce comparable utterances, coordinated actions, or consistent responses to a rule while relying on different histories of learning, physical dependencies, feasible actions, interpretive resources, and institutional permissions. These differences matter when a theory attempts to explain the formation of norms through interaction. A model restricted to homogeneous agents and a fixed action set may be useful within its proper domain; applying it to a heterogeneous encounter requires evidence that the relevant differences have been adequately represented or shown to be immaterial to the question being studied.

The GR vocabulary developed here provides a way of organizing those differences. It treats participants' practices, local normative repertoires, available interpretive relations, and the material or insti-

tutional conditions of interaction as analytically distinguishable, potentially changing components of an encounter. In this formulation, a difference in local repertoires does not automatically indicate a disagreement between propositional beliefs. Participants may lack the same perceptual access, opportunities for communication, or capacity to register a deviation as a violation. The corresponding evidence ranges from articulated human expectations and contestation to observed social learning, interaction logs, and documented changes in system permissions. These evidence types require domain-specific methods, as the discussions of human, animal, and artificial interaction have emphasized.

The proposed category of an interacting *subject* is correspondingly provisional. It makes possible comparative questions about participation in relationships without settling questions of consciousness, intentional agency, legal personality, or moral status. In particular, a computational system's successful coordination with a human user provides evidence about its observed behavior under specified conditions. The system's model architecture, training history, deployment policies, and external control mechanisms require separate investigation. For nonhuman living systems, the relevant social and ecological evidence likewise constrains any inference from regularity to expectation, practical commitment, or explicitly moral appraisal. This calibrated vocabulary aims to preserve comparison while allowing genuine differences in kind.

Established accounts of social conventions, normative expectations, norm-emergence simulations, and institutional grammars supply substantive explanatory resources for parts of this program [52, 8, 43, 17]. The proposed GR extension contributes a particular analytical question: how do the means of recognizing, communicating, and revising a practice change during the interaction whose normative consequences investigators are attempting to explain? Answering that question may require tracing changes in the encounter's organization as well as changes in each participant's locally observable behavior. The extension earns explanatory standing where it supports distinctions, predictions, or counterfactual comparisons that simpler accounts leave unresolved.

14.2 Initialization, Translation, and Divergent Evolutionary Outcomes

The problem of initialization exposes the situated character of normative action. At an early encounter, participants may have limited information about another system's interpretations, acceptable signals, practical vulnerabilities, or means of expressing a response. Communication, observation, withdrawal, and delay can each carry consequences under some plausible descriptions of the encounter. Under explicitly stipulated and incompatible sets of admissible actions, a common admissible choice may be absent. The corresponding conditional argument identifies a constraint on protocol selection; it establishes neither the inevitability of conflict in actual encounters nor a universal prescription for first contact.

Provisional normative anchoring offers one possible response to practical indeterminacy. An actor can use a revisable working rule to organize interaction, record disagreement, or preserve an opportunity for later correction. The capacity to act on such a rule should be distinguished from identification with a community, inward endorsement, and acceptance of the rule's authority. A working arrangement may derive its stability from a shared expectation, a negotiated procedure, an external incentive, or restricted alternatives. The investigation therefore also retains multiple local practices, contextual switching, temporary restraint, and jointly created procedures as possible forms of sustained interaction. Their adequacy depends upon the participants' capacities and the consequences of the practice in question.

Normative translation introduces a second difficulty. A pair of participants may coordinate successfully even when their internal representations, learned histories, or reasons for compliance differ. Conversely, repeated failures of coordination may reveal a restricted communication interface or unequal opportunities to correct an interpretation, with limited implications for any presumed incompatibility of their underlying values. The contextual, potentially partial interpretive relations proposed in this paper are intended to describe observable correspondences and their limitations. They permit the formation of new signs and practices during contact while leaving the question of shared internal meaning to further evidence. A successful translation claim consequently needs an identified domain of use, repeatable or independently assessed behavioral and communicative evidence, and scrutiny of who controls the

categories of interpretation.

The possible outcomes of normative interaction are diverse. Negotiated coordination, asymmetric fixation, contextual pluralism, and continued alternation can each be generated by the simple conditional examples examined in the formal section. These examples demonstrate compatibility between specified update assumptions and different kinds of outcomes. They do not estimate the relative prevalence of those outcomes in living societies, animal communities, or deployed AI systems. Likewise, persistence of a convention, attraction to a state in a mathematical model, and legitimate commitment to a rule remain analytically separate. Particular descriptions of path dependence, switching barriers, or critical transitions require the evidence and parameters appropriate to the model employed. The term *evolutionary* here describes transformation within specified biological, cultural, institutional, learning, or computational accounts; it provides no presumption of progress, necessary convergence, or a fundamental ontological direction of time.

14.3 Power, Practical Dependence, and Normative Authority

The formation and stabilization of norms also depend upon access to the conditions in which practices can be carried out. Control over sanctions, communication interfaces, agenda setting, infrastructure, and rule amendment can alter the feasible options available to participants. Such arrangements may produce highly regular compliance while leaving the distribution of understanding, endorsement, and authority unresolved. The simple observational-equivalence result in the formal framework illustrates this inferential limitation under specified model assumptions: distinct underlying mechanisms can generate the same recorded pattern of compliance. Identifying a participant's reasons for acting, where such reasons are meaningfully attributable and empirically accessible, therefore calls for additional evidence.

Human–AI interaction illustrates the importance of locating power in the relevant socio-technical arrangement. Users, developers, model providers, deployment organizations, regulators, and other affected parties may exercise different forms of control over the visible protocol and its revision. An apparent conversation between two agents can be conditioned by model-version updates, access restrictions, logging policies, default interface settings, and institutional decisions made outside the exchange. A research design that tracks these conditions can test whether an observed adaptation reflects mutual adjustment, one-sided accommodation, or a change imposed by the infrastructure operator. Such a finding would concern the organization of interaction; it would provide no automatic verdict on an artificial system's consciousness or normative standing.

Similar considerations extend to comparisons involving human and nonhuman life. The capacity to restructure habitat, mediate a species' opportunities for communication, or determine the conditions of conservation can create asymmetric effects upon available practices. A technically powerful participant may be capable of protecting or enlarging another system's opportunities while simultaneously introducing structural dependence. Assessing those relationships requires ecological and welfare evidence together with careful ethical analysis. Hypothetical contact with a technologically superior extraterrestrial civilization offers a power-reversal exercise for examining the consistency of human justifications for intervention. Its contribution is conceptual: there is no empirical extraterrestrial case to which the paper's proposed mechanisms have been fitted.

This analysis supplies an interface with the separate political economy of heterogeneous GR subjects. The present inquiry studies the effects of material and institutional asymmetry on the *formation and persistence* of norms. A political-economic account can investigate the production, ownership, circulation, and control of the conditions upon which those processes depend. Their intersection is significant because an ability to make a rule practically effective can be mistaken for authority to make that rule binding. Yet descriptions of capability, observed acceptance, procedural legitimacy, and substantive justification address distinct questions. Accordingly, the proposed framework leaves the ethical evaluation of coercion, dependency, negotiated consent, representation, and remedies open to independent argument.

14.4 Empirical Evaluation and the Limits of the Proposed Framework

The comparative and formal analyses have identified candidate mechanisms. Their evidential status remains preliminary. The typed state representations and coupled updates permit heterogeneous local states, observation processes, interpretation mechanisms, and changing relational conditions. They provide a disciplined vocabulary for constructing models and comparing assumptions, rather than a completed empirically estimated account of normative development. Their value should be assessed against simpler and established alternatives, including individual reward-sensitive learning, imitation, fixed conventions, selective exposure, centralized filtering, and institutionally enforced compliance. Different explanatory mechanisms may coexist in a single setting, and the relevant study must identify the additional effect attributed to relational coevolution.

An empirical program could begin with reproducible, low-stakes human–AI interactions in which amendment rights, clarification channels, and available alternatives are recorded and varied under approved research protocols. The resulting observations should distinguish behavioral compliance, the ability to propose and implement revisions, repair of misunderstood instructions, and retention of a practice when particular constraints are removed. Longitudinal organizational studies could then examine how changing governance, access, and interpretive resources affect the practical revision of shared procedures. Research involving nonhuman organisms would require species-appropriate observational evidence and independently justified welfare safeguards. These studies should specify their identification assumptions, competing models, measurement limits, and the conditions under which apparently informative results remain inconclusive.

The proposed framework becomes more accountable when its claims are exposed to informative negative findings. If fixed-relation learning models predict behavioral coordination and rule change as successfully as models with additional coevolving relational variables, the explanatory case for those variables weakens in that domain. If a purported revision right has no observable implementation or accessible alternative, a null result may instead reveal a failed intervention. If investigators lack records of operative rules or changes in platform configuration, changes attributed to mutual agreement may remain unidentifiable. Such distinctions are central to causal evaluation, which requires clearly specified interventions and justified comparisons [28]. They also caution against interpreting a successful demonstration of computational coordination as a demonstration of common moral understanding.

Several limits persist even after improved measurement. Reliable cross-system recognition of normative expectations may be unavailable in some cases; a common measure of generative potential or normative value has not been defined or validated; partial translations may remain dependent upon the investigator’s categories; and experiments can expose participants to the very asymmetries they are designed to analyze. An increase in the precision of the model does not dissolve these methodological and ethical constraints. Careful reporting should make visible the researcher or institution that chose the descriptors, the parties able to contest them, and the empirical scope of each inference. The GR formalism itself remains an optional representation whose suitability is to be established locally, with ontological commitments held in suspension.

14.5 Prospects for Evolving Normative Relations

The practical relevance of the research program lies in its approach to interaction under persistent uncertainty and unequal power. Where a shared standard is unavailable or contested, a provisional protocol can make its own assumptions explicit, identify the agents or institutions authorized to apply it, provide means for recording divergent interpretations, and preserve opportunities for review. In an appropriate setting, reciprocal clarification, appeal, reversible amendment, and independent assessment of consequences may support cooperation while leaving substantive disagreement visible. These are candidate institutional practices, with context-specific costs and limits. Their selection requires a further normative argument and attention to parties who may be unable to participate in familiar forms of deliberation.

For diplomacy and peacemaking, the proposed inquiry invites examination of changes in interpretive access and practical options alongside changes in declared positions. For human–AI governance, it

encourages versioned documentation of rule changes and auditability of who can revise them. For ecological coexistence, it directs attention to the conditions under which human interventions alter another living system’s feasible practices, with due regard for uncertainty in attributing norms. Each application demands its own empirical and normative standards. An analytically useful shared vocabulary can facilitate comparison while permitting historically and biologically specific forms of agency, dependence, and value to remain distinct.

Heterogeneous contact may change the capacities through which participants recognize one another’s practices as meaningful and potentially normative. Interactions can generate new interpretive resources as well as new patterns of coordination, disagreement, and institutional control. Understanding these transformations requires a sustained relationship among formal modeling, historical and interpretive inquiry, empirical comparison, and normative criticism. In the present framework, documented co-change between practices and relational conditions is an initial observation. A stronger coevolutionary account requires evidence of an interaction-dependent feedback pathway; its presence, direction, and asymmetry remain open to empirical testing. The formation of an enduring practice remains a question about its history and mechanisms; its claim upon participants remains a question of justification. The generative-relational perspective remains a revisable framework for studying evolving relations and the conditions under which rules are interpreted, challenged, and changed.

Research Ethics and Source Provenance

Scope. This revised full draft contains Sections 1–14, including the preliminary formal framework, bounded cross-system comparison, and a proposed program of causal, comparative, and ethically constrained empirical evaluation. Original empirical investigations, new agent-based experiments, and comprehensive normative evaluations remain planned; no original empirical findings are reported. Cross-species and hypothetical first-contact cases are illustrative and carry no assumptions of identical moral status or subjectivity.

Sources. Cited bibliographical records and their intended uses were checked against accessible publisher, institutional, or open-access source materials; the verification ledger records the level of access for each entry. Many books and articles were checked through metadata and abstracts rather than full-text examination. The ledger does not constitute an exhaustive claim-by-claim audit of every work.

Authorship and status. This independent draft was prepared with AI assistance under the human author’s research direction. It has not been peer reviewed. The author remains responsible for interpretation, citations, revision, and dissemination.

References

- [1] Saleema Amershi et al. “Guidelines for Human–AI Interaction”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, pp. 1–13. DOI: 10.1145/3290605.3300233.
- [2] Kristin Andrews, Simon Fitzpatrick, and Evan Westra. “Human and Nonhuman Norms: A Dimensional Framework”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 379.1897 (2024), p. 20230026. DOI: 10.1098/rstb.2023.0026.
- [3] Masahiko Aoki. *Toward a Comparative Institutional Analysis*. Cambridge, MA: MIT Press, 2001. ISBN: 9780262011877. DOI: 10.7551/mitpress/6867.001.0001.
- [4] W. Brian Arthur. “Competing Technologies, Increasing Returns, and Lock-In by Historical Events”. In: *The Economic Journal* 99.394 (1989), pp. 116–131. DOI: 10.2307/2234208.
- [5] Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. “Emergent Social Conventions and Collective Bias in LLM Populations”. In: *Science Advances* 11.20 (2025), eadu9368. DOI: 10.1126/sciadv.adu9368.

- 3230 [6] Peter Bachrach and Morton S. Baratz. "Two Faces of Power". In: *American Political Science Review* 56.4
3231 (1962), pp. 947–952. DOI: 10.2307/1952796.
- 3232 [7] Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde. "Defining Agency: Individuality, Norma-
3233 tivity, Asymmetry, and Spatio-temporality in Action". In: *Adaptive Behavior* 17.5 (2009), pp. 367–386.
3234 DOI: 10.1177/1059712309343819.
- 3235 [8] Cristina Bicchieri. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford
3236 University Press, 2017. DOI: 10.1093/acprof:oso/9780190622046.001.0001.
- 3237 [9] Guido Boella, Leendert van der Torre, and Harko Verhagen. "Introduction to Normative Multiagent Sys-
3238 tems". In: *Computational and Mathematical Organization Theory* 12.2-3 (2006), pp. 71–79. DOI: 10.
3239 1007/s10588-006-9537-7.
- 3240 [10] Robert Boyd and Peter J. Richerson. *Culture and the Evolutionary Process*. Chicago: University of Chicago
3241 Press, 1985.
- 3242 [11] Philippa Brakes et al. "Animal Cultures Matter for Conservation". In: *Science* 363.6431 (2019), pp. 1032–
3243 1034. DOI: 10.1126/science.aaw3557.
- 3244 [12] Michael E. Bratman. *Intention, Plans, and Practical Reason*. Planning theory of intention and defeasible
3245 practical commitments. Harvard University Press, 1987. ISBN: 9780674458192.
- 3246 [13] L. L. Cavalli-Sforza and M. W. Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*.
3247 Princeton: Princeton University Press, 1981. DOI: 10.2307/j.ctvx5wbt8.
- 3248 [14] Damon Centola et al. "Experimental Evidence for Tipping Points in Social Convention". In: *Science*
3249 360.6393 (2018), pp. 1116–1119. DOI: 10.1126/science.aas8827.
- 3250 [15] Robert B. Cialdini, Carl A. Kallgren, and Raymond R. Reno. "A Focus Theory of Normative Conduct:
3251 A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior". In: *Advances in*
3252 *Experimental Social Psychology*. Vol. 24. Academic Press, 1991, pp. 201–234. DOI: 10.1016/S0065–
3253 2601(08)60330-5.
- 3254 [16] Herbert H. Clark and Susan E. Brennan. "Grounding in Communication". In: *Perspectives on Socially*
3255 *Shared Cognition*. Ed. by Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley. Washington, DC:
3256 American Psychological Association, 1991, pp. 127–149. DOI: 10.1037/10096-006.
- 3257 [17] Sue E. S. Crawford and Elinor Ostrom. "A Grammar of Institutions". In: *American Political Science Review*
3258 89.3 (1995), pp. 582–600. DOI: 10.2307/2082975.
- 3259 [18] Donald Davidson. "On the Very Idea of a Conceptual Scheme". In: *Proceedings and Addresses of the*
3260 *American Philosophical Association* 47 (1974), pp. 5–20.
- 3261 [19] Miranda Fricker. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.
3262 DOI: 10.1093/acprof:oso/9780198237907.001.0001.
- 3263 [20] Karl Friston. "The Free-Energy Principle: A Unified Brain Theory?" In: *Nature Reviews Neuroscience* 11
3264 (2010), pp. 127–138. DOI: 10.1038/nrn2787.
- 3265 [21] Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. Cambridge, MA: MIT Press,
3266 1998.
- 3267 [22] Bruno Galantucci. "An Experimental Study of the Emergence of Human Communication Systems". In:
3268 *Cognitive Science* 29.5 (2005), pp. 737–767. DOI: 10.1207/s15516709cog0000_34.
- 3269 [23] Michele J. Gelfand et al. "Differences between Tight and Loose Cultures: A 33-Nation Study". In: *Science*
3270 332.6033 (2011), pp. 1100–1104. DOI: 10.1126/science.1197754.
- 3271 [24] James J. Gibson. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.
- 3272 [25] Mark Granovetter. "Threshold Models of Collective Behavior". In: *American Journal of Sociology* 83.6
3273 (1978), pp. 1420–1443. DOI: 10.1086/226707.
- 3274 [26] Brian Hare et al. "The Domestication of Social Cognition in Dogs". In: *Science* 298.5598 (2002), pp. 1634–
3275 1636. DOI: 10.1126/science.1072702.
- 3276 [27] Joseph Henrich et al. "'Economic Man' in Cross-Cultural Perspective: Behavioral Experiments in 15
3277 Small-Scale Societies". In: *Behavioral and Brain Sciences* 28.6 (2005), pp. 795–855. DOI: 10.1017/
3278 S0140525X05000142.

- 3279 [28] Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Continuously updated online edition;
3280 2020 citation recommended by authors. Boca Raton, FL: Chapman & Hall/CRC, 2020. URL: [https://](https://miguelhernan.org/whatifbook)
3281 miguelhernan.org/whatifbook.
- 3282 [29] International Academy of Astronautics. *Declaration of Principles Concerning the Conduct of the Search for*
3283 *Extraterrestrial Intelligence (SETI): June 2026 Update*. Position paper; applicability limited to astronomy-
3284 based SETI and post-detection procedures. June 1, 2026. URL: [https://www.seti.org/research/](https://www.seti.org/research/seti-101/protocols-for-an-eti-signal-detection/)
3285 [seti-101/protocols-for-an-eti-signal-detection/](https://www.seti.org/research/seti-101/protocols-for-an-eti-signal-detection/).
- 3286 [30] Michihiro Kandori, George J. Mailath, and Rafael Rob. “Learning, Mutation, and Long Run Equilibria in
3287 Games”. In: *Econometrica* 61.1 (1993), pp. 29–56. DOI: 10.2307/2951777.
- 3288 [31] Christian Kuehn. “A Mathematical Framework for Critical Transitions: Bifurcations, Fast–Slow Systems
3289 and Stochastic Dynamics”. In: *Physica D: Nonlinear Phenomena* 240.12 (2011), pp. 1020–1035. DOI:
3290 10.1016/j.physd.2011.02.012.
- 3291 [32] Timur Kuran. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge,
3292 MA: Harvard University Press, 1995. ISBN: 9780674707580.
- 3293 [33] Robert J. Lempert et al. “A General, Analytic Method for Generating Robust Strategies and Narrative
3294 Scenarios”. In: *Management Science* 52.4 (2006), pp. 514–528. DOI: 10.1287/mnsc.1050.0472.
- 3295 [34] Michael L. Littman. “Markov Games as a Framework for Multi-Agent Reinforcement Learning”. In:
3296 *Proceedings of the Eleventh International Conference on Machine Learning*. Morgan Kaufmann, 1994,
3297 pp. 157–163. DOI: 10.1016/B978-1-55860-335-6.50027-1.
- 3298 [35] National Research Council. *Valuing Ecosystem Services: Toward Better Environmental Decision-Making*.
3299 Washington, DC: National Academies Press, 2005. DOI: 10.17226/11139.
- 3300 [36] Brian A. Nosek et al. “The Preregistration Revolution”. In: *Proceedings of the National Academy of Sci-*
3301 *ences* 115.11 (2018), pp. 2600–2606. DOI: 10.1073/pnas.1708274114.
- 3302 [37] F. John Odling-Smee, Kevin N. Laland, and Marcus W. Feldman. *Niche Construction: The Neglected*
3303 *Process in Evolution*. Monographs in Population Biology 37. Princeton: Princeton University Press, 2003.
3304 ISBN: 9780691044378. URL: <https://www.jstor.org/stable/j.ctt24hqpd>.
- 3305 [38] Hisashi Ohtsuki et al. “A Simple Rule for the Evolution of Cooperation on Graphs and Social Networks”.
3306 In: *Nature* 441 (2006), pp. 502–505. DOI: 10.1038/nature04605.
- 3307 [39] Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge:
3308 Cambridge University Press, 1990. ISBN: 9780521405997.
- 3309 [40] Elinor Ostrom. *Understanding Institutional Diversity*. Princeton University Press, 2005. DOI: 10.1515/
3310 9781400831739.
- 3311 [41] Willard Van Orman Quine. *Word and Object*. Cambridge, MA: The MIT Press, 1960. ISBN:
3312 9780262170017.
- 3313 [42] Inioluwa Deborah Raji et al. “Closing the AI Accountability Gap: Defining an End-to-End Framework
3314 for Internal Algorithmic Auditing”. In: *Proceedings of the 2020 Conference on Fairness, Accountability,*
3315 *and Transparency*. Association for Computing Machinery, 2020, pp. 33–44. DOI: 10.1145/3351095.
3316 3372873.
- 3317 [43] Bastin Tony Roy Savarimuthu and Stephen Cranefield. “Norm Creation, Spreading and Emergence: A
3318 Survey of Simulation Models of Norms in Multi-Agent Systems”. In: *Multiagent and Grid Systems* 7.1
3319 (2011), pp. 21–54. DOI: 10.3233/MGS-2011-0167.
- 3320 [44] Yoav Shoham and Moshe Tennenholtz. “On the Emergence of Social Conventions: Modeling, Analysis,
3321 and Simulations”. In: *Artificial Intelligence* 94.1-2 (1997), pp. 139–166. DOI: 10.1016/S0004-3702(97)
3322 00028-3.
- 3323 [45] Susan Leigh Star and Karen Ruhleder. “Steps Toward an Ecology of Infrastructure: Design and Access for
3324 Large Information Spaces”. In: *Information Systems Research* 7.1 (1996), pp. 111–134. DOI: 10.1287/
3325 isre.7.1.111.
- 3326 [46] Luc Steels. “A Self-Organizing Spatial Vocabulary”. In: *Artificial Life* 2.3 (1995), pp. 319–332. DOI: 10.
3327 1162/artl.1995.2.3.319.
- 3328 [47] Luc Steels. “The Origins of Syntax in Visually Grounded Robotic Agents”. In: *Artificial Intelligence* 103.1-
3329 2 (1998), pp. 133–156. DOI: 10.1016/S0004-3702(98)00066-6.

- 3330 [48] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. 2nd ed. Cambridge,
3331 MA: MIT Press, 2018.
- 3332 [49] United Nations. *Treaty on Principles Governing the Activities of States in the Exploration and Use of*
3333 *Outer Space, Including the Moon and Other Celestial Bodies*. Article IX; treaty obligations of states par-
3334 ties. Jan. 27, 1967. URL: [https://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/](https://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/outerspacetreaty.html)
3335 [outerspacetreaty.html](https://www.unoosa.org/oosa/en/ourwork/spacelaw/treaties/outerspacetreaty.html).
- 3336 [50] Evan Westra et al. “In Search of Animal Normativity: A Framework for Studying Social Norms in Non-
3337 Human Animals”. In: *Biological Reviews* 99.3 (2024), pp. 1058–1074. DOI: 10.1111/brv.13056.
- 3338 [51] H. Peyton Young. “The Evolution of Conventions”. In: *Econometrica* 61.1 (1993), pp. 57–84. DOI: 10.
3339 2307/2951778.
- 3340 [52] H. Peyton Young. “The Evolution of Social Norms”. In: *Annual Review of Economics* 7 (2015), pp. 359–
3341 387. DOI: 10.1146/annurev-economics-080614-115322.