V1.0.0

A Minimal Generative Grammar of Coordination: Making the Enactive Theory of Language Falsifiable in Multi-Agent Simulation

Rowan Brad Quni-Gudzinas

QNFO-ULA

ABSTRACT

The enactive thesis that language is an interaction protocol which generates, rather than describes, social reality is widely cited but rarely tested, because no operational specification of the thesis exists. This gap blocks both a research programme — nobody can say which simulation outcome would confirm or refute the thesis — and an engineering goal, namely autonomous agents that bootstrap shared meaning without a pre-installed world model. We convert the thesis into a design specification by defining a minimal grammar of coordination: a finite inventory of form-meaning constructions, an alignment update rule, and a norm-codification threshold, implemented as the only communicational endowment of agents that share no prior ontology. We derive analytical predictions for four measurable outputs: a lexicon-sufficiency bound $L \geq S^{1/m}$ giving 22 constructions for 10^4 situation types at message length 3; a normative-complexity ceiling of 66 coordination patterns for a 50-agent community, an 18.6-fold compression over unstructured pairwise norms; a perturbation law under which a 30% bot-message fraction reduces the stable shared lexicon from 52.8 to 39.0 constructions, a 26.1% drop; and a spectrum-retention result showing that a strict threshold discards 95.02% of pairwise coordination information, a 20.1-fold loss relative to spectrum-based measurement. All quantities are computed from stated assumptions, not simulated. The derivations yield falsifiable predictions that distinguish grammar variants, specify the simulation study that would test them, and give trust-and-safety analysts a spectrum-based rather than threshold-based target for measuring coordination in live platforms.

PUBLISHER CLASSIFICATION High-quality scholarly research (QNFO publisher classification for open-access, versioned research with a public revision history and reader correction channel; this is not a claim of independent peer review.)

PEER-REVIEW STATUS Not independently peer-reviewed

CITE AS Quni-Gudzinas, R. B. (n.d.). A Minimal Generative Grammar of Coordination: Making the Enactive Theory of Language Falsifiable in Multi-Agent Simulation. QNFO Preprint v1.0.0.

1. Introduction

Two communities face the same unsolved problem from opposite sides. The first is the trust-and-safety and platform-integrity community: coordinated inauthentic behaviour on social media — astroturf networks, troll farms, bot-driven spam — manipulates public discussion, and the analytical tools available to detect it impose arbitrary binary thresholds that discard most of the coordination signal present in the data [1]. The second is the multi-agent AI community: as autonomous agents and generative systems enter social contexts whose societal impacts are still being assessed [7], systems are deployed that communicate through human language acquired from data, with no specification of how a community of artificial agents could generate its own shared meanings from interaction alone. Both communities lack a theory of coordination precise enough to engineer: the first cannot predict how a platform's coordination structure responds to interventions, and the second cannot specify the communicational protocol an agent population needs to sustain a stable social reality. What would change if the result of this paper holds: the claim that language generates social reality would become a testable design specification, and a designer could tune a named parameter of an agent protocol and predict the resulting lexicon size, normative complexity, and resilience to perturbation before running anything.

The philosophical source of the blockage is the enactive model of language, in which language is not a code that transmits descriptions of a pre-existing world but an interaction protocol that brings a shared social world into being between agents. The thesis is widely repeated and almost never tested, because it has been stated without an operational form: no definition says what the minimal protocol is, what outputs count as an "enacted world," or which observation would refute the claim. Critics can therefore dismiss the thesis as unfalsifiable, and engineers cannot use it, because a thesis with no parameters is not a design. The present paper removes that blockage by treating the enactive thesis the way a physicist treats a symmetry principle: as a constraint from which measurable consequences are derived.

Our approach is to write down the smallest protocol that could plausibly generate a social reality between agents with no pre-shared world model, and to call it a grammar of coordination. The protocol has three components, each chosen to be minimal: (i) a finite inventory of constructions — form-meaning mappings that agents imitate, align, and reuse, in the sense in which construction grammar treats all linguistic knowledge as freely combinable constructions on a lexicon-grammar continuum [3]; (ii) an alignment update rule that shifts each agent's constructions toward those of successful interaction partners; and (iii) a norm-codification threshold at which a repeatedly successful coordination pattern becomes a norm that constrains future interactions. We then derive, analytically and with all arithmetic shown, four quantities that a simulation of this protocol must reproduce if the enactive thesis is operational: a lexicon-sufficiency bound, a normative-complexity ceiling, a perturbation-response law for injected inauthentic messages of the kind that bot networks produce in real platforms [6], and a spectrum-retention quantity for the measurement layer.

The derivations use a coordination-weight model chosen to mirror the spectrum view of coordination argued for in the platform-integrity literature: coordination is a matter of degree present to some extent in every community, not a binary property cut at a fixed threshold [1]. The perturbation model uses a bot-fraction parameter motivated by documented bot behaviour in content-spreading dynamics and election-period manipulation [6], [5]. The falsifiability criterion for the whole programme follows the requirement that a hypothesis posit cause and effect and be testable at bounded resource cost [8].

The contributions of this paper are:

- A formal definition of the minimal grammar of coordination as a three-component protocol (construction inventory, alignment rule, codification threshold) for agents with no pre-shared world model, given in Section 3.
- A derivation of the lexicon-sufficiency bound $L \geq S^{1/m}$, giving $L = 22$ constructions for $S = 10^4$ situation types at message length $m = 3$, in Section 4.1.
- A derivation of the normative-complexity ceiling $n = \min\{\binom{c}{2}, Lm\} = 66$ norms for a 50-agent community, with a compression ratio of 18.6 relative to unstructured pairwise norms, in Section 4.2.
- A derivation of the perturbation-response law: at bot fraction $b = 0.3$ the stable shared lexicon falls from $L^* = 52.8$ to $L^* = 39.0$ constructions, a 26.1% reduction, in Section 4.3.
- A derivation of the coordination-spectrum retention result: a strict threshold at $t = 0.15$ on an exponential coordination-weight distribution with mean 0.05 retains 24,868 of 499,500 edges and discards 95.02% of the graded coordination information, a 20.1-fold loss, in Section 4.4.
- A falsifiability analysis stating which simulation outcomes would confirm and which would refute the operationalised enactive thesis, in Section 5 and Section 6.
- A translation of the results into an engineering pathway — who would use the grammar, at what readiness level, and what the smallest decisive experiment is — in Section 7.

Beyond multi-agent AI and platform integrity, the same protocol-level view of coordination bears on any field that models how shared structure grows out of repeated interaction, including the layered relational data extracted from online sharing systems [2] and the study of coordinated behaviour in collective online settings [1].

2. Background and Related Work

Coordination as a measurable spectrum. The detection of coordinated behaviour on social media has been dominated by manual investigation and by computational approaches that make simplifying assumptions to become tractable; the network-based framework of [1] instead embeds the intrinsic properties of coordination in its methodology, using a six-step pipeline that begins by selecting a starting set of users and proceeds through network construction and filtering to coordination-aware community detection. Its comparative evaluation shows the decisive methodological point for our purposes: by refusing to enforce a binary separation between coordinated and uncoordinated behaviour, the framework studies coordination across its entire spectrum, measuring the intrinsic coordination of each emerging community, whereas threshold-based approaches cut all edges below an arbitrary strict threshold and thereby discard the graded structure. The conclusions of [1] state the finding explicitly: there is a certain extent of coordination in every online community, but not all coordinated groups are equally interesting, and per-community coordination degree enables finer-grained downstream analysis than fixed thresholds. We adopt exactly this spectrum view as the measurement layer of our simulation programme: the coordination structure that our agent grammar generates is a weighted network whose edge weights are degrees of alignment, not a binary graph.

Layered relational structure. The data that online sharing systems accumulate reflects users' collective behaviour and shared activities, and from it multiple kinds of relationships can be extracted and grouped into layers — direct and oblique relations — forming a multidimensional social network [2]. This matters for our protocol design because a community of agents generates exactly such layers endogenously: a layer of direct interaction history, a layer of imitated or second-hand alignment, and a layer of codified norms. Our grammar treats these layers as outputs to be measured, and the multidimensional-network result of [2] supplies the representational format in which the outputs are compared across simulation runs and against real platform data.

Constructions as the unit of coordination. Construction grammar and artificial intelligence share a long-standing attitude toward communication: language serves the communicative needs of its users, production expresses an idea as an utterance, comprehension reconstructs the intention behind an observed utterance, and communication is bidirectional [3]. The basic tenets that [3] maps onto AI data structures are the ones our grammar adopts verbatim: all linguistic knowledge is captured in constructions, form-meaning mappings that can freely combine as long as no conflicts occur, and there is no lexicon-grammar divide — words and grammar rules are points on one continuum. This gives our protocol its minimal unit: an agent's entire communicational endowment is a set of constructions, and "learning to coordinate" is the growth and alignment of that set. [3] further documents that AI methods have historically been the route by which constructionist linguistics was made computationally operational, which is precisely the move we make for the enactive thesis.

High-automation analysis of social patterns. GeoAI in social science emphasises that high-automation AI, supported by massive computational resources, enables fine-grained analysis of social patterns that earlier methods could not reach [4]. Our combinatorial derivations and the proposed simulation programme take the same stance at the protocol level: the expressive capacity of an interaction grammar is a quantity to be computed and measured at scale, not only argued about qualitatively. Where [4] applies high-automation analysis to spatial social data, we apply the same computational posture to the space of possible coordination protocols.

Perturbation by inauthentic agents. Social bots are efficient automatic tools for generating content and diffusing information to human populations, and their interplay with information diffusion has been documented in two domains: political manipulation, as in the use of social bots during the 2016 US Presidential Election to sway discussion around candidates and frame agendas, and social spam campaigns [6]. The same work reviews the detection toolbox — graph-centric detection, crowdsourcing, and feature-based supervised learning — and the tools used in its case studies, including Bot Or Not for bot detection and SentiStrength for sentiment analysis. The concrete mechanism of manipulation is documented for the 2016 case in which a network of accounts animated by Russian operatives attempted to di-

vert American political discourse, creating an online astroturf in direct contact with the public [5]. For our programme these works fix the perturbation model: an exogenous population of agents that emits well-formed-looking messages without participating in the alignment dynamics. Section 4.3 derives the predicted effect of such a population on the shared lexicon, and Section 7 proposes the simulation experiment that would test the prediction.

Generative AI entering social reality. The social impact of generative models has become a live assessment problem, with documented benefits such as enhanced customer service and automation of repetitive tasks, and documented areas of concern concentrated on privacy and bias [7]. That review's conclusion — that generative AI reflects the social context in which it was created, so that repairing the model alone is insufficient without examining the social factors around it — is an argument for systems whose social context is generated endogenously and therefore inspectable at the protocol level. Our grammar supplies exactly that: a generative model whose "social context" is the construction population its own interactions produce, making the analogue of bias and norm formation a measurable quantity rather than an emergent mystery.

Falsifiability and experiment design as constraints on the programme. The Artificial Scientist framework requires that a hypothesis be falsifiable, positing cause and effect, and that under finite computational resources the search over hypotheses be guided by an explicit heuristic weighing plausibility and expected profit [8]. Its treatment of experimental design — isolating the hypothesised cause, controlling other variables, and evaluating the information gain of an experiment against its resource cost and risk, with geometric-mean utility recommended when future capability depends on preceding outcomes — is the standard we use in Section 7 when specifying the smallest decisive experiment. [8] also supplies a cautionary negative result that shapes our protocol design: in its binary-task example, observational situation-response data alone cannot induce two concurrent rules, so an agent restricted to passive observation fails where interactive experimentation would succeed; this is a direct argument that our agents must act and align, not merely observe.

Together, these works supply the measurement philosophy [1], the representational format [2], the minimal linguistic unit [3], the computational posture [4], the perturbation model [5], [6], the deployment context [7], and the methodological standard [8] for the grammar defined next.

3. Methods

This paper's method is analytical model-building followed by explicit derivation; no simulation, experiment, or data collection was performed, and every number in Section 4 is computed from stated assumptions by shown arithmetic. A study that would test the derivations is specified as a proposal in Section 7.

Choice of protocol formalism. Two formalisms are available for specifying a minimal coordination protocol: an enumerative rule-triple model, in which a protocol is a sequence of rules each combining a primitive action, a coordination operator, and a lexical slot; and a construction-inventory model, in which each agent carries a population of form-meaning constructions that align and codify through use. We take the construction-inventory model, under the convention that the enactive thesis concerns the endogenous growth and alignment of shared form-meaning mappings rather than the enumeration of a fixed rule space; the enumerative model fixes the rule inventory before interaction, which contradicts the generative character the thesis asserts. All derivations below use the construction-inventory convention.

The protocol. A population of N agents interacts in discrete rounds. Agent i carries a construction inventory $C_i = \{(f_k, \mu_k)\}$ of form-meaning pairs, following the construction-grammar tenet that all linguistic knowledge is captured in freely combinable constructions [3]. No agent is given a world model, an ontology, or a shared codebook: the inventory is the agent's entire communicational endowment, and the meanings μ_k are grounded only in the agent's own interaction history. In round t, a pair of agents (i, j) engages in a joint task whose success is defined operationally (a coordinated action succeeds if both agents' subsequent actions match). Agent i selects constructions from C_i, combines

them without conflict as construction grammar permits [3], and transmits; agent j attempts to reconstruct the intention, in the bidirectional production-comprehension loop of [3]. On success, both agents shift the used constructions toward each other's variants by an alignment rate $\alpha \in (0, 1]$; on failure, each agent invents or mutates a candidate construction with novelty rate ν. When a construction has participated in successful coordination in a fraction θ of a sliding window of interactions across distinct partners, it is codified: it becomes a norm that constrains future interactions and is resistant to mutation. The triple (C, α, θ), together with (ν), is the grammar of coordination.

Measurement layer. Following the spectrum view of coordination [1], we do not threshold interactions into coordinated/not-coordinated. Instead, a simulation implementing the protocol would maintain a weighted interaction network in which the weight $w_{ij} \in [0, 1]$ is the success rate of the pair (i, j), and would record per-community distributions of w. Following the multidimensional-network format of [2], we record three layers: the direct interaction layer (w_{ij}), the imitation layer (constructions adopted from third parties), and the norm layer (codified constructions).

Viability outputs. The enactive thesis, operationalised, claims that a stable social reality emerges endogenously. We define three measurable viability outputs, chosen to instantiate a trilemma of adaptability, efficiency, and persistence:

- L^*: the equilibrium size of the shared construction inventory (efficiency/parsimony);
- n: the number of codified norms (normative complexity);
- $\rho(b)$: the resilience of L^* under an exogenous bot fraction b (persistence).

Adaptability is measured as the rate at which new constructions enter the inventory and is left as a simulation-measured quantity; the derivations here concern L^*, n, and ρ.

Assumptions. The derivations require input values that no cited source provides; each is stated as an assumption and its sensitivity is examined in Section 6. We assume: $S = 10^4$ distinct situation types the community must coordinate over; message length $m = 3$ constructions; community size $c = 50$ for the norm derivation; pairwise coordination weights approximately exponential with mean $\mu_w = 0.05$; bot fraction $b = 0.3$; construction reinforcement rate $r = 1.0$ per round; construction decay rate $\lambda = 0.2$ per round without reinforcement; inventory carrying capacity $K = 66$ from the norm derivation of Section 4.2; and a measurement population of $N = 1000$ agents for the spectrum-retention derivation.

4. Analysis

4.1 Lexicon-sufficiency bound

Input numbers. $S = 10^4$ situation types (assumption, Section 3); message length $m = 3$ constructions (assumption, Section 3).

Derivation. A message of m constructions drawn from a lexicon of size L can distinguish at most L^m situation types, since each of the m slots admits L choices and constructions combine freely [3]. Sufficiency of the lexicon requires

$$L^m \geq S \quad \Longrightarrow \quad L \geq S^{1/m}.$$

With $S = 10^4$ and $m = 3$:

$$L \geq 10^{4/3} = 10^{1.3333\ldots} = 21.5443\ldots$$

Since L is an integer count of constructions, we round up:

$$L_{\min} = 22.$$

Check. $22^3 = 10648 \geq 10^4$; $21^3 = 9261 < 10^4$. The bound is tight at 22.

Interpretation. A community of agents with no pre-shared world model needs only 22 aligned constructions to coordinate over ten thousand situation types at message length three. This is the efficiency claim of the operationalised enactive thesis: shared meaning is cheap because constructions compose. It also yields a falsifiable prediction: a simulation in which the grammar is complete should converge to a shared inventory of order $10^{4/m}$ constructions, and a grammar variant that forbids free combination (a fixed-phrase protocol) would need $L \geq S = 10^4$ instead — a difference of three orders of magnitude that Section 5 turns into a decisive test.

4.2 Normative-complexity ceiling

Input numbers. Community size $c = 50$ (assumption, Section 3); $L = 22$ (derived, Section 4.1); $m = 3$ (assumption, Section 3).

Derivation. Without constructional reuse, every distinct pair of agents that must adjudicate a recurring interaction type would require its own norm, giving an unstructured norm count of

$$\binom{c}{2} = \frac{50 \times 49}{2} = \frac{2450}{2} = 1225.$$

With constructional reuse, each norm is a codified coordination pattern expressible as a construction-sequence of length up to m over the lexicon, so the number of distinct expressible patterns is bounded by

$$L \cdot m = 22 \times 3 = 66.$$

The codified norm count is therefore

$$n = \min\{\binom{c}{2},\ L \cdot m\} = \min\{1225,\ 66\} = 66.$$

The compression ratio of constructional reuse over pairwise norms is

$$\frac{1225}{66} = 18.5606\ldots \approx 18.6.$$

Interpretation. The grammar caps normative complexity at 66 for a 50-agent community, an 18.6-fold reduction relative to unstructured pairwise norm formation. This supplies the second falsifiable prediction: norm counts in simulation should saturate near Lm, not near $\binom{c}{2}$, and should scale as $O(Lm)$ rather than $O(c^2)$ as the community grows.

4.3 Perturbation-response law

Input numbers. Bot fraction $b = 0.3$ (assumption, Section 3, motivated by documented bot-driven diffusion and manipulation in platforms [6], [5]); reinforcement rate $r = 1.0$ per round; decay rate $\lambda = 0.2$ per round; carrying capacity $K = 66$ (derived, Section 4.2).

Derivation, step 1: effective alignment. A bot emits well-formed messages but does not participate in the alignment dynamics: its constructions are not updated by success, and its messages fail coordination with aligned partners at an elevated rate. We model the bot's effect as noise on both sides of each participating interaction: a message is useful for alignment only if it originates from a participating agent (probability $1 - b$) and is received in a context where the partner's participating message dominates (probability $1 - b$). The effective alignment rate is therefore

$$\alpha_{\text{eff}} = \alpha(1 - b)^2.$$

With $b = 0.3$:

$$(1 - b)^2 = 0.7^2 = 0.49, \qquad \alpha_{\text{eff}} = 0.49\,\alpha.$$

Equivalently, the effective decay rate of unreinforced constructions rises by the inverse factor:

$$\lambda_{\text{eff}} = \frac{\lambda}{(1-b)^2} = \frac{0.2}{0.49} = 0.40816\ldots$$

Derivation, step 2: equilibrium lexicon. We model the shared inventory size $L(t)$ as growing logistically toward the capacity K through reinforcement and decaying at rate λ_{eff}:

$$\frac{dL}{dt} = r\left(1 - \frac{L}{K}\right) - \lambda_{\text{eff}} L.$$

Setting $dL/dt = 0$:

$$r\left(1 - \frac{L^*}{K}\right) = \lambda_{\text{eff}} L^* \quad \Longrightarrow \quad L^* = K\left(1 - \frac{\lambda_{\text{eff}}}{r}\right).$$

Clean (unperturbed) case, $\lambda_{\text{eff}} = \lambda = 0.2$, $r = 1.0$, $K = 66$:

$$L^*_{\text{clean}} = 66 \times (1 - 0.2) = 66 \times 0.8 = 52.8.$$

Perturbed case, $\lambda_{\text{eff}} = 0.40816$:

$$L^*_{\text{bot}} = 66 \times \left(1 - \frac{0.40816}{1.0}\right) = 66 \times 0.59184 = 39.0614\ldots \approx 39.0.$$

Derivation, step 3: resilience. The fractional loss of the stable shared lexicon is

$$\rho = \frac{L^*_{\text{clean}} - L^*_{\text{bot}}}{L^*_{\text{clean}}} = \frac{52.8 - 39.0614}{52.8} = \frac{13.7386}{52.8} = 0.26020\ldots \approx 26.1\%.$$

Interpretation. A 30% bot-message fraction, because it degrades alignment on both the sending and receiving side, cuts the sustainable shared lexicon by 26.1% — more than the naive $b = 30\%$ because of the quadratic factor $(1-b)^2 = 0.49$. This quadratic amplification is the third falsifiable prediction: a simulation should show lexicon loss scaling as $(1-b)^2$ in the moderate-b regime, not linearly in b. It also gives platform analysts a quantitative stake in the spectrum view of [1]: the damage of an inauthentic population is a measurable, nonlinear function of its share of message volume, estimable per community from the weighted coordination network rather than from a binary bot/not-bot classification of the kind whose limitations [6] documents in its review of graph-centric, crowdsourcing, and feature-based detection.

4.4 Coordination-spectrum retention

Input numbers. Exponential coordination-weight distribution with mean $\mu_w = 0.05$ (assumption, Section 3); threshold $t = 0.15$, chosen as a strict cut in the style criticised in [1]; population $N = 1000$ agents (assumption, Section 3).

Derivation. For an exponential distribution with mean μ_w, the fraction of pairwise coordination weights above threshold t is

$$P(w > t) = e^{-t/\mu_w} = e^{-0.15/0.05} = e^{-3} = 0.049787\ldots \approx 4.98\%.$$

The complete interaction graph on $N = 1000$ agents has

$$\binom{1000}{2} = \frac{1000 \times 999}{2} = 499500$$

edges. A threshold-based analysis retains

$$499500 \times 0.049787 = 24868.4\ldots \approx 24868$$

edges and discards the remaining $499500 - 24868 = 474632$ edges, i.e. $474632/499500 = 0.95022\ldots \approx 95.02\%$ of the pairwise coordination information. The spectrum-based layer of Section 3 retains all 499500 weighted edges. The retention ratio is

$$\frac{499500}{24868} = 20.087\ldots \approx 20.1\text{-fold}.$$

Interpretation. At a strict threshold, twenty-fold more graded coordination structure is available to a spectrum-based measurement than to a binary cut. This quantifies, for the simulation setting, the methodological argument that [1] makes empirically — that threshold approaches enforce a binary separation and lose the graded structure that per-community coordination degrees reveal — and it defines what the simulation's measurement layer must record so that per-community coordination degrees, rather than fixed thresholds, can be tested as predictors of resilience under the perturbation law of Section 4.3.

5. Results

All values below are computed in Section 4 from the stated assumptions; none is measured or simulated.

Quantity	Derived value	Source
Minimal lexicon size $L_{\min}$ for $S = 10^4$, $m = 3$	22 constructions	Section 4.1
Unstructured pairwise norm count for $c = 50$	1225	Section 4.2
Normative-complexity ceiling $n = \min\{1225, 66\}$	66 norms	Section 4.2
Compression ratio $1225/66$	18.6	Section 4.2
Effective alignment factor $(1 - 0.3)^2$	0.49	Section 4.3
Effective decay rate $\lambda_{\text{eff}} = 0.2/0.49$	0.408 per round	Section 4.3
Clean equilibrium lexicon L^*_{clean}	52.8 constructions	Section 4.3
Perturbed equilibrium lexicon L^*_{bot}	39.0 constructions	Section 4.3
Lexicon loss ρ at $b = 0.3$	26.1%	Section 4.3
Edges above threshold $t = 0.15$ (of 499500)	24868 (4.98%)	Section 4.4
Coordination information discarded by the threshold	95.02%	Section 4.4
Spectrum-to-threshold retention ratio	20.1-fold	Section 4.4

Three falsifiable predictions follow. First, a simulation with the complete grammar should converge to a shared inventory of order $10^{4/m} \approx 22$ constructions, while a fixed-phrase variant should require of order 10^4; observation of the latter under the complete grammar, or of the former under a non-combinatorial variant, would contradict the lexicon bound. Second, codified norm counts should saturate near $Lm = 66$ and scale as $O(Lm)$, not $O(c^2)$; saturation near $\binom{c}{2}$ would contradict the ceiling. Third, lexicon loss under bot injection should scale as $(1 - b)^2$ in the moderate-b regime; linear scaling in b would contradict the perturbation law.

6. Discussion

What the derivations do and do not establish. The four quantities of Section 4 are consequences of stated assumptions, not observations: they show that the operationalised enactive thesis, once given a protocol, yields numerical predictions, and they do not show that any real or simulated agent population satisfies those predictions. The value of the

exercise is that each derivation names the parameter whose value a simulation must reproduce, so the thesis now has a failure condition. This section examines the assumptions, the failure modes, and what would falsify the central claims.

Falsification of the main result. The main result is that a three-component protocol — construction inventory, alignment rule, codification threshold — suffices as an operational form of the enactive thesis, with measurable consequences L^*, n, and $\rho(b)$. The observation that would falsify it is a simulation in which the complete grammar, run to equilibrium under the stated assumptions, fails to produce a stable shared inventory of order $10^{4/m} \approx 22$ constructions, or produces one whose norm count saturates near $\binom{c}{2} = 1225$ rather than near $Lm = 66$, or whose lexicon loss under bot injection scales linearly in b rather than as $(1-b)^2$. Any one of these three outcomes, obtained with the grammar implemented as specified and the measurement layer recording the full weighted spectrum, would show that the derivation's model of the protocol is wrong, and the specific mismatch would indicate which component — combination freedom, alignment dynamics, or codification — carries the error. Conversely, a simulation matching all three predictions would not prove the enactive thesis; it would show only that this operationalisation is internally coherent and empirically distinguishable from its alternatives, which is the strongest claim an analytical treatment can support.

Limitations. The derivations rest on assumptions whose sensitivity the following list states explicitly:

- $S = 10^4$ situation types: the lexicon bound scales as $S^{1/m}$, so a tenfold error in S shifts $L_{\min}$ by a factor $10^{1/3} \approx 2.15$; the qualitative claim that combinatorial protocols need lexicons orders of magnitude smaller than fixed-phrase protocols survives any plausible S, but the specific value 22 does not.
- Message length $m = 3$: the bound is exponentially sensitive to m; at $m = 2$ the requirement rises to $L \geq 100$, at $m = 4$ it falls to $L \geq 10$. The prediction "of order $10^{4/m}$" is therefore only as good as the assumption that agents use three-construction messages.
- The logistic model in Section 4.3, with $r = 1.0$ and $\lambda = 0.2$: the equilibrium $L^* = K(1 - \lambda_{\text{eff}}/r)$ is linear in the ratio λ_{eff}/r, so the 26.1% loss figure moves proportionally with any error in the decay-to-reinforcement ratio. If real alignment dynamics are not logistic — for example, if decay acts only on unused constructions rather than proportionally on the whole inventory — the perturbation law changes form, and the $(1-b)^2$ prediction fails with it.
- The quadratic noise model $\alpha_{\text{eff}} = \alpha(1-b)^2$: this assumes bot messages are indistinguishable noise on both the sending and receiving side. If bots instead imitate participating agents well enough to pass alignment occasionally, the effective factor lies between $(1-b)^2$ and $(1-b)$, and the predicted loss at $b = 0.3$ lies between 26.1% and 10.7% (the latter from $L^*_{\text{bot}} = 66 \times (1 - 0.2/0.7) = 66 \times 0.71428\ldots = 47.14$, giving $\rho = (52.8 - 47.14)/52.8 = 0.1071\ldots \approx 10.7\%$; recomputing: $0.2/0.7 = 0.28571\ldots$, $1 - 0.28571 = 0.71429$, $66 \times 0.71429 = 47.1429$, $\rho = 5.6571/52.8 = 0.10714 \approx 10.7\%$). The falsifiable content is the quadratic scaling, not the point value 26.1%.
- The exponential weight distribution with mean $\mu_w = 0.05$ and the threshold $t = 0.15$: the 95.02% discard figure depends on the ratio $t/\mu_w = 3$; a heavier-tailed weight distribution would raise the retained fraction and lower the 20.1-fold ratio, though any threshold strictly above the mean discards the majority of an exponential distribution's mass.
- The norm ceiling $n = \min\{\binom{c}{2}, Lm\}$: this treats every construction-sequence of length up to m as a candidate norm and assumes codification saturates at the expressible-pattern count. If norms can be compositional beyond length m, or if distinct norms share surface forms, the ceiling is an underestimate; if many expressible patterns are functionally equivalent, it is an overestimate.
- The carrying capacity $K = 66$ imported from Section 4.2 into Section 4.3: this identifies the inventory capacity with the norm ceiling, a convention chosen for parsimony; a simulation in which inventories exceed codified norms would break the identification and change L^* proportionally.

Failure modes of the protocol itself. Three endogenous failure modes are visible in the model. First, communicational collapse: if the decay rate exceeds the reinforcement rate ($\lambda_{\text{eff}} \geq r$), the equilibrium $L^* = K(1 - \lambda_{\text{eff}}/r)$ reaches zero, and the derivation predicts that a bot fraction b satisfying $(1-b)^2 \leq \lambda/r = 0.2$ — i.e. $b \geq 1 - \sqrt{0.2} = 1 - 0.44721\ldots = 0.55279\ldots \approx 0.553$ — collapses the shared lexicon entirely. This is a sharp, testable threshold: partial degradation below $b \approx 0.553$, total collapse above it. Second, norm ossification: codified constructions resist mutation, so a community that codifies too early locks in a lexicon too small for its situation space; the model has no term for this, and the simulation would need to measure whether codification timing controls it. Third, meaning drift without anchoring: because meanings are grounded only in interaction history, nothing in the protocol prevents the shared inventory from drifting to coordinate actions without tracking any external structure; whether that counts as an "enacted world" or as empty convention is exactly the question the simulation is designed to sharpen, not to settle.

What would strengthen the case. The analytical predictions gain force only by contact with the proposed simulation and, eventually, with platform data of the kind the spectrum framework of [1] was built to analyse. The strongest internal check available now is consistency among the derivations: the collapse threshold $b \approx 0.553$ derived above and the 26.1% loss at $b = 0.3$ come from the same equation, so a simulation measuring both would either confirm them jointly or refute the shared model — a two-for-one test that Section 7 places first in the experimental sequence.

7. Practical Implications and Engineering Pathways

(a) Who would use the result and what changes. Two user groups follow from the two communities of Section 1. First, trust-and-safety analysts and platform-integrity teams: if the perturbation law holds, the decision that changes is the choice of measurement target — instead of classifying accounts as bot/not-bot at a threshold and counting them, an analyst would estimate the community's coordination-weight distribution and predict lexicon and norm degradation as a function of the inauthentic message fraction b, using the spectrum-based per-community coordination degrees that [1] shows to be finer-grained than fixed thresholds. The operational change is from binary account labelling to distribution-level monitoring of the weighted interaction network, with the quadratic law $\rho \propto (1-b)^2$ as the early-warning signature: a community whose coordination-weight distribution shifts in the way the model predicts is absorbing inauthentic volume even before any individual account is classified. Second, multi-agent system designers: if the lexicon bound and norm ceiling hold, the design parameter that changes is the communicational endowment given to agents — a designer can provision a construction inventory of order $10^{4/m}$ rather than a per-situation codebook of order S, and can set the codification threshold θ knowing that normative complexity is capped near Lm rather than growing as $O(c^2)$. For both groups the work is currently foundational: no simulation has yet confirmed the derivations, so no deployment decision should rest on them; what the results change today is the specification of the experiments that would license such decisions.

(b) Current readiness. Against the NASA Technology Readiness Level definitions, this work is analytical: the derivations have not been tested in simulation, and the paper reports no experimental validation. The appropriate level is TRL 2 (concept and application formulated, analytical predictions derived, no proof of concept). It does not exceed TRL 3 even after the proposed simulation, since a multi-agent simulation is an analytical proof of concept, not a demonstrated capability in an operational environment.

(c) The smallest decisive experiment. The smallest test that would de-risk or falsify the central claims is a single multi-agent simulation run with the following specification, proposed here and not yet performed: a population of $N = 50$ agents carrying the three-component grammar of Section 3, interacting over a situation space of $S = 10^4$ types at message length $m = 3$, run to equilibrium with the measurement layer recording the full weighted interaction network and the three layers of Section 3. The measured quantities are the equilibrium shared inventory L^*, the codified norm count n, and the coordination-weight distribution. The refuting outcomes are stated in Section 6: L^* far from

order 22, n saturating near 1225, or a weight distribution inconsistent with the assumed exponential form. A second, equally small run adds a bot fraction $b = 0.3$ and measures the lexicon loss; a loss near 26.1% supports the quadratic law, a loss near 10.7% supports the linear-im imitation alternative of Section 6, and a loss near 30% refutes the two-sided noise model. Following the experimental-design standard of [8] — isolating the hypothesised cause (the bot fraction), controlling other variables, and evaluating information gain against resource cost — the two runs together test three predictions at the cost of two simulations, which is the highest information gain per run available in this design space.

(d) Binding constraints. The constraints that would bind first in building the simulation, in order: (i) the situation-space specification — the model requires 10^4 distinct coordination tasks with defined success, and generating them without smuggling in a shared ontology is the hardest design problem, since a task generator that encodes a world model defeats the purpose; (ii) compute for the measurement layer — the full weighted graph on N agents requires $O(N^2)$ edge storage per round, which at $N = 1000$ is 499500 weights per snapshot, tractable but the dominant memory cost; (iii) the codification window — the threshold θ is defined over a sliding window of interactions across distinct partners, and the window length is a free parameter the derivations do not fix, so the first simulations must sweep it; (iv) validation data — comparing simulated coordination-weight distributions against real platform distributions requires access to interaction data of the kind used in [1], which is governed by platform data-sharing terms and is an open question, not a stated constraint of this paper. Cost drivers beyond the simulation itself are not derivable from the stated assumptions and are left open.

(e) Order-of-magnitude bounds. The only performance bound derivable from Section 4 without new assumptions is the collapse threshold: the shared lexicon reaches zero when $(1-b)^2 \leq \lambda/r$, giving $b_{\text{collapse}} = 1 - \sqrt{\lambda/r}$. As a projection with stated assumptions (logistic dynamics, $r = 1.0$, $\lambda = 0.2$, two-sided noise model), this gives $b_{\text{collapse}} \approx 0.553$: a platform community modelled by these parameters tolerates inauthentic message fractions up to roughly 55% before predicted communicational collapse, and the tolerance scales as $1 - \sqrt{\lambda/r}$, so halving the decay rate raises the tolerance to $1 - \sqrt{0.1} = 1 - 0.31623\ldots = 0.68377\ldots \approx 0.684$. No cost figures are derivable from the stated inputs and none is offered.

(f) Risks and misuse paths. The same protocol that lets a designer bootstrap shared meaning lets an operator engineer it: a coordinated population emitting well-formed messages at a controlled fraction b is exactly the perturbation model of Section 4.3, and the documented manipulation cases — bot-driven diffusion and election-period manipulation [6], and the 2016 astroturf operation in direct contact with the public [5] — are instances of the mechanism the law quantifies. A party holding the perturbation law could tune b to maximise lexicon degradation while staying below detection thresholds; the mitigation is the same spectrum-based monitoring the law motivates, since the quadratic signature is visible in the weight distribution before individual accounts are classifiable. The dual-use character is intrinsic and is stated here rather than deferred.

8. Conclusion

The enactive thesis that language generates social reality has been converted from a philosophical claim into a design specification: a three-component protocol — a construction inventory, an alignment rule, and a codification threshold — from which four numerical consequences are derived with all arithmetic shown. A community of agents with no pre-shared world model needs $L_{\min} = 22$ constructions to coordinate over 10^4 situation types at message length 3; its normative complexity is capped at $n = 66$ for a 50-agent community, an 18.6-fold compression over unstructured pairwise norms; a 30% inauthentic message fraction cuts its stable shared lexicon by 26.1% through the quadratic factor $(1-b)^2 = 0.49$, with total collapse predicted above $b \approx 0.553$; and a strict measurement threshold at $t = 0.15$ on an exponential coordination-weight distribution with mean 0.05 discards 95.02% of the graded coordination information, a 20.1-fold loss against spectrum-based measurement. Each number is a prediction a simulation can confirm or refute,

and Section 6 states the specific outcomes that would falsify the operationalisation. The practical yield, at the current analytical readiness level of TRL 2, is a specification for the smallest decisive experiment — two simulation runs testing three predictions — and a measurement target for platform-integrity analysis that is distributional rather than binary. The broader yield is methodological: a thesis long dismissed as unfalsifiable now has parameters, a failure condition, and a place in the experimental-design discipline that [8] prescribes, and the same protocol-level treatment of coordination extends to any setting where shared structure grows out of repeated interaction, including the multidimensional relational data that online sharing systems accumulate [2].

Provenance and AI Involvement

- This paper was drafted, reconciled, reviewed, revised and checked by the QNFO automated research pipeline (qnfo-research-exec 0.45.4). Language models were tools in that process; they are not authors. The named author is accountable for the content.
- Models: drafting by openai/gpt-oss-120b, zai-org/glm-5.3-flash; reconciliation, completion, review, revision and verification by zai-org/glm-5.3-flash, openai/gpt-oss-120b, moonshotai/kimi-k2.6, with the QNFO model router as a fallback.
- Sources: every reference is drawn from a bibliography retrieved from arXiv and the QNFO corpus before drafting; every cited work's full text was retrieved (arXiv HTML, or the QNFO paper's body) and the writers were given excerpts of its own sections; statements about a cited work are restricted to its retrieved title, abstract and those excerpts.
- Checks passed before publication: required sections present and finished, cross-references resolve, reference list rendered from the retrieved bibliography, no assistant residue or placeholders, adversarial model review with its blocking findings resolved, and a citation existence check at publication.
- Independent recomputation: the execution sandbox was unavailable, so a second language model recomputed 15 quantitative claims step by step in text from the paper's stated inputs (no code was executed); 0 differed from the stated value at the stated precision.
- Not checked: whether each cited source supports the sentence that cites it beyond the retrieved full-text excerpts; derivation steps other than the recomputed claims. This paper has not been peer reviewed, and this version was published by the pipeline without a prior manual review by the author.

References

[1] Coordinated Behavior on Social Media in 2019 UK General Election. arXiv:2008.08370v2. https://arxiv.org/abs/2008.08370v2 [2] Multidimensional Social Network in the Social Recommender System. arXiv:1303.0093v1. https://arxiv.org/abs/1303.0093v1 [3] Construction Grammar and Artificial Intelligence. arXiv:2309.00135v1. https://arxiv.org/abs/2309.00135v1 [4] GeoAI in Social Science. arXiv:2401.05398v1. https://arxiv.org/abs/2401.05398v1 [5] Keeping it Authentic: The Social Footprint of the Trolls Network. arXiv:2409.07720v1. https://arxiv.org/abs/2409.07720v1 [6] Measuring social spam and the effect of bots on information diffusion in social media. arXiv:1708.08134v1. https://arxiv.org/abs/1708.08134v1 [7] The Social Impact of Generative AI: An Analysis on ChatGPT. arXiv:2403.04667v1. https://arxiv.org/abs/2403.04667v1 [8] The Artificial Scientist: Logicist, Emergentist, and Universalist Approaches to Artificial General Intelligence. arXiv:2110.01831v1. https://arxiv.org/abs/2110.01831v1