# Annealed Social-Hypergraph Bootstrapping for Comparative Animal Communication

Whale Conversations Project

July 3, 2026

**Abstract**

We propose a mathematical scaffold for comparative animal communication: moving from observed signals and social context to cautious, progressively grounded models of communicative function. Sperm-whale codas are the first full case study, but the method is species-general. The central change in viewpoint is that a signal should not first be modeled as a word-object label. It should be modeled as an operator on an interaction field: a typed transformation of dyads, group frames, role memberships, attention, uncertainty, spatial or ecological targets, and expected future value. The comparative claim is not that human language and whale communication share words, or that one translates directly into the other. It is that communicative systems can share coordination-value signatures: repeatable patterns of social-state update, value-gradient displacement, repair, attention, and future-oriented action. Human-language and emoji interfaces are then late projections from this operator space, constrained by structural regularities, acoustic evidence, behavioural grounding, and weak human glosses. The method combines a temporal social hypergraph, value-field-style rNPV value coordinates, derivative state fields, sheaf consistency across contexts, gauge-like role/register transformations, Koopman/Perron-Frobenius operator semantics, optimal-transport alignment to human phrases, and an annealed posterior over proto-meanings. Current whale evidence supports structural proto-meanings for contact assembly, pair identity, repair/boundary management, path/probe operations, and scalar probe ladders. Emotion is treated as value-field dynamics rather than as anthropomorphic labeling: emotion terms name partial derivatives of nested rNPV over body, world, memory, future, and social hypergraph state. Consciousness-related language is bounded to the theoretical and operational limit of this same framing: integrated self/other/world updating that compresses multiscale collaborative value into a present control state, not private subjective access. The updated pipeline also includes registered comparative readers and readiness payloads for non-whale species, plus a Homo sapiens coordination-control layer. The human layer uses DeliData group decisions, AMI/AMI-ME meeting effectiveness, CHILDES caregiver-child transcripts, public GitHub coordination threads, and aggregate emoji/emotion calibration to test whether human language is itself an operator on shared state, value pressure, repair, affiliation, and commitment, not the privileged meaning layer. These readers and controls test shared schema, social-pattern, ecological-grounding, care-context, affect-context, acoustic-context, rhythm, welfare, partner-context, and human coordination mappings, but they do not treat comparative animal data or human text as whale-English supervision or as literal translation evidence.

## 1 Motivation

Traditional language modeling begins with token prediction. That is useful, but it is the wrong primitive for a non-human social signal system when aligned semantics are sparse. The available whale corpus gives us timed codas, speakers, registers, gaps, tempo, response structure, and weak action glosses. It gives much less direct evidence for depth, prey, heading, body state, range, calf presence, or controlled intervention. A good mathematical model must therefore infer meaning as a constrained latent field, not as a dictionary.

The working interpretation is:

$$\text{coda} \approx \text{operator on social/value context},$$

and only later:

$$\text{operator} \xrightarrow{\pi_{\text{English}}} \text{English phrase}.$$

This turns translation into a projection problem under uncertainty, rather than a lookup problem.

## 2 Central Reframing: Coordination-Value Signatures

The paper's central object is not a dictionary between whale signals and human words. It is a measurable signature of how a signal changes an interaction. A coordination-value signature is the tuple

$$\Sigma_u = (\Delta H_u, \Delta\nabla\mathcal{V}_u, R_u, A_u, G_u, \kappa_u),$$

where $\Delta H_u$ is the social-hypergraph update induced by signal unit $u$, $\Delta\nabla\mathcal{V}_u$ is the value-gradient displacement, $R_u$ is the receiver-response kernel, $A_u$ is the acoustic or signal-form family, $G_u$ is the available ecological or behavioural grounding, and $\kappa_u$ records the guardrails and unresolved blockers.

This makes the comparison with human language precise. Human speech, gesture, emoji, and whale codas do not need to share the same surface forms. They can be compared when they show homologous or analogous roles in a coordination field: opening contact, holding a dyad, repairing a rupture, probing uncertainty, summoning attention, stabilizing a group room, or changing expected future value. The word or emoji shown in the interface is therefore a projection from $\Sigma_u$ into a human-readable inspection layer. It is useful when it preserves the coordination-value structure; it is misleading if it is read as a literal lexical translation.

**Principle 1** (Projection, not equivalence)**.** *Shared coordination-value signatures do not imply that whale communication is human language in another code. They imply that both systems can be studied as operators over interaction, value, attention, memory, and future action. Surface labels enter only after the operator signature is estimated and bounded.*

## 3 Language, Emotion, And Nested Collaborative Value

The same object defines language and emotion. Let the World Loom state be

$$H_t = (X_t, B_t, F_t, R_t, K_t, M_t),$$

where $X$ is world or task state, $B$ is embodied value need, $F$ is information flow, $R$ is the social hypergraph, $K$ is credit, trust, repair, status, and memory, and $M$ is shared symbolic memory. For an individual or group $i$, define nested rNPV as

$$\mathcal{V}_{i,t} = \text{rNPV}_i(H_t) = I_t\Big[V_{i,t}^{body}, V_{i,t}^{world}, V_{i,t}^{other}, V_t^{group}, K_{i,\cdot,t}, M_t, \mathbb{E}_t V_{i,t+\tau}\Big],$$

where $I_t$ is the integration operator that compresses body, action, memory, future, other, group, and symbolic context into one present value field. A signal $u$ is language-like when it acts as a controlled operator

$$L_u : H_t \to H_{t+1}$$

and the induced update predicts response, action closure, uncertainty reduction, repair, or value movement above controls:

$$\Delta H_u = H_{t+1} - H_t, \qquad \Delta \mathcal{V}_u = \mathcal{V}(H_{t+1}) - \mathcal{V}(H_t).$$

Thus language is not defined by English-like words. It is defined by reliable state update in a shared world/social/value field.

Emotion is the local geometry of the same field:

$$\text{emotion}_{i,t} = P\left(\nabla_H \mathcal{V}_{i,t}, \frac{D}{Dt}\nabla_H \mathcal{V}_{i,t}\right),$$

for a projection $P$ appropriate to the species and evidence channel. Basic affect corresponds to derivatives over safety, uncertainty, novelty, and future value. Social emotion corresponds to derivatives through $R$, $K$, and $M$: relationship edges, social credit, repair debt, recognition, status benchmarks, and shared symbolic norms.

| Emotion family | Mathematical form | Operational interpretation |
|---|---|---|
| Curiosity | $\partial\mathcal{V}/\partial F > 0$ under uncertainty | information gain has positive expected value |
| Fear/risk | downside-weighted $\partial\mathcal{V}/\partial X < 0$ | future viability is threatened |
| Jealousy/envy | $\partial\mathcal{V}_i/\partial B_{ij} < 0$ | another becomes a value or status benchmark |
| Shame | $\text{regret}(a_i)\,\partial K_i/\partial M_G < 0$ | action lowers public credit under a shared norm |
| Guilt | $\partial\mathcal{V}_j/\partial a_i < 0$ or $\partial\mathcal{V}_G/\partial a_i < 0$ | own action lowers another or group value |
| Resentment | $D_{ij,k} > 0$ after repair | a subvalue debt does not zero after nominal repair |
| Pride | $\partial K_i/\partial C_i > 0$ with recognition | capability gain is credited by others |
| Group fusion | $\text{corr}(\mathcal{V}_i, \mathcal{V}_G) \uparrow$ | self-value couples to group value |

This table is a theory of measurable operators, not a claim of private access. It is useful because it turns high-level terms such as jealousy, shame, pride, or resentment into falsifiable social-value ledger dynamics. The pipeline's current human value/affect audit tests the first derivative layer–question, repair, affiliation, commitment, response density, social breadth, symbolic diversity, interaction mass, and their mixed partials. It now also writes `analysis/human_social_emotion_derivative_audit.csv`, a 41-row social-ledger audit over information gain, downside risk, benchmark, status-credit, regret, group-value loss, residual-debt, capability-recognition, group fusion, and group-versus-individual rNPV fields. It also writes `analysis/human_whale_emotion_signature_alignment.csv`, a 9-row human–whale map that distinguishes a word/operator object from an emotion-geometry object. A word/operator object is a discrete signal, coda-family, or projected phrase $L_u : H_t \to H_{t+1}$. An emotion-geometry object is a derivative vector such as $\partial\mathcal{V}/\partial F$, $\partial\mathcal{V}/\partial\text{risk}$, or $\partial\mathcal{V}/\partial K$ on the same nested-rNPV state. Five rows are structure-preserving matches where both human and whale effects beat controls; two are candidate social-ledger matches; and two are blocked because the whale artifact set does not yet contain public-credit or recognized-capability controls. The claim is shared derivative geometry, not shared English words, private feeling, consciousness proof, or literal translation.

## 3.1 Multimodal Channel Ecology

The same distinction must also hold across sound, body, gaze, facial movement, tail or ear position, dance, touch, scent, written words, emoji, and prosody. A signal event is therefore a channel-weighted signal bundle

$$u_t = \{u^{sound}, u^{body}, u^{gaze}, u^{face}, u^{appendage}, u^{movement}, u^{lexical}, u^{tone}, u^{symbolic}, u^{social}\}_t$$

with a species-, individual-, source-, and context-dependent allocation

$$\Delta H_u = \sum_c \alpha_{s,i,k,c}\,\Phi_c(u_t^c), \qquad \sum_c \alpha_{s,i,k,c} \le 1.$$

The missing mass is not ignored: it records channels the current source cannot observe. Thus whales may be sound/timing/social-context dominant in the current repo, dogs may require body, tail, ear, gaze, action, and human-speech channels before public chat is allowed, bees may be movement/spatial dominant, and human text controls may be lexical/timing/emoji dominant while tone, face, gaze, and body are absent. This is channel-weighted operator geometry: species differ in channel ecology while the operator object and emotion-geometry object remain typed separately.

The executable witness is `analysis/multimodal_channel_operator_geometry.csv`, a 14-row audit over currently observable channel allocations. It keeps the website mapping aligned: bubble text is the guarded operator phrase, weighted word colors show the message-level observable evidence-channel mix, a small channel corner badge gives the leading percentage weight, dots show the channel mix, and emoji/mood stays inside the message as the emotion-geometry projection. Evidence metadata carries the full channel split and missing-channel mass. Thus blue sound text, green body text, or purple lexical text are not competing translations; they are the measured route by which the same typed operator may be observed. `analysis/channel_token_attribution_boundary.csv` is the companion channel-token attribution boundary: the colors are not token-level decoding of individual English words into raw channels. They are a message-level chart of the coefficients $\alpha_{s,i,k,c}$. A true token-level claim would require independent token/channel ablation or raw multimodal alignment evidence, and is blocked for the current artifacts.

**Principle 2** (Integrated collaborative value boundary)**.** *Consciousness-related language enters only as a theoretical limit of collaborative value: the integration of nested rNPV over body, world, memory, future, other, group, and social hypergraph into a present control state. The current data can test bounded signatures of such integration; it cannot certify private experience, qualia, or human-style self-report.*

## 4 Comparative Scope Across Species

The framework is intended for many communicative species, not only whales. A species enters the pipeline through an evidence adapter

$$A_s : \mathcal{D}_s \to \mathcal{E}_s,$$

where $\mathcal{D}_s$ is the raw source dataset and $\mathcal{E}_s$ is a typed event stream in a shared comparative schema. The adapter is species-specific: whales may provide codas, registers, inter-coda intervals, acoustic features, and social identities; bats may provide annotated calls, caller/receiver fields, and behavioural contexts; elephants may provide caller, receiver, addressee, and rumble acoustics; honey bees may provide dance vectors and spatial foraging targets rather than vocal tokens; mouse pups may provide ultrasonic calls aligned to caregiver distance and pup state; pigs may provide call-type, valence, context, and acoustic-feature rows. The common layer is not the signal type. The common layer is interaction:

$$e_t^s = (\text{species}, \text{signal}, \text{caller}, \text{receiver}, \text{time}, \text{context}, \text{response}, \text{grounding}, \text{guardrail}).$$

Every species must then declare a capability vector

$$\gamma_s = (\gamma_{\text{schema}}, \gamma_{\text{reader}}, \gamma_{\text{acoustic}}, \gamma_{\text{value}}, \gamma_{\text{translation}}),$$

with each coordinate marked present or absent. Present capabilities require evidence tables. Absent capabilities require a boundary reason. This prevents a common failure mode: treating a polished reader projection as if it were a validated translation system.

**Principle 3** (Species-specific evidence classes)**.** *The same mathematical scaffold can compare species, but claims do not transfer automatically across species. A bat annotation row, an elephant addressee label, a dolphin signature whistle, a marmoset turn-taking bout, and a honeybee dance vector are different evidence classes. They may share a schema and social operator vocabulary, but each needs its own acoustic, behavioural, ecological, and translation gates.*

The correct scientific question therefore becomes:

What interaction operator does this signal support in this species?

Only after that operator is stable should the project ask whether a human phrase is a useful projection.

# 5 Observed Event Stream

For a species $s$, let the observed event stream be

$$y_t^s = (w_t, u_t, r_t, q_t, \Delta_t, \rho_t, a_t, g_t),$$

where $w_t$ is the emitter, $u_t$ is the signal unit, $r_t$ is a register, call type, role, or modality tag when available, $q_t = (u_t, r_t)$ is the signal-state, $\Delta_t$ is the inter-event gap, $\rho_t$ is a tempo, frequency, duration, or other acoustic/kinematic feature, $a_t$ is a structural action or context gloss, and $g_t$ is any available grounding or annotation. For the whale case study, $u_t = c_t$ is a coda template and $q_t = (c_t, r_t)$ is the coda-register state.

In the current project corpus:

$$T = 1139, \quad |\mathcal{C}| = 23, \quad |\mathcal{S}| = 33,$$

with 89 conversations and 13 speakers. After the verified Zenodo extraction, the processed layer contains 91 acoustic feature rows and 10 annotation-linked candidate events. The strongest present evidence is still structural; the new acoustic and annotation evidence is useful for grounding pressure, but broad "clicks" labels are not literal coda translations.

# 6 Temporal Social Hypergraph

**Definition 1** (Social hypergraph)**.** *At event time $t$, let*

$$H_t = (V, E_t, \tau, \omega, \mu_t)$$

*be a typed weighted hypergraph. $V$ is the set of whale identities or provisional speaker labels. Each hyperedge $e \in E_t$ is a conversation, dyad, role frame, repair loop, contact formation, or path/probe episode. The map $\tau(e)$ gives its type, $\omega(e)$ its strength, and $\mu_t(e)$ its current role memberships.*

The hypergraph is not metadata around the language model. It is part of the semantic state. For example, the same coda in a pair-identity edge and a contact-formation edge should not be forced to share one flat meaning. It may share an operator family while acting on different local social coordinates.

Empirically, the current hypergraph pass finds:

$$\text{sameC}_{\text{SAM,TBB}} = 0.97872, \quad \text{roleSplit}_{\text{SAM,TBB}} = 0.48936,$$

and the strongest pair-identity hyperedge is conversation 18, dominated by $5R1$ at rate 0.92683. This supports treating $5R1$ as a pair-frame stabilizer rather than as a simple object label.

# 7 rNPV Value Field

Let the latent social-world state be

$$\sigma_t = (H_t, x_t, b_t, u_t),$$

where $x_t$ are acoustic/prosodic variables, $b_t$ are inferred behavioural/ecological variables, and $u_t$ is the active coda operator.

Define an rNPV-style potential:

$$\mathcal{V}_\theta(\sigma_t) = \sum_{e \in E_t} \alpha_{\tau(e)} \Psi_{\tau(e)}(\mu_t(e), x_t, b_t) + \sum_{v \in V} \Phi_v(\text{role}_t(v), \text{history}_t(v)) - \Omega(\sigma_t).$$

Here $\Psi$ scores typed social edge value, $\Phi$ scores individual role/history value, and $\Omega$ penalizes uncertainty, threat, repair burden, or contradiction.

Value-field emotion coordinates are not anthropomorphic labels. They are derivative coordinates on $\mathcal{V}_\theta$:

$$m_t = B \begin{bmatrix} \partial \mathcal{V}/\partial \text{survival} \\ \partial \mathcal{V}/\partial \text{uncertainty} \\ \partial \mathcal{V}/\partial \text{future} \\ \partial \mathcal{V}/\partial \text{benchmarkGap} \\ \partial \mathcal{V}/\partial \text{reputation} \\ \partial \mathcal{V}/\partial \text{autonomy} \\ \partial \mathcal{V}/\partial \text{fusion} \\ \partial \mathcal{V}/\partial \text{novelty} \end{bmatrix}_t .$$

In code these appear as fear-like, anxiety-like, hope-like, anger-like, curiosity-like, collaboration/fusion, and benchmark-gap coordinates. They are mathematical gradients, not claims about human emotions in whales.

This is also the boundary around consciousness language. The pipeline can test whether a signal participates in integrated self/other/world updating: whether it changes an individual's relation to a partner, a group, an uncertainty, a repair burden, a resource, or a future trajectory. It cannot, from the current evidence, read out private subjective experience. Consciousness-related claims therefore remain operational: integrated update, memory, attention, value, and counterfactual future sensitivity are measurable signatures; qualia, self-report, and human-style emotion terms are not inferred.

**Principle 4** (Operational emotion and consciousness boundary)**.** *Emotion coordinates are value-field derivatives over interaction state. Consciousness-related evidence is limited to testable signatures of integrated self/other/world updating. Neither layer licenses a claim that a whale has a particular human feeling, belief, or inner sentence.*

Current regularities match all seven tested derivative predictions: NOISE raises fear by 0.76749, NOISE raises anxiety by 0.27592, role-split raises benchmark gap by 0.60735, same-C cross replies raise collaboration by 0.33917, loss actions raise anger by 0.23637, path families raise curiosity by 0.23229, and hold actions raise hope by 0.18759.

## 8 rNPV Emotion Field As Differential Equations

The value-field interpretation is stronger than assigning emotion labels to events. Let the rNPV potential be a time-dependent field

$$\mathcal{V} = \mathcal{V}(t, S, E, U, B, R, G, C, M, T, \Phi, H),$$

where $S$ is survival/safety, $E$ expected future value, $U$ uncertainty or variance, $B$ benchmark or role gap, $R$ reputation/status, $G$ group welfare, $C$ collaboration/fusion, $M$ moral or repair cost, $T$ transcendent continuity, $\Phi$ flow/novelty, and $H$ the social hypergraph. The total derivative along a whale conversation trajectory is

$$\frac{D\mathcal{V}}{Dt} = \frac{\partial \mathcal{V}}{\partial t} + \nabla_x \mathcal{V} \cdot \dot{x}_t + \left\langle \frac{\delta \mathcal{V}}{\delta H}, \dot{H}_t \right\rangle, \qquad x_t = (S, E, U, B, R, G, C, M, T, \Phi)_t.$$

Thus a coda is not merely associated with an emotion; it perturbs the social-value field and changes the directional derivatives of rNPV.

In this framing, emotion coordinates are projections of the value-gradient and its local temporal movement:

$$m_t = B_1 \nabla_x \mathcal{V}(t, x_t, H_t) + B_2 \frac{D}{Dt} \nabla_x \mathcal{V}(t, x_t, H_t) + B_3 \dot{H}_t.$$

The implementation uses finite differences against each speaker's exponential moving baseline, but the limiting object is this differential field. The axes used in the code correspond to signed partial derivatives:

$$\begin{aligned}
\text{hope}_t &\approx \left[\frac{\partial \mathcal{V}}{\partial E} \dot{E}_t\right]_+, \\
\text{anxiety}_t &\approx \left[-\frac{\partial \mathcal{V}}{\partial U} \dot{U}_t\right]_+ \quad \text{or high local variance}, \\
\text{fear}_t &\approx \left[-\frac{\partial \mathcal{V}}{\partial S} \dot{S}_t\right]_+ \quad \text{plus a low-safety barrier}, \\
\text{anger/friction}_t &\approx \left[-\frac{\partial \mathcal{V}}{\partial C} \dot{C}_t\right]_+, \\
\text{role tension}_t &\approx \left[\frac{\partial \mathcal{V}}{\partial B} \dot{B}_t\right]_+, \\
\text{curiosity}_t &\approx \left[\frac{\partial \mathcal{V}}{\partial \Phi} \dot{\Phi}_t\right]_+.
\end{aligned}$$

Here $[z]_+ = \max(z, 0)$. Collaboration and expectation are therefore not side variables. They are explicit coordinates of the rNPV field. A same-coda cross-speaker reply can raise $C$ and lower uncertainty; a loss/reset coda can lower $E$ and $C$ while raising $U$; a role-split exchange can increase $B$ while preserving a shared dyadic frame. These partial derivatives are what the web interface later renders as small emotion/social/contract chips.

The social hypergraph contributes an additional partial derivative:

$$\partial_H \mathcal{V}[e] = \frac{\delta \mathcal{V}}{\delta H(e)},$$

so that a signal can be evaluated by how it changes the value of a pair edge, group-room edge, repair edge, or path/probe edge. This is the formal link between rNPV, emotion, collaboration, expectation, and the social hypergraph.

## 9 Meaning As Operator, Not Label

**Definition 2** (Signal operator)**.** *Each signal unit or signal-state $u$ denotes a stochastic operator*

$$K_u : \mathcal{P}(\Sigma) \to \mathcal{P}(\Sigma),$$

*mapping a distribution over latent social-world states before the signal to a distribution after the signal:*

$$p(\sigma_{t+1} \mid y_{\leq t}) = K_{u_t} p(\sigma_t \mid y_{<t}).$$

**Definition 3** (Proto-meaning)**.** *The proto-meaning of $u$ is the equivalence class*

$$\mathcal{M}(u) = \left[K_u,\ \Delta\nabla\mathcal{V}_u,\ R_u,\ A_u,\ G_u\right],$$

*where $\Delta\nabla\mathcal{V}_u$ is the value-gradient displacement caused by $u$, $R_u$ is the receiver-response kernel, $A_u$ is the acoustic signature family, and $G_u$ is the ecological/behavioural grounding distribution.*

This immediately explains why the current whale labels should be phrases like "contact assembly", "pair identity", "repair/boundary operator", and "scalar probe". For other species the phrase inventory may differ: elephants may emphasize addressee and identity operators, marmosets turn-taking and coordination operators, and honeybees spatial recruitment operators. In every case the phrase describes a transformation in the interaction field, not a direct English noun.

## 10 Sheaf Consistency Across Contexts

Let $\mathcal{U} = \{U_i\}$ be a cover of the corpus by local contexts: dyads, speaker repertoires, conversation starts, repair loops, scalar probe episodes, contact-formation runs, and path-family episodes. In each context $U_i$, the unit $u$ has a local semantic section

$$\phi_i(u) \in F(U_i),$$

where $F$ is a semantic sheaf over the hypergraph.

Meaning is stable when local sections glue:

$$\delta\phi(u)_{ij} = \phi_i(u)|_{U_i \cap U_j} - \phi_j(u)|_{U_i \cap U_j}$$

is small over overlapping contexts. The sheaf energy is

$$E_{\text{sheaf}}(u) = \sum_{(i,j)} \left\| \phi_i(u)|_{U_i \cap U_j} - \phi_j(u)|_{U_i \cap U_j} \right\|^2.$$

Units with low sheaf energy are good candidates for stable proto-meanings. Units with high sheaf energy are polysemous, context-bound, or under-grounded.

Role splits can be represented as non-zero 1-cocycles. For example, $5R1|a$ and $5R1|i$ may share the same base pair-identity operator while carrying different local role orientations. That is more expressive than treating register variants as independent words.

## 11 Annealed Translation Bootstrap

Translation should be annealed. Early in training, the model trusts structural regularities because those are dense. Later, as audio, identity, ecology, and intervention data arrive, the model increases acoustic and grounded constraints. Human glosses are boundary conditions, not truth.

Let $z_u$ be the latent proto-meaning of unit $u$, and let $e_u$ be an English phrase candidate. Define the energy

$$\begin{aligned} E_T(z,e) =& \lambda_s(T) D_{\text{struct}}(z;y) + \lambda_h(T) D_{\text{hyper}}(z;H) + \lambda_v(T) D_{\text{value}}(z;\nabla\mathcal{V}) \\ &+ \lambda_a(T) D_{\text{acoustic}}(z;x) + \lambda_g(T) D_{\text{ground}}(z;b,g) + \lambda_e(T) D_{\text{English}}(e;\pi(z)). \end{aligned}$$

The posterior is

$$q_T(z,e \mid y,H,x,g) \propto \exp\left(-E_T(z,e)/T\right).$$

At high temperature, $q_T$ explores broad operator families. At lower temperature, it sharpens only when independent channels agree. A useful schedule is:

$$\lambda_s,\lambda_h,\lambda_v \text{ high early}; \qquad \lambda_a,\lambda_g,\lambda_e \text{ increase only with evidence.}$$

**Principle 5** (No premature literalization)**.** *An English phrase is not a translation until the posterior is stable under structural, acoustic, and grounded perturbations:*

$$\mathrm{Var}_{T\downarrow 0}[\pi_{\text{English}}(z_u)] \approx 0$$

*and acoustic/grounded confidence are both high.*

This is exactly why the current reports separate

$$(C_{\text{struct}}, C_{\text{acoustic}}, C_{\text{grounded}}).$$

For example, $1+1+3$ and $5R1$ currently have structural confidence near 0.68, while grounded confidence is only 0.18 after adding broad Zenodo click annotations. They are good modeling targets, not literal translations.

## 12 Harmonic Extension From Weak Glosses

Let $G_L$ be a lexicon graph whose nodes are coda units, coda-register states, acoustic prototypes, social hyperedge types, and English gloss candidates. Edges encode distributional similarity, response similarity, value-gradient similarity, acoustic similarity, and weak gloss links.

Weak translations seed boundary values $f_B$ on a small subset of nodes. Unknown meanings are inferred by minimizing a hypergraph Dirichlet energy:

$$\min_f \sum_{(i,j)\in E_L} w_{ij}\|f_i - f_j\|^2 + \eta \sum_{i\in B} \|f_i - f_{B,i}\|^2 + \gamma E_{\text{sheaf}}(f).$$

This is harmonic extension with sheaf regularization. It lets glosses bootstrap unknowns, but only through the geometry of the corpus. If a gloss conflicts with social-role, response, or value-gradient evidence, the energy exposes that conflict instead of silently copying the gloss.

## 13 Conversation As Semantic Context

The richest available context is not a fixed token window. It is the whole conversation as a structured social event. Let each conversation $C_j$ define a local semantic chart with features

$$\chi_j(u) = [\text{position}, \text{edgeType}, \text{speakerRole}, \text{replyKernel}, \text{runStructure}, \nabla\mathcal{V}, \text{repair/reset}]_{u,C_j}.$$

The global proto-meaning of $u$ is then a gluing problem over all conversations in which $u$ appears:

$$\mathcal{M}_{\text{conv}}(u) = \arg\min_m \sum_{j:u\in C_j} d(m, \chi_j(u))^2 + \lambda E_{\text{sheaf}}(u).$$

This is stronger than local distributional semantics. A coda is not defined by nearby codas alone, but by the stable operation it performs over many social episodes.

In the current corpus, this conversation-context bootstrap gives high structural-context confidence to $1+1+3$ as a contact-field stabilizer across 59 conversations, $5R1$ as a dyadic identity and role-alignment operator across 23 conversations, $5R2$ as within-speaker frame maintenance, and NOISE forms as repair/boundary regulators. These are still proto-meanings, but they are better proto-meanings because they are learned from the whole social manifold of conversations.

**Principle 6** (Conversation before dictionary)**.** *Before asking for an English translation of $u$, first learn the function*

$$u \mapsto \{\chi_j(u)\}_{j:u\in C_j}.$$

*Only after the conversation-role distribution is stable should acoustic, ecological, and English constraints be annealed in.*

## 14 Hypergraph Differential Calculus

To make the social scaffold operational, the hypergraph should have a calculus. Let $B_t$ be the vertex-hyperedge incidence operator,

$$(B_t)_{ve} = \begin{cases} \mu_t(v,e) & v \in e, \\ 0 & v \notin e, \end{cases}$$

and let $W_t$ be a diagonal matrix of hyperedge weights. A typed hypergraph Laplacian is

$$L_t = B_t W_t B_t^\top + \sum_k \beta_k L_t^{(k)},$$

where $L_t^{(k)}$ are type-specific Laplacians for pair identity, contact formation, repair, path/probe, and role-split edges.

A coda operator $u$ then induces a vector field on the social state:

$$\dot{h}_t = F_u(h_t, x_t, b_t) + \epsilon_t, \qquad h_t = \text{vec}(H_t).$$

The structural meaning of $u$ can be estimated by the displacement it produces in low-energy modes of $L_t$:

$$\Delta_u^{(\ell)} = \langle \varphi_\ell(L_{t+1}) - \varphi_\ell(L_t),\ \varphi_\ell(L_t) \rangle,$$

where $\varphi_\ell$ is a hypergraph eigenmode. This gives a non-lexical semantics: a coda can be a stabilizer of a dyad mode, an injector of contact energy, a repair dissipator, or a path/probe perturbation.

## 15 Gauge View Of Register And Role

Registers such as $|a$ and $|i$ should not be treated merely as suffixes. They can be modeled as local gauge choices on the social bundle. Let $P \to H$ be a principal bundle whose fiber at a hyperedge is the set of role frames available to the participants. A coda-register state $s = (c, r)$ acts by

$$K_{(c,r)} = G_r^{-1} K_c G_r,$$

where $G_r$ changes the local role frame. Two registers are semantically equivalent when they are gauge transforms preserving observables:

$$\mathcal{O}(K_{(c,a)}) = \mathcal{O}(K_{(c,i)}),$$

but role-bearing when their connection curvature is non-zero:

$$\mathcal{F} = dA + A \wedge A \neq 0.$$

In this view $5R1|a$ and $5R1|i$ can share a pair-identity base operator while carrying different speaker/receiver role orientations. Role-split exchanges become curvature events, not separate dictionary entries.

## 16 Koopman Semantics

The phrase "meaning is an operator" can be made precise with Koopman and Perron-Frobenius operators. Let $f : \Sigma \to \mathbb{R}$ be an observable: future same-C response, repair success, collaboration derivative, depth change, click-band energy, or English gloss compatibility. A coda $u$ induces a Koopman operator

$$(\mathcal{K}_u f)(\sigma) = \mathbb{E}[f(\sigma_{t+1}) \mid \sigma_t = \sigma, u_t = u],$$

and a dual Perron-Frobenius operator $\mathcal{P}_u$ moving probability mass over social-world states.

**Definition 4** (Operational meaning spectrum)**.** *The operational spectrum of a coda is*

$$\Lambda(u) = \mathrm{spec}(\mathcal{K}_u|_{\mathcal{F}_{\mathrm{social}}}),$$

*restricted to observables over social response, value derivatives, acoustic families, and ecological covariates.*

Stable social operators should have persistent spectral signatures across contexts. Repair operators should damp unstable modes; identity operators should preserve dyad modes; probe operators should move mass toward high-novelty or high-uncertainty regions. This is a stronger target than next-token prediction because it asks what invariant dynamics the token induces.

## 17 Optimal Transport To Human Phrases

Human-readable phrasing should be a late-stage barycentric projection from species-specific operator space, not the training primitive. Let $\mathcal{Z}_s$ be the space of proto-meanings for species $s$ and $\mathcal{E}$ a human phrase manifold. Define a cost

$$c(z, e) = \alpha c_{\mathrm{role}}(z, e) + \beta c_{\mathrm{value}}(z, e) + \chi c_{\mathrm{ground}}(z, e) + \delta c_{\mathrm{pragmatic}}(z, e).$$

The translation plan is an entropic optimal transport coupling

$$\Pi_T^\star = \arg \min_{\Pi \in U(p_Z, p_E)} \langle C, \Pi \rangle T \, \mathrm{KL}(\Pi \| p_Z p_E).$$

At high temperature, English labels remain diffuse. At low temperature, a phrase becomes defensible only if its coupling concentrates without violating sheaf, acoustic, and grounding constraints.

**Proposition 1** (Translation as low-temperature barycenter)**.** *For a unit $u$, define*

$$\hat{e}_u(T) = \mathrm{bar}_{\mathcal{E}} \left( \Pi_T^\star(\cdot \mid z_u) \right).$$

*The phrase $\hat{e}_u$ is a literal translation only if the barycenter is stable under temperature lowering and under removal of any one evidence channel:*

$$d_{\mathcal{E}}(\hat{e}_u(T), \hat{e}_u(T/2)) < \varepsilon \quad \textit{and} \quad \max_j d_{\mathcal{E}}(\hat{e}_u, \hat{e}_u^{(-j)}) < \varepsilon.$$

This prevents the common failure mode where a plausible English gloss gets copied through the model and mistaken for field evidence.

# 18 Thermodynamic And Renormalization View

The bootstrap can also be read as simulated annealing over a rugged semantic landscape. Let

$$Z(T) = \sum_{z,e} \exp(-E_T(z,e)/T), \qquad \mathcal{F}(T) = -T \log Z(T).$$

As $T$ decreases, proto-meanings condense from broad operator families to sharper social roles. Phase transitions are scientifically useful: if $5R1$ remains stable across a wide temperature interval while $8i$ splits into multiple local modes, the model has learned that pair identity is more globally stable than scalar probing.

There is a second scale parameter $\ell$ for temporal/social resolution:

$$\mathcal{R}_\ell : \{y_t\}_{t=1}^T \mapsto \{\text{runs, motifs, dyads, conversations, clans}\}.$$

A meaning is robust if it is invariant under coarse-graining:

$$\mathcal{M}_\ell(u) \approx \mathcal{M}_{\ell+1}(u).$$

This gives a renormalization criterion for culture: a coda pattern that survives from events to runs to dyads to groups is more likely to be a cultural operator than a local acoustic accident.

# 19 Active Evidence And Experiment Design

The framework also tells us what data to seek next. Let $\mathcal{A}$ be possible data acquisitions: more deployments, depth alignment, photo-ID, prey field, calf presence, controlled playback, or better annotation. The value of an acquisition is expected free-energy reduction:

$$\mathrm{EVI}(a) = \mathbb{E}_{o \sim p(o|a)} \left[ \mathcal{F}(q) - \mathcal{F}(q \mid o, a) \right].$$

The best next dataset is not the largest one. It is the one that most reduces uncertainty in the weakest channel of the annealing objective. At present, structural evidence is dense, acoustic evidence is improving, and ecological/behavioural grounding is the bottleneck.

## 20 Algorithmic Sketch

1. Build a species-specific event stream $y_t^s$ and temporal hypergraph $H_t^s$ from emitters, signals, modalities/registers, gaps, responses, and interaction context.

2. Estimate value-field value coordinates $m_t$ as derivatives of $\mathcal{V}_\theta(\sigma_t)$.

3. Learn signal operators $\mathcal{K}_u$ over social, acoustic, spatial, behavioural, and value observables.

4. Compute sheaf inconsistency and gauge curvature to separate stable meanings from role/register variants.

5. Add weak human glosses or behavioural labels as boundary conditions in the lexicon/evidence graph, never as hard labels.

6. Anneal the posterior $q_T(z, e)$ while increasing acoustic and grounded weights only when evidence exists.

7. Project to human-readable phrases by low-temperature optimal transport and reject any phrase that fails channel-ablation stability.

## 21 Detailed Methods

### 21.1 Data Sources And Storage Layout

The project uses a two-tier data layout. Small reproducible artifacts are committed in the repository under `analysis/`, `paper/`, `scripts/`, and `web/`. Large public source data and generated comparative exports are staged on an external SSD:

`/Volumes/Extreme SSD/whale-language-data`.

The pipeline contract is:

$$\text{SSD raw data} \to \texttt{src/whale\_language\_analysis.py} \to \texttt{analysis/*.csv}$$
$$\to \texttt{web/build\_site.py} \to \texttt{web/data.json} \to \text{static reader.}$$

Hosting is deliberately outside this contract. Cloudflare Pages distributes the static `web/` directory, but it does not ingest data, rebuild analysis tables, or validate the scientific claims.

### 21.2 Species Adapter Protocol

Every new animal source is added through the same adapter protocol:

$$\mathcal{D}_s \to \text{raw/metadata staging} \to \text{normalized events}$$
$$\to \text{species registry} \to \text{reader projection} \to \text{capability gates.}$$

The adapter must write a source manifest, an event table, a schema-coverage manifest, summary tables, and model-ready comparative exports. The registry entry must state raw SSD roots, metadata roots, analysis tables, web data key, reader contract, and math-capability boundary. A species may be excellent for one capability and weak for another: honeybees can be strong for spatial grounding without being vocal; elephants can be strong for addressee tests without having whale-style coda sequences; marmosets can be strong for turn-taking rhythm while lacking broad ecological grounding.

The required comparative fields are:

$$\text{species}, \text{dataset}, \text{recording}, \text{event}, \text{start}, \text{end}, \text{caller}, \text{receiver},$$
$$\text{signal}, \text{features}, \text{sequenceContext}, \text{behaviourContext}, \text{ecologicalContext},$$
$$\text{response}, \text{latency}, \text{interactionType}, \text{groundingConfidence}, \text{guardrail}.$$

Species-specific fields can be added, but the shared fields are what make cross-species comparison possible.

The whale core begins with `whale_conversation_translation_side_by_side.pdf`. The extraction stage writes a text layer, then normalizes each observed coda event into

$$y_t = (w_t, c_t, r_t, s_t, \Delta_t, \rho_t, a_t, g_t),$$

where $w_t$ is speaker identity, $c_t$ is coda family, $r_t$ is register, $s_t = (c_t, r_t)$ is the coda-register state, $\Delta_t$ is the inter-event gap, $\rho_t$ is tempo or ICI information when available, $a_t$ is a structural action code, and $g_t$ is any external grounding or annotation. The committed whale event table contains 1,139 events in 89 original conversations.

External whale enrichment data are read only when the SSD is mounted and the pipeline is run with `--external-data-root`. Those runs add source inventories, acoustic feature rows, alignment summaries, contamination audits, and DSWP-derived probable conversation windows. The no-SSD core run remains available for reproducibility and separates the paper's mathematical scaffold from local mounted data availability.

### 21.3 Event Normalization

Each raw whale event is parsed into canonical fields before any semantic projection. Register suffixes are split from coda templates, speaker labels are mapped to stable participant identifiers, gaps are binned into local response classes, and coda families are assigned action priors such as contact, hold, repair, path/probe, loss/reset, or boundary. No English phrase is introduced at this stage. The normalized event table is the only source for later structural estimators, so every downstream row can be traced back to an event index, conversation index, speaker, coda, register, and local context.

For a conversation $C_j$ with ordered events $t \in C_j$, the implementation stores both event-local features and conversation-local response features:

$$\text{cross}_t = \mathbf{1}[w_t \neq w_{t-1}], \qquad \text{sameC}_t = \mathbf{1}[c_t = c_{t-1}],$$
$$\text{roleSplit}_t = \mathbf{1}[c_t = c_{t-1},\, r_t \neq r_{t-1}].$$

These variables are not treated as translations. They are structural observables used to estimate whether a coda acts as a dyadic stabilizer, a repair/boundary operator, a public-room/contact operator, or a probe/path operator.

### 21.4 Whale Structural And Value-Field Estimators

The social hypergraph estimator groups events into typed conversation and dyad edges. For each edge $e$, participant membership, dominant coda family, reply pattern, same-coda response rate, role/register asymmetry, and value-field coordinates determine a type

$$\tau(e) \in \{\text{pairIdentity}, \text{contactFormation}, \text{portalPath}, \text{repairBoundary}, \text{soloFrame}\}.$$

The typed edge table is then used as the social state for operator inference.

The value-field estimator computes a deterministic feature vector

$$\xi_t \in [0,1]^{10}$$

for survival/safety, future value, uncertainty, benchmark gap, reputation, group welfare, collaboration, moral or repair cost, transcendent continuity, and flow/novelty. A scalar rNPV proxy $V_t$ is a fixed weighted projection of $\xi_t$, while emotion-like axes are finite differences against each speaker's exponential moving baseline:

$$\bar{\xi}_{w,t} = (1-\alpha)\bar{\xi}_{w,t-1} + \alpha\xi_t, \qquad \alpha = 0.18,$$

$$d_t = \xi_t - \bar{\xi}_{w,t-1}.$$

Fear-like, anxiety-like, hope-like, anger/friction-like, collaboration/fusion-like, role-tension-like, and curiosity-like scores are functions of $d_t$, local variance, social response, and hypergraph context. They are derivative coordinates of the implemented value field, not claims that human emotions have been read from whale calls.

The contract ledger projects reply history, hyperedge type, value derivatives, and complex-emotion axes into six pragmatic coordinates:

$$\kappa_t = [\kappa^T, \kappa^R, \kappa^O, \kappa^K, \kappa^B, \kappa^P]_t,$$

standing for trust credit, repair debt, role obligation, common knowledge, private bond, and public room. The primary contract state is the largest coordinate after clipping. These axes are used for guarded reader summaries and for coda-level operator aggregation.

## 21.5 Coda-Level Operator Aggregation

For each coda or coda-register state $u$, the pipeline aggregates event-level evidence over all occurrences:

$$\bar{m}_u = \mathbb{E}[m_t \mid u_t = u], \qquad \bar{\kappa}_u = \mathbb{E}[\kappa_t \mid u_t = u], \qquad \bar{R}_u = \mathbb{E}[R_t \mid u_t = u],$$

where $R_t$ includes receiver response, same-coda reply, role split, repair recovery, and conversation-position observables. These aggregates form the operator semantic basis. A proposed English phrase is attached only after this operator basis exists, and only as a bounded phrase such as "contact-field stabilizer" or "repair/boundary operator." The paper does not claim a literal dictionary unless the structural, acoustic, and grounded channels all pass promotion gates.

## 21.6 Annealing And Promotion Gates

Every projected phrase carries three confidence channels:

$$(C_{\text{struct}}, C_{\text{acoustic}}, C_{\text{grounded}}).$$

$C_{\text{struct}}$ is computed from conversation structure, response regularity, hypergraph stability, coda-level aggregation, and value/contract regularities. $C_{\text{acoustic}}$ is allowed to rise only from acoustic feature or alignment evidence. $C_{\text{grounded}}$ is allowed to rise only from ecological, behavioural, identity, or intervention-like evidence. The current promotion rule is:

$$C_{\text{struct}} \geq 0.70, \qquad C_{\text{acoustic}} \geq 0.55, \qquad C_{\text{grounded}} \geq 0.55.$$

If any channel remains below threshold, the English string is kept as a structural or operator phrase. This is the operational version of the no-premature-literalization rule.

## 21.7 Comparative Species And Human-Control Layers

The repository now contains 22 registered entries including the sperm-whale case study, promoted event readers, source-intake candidates, and Homo sapiens comparative controls. The comparative source data are staged under the external SSD root

`/Volumes/Extreme SSD/whale-language-data`

and normalized into `analysis/<species>_*.csv` tables plus `analysis/model_ready/comparative/` exports. Each species has a row in `data/comparative_species_registry.json`; the registry is the authority for which mathematical capabilities are present and which remain blocked.

| Species | Primary evidence | Current reader artifact | Claim boundary |
|---|---|---|---|
| Egyptian fruit bat | 91,080 annotated social-call rows | 300 web threads | annotation reader only |
| African elephant | 469 caller/receiver/context rows plus acoustic features | 80 web threads | no literal name translation |
| Bottlenose dolphin | 344 acoustic windows and 384 reader turns | compact boundary payload | no identity/addressee labels |
| Common marmoset | 871,045 MarmAudio calls plus 28,698 kinship calls | compact boundary payload | no receiver-labelled turn-taking |
| Homo sapiens | DeliData, AMI/AMI-ME, CHILDES, and public GitHub coordination | 6 comparative control threads | no animal-translation supervision |

The comparative readers share the whale interface discipline but not the whale translation claim. Bats contribute emitter/addressee/context/action evidence; elephants contribute directed caller/receiver/addressee and social-pattern evidence; dolphins contribute whistle-contour and timing evidence; marmosets contribute call-type rhythm, adjacent-call reconstruction, and caller-identity timing evidence. Humans contribute a control layer, not a template: DeliData tests group-decision movement, AMI/AMI-ME tests meeting-objective coordination, CHILDES tests caregiver-child developmental turn structure, and GitHub tests public task closure. The human triad currently contains 92,452 derived turns, 9,963 coordination windows, shared viability 0.540, and trans-individual excess 0.381. A second human value/affect signature audit gives 60 rows over DeliData, AMI/AMI-ME, CHILDES, and the combined triad: 24 first-order partial derivatives, 18 mixed partial derivatives, 14 triad aggregate partials, 3 source summaries, and 1 triad summary. These rows estimate operational rNPV/value movement over question pressure, repair pressure, affiliative agreement, commitment/decision, response density, social breadth, annotation diversity, interaction mass, and mixed coordination axes. The triad summary has response density 0.769, observed value-delta rate 0.861, random-label null rate 0.382, and effect above null 0.479. A 41-row human social-emotion derivative audit adds operational ledger rows for curiosity, fear/risk, jealousy/envy, shame, guilt, resentment, pride, status benchmarking, group fusion, and individual-group rNPV conflict; the combined triad summary has mean social-ledger score 0.137 and effect above null 1.130. The cross-species Sigma audit then checks six shared components against the animal comparative emotion/value proxy tables: $\Delta H$, $\Delta\nabla\mathcal{V}$, $R$, $A$, $G$, and $\kappa$. The new 9-row human–whale emotion-signature alignment makes the same boundary more explicit: information gain/curiosity, downside risk/fear, repair-boundary pressure, affiliation/social trust, and public group coordination are structure-preserving derivative matches; benchmark role asymmetry and residual repair debt are candidate social-ledger matches; public credit/shame and recognized capability/pride are blocked overclaims. It maps emotion-geometry objects, not word/operator objects. Raw human text remains on the SSD. Private emotion and consciousness are not inferred, the human row is labelled comparative-control-only, and none of these rows are allowed to supervise whale translation examples or raise literal English confidence.

A separate multimodal channel audit now checks the channel ecology itself. The 14-row `analysis/multimodal_channel_operator_geometry.csv` table records whether a source is sound-dominant, movement/spatial-dominant, text-symbolic, or partial because important body, tone, action, receiver, or addressee channels are missing. This prevents a reviewer-facing confusion: an animal may share an operator or emotion geometry with another species without sharing the same expressive channel.

### 21.8 Social-Pattern And Reader Projection

For species with repeated individuals or addressees, the adapter computes a guarded social-pattern layer. For bats and elephants this has three scopes:

$$\text{individual node}, \qquad \text{directed edge}, \qquad \text{recording group.}$$

The implementation records support count, partner count, context lift, role asymmetry, bridge score, and a reader-facing effect. The elephant table `elephant_social_hypergraph_insights.csv`, for example, contains 619 rows: 151 individual nodes, 218 directed edges, and 250 recording-level group moments. These rows are used to color the reader with phrases such as "steady care presence," "familiar contact pattern," or "contact-focused group moment." The public interface does not expose raw social-hypergraph terminology.

For a window $W$, the comparative reader uses a conversation-likeness score

$$\begin{aligned}\text{score}_s(W) =& a_1\text{directed}(W) + a_2\text{turnChanges}(W) + a_3\text{distinctCallers}(W)\\ &+ a_4\text{contextDiversity}(W) + a_5\text{motifDiversity}(W) + a_6\text{socialSupport}(W),\end{aligned}$$

where unsupported terms are set to zero for that species. The selected windows are projected to `web/data.json` with scene summary, mood/emoji state, friendly bubble text, evidence metadata, and a guardrail. The projection is guarded: it is a human-readable map of supported distinctions, not a word-for-word translation.

### 21.9 Human-Language Mapping For Comparative Readers

The updated reader uses a four-stage human-language map:

$$\text{evidence} \to \text{communicative act} \to \text{affective/social stance} \to \text{English phrase.}$$

This is intentionally weaker than translation. The act inventory is species-specific. For bats, annotated context and movement map to boundary, roost, feeding, contact, distance, and arrival phrases. For elephants, caller/receiver context and repeated social patterns map to care, greeting, reunion, separation repair, group gathering, and familiar pair contact. For dolphins, whistle contour and local timing map to contact reach, quick check, softening drop, repair, hold, and close. For marmosets, call type and sequence position map to long contact, small check, edge cue, tempo lift, young-call focus, and guarded rhythm.

Formally, for each comparative species $s$ the renderer computes

$$\pi_s(e_t) = \rho_s(\alpha_s(e_t),\, \eta_s(e_t),\, \chi_s(e_t),\, G_s(e_t))\,,$$

where $\alpha_s$ is the communicative act class, $\eta_s$ is the mood or value-proxy coordinate, $\chi_s$ is the social-pattern context, and $G_s$ is the guardrail. The literal translation confidence for comparative species is fixed at

$$C^s_{\text{literal}} = 0$$

unless a species-specific promotion gate is added and passed. The English phrase is therefore an empathy-oriented inspection surface for researchers and conservation communication, not a dictionary entry.

### 21.10 Website Projection And Layout Controls

The website is a generated static reader with a species-selection entry page. The public conversation entry currently exposes whales, Egyptian fruit bats, and African elephants. Additional comparative species remain in `web/data.json` as bounded reader or source-readiness payloads until their species-specific chat gates pass. The promoted pages use the same reader layout: left list pane, filters/search, right chat pane, mood chips, evidence notes, and a return control for moving back to species selection. The shared layout is part of the method because it prevents apparent quality differences from being introduced by unrelated interface changes.

The chat pane also carries a small channel key. Weighted word colors show the message-level source-channel mix–for example sound, body/posture, face/gaze, movement, tone, words/text, or emoji/symbol–and not token-level decoding, while a small channel corner badge shows the leading percentage weight and dots show the channel mix. Emoji/mood stays inside the message as the emotion-geometry projection, preserving the feel of a real chat while keeping the mathematical boundary visible. This is the interface version of the paper's main distinction: a word-like object is the operator phrase $L_u : H_t \to H_{t+1}$; an emotion-like object is a derivative coordinate $\nabla_H \mathrm{rNPV}$; and the species channel weights describe how the source lets us observe the operator.

The web renderer is tested for two classes of errors. First, it must not expose raw normalized row syntax as chat text. Second, it must show species-specific guardrails. For comparative readers, the bubble body should be a friendly interaction phrase, while evidence notes carry the source fields, acoustic boundary, identity/addressee boundary, or nonliteral status.

### 21.11 Validation, Regression Tests, And Failure Conditions

The main project gate is:

`npm run check:pipeline`.

This command validates the SSD-to-analysis-to-web contract, the comparative species registry, the reader projection layer, and the comparative maths boundary. It fails if required SSD roots or analysis tables are missing; if a species appears on the website without registry declaration; if normalized rows reach the chat reader without readable projection and evidence metadata; if comparative rows leak into whale model-ready examples; if comparative annotations can raise literal translation confidence; or if a non-whale species silently claims whale-level acoustic, value-field, or translation capability.

The public website deployment witness is:

`npm run deploy:web` → `npm run verify:public`.

It verifies the only public website, the canonical Cloudflare Pages URL `https://animal-chats.pages.dev`. The verification checks that the shared reader loads, that whale, bat, and elephant pages use the same layout class, that bubbles contain mood state and conversational projection, that the Species return control exists, and that raw row strings such as `caller:` or `present -> stay` do not appear as primary bubble text.

## 22 Current Whale Findings In This Framework

| Unit | Operator reading | Current evidence | Literal status |
|---|---|---|---|
| $1 + 1 + 3$ | contact assembly/greeting alignment | structural strong | not literal |
| $5R1$ | pair identity/social memory | structural strong | not literal |
| $5R1\|i$ | marked answering role | dyad/role-split | not literal |
| 6-NOISE | repair/boundary management | value derivative | not literal |
| $8i, 9i, 10i, 7i$ | scalar probe/path ladder | transition asymmetry | not literal |
| $6i$ | in-frame try/adjust insert | local context | not literal |
| $5R2$ | private/local frame maintenance | speaker/run structure | not literal |

The strongest cultural claim is that the language appears organized around relationship frames. Codas maintain, open, perturb, repair, split, and close social fields. A future full language model should therefore learn not only token sequences, but transformations of social state.

## 23 Training Program

The next model should have four coupled layers:

$$\text{audio field} \to \text{event operators} \to \text{social hypergraph dynamics} \to \text{English projection}.$$

The training losses should include:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{nextEvent}} &+ \mathcal{L}_{\text{receiverResponse}} + \mathcal{L}_{\text{hyperedgeUpdate}} + \mathcal{L}_{\text{valueGradient}} \\ &+ \mathcal{L}_{\text{acousticContrast}} + \mathcal{L}_{\text{grounding}} + \mathcal{L}_{\text{sheaf}} + \mathcal{L}_{\text{gauge}} \\ &+ \mathcal{L}_{\text{OT}} + \mathcal{L}_{\text{EnglishProjection}}.\end{aligned}$$

Only the final term should directly care about human-readable wording, and it should be low-weight until the underlying operator is stable. The sheaf term penalizes inconsistent local meanings, the gauge term separates role/register transforms from new lexical roots, and the optimal-transport term keeps phrase projection diffuse until the species-side operator geometry has cooled.

The confidence columns in the current reports can be treated as observable annealing coordinates:

$$\lambda_s \propto C_{\text{struct}}, \qquad \lambda_a \propto C_{\text{acoustic}}, \qquad \lambda_g \propto C_{\text{grounded}}.$$

The current state is therefore an early-to-mid annealing regime: $C_{\text{struct}}$ is meaningful for several units, $C_{\text{acoustic}}$ is now supported by 91 decoded feature rows, and $C_{\text{grounded}}$ remains low because annotation labels are broad clicks rather than behaviourally rich events.

## 24 Implemented Mathematical Workflow

The CSV files are not the method; they are snapshots of estimators in the mathematical workflow. For each event $t$, the code first constructs a structural feature vector

$$\xi_t = [\xi^S, \xi^F, \xi^U, \xi^B, \xi^R, \xi^G, \xi^C, \xi^M, \xi^T, \xi^\Phi]_t \in [0,1]^{10}.$$

The coordinates denote survival, future value, uncertainty/variance, benchmark gap, reputation, group welfare, collaboration, moral cost, transcendent continuity, and flow. These coordinates are deterministic functions of coda family, action code, register, gap class, tempo, previous event, and next event. For example, same-coda cross-speaker replies increase collaboration and transcendent continuity; NOISE, portal, and loss actions increase variance/threat; role-split actions and $|i$ registers increase benchmark-gap coordinates.

The scalar rNPV proxy written to `value_field_rnpv_emotions.csv` is the implemented linear potential

$$\begin{aligned} V_t = {} & 2.0\xi_{t,\text{survival}} + 1.5\xi_{t,\text{future}} - 1.2\xi_{t,\text{variance}} - 0.6\xi_{t,\text{benchmarkGap}} \\ & + 0.8\xi_{t,\text{reputation}} + 1.2\xi_{t,\text{groupWelfare}} + 1.4\xi_{t,\text{collaboration}} - 0.7\xi_{t,\text{moralCost}} \\ & + 0.4\xi_{t,\text{transcendent}} + 0.3\xi_{t,\text{flow}}. \end{aligned}$$

For each speaker $w$, the model maintains an exponential moving average

$$\bar{\xi}_{w,t} = (1-\alpha)\bar{\xi}_{w,t-1} + \alpha\xi_t, \qquad \alpha = 0.18,$$

and computes derivative emotion coordinates from

$$d_t = \xi_t - \bar{\xi}_{w,t-1}.$$

Thus the emotion rows are finite-difference value-gradient coordinates, not free text labels. Schematically,

$$m_t = \Gamma(d_t, \text{Var}_{w,t-1}, NE_t, OT_t),$$

where $NE_t$ is a normalized entropy/uncertainty proxy and $OT_t$ is an other/self collaboration proxy. In the implementation, fear-like coordinates rise under survival drops or low survival, anxiety-like coordinates rise under variance, hope-like coordinates rise under future-value increases, anger-like coordinates rise under collaboration drops, and curiosity-like coordinates rise under flow/probe energy.

The social hypergraph estimator then groups events into typed edges. A conversation edge $e$ receives type

$$\tau(e) \in \{\text{pairIdentity}, \text{contactFormation}, \text{portalPath}, \text{repairBoundary}, \text{soloFrame}\}$$

from dominant coda, participant set, response structure, and value coordinates. Two response observables are central:

$$\text{sameC}(e) = \frac{\#\{t \in e : w_t \neq w_{t-1},\, c_t = c_{t-1}\}}{\#\{t \in e : w_t \neq w_{t-1}\}},$$

$$\text{roleSplit}(e) = \frac{\#\{t \in e : w_t \neq w_{t-1},\, c_t = c_{t-1},\, r_t \neq r_{t-1}\}}{\#\{t \in e : w_t \neq w_{t-1},\, c_t = c_{t-1}\}}.$$

These are written to `value_field_social_hyperedges.csv` and related dyad/history tables. They are the reason a coda is interpreted as an operator on a social edge rather than as an isolated token.

The contract ledger is the bridge from social calculus to conversational pragmatics. For each event, the code constructs

$$\kappa_t = [\kappa^T, \kappa^R, \kappa^O, \kappa^K, \kappa^B, \kappa^P]_t \in [0,1]^6.$$

The six coordinates are trust credit, repair debt, role obligation, common knowledge, private bond, and public room. Each axis is a clipped weighted projection from reply history, hyperedge type, rNPV derivatives, and complex-emotion coordinates. For example, the implemented trust-credit and repair-debt axes are:

$$\kappa_{t,\text{trust}} = \text{clip}(0.30 \text{ sameC}_t + 0.22 \max(0, \Delta V_t^{\text{collab}}) + 0.20\, m_{t,\text{collaboration}} + 0.15\, m_{t,\text{socialTrust}} + 0.13 \max(0, m_{t,\text{hope}})),$$

$$\kappa_{t,\text{repair}} = \text{clip}(0.30\, \mathbf{1}[\text{NOISE} \vee A4 \vee A9] + 0.22\, m_{t,\text{anxiety}} + 0.18\, m_{t,\text{fear}} + 0.15\, \xi_{t,\text{variance}} + 0.15 \max(m_{t,\text{vigilantRepair}}, m_{t,\text{boundaryNegotiation}})).$$

The remaining axes follow the same form: role obligation weights role splits, benchmark gap, $|i$ register, and prior dyad replies; common knowledge weights same-C replies, collective focus, group welfare, PLUS codas, and contact-formation edges; private bond weights $5R$ codas, attachment continuity, social trust, pair-identity edges, and dyad history; public room weights PLUS codas, collective focus, group welfare, contact formation, and common knowledge. The primary contract state is

$$\arg\max_j \kappa_{t,j},$$

with phase thresholds such as repair due, role terms, shared record, private bond, and public room. These rows are `value_field_contract_ledger.csv`.

Citing these axes in an academic paper requires a second test: they must separate observed language-use contexts, not merely exist as hand-designed coordinates. The repository therefore writes `value_field_language_use_validation.csv`. For each candidate context $Y_k(t) \in \{0, 1\}$ and axis score $s_k(t)$, the implemented audit reports

$$\Delta_k = \mathbb{E}[s_k(t) \mid Y_k(t) = 1] - \mathbb{E}[s_k(t) \mid Y_k(t) = 0], \qquad \text{AUC}_k = \Pr[s_k(t^+) > s_k(t^-)].$$

This is a use test, not a feeling test: it asks whether the value-field coordinates predict where the relevant coda functions appear.

| Language-use context | Positive events | Effect Δ | AUC |
|---|---|---|---|
| same-coda collaborative reply | 60 | 0.62157 | 1.00000 |
| role/register asymmetry | 408 | 0.28165 | 0.99869 |
| repair/boundary language | 126 | 0.58049 | 0.99860 |
| public-room/PLUS language | 709 | 0.33012 | 0.98919 |
| private-pair/5R language | 259 | 0.19427 | 0.90287 |
| path/probe language | 234 | 0.12674 | 0.67899 |

The first five rows support using the axes as latent language-use scaffolds in the paper. The path/probe row remains below the promotion threshold and should be treated as an exploratory direction until acoustic trajectory, prey, or movement grounding improves.

Coda-level proto-meanings are then aggregations over all events containing unit $u$:

$$\bar{\kappa}_u = \mathbb{E}[\kappa_t \mid c_t = u], \qquad \bar{m}_u = \mathbb{E}[m_t \mid c_t = u], \qquad \bar{R}_u = \mathbb{E}[R_t \mid c_t = u],$$

where $R_t$ includes receiver-response, same-C, role-split, and repair/recovery observables. These aggregated axes become `whale_operator_semantic_basis.csv`. This is the point at which the model can say that $5R1$ is structurally a pair-identity/social-memory operator or that $1 + 1 + 3$ is a contact-field stabilizer.

Finally, English is a guarded projection, not an estimator of ground truth. The final projection chooses an operator phrase $e_t$ from the proto-meaning basis and assigns

$$C_{\min,t} = \min(C_{\text{struct},t}, C_{\text{acoustic},t}, C_{\text{grounded},t}).$$

The implementation promotes a phrase only to a grounded candidate when

$$C_{\text{struct}} \geq 0.70, \qquad C_{\text{acoustic}} \geq 0.55, \qquad C_{\text{grounded}} \geq 0.55.$$

Otherwise the phrase remains an operator phrase or structural phrase only. The WhatsApp message is therefore the rendering

$$\text{message}_t = \rho(e_t, \text{emoji}_t, m_t, \kappa_t, H_t, C_{\min,t}),$$

where $\rho$ is the web renderer. This makes the interface an inspectable projection from the mathematical state, not an independent translation.

## 25 Reproducibility And Artifact Map

The analysis is designed to be reproduced as an ordered artifact chain rather than as a manually curated narrative. The minimal PDF-derived core can be rebuilt without mounted external data:

`python3 src/whale_language_analysis.py --no-external-data`.

The public reader is then rebuilt by

`python3 web/build_site.py`.

The repository also provides an artifact audit:

`python3 scripts/reproduce_paper_outputs.py --check`.

| Mathematical object | Primary artifact | Current rows |
|---|---|---|
| event stream $y_t$ | `analysis/events.csv` | 1139 |
| social hypergraph $H_t$ | `analysis/value_field_social_hyperedges.csv` | 89 |
| rNPV/value derivatives | `analysis/value_field_rnpv_emotions.csv` | 1139 |
| complex emotion field | `analysis/value_field_complex_emotions.csv` | 1139 |
| contract ledger | `analysis/value_field_contract_ledger.csv` | 1139 |
| axis validation | `analysis/value_field_language_use_validation.csv` | 6 |
| coda operator basis | `analysis/whale_operator_semantic_basis.csv` | 23 |
| English/emoji projection | `analysis/emoji_translation_projection.csv` | 1139 |
| bootstrap examples | `analysis/bootstrap_training_examples.csv` | 6058 |
| derived chat windows | `analysis/dswp_derived_conversation_windows.csv` | 165 |
| bat comparative annotations | `analysis/bat_annotation_events.csv` | 91080 |
| bat schema manifest | `analysis/bat_comparative_schema_manifest.csv` | 27 |
| elephant social-pattern layer | `analysis/elephant_social_hypergraph_insights.csv` | 619 |
| dolphin reader turns | `analysis/dolphin_reader_turns.csv` | 384 |
| marmoset reconstructed turns | `analysis/marmoset_reconstructed_exchange_turns.csv` | 1080 |
| marmoset kinship turns | `analysis/marmoset_kinship_sequence_turns.csv` | 642 |
| species capability registry | `data/comparative_species_registry.json` | 5 species |
| web projection | `web/data.json` | 254 whale chats; 300 bat threads; 80 elephant threads |

The full artifact-to-section mapping is recorded in `paper/artifact_table.csv`. This makes the paper falsifiable in a practical sense: a reader can inspect which table supports each mathematical claim, rerun the core pipeline, and verify that the WhatsApp-style interface is a projection from the tables rather than a separate hand-written translation layer.

**Principle 7** (Reproducible projection)**.** *Every displayed phrase should be traceable to a source event, a signal operator, a social-hypergraph context, value-gradient coordinates when available, contract-ledger or interaction-state coordinates when available, and confidence gates. If that trace is absent, the phrase is an interface gloss, not a model claim.*

## 26 Code And Interface Availability

The full codebase is part of the research artifact. The source code, generated tables, model-ready exports, LaTeX source, paper PDF, reproducibility audit, and WhatsApp-style reader should be archived under one immutable release tag. The interactive reader is currently deployed at `https://animal-chats.pages.dev`. It is not a separate interpretive product: it is generated from `analysis/*.csv` through `web/build_site.py`, and its data payload is `web/data.json`. Thus a reader can move in both directions:

$$\text{paper claim} \leftrightarrow \text{artifact table} \leftrightarrow \text{CSV evidence} \leftrightarrow \text{web message.}$$

For publication, the repository should be released with a fixed git tag and DOI-backed archive. The required pre-release checks are the unit test suite, the paper artifact audit, the web rebuild, and archival of the exact `web/` directory used by the public interface. The interface should be cited as an interactive visualization of guarded operator-phrase projections, not as a literal animal-to-human dictionary.

## 27 Predictions

**Hypothesis 1.** *If $5R1$ is a pair-identity operator, partner-label permutation will destroy its same-$C$ response advantage faster than it destroys ordinary coda frequency.*

**Hypothesis 2.** *If final-i codas form a scalar probe ladder, numeric transitions should preserve order-sensitive future distributions; $8i \to 9i$ and $9i \to 8i$ should not be exchangeable.*

**Hypothesis 3.** *If NOISE codas are repair/boundary operators, post-NOISE states should have lower entropy than matched non-NOISE perturbations when the repair succeeds, and higher entropy when it fails.*

**Hypothesis 4.** *If translation is a projection from operator space, English labels should become stable only after hypergraph role, acoustic embedding, and ecological variables are jointly conditioned.*

**Hypothesis 5.** *If emotion is a value-field derivative rather than a label, emotion-like reader states should predict coordination outcomes such as repair success, same-partner return, role stabilization, or uncertainty reduction better than raw affect words alone.*

**Hypothesis 6.** *If consciousness-related evidence is operationally present, integrated self/other/world update signatures should persist across time, partner, and context controls without being reducible to a single caller identity, recording artifact, or context label.*

**Hypothesis 7.** *If register is gauge-like, many $|a/|i$ contrasts should preserve base operator spectra while changing receiver-role curvature. This predicts spectral similarity between $5R1|a$ and $5R1|i$ but different cross-speaker response geometry.*

**Hypothesis 8.** *If cultural operators are renormalization-stable, codas such as $5R1$ should remain identifiable after coarse-graining from events to runs to dyads, while more local repair/probe variants should split or merge under scale changes.*

## 28 Conclusion

The boundary-pushing move is to stop asking, “What English word does this animal signal mean?” and instead ask, “What transformation does this signal perform on this species’ interaction field?” The social hypergraph supplies context, value-field derivatives supply value geometry where evidence supports them, weak annotations supply boundary conditions, and annealing supplies the path from structural proto-meaning to guarded human-readable projection. This produces a principled route toward comparative animal communication research while keeping the model honest about what is known, what is inferred, and what remains ungrounded.