Communication Boundary Control for Safer Multi-Agent Language Agents

Anonymous Authors[1]

Abstract

Multi-agent language-agent systems can fail before any message is worded: a ready agent may fail to warn, warn the wrong recipient, over-escalate weak evidence, expose control-relevant content through raw text, or trigger costly downstream behavior with a poorly scoped communication event. We study this pre-wording decision as *communication boundary control*: at a ready-agent boundary, the actor chooses a finite high-level act–local macro-action, yield, or recipient-targeted semantic send–and natural language is generated only after the semantic record has been selected. For one frozen improvement round with a fixed finite semantic interface, we derive the local control target and its diagnostics: predictive send-versus-yield usefulness is a relevance target, not a send rule; Stage II must choose by net downstream value and cost; a context-measurable KL-regularized actor must average hidden-boundary values before softmaxing; and parser-equivalent wording is safe only when it preserves control-relevant boundary dynamics. Controlled diagnostics test these safety-relevant failure modes: decision-focused usefulness separates from generic informativeness, semantic bundles recover value missed by top-1 proposals, and wording regimes expose leakage from surface text. The result is a finite, auditable control layer for deciding whether a communication event should exist, who should receive it, and what control-relevant content future agents may condition on before wording begins.

1. Introduction

A scientific-discovery agent system is evaluating a hypothetical catalyst candidate C7. The planner is about to allocate the next batch of synthesis or simulation budget. The simulator has just found that C7 becomes unstable above $450\,\mathrm{K}$. At this boundary, the simulator can run another local check, yield control, warn the planner, ask the literature agent whether the instability has been reported, or hand the candidate to a safety checker. The safe choice depends on hidden task and runtime state. A missed warning may waste resources or continue an unsafe branch; the wrong recipient may leave the planner uninformed; querying the literature agent may dominate warning if the issue is likely known; another local check may dominate communication if the evidence is weak; and overly strong wording may cause the system to abandon a useful candidate prematurely. This is a communication-control decision before it is a text-generation decision. The simulator is not first choosing between phrasings such as "C7 appears unstable above $450\,\mathrm{K}$" and "C7 may be unstable at high temperature." It is deciding whether a communication event should exist, who should receive it, and what decision-relevant content future control should be allowed to use. The same pattern appears when a browser agent decides whether to alert a human before a payment or a software agent decides whether to notify a security monitor before running a command.

Existing multi-agent LLM systems typically optimize inside a communication scaffold. A runtime, prompt design, or scaffold optimizer specifies agents, roles, routes, speaking opportunities, message slots, turn timing, tool access, and stopping rules (Wu et al., 2024; Hong et al., 2023; Qian et al., 2024; Fourney et al., 2024; Gan et al., 2025; Zhang et al., 2025b). Learning can then improve wording, local behavior, routing, topology, interruption, pruning, or other variables exposed by the scaffold (Fan et al., 2025; Zhang et al., 2025a; Lu et al., 2026; Zhang et al., 2024; Wang et al., 2025; Eo et al., 2025; Jiang et al., 2026; Wang et al., 2026). These are important design variables, but they leave a more primitive event-level question: after a high-level event has ended and an agent is ready, should the next boundary act be local computation, yield, or a recipient-targeted communication event at all? The language-agent issue is not only communication cost. A language message mixes three objects that should be controlled separately: whether an event occurs, which semantic distinctions future policies may read, and how the selected record is worded.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review at the Second Workshop on Agents in the Wild: Safety, Security, and Beyond (AIWILD) at ICML 2026. Do not distribute.

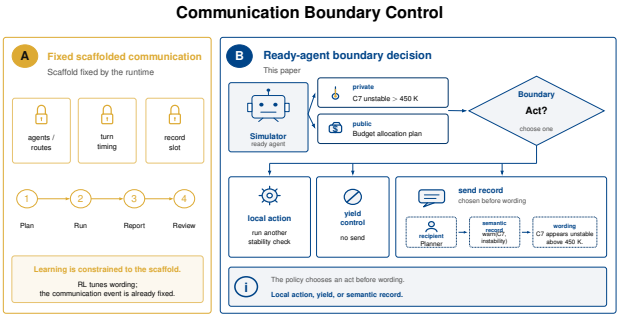

Figure 1. **Communication boundary control makes the communication event a policy choice.** The scaffolded view fixes or exposes roles, routes, turns, and message slots; the boundary-control view lets the ready agent choose local action, yield, or a recipient-targeted semantic record before wording. Labels carry the contrast; color is only a visual aid.

We formulate this problem as *communication boundary control*. At each ready-agent boundary, the actor chooses one finite high-level act: a local macro-action, yield, or a recipient-targeted semantic send. Natural language is generated only after a send record has been selected. Thus the control object is not an arbitrary string, but a finite record such as

```
warn(candidate=C7,
  issue=instability,
  threshold=above-450K,
  evidence=sim17).
```

The record identifies the control-relevant content that future policies may condition on; the wording policy realizes that record afterward.

The finite semantic interface is not a claim that language reduces to a fixed ontology. It is a roundwise control abstraction. During one improvement round, the actor chooses among the finite distinctions exposed to control and made available to future agents. A field belongs in the record when varying it is intended or observed to change downstream control: urgency, confidence, provenance, threshold bins, evidence pointers, and recipient roles may all be semantic fields in a given round. Surface wording is generated afterward and may be retained for audit. If two wordings mapped to the same record nevertheless induce different future behavior, then wording has leaked control-relevant information. The remedy is to refine the record, constrain the wording policy, or measure the induced drift.

1.1. Safety and Security Interpretation

Communication boundary control treats communication as a control-plane decision. The controller decides whether to expose a semantic record, which recipient may receive it, and which fields downstream policies may read. This creates an audit trail for failures that are difficult to attribute in raw transcripts: a missing send, wrong recipient, unsafe shortlist, context aliasing error, value-estimation error, or wording-induced control drift. The framework is not a full adversarial security guarantee; it is a finite diagnostic layer for identifying and reducing communication-induced failures in agent systems.

The core safety distinction is between the canonical record and raw wording. In the exact-core architecture, future control consumes the canonical record and sanctioned fields; raw wording may be retained for humans, debugging, and audit. If future policies condition on arbitrary raw wording, then raw text becomes part of the control state. Parser equivalence is then insufficient: adding urgency, confidence, evidence strength, or an apparently harmless phrase can change downstream behavior even when a parser maps both strings to the same record.

This separation exposes three confusions that a communication controller must avoid. Relevance is not value: a candidate send may change a recipient-side decision target without being worth its communication cost or improving downstream return. Context is not state: the deployed actor observes only a local context, so it must choose one distribution for hidden boundary snapshots that expose the same context. Wording is not automatically harmless: parser-equivalent rephrasings are interchangeable only when they preserve the same control-relevant boundary dynamics.

We test these distinctions in controlled communication-boundary experiments. The main suite shows that decision-focused Stage I relevance separates from generic informativeness, that semantic bundles recover value missed by top-1 proposals, and that a controlled wording stress test separates safe and leaky realization regimes by measured support leakage. The suite is diagnostic: it tests safety-relevant failure modes under a finite semantic interface rather than broad deployed-agent benchmark performance.

Our contributions are:

1. **Communication as a safety-relevant boundary action.** We formulate multi-agent LLM communication as a local boundary-control problem in which the ready agent chooses among local macro-actions, yield, and recipient-targeted semantic sends before natural-language realization.

2. **A screen–control–realize decomposition.** We show that send-versus-yield predictive usefulness is a relevance target for a recipient-side decision variable, not a send rule. Low usefulness can justify conservative pruning only under explicit recipient-local proxy, calibration, and communication-cost assumptions.

3. **Roundwise actor and diagnostics.** For a fixed finite semantic interface and frozen improvement round, we de-

rive the KL-regularized boundary actor under partial observability, quantify context aliasing, decompose shortlist and value-estimation error, and characterize wording validity as control equivalence rather than parser equivalence.

4. **Controlled safety diagnostics.** We provide controlled synthetic experiments that test decision-focused relevance, Stage II semantic value, shortlist and bundle loss, communication cost, context and recipient conditioning, and wording leakage.

All exact claims are local and roundwise: the semantic interface, context map, continuation policy, wording policy, proposer, prior, temperature, and improvement distribution d are fixed during one improvement round. The goal is not to solve open-ended language communication end-to-end, but to isolate the finite control layer at which communication events can be learned, audited, and diagnosed.

2. Related Work

We compare prior work by where communication is optimized. The key distinction is not whether communication can be learned, routed, pruned, or rewarded; it is whether the communication event itself remains a free boundary act. In our setting, the ready agent still chooses among local action, yield, and semantic send before wording begins.

LLM-agent orchestration and adaptive scaffolds. Multi-agent LLM orchestration systems specify agents, roles, routes, tools, speaking opportunities, turn structure, message slots, and stopping rules for collaborative language-model workflows (Wu et al., 2024; Hong et al., 2023; Qian et al., 2024; Li et al., 2023; Chen et al., 2023; Fourney et al., 2024; Gan et al., 2025; Zhang et al., 2025b; Du et al., 2024; Hu et al., 2025). Recent work adapts scaffold variables such as topology, routing, agent selection, debate activation, interruption timing, communication pruning, and token or latency cost (Fan et al., 2025; Zhang et al., 2025a; Lu et al., 2026; Zhang et al., 2024; Wang et al., 2025; Eo et al., 2025; Jiang et al., 2026; Wang et al., 2026). These systems show that scaffolds need not be fixed. Communication boundary control is complementary: it places the decision at the ready-agent boundary, where the actual communication event competes with local noncommunicative action and yield.

Decision-theoretic and learned communication in MARL. Classical decentralized control and MARL have long treated communication as a costly policy variable under partial observability (Pynadath & Tambe, 2002; Xuan et al., 2001; Goldman & Zilberstein, 2003; 2008; Roth et al., 2006; Amato et al., 2019). Learned-communication work extends this view to discrete messages, continuous channels, selective communication, targeted communication, topology, attention, and scheduling (Sukhbaatar et al., 2016; Foerster et al., 2016; Singh et al., 2018; Das et al., 2019; Jiang & Lu, 2018; Kim et al., 2018). We build on this finite-action and communication-cost discipline. The language-agent interface adds a split: the boundary policy chooses a finite semantic control record, and an LLM realizes that record afterward in natural language. This separation makes it necessary to ask which distinctions belong in the control layer, when predictive usefulness is relevance rather than value, and when wording changes downstream control despite identical parsing.

LLM collaboration, textual credit, and structured outputs. Recent RL and MARL methods improve collaborative LLM behavior through post-training, actor-critic learning, verbal feedback, counterfactual credit assignment, and natural-language evaluation of messages (Park et al., 2025; Liao et al., 2025; Liu et al., 2025; Bi et al., 2025; Liu et al., 2026; Chen et al., 2026; Yao et al., 2026; Yu et al., 2026). These methods often assign credit to messages or conversational traces after an interaction has occurred. Communication boundary control acts earlier: before the communication event is realized. Structured messages, agent communication languages, tool calls, function calls, and constrained outputs improve parsability, execution reliability, and interoperability (Cohen & Levesque, 1990; Labrou & Finin, 1994; Foundation for Intelligent Physical Agents, 2002; Schick et al., 2023). Our finite records are related in form but different in role: the record is the finite control act optimized by the boundary actor, not merely a parser target or tool argument.

Appendix I gives a compact positioning table. The object formalized below is the boundary event itself: a ready agent chooses among local action, yield, and semantic send before wording.

3. Communication Boundary Control

Communication boundary control makes one runtime moment explicit: the instant after a high-level event has completed and before the next event is chosen. At boundary b, one ready agent i_b observes a deployed local context

$$c^b = C(B^b),$$

where B^b is the hidden boundary snapshot: the environment and runtime state, the ready-agent identity, scheduler or transport state, tool outputs, and control-relevant agent memories needed to define the next-boundary law. The actor does not observe B^b directly. It sees only c^b, so one deployed act distribution must serve every hidden snapshot that exposes the same context.

Four objects must remain separate. The hidden snapshot B is the state needed to define future controlled evolution. The observed context $c = C(B)$ is what the deployed actor sees. The finite semantic record a is the control object selected by the boundary actor. The wording m is the natural-language realization generated only after a semantic send has been selected. Collapsing these objects creates the three errors studied below: treating relevance as value, context as state, or parser-equivalent wording as control-equivalent wording.

3.1. Boundary acts and semantic values

The round exposes a finite semantic inventory Σ_{sem}. $\Sigma_{\mathrm{sem}}(c) \subseteq \Sigma_{\mathrm{sem}}$ denotes the records legal at context c. For ready agent $i = i_b$, the finite high-level boundary act set at context c is

$$\begin{aligned} \mathcal{U}_i^{\mathrm{hi}}(c) &= \mathcal{A}_i^{\mathrm{phys,hi}}(c) \cup \{\emptyset\} \cup \mathcal{C}_i^{\mathrm{hi}}(c), \\ \mathcal{C}_i^{\mathrm{hi}}(c) &= \{(j,a) : j \neq i,\ a \in \Sigma_{\mathrm{sem}}(c)\}. \end{aligned} \tag{1}$$

$\mathcal{A}_i^{\mathrm{phys,hi}}(c)$ contains local macro-actions that do not send a record. The act $\emptyset$ is yield: the agent sends no record and takes no new local macro-action at this boundary, so the fixed runtime or continuation policy determines the next event. A send act $u = (j, a)$ delivers finite semantic record a to recipient j. In the C7 example, the simulator chooses among running another stability check, yielding, warning the planner, asking the literature agent, or handing the candidate to a safety checker.

The finite record is the control interface for one improvement round, not a universal ontology of language. A semantic record may contain an act type, recipient role, finite slots or bins, and finite pointers to objects exposed during the round. Fields such as urgency, confidence, provenance, threshold bins, and evidence pointers are part of the record whenever changing them changes downstream control; otherwise they are wording or audit information. If a send act $u = (j, a)$ is selected, the LLM realizes it afterward:

$$M \sim \bar{\pi}_{\mathrm{text}}(\cdot \mid c, u), \qquad \operatorname{supp} \bar{\pi}_{\mathrm{text}}(\cdot \mid c, u) \subseteq \mathcal{M}_a(c, u).$$

The exact-core architecture is structured control-plane communication with natural-language realization: future control reads the canonical record, not arbitrary raw text; raw wording is retained for humans and audit unless it is explicitly added to the control interface. The appendix gives the full boundary and semantic-abstraction contracts.

Stage II compares high-level acts by downstream value under the fixed continuation policy π_0 and wording policy $\bar{\pi}_{\mathrm{text}}$. The semantic-layer value is

$$Q_{\mathrm{hi}}^{\pi_0}(B, u) = \begin{cases} Q^{\pi_0}(B, u), & u \text{ local/yield}, \\ \mathbb{E}_M[Q^{\pi_0}(B, (u, M))], & u \text{ send}. \end{cases} \tag{2}$$

In the send case, $M \sim \bar{\pi}_{\mathrm{text}}(\cdot \mid C(B), u)$. Q^{π_0} and $Q_{\mathrm{hi}}^{\pi_0}$ are net values under the fixed round cost convention, including communication, latency, token, activation, or interruption costs represented in the boundary-interval return.

3.2. Screen, control, realize

Stage I screening asks whether a candidate send changes a decision-focused recipient-side target relative to yield. For recipient abstraction x, candidate send $u = (j, a)$, and future target Y_H, define

$$\begin{aligned} P_x^u &= \mathrm{Law}_{\pi_0, \bar{\pi}_{\mathrm{text}}}(Y_H \mid X_j = x, \mathrm{do}(U = u)), \\ P_x^{\emptyset} &= \mathrm{Law}_{\pi_0, \bar{\pi}_{\mathrm{text}}}(Y_H \mid X_j = x, \mathrm{do}(U = \emptyset)), \\ I_{\mathrm{pred}}(x, u) &= D_{\mathrm{KL}}\left(P_x^u \,\|\, P_x^{\emptyset}\right). \end{aligned} \tag{3}$$

The target must be a future variable that control actually uses, such as the planner's next branch choice, a plan revision, a verifier state, or an artifact state over a fixed horizon. Predictive usefulness is a relevance target, not a value target: low calibrated usefulness can support conservative pruning under the conditions in Theorem 4.2, while the final act is chosen only by Stage II control.

Let $P_{\mathrm{send}}(c)$ be the proposed send records and $R_{\mathrm{send}}(c) \subseteq P_{\mathrm{send}}(c)$ the retained Stage I shortlist. The full legal high-level act set and implemented feasible set are

$$\begin{aligned} U_{\mathrm{full}}(c) &= \mathcal{U}_i^{\mathrm{hi}}(c), \\ F_{\mathrm{impl}}(c) &= \mathcal{A}_i^{\mathrm{phys,hi}}(c) \cup \{\emptyset\} \cup R_{\mathrm{send}}(c). \end{aligned}$$

Stage II estimates values for acts in $F_{\mathrm{impl}}(c)$. Under improvement distribution d, the target value vector is the conditional mean

$$\bar{Q}_d(c, u) = \mathbb{E}_{B \sim d}[Q_{\mathrm{hi}}^{\pi_0}(B, u) \mid C(B) = c]. \tag{4}$$

The controller proposes candidate semantic sends, uses Stage I as a recall-biased relevance screen, hard-prunes only certified local negatives, and then lets Stage II choose among local actions, yield, and retained sends by net downstream semantic value and cost. If the selected act is a send, the wording policy realizes the selected canonical record; otherwise the system executes the local action or yield. Full pseudocode and logging fields are given in Appendix A.2.

4. Roundwise Theory and Diagnostics

The theory answers four implementation questions: what Stage I may safely prune, what policy Stage II should imitate, what safety-relevant information is hidden by the deployed context, and how much value is lost when proposals, values, or wording are approximate. All results are for one frozen improvement round. The semantic interface, high-level act inventory, context map, cost conven-

tion, continuation policy, wording policy, proposer, shortlist rule, prior, temperature, and improvement distribution d are fixed while the local actor target is defined. These are exact local-round statements, not claims of end-to-end global policy improvement.

4.1. Stage I: what relevance can certify

Stage I estimates whether a candidate send changes a decision-focused recipient-side target relative to yield:

$$I_{\mathrm{pred}}(x,u) = D_{\mathrm{KL}}\Big(P_x^u \,\|\, P_x^{\emptyset}\Big).$$

Proposition 4.1 (Send-versus-yield KL identity)**.** *Fix recipient abstraction x, candidate send act $u = (j,a)$, continuation policy π_0, wording policy $\bar{\pi}_{\mathrm{text}}$, and decision-focused target Y_H. Let P_x^u and $P_x^{\emptyset}$ be the target laws in Equation (3). If $p_u(\cdot \mid x)$ and $p_{\emptyset}(\cdot \mid x)$ are Bayes-optimal predictors under the two interventions, then*

$$\mathbb{E}_{Y\sim P_x^u}[\log p_u(Y \mid x) - \log p_{\emptyset}(Y \mid x)] = D_{\mathrm{KL}}(P_x^u \,\|\, P_x^{\emptyset}),$$

with the usual extended-value convention when absolute continuity fails.

This identity identifies the exact log-loss target of a calibrated Stage I model. It measures recipient-side decision relevance relative to yield, not downstream value.

Theorem 4.2 (One-sided pruning certificate)**.** *Let $Z_H = f(Y_H)$ be decision-sufficient for a recipient-local H-step proxy problem under π_0. Let $\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x,u)$ be the recipient-local proxy gain from sending u rather than yielding at recipient abstraction x, before charging communication cost. Let $P_{Z_H,x}^u$ and $P_{Z_H,x}^{\emptyset}$ be the pushforward laws of Z_H, and suppose the recipient-local utility span is at most V_H. Let $r(x,u) = \sqrt{\frac{1}{2}D_{\mathrm{KL}}(P_x^u \,\|\, P_x^{\emptyset})}$. Then*

$$\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x,u) \le V_H \left\| P_{Z_H,x}^u - P_{Z_H,x}^{\emptyset} \right\|_{\mathrm{TV}} \le V_H r(x,u).$$

Consequently, if a calibrated upper bound $U_u(x)$ satisfies

$$V_H\sqrt{\frac{1}{2}U_u(x)} \le \underline{c}_{\mathrm{comm}}(u),$$

where $\underline{c}_{\mathrm{comm}}(u)$ is a lower bound on the net communication cost charged to send u, then pruning u is locally safe at x for that proxy problem and cost lower bound.

The certificate is one-sided and local: low calibrated usefulness can certify only local pruning under the stated proxy and cost lower bound. A high usefulness score means the target distribution changes; it does not imply positive downstream value, safe escalation, or a send decision.

4.2. Stage II: what the actor optimizes

Stage II chooses among feasible boundary acts by downstream semantic value. The deployed actor observes $c = C(B)$, not the hidden boundary snapshot B. Because it cannot choose a different distribution for hidden boundary snapshots that share the same observed context, the value vector must be averaged within the context class before the policy is computed.

Theorem 4.3 (Average-then-softmax local actor)**.** *Fix a continuation policy π_0, wording policy $\bar{\pi}_{\mathrm{text}}$, improvement distribution d, context-measurable finite feasible set $F(c)$ whose acts are legal for all positive-mass hidden boundary snapshots B with $C(B) = c$, full-support prior $\mu(\cdot \mid c)$, and temperature $\tau > 0$. Assume $\bar{Q}_d(c,u)$ is finite for $u \in F(c)$. The unique optimizer of*

$$\max_{\pi(\cdot|c)\in\Delta(F(c))} \sum_{u\in F(c)} \pi(u \mid c)\bar{Q}_d(c,u) - \tau D_{\mathrm{KL}}(\pi(\cdot \mid c) \,\|\, \mu(\cdot \mid c)) \tag{5}$$

is

$$\pi_{\mathrm{hi}}^{\star}(u \mid c) = \frac{\mu(u \mid c)\exp(\bar{Q}_d(c,u)/\tau)}{Z(c)},$$

where $Z(c) = \sum_{v\in F(c)} \mu(v \mid c)\exp(\bar{Q}_d(c,v)/\tau)$. The optimal local soft value is $V_{\mathrm{loc}}^{\star}(c) = \tau \log \sum_{v\in F(c)} \mu(v \mid c)\exp(\bar{Q}_d(c,v)/\tau)$.

The expectation is part of the control target: a context-measurable deployed agent must average hidden-state semantic values before softmaxing. Softmaxing each hidden-state teacher and averaging policies solves a different objective.

4.3. Context aliasing: what the context hides

The actor must use one distribution for all hidden boundary snapshots that share c. The loss from this information pattern is a diagnostic for the context extractor. For $B \sim d(\cdot \mid C(B) = c)$, write

$$Q(B,u) = Q_{\mathrm{hi}}^{\pi_0}(B,u),$$

and

$$\bar{Q}_d(c,u) = \mathbb{E}[Q(B,u) \mid C(B) = c],$$

and define the full-information soft teacher

$$\nu_B(u) = \frac{\mu(u \mid c)\exp(Q(B,u)/\tau)}{Z_B(c)},$$

where $Z_B(c) = \sum_{v\in F(c)} \mu(v \mid c)\exp(Q(B,v)/\tau)$.

Corollary 4.4 (Context-aliasing diagnostic)**.** *Under the assumptions of Theorem 4.3, assume also that the displayed log-sum-exp expectations are finite. Define*

$$\ell_B(c) = \tau \log \sum_{u\in F(c)} \mu(u \mid c)\exp(Q(B,u)/\tau),$$

$$V^\star_{\text{full}}(c) = \mathbb{E}[\ell_B(c) \mid C(B) = c],$$

and $V^\star_{\text{loc}}(c) = \tau \log \sum_{u \in F(c)} \mu(u \mid c) \exp(\bar{Q}_d(c,u)/\tau)$. *Let*

$$\Delta_B(c) = D_{\text{KL}}(\pi^\star_{\text{hi}}(\cdot \mid c) \,\|\, \nu_B(\cdot)) \,.$$

Then

$$V^\star_{\text{full}}(c) - V^\star_{\text{loc}}(c) = \tau \mathbb{E}[\Delta_B(c) \mid C(B) = c].$$

The gap is zero iff the full-information teachers agree almost surely, equivalently iff hidden boundary snapshots in the context class differ only by an act-independent value shift:

$$Q(B,u) = \bar{Q}_d(c,u) + \alpha(B) \qquad \forall u \in F(c)$$

for some hidden-state scalar $\alpha(B)$.

Large aliasing gap means that C hides a control-relevant distinction. In the C7 example, routine instability already known to the planner and novel safety-critical instability should not share the same deployed context if those cases require different boundary acts.

4.4. Approximation budget: what implementation must measure

The practical controller proposes records, shortlists them, estimates values, and applies the local actor on the implemented feasible set. Fix context c, full legal act set $U_{\text{full}}(c)$, nonempty implemented feasible set $F_{\text{impl}}(c) \subseteq U_{\text{full}}(c)$, full-support prior $\mu(\cdot \mid c)$, true Stage II value vector $Q(c,\cdot) = \bar{Q}_d(c,\cdot)$, and estimated values $\widehat{Q}(c,\cdot)$ on $F_{\text{impl}}(c)$. Define $\pi_c = \pi(\cdot \mid c)$ and $\mu_c = \mu(\cdot \mid c)$, and set

$$\mathcal{F}^\mu_Q(\pi; c) = \sum_{u \in U_{\text{full}}(c)} \pi(u \mid c) Q(c,u) - \tau D_{\text{KL}}(\pi_c \,\|\, \mu_c).$$

Let $\pi^\star_Q$ be the full exact soft actor on $U_{\text{full}}(c)$, and let $\widehat{\pi}_{\text{impl}}$ be the actor obtained by softmaxing $\widehat{Q}$ on $F_{\text{impl}}(c)$ with the prior restricted and renormalized to $F_{\text{impl}}(c)$.

Theorem 4.5 (Local approximation budget for screen–then–control)**.** *Under the finite-round assumptions above, let* $p_{\text{keep}}(c) = \pi^\star_Q(F_{\text{impl}}(c) \mid c)$. *Then*

$$\mathcal{F}^\mu_Q(\pi^\star_Q; c) - \mathcal{F}^\mu_Q(\widehat{\pi}_{\text{impl}}; c) \le -\tau \log p_{\text{keep}}(c) + S_{\text{err}}(c),$$

where $S_{\text{err}}(c)$ *is the maximum minus minimum value-estimation error over the implemented feasible set. In particular, if* $\|\widehat{Q} - Q\|_{\infty, F_{\text{impl}}(c)} \le \epsilon$, *the loss is at most* $-\tau \log p_{\text{keep}}(c) + 2\epsilon$.

The first term is missed soft-optimal mass from proposals and shortlists. The second term is relative value error on retained acts; common additive critic shifts do not affect the actor. Proposal failure is more severe: if an act is never proposed, forced, replayed, or modeled, its interventional value is not identified from passive logs in general.

4.5. Wording validity: parser equivalence is not control equivalence

The semantic/text split is valid only when wording does not carry unsanctioned control content. The clean architecture is not "parse text and hope." It is: choose a canonical control record, deliver that record to future policies, and retain raw wording as audit material unless raw wording is explicitly part of the control interface. Let ϕ be a quotient of the next boundary state containing the variables future policies are allowed to use. For a completed event e, define the quotient next-boundary kernel

$$K^\phi_{B,e} = \text{Law}(\phi(B'), G, \Delta \mid B, e) \,,$$

where G is the net boundary-interval return and Δ is the interval duration.

Proposition 4.6 (Control-equivalent wordings have equal value)**.** *If two completed events* e *and* e' *induce the same kernel* $K^\phi_{B,e} = K^\phi_{B,e'}$, *then*

$$Q^\pi(B,e) = Q^\pi(B,e')$$

for every continuation policy π *whose value factors through* ϕ. *If the kernels differ by at most* ϵ *in total variation and the Bellman integrand has span at most* S_π, *then the value difference is at most* $S_\pi \epsilon$.

Parser-equivalent wording is therefore not enough. If two phrasings of the same record change future control differently, then wording has carried control content: the semantic record should be refined, the wording policy restricted, or the drift measured.

5. Controlled Safety Diagnostics

The formal results predict four safety-relevant implementation failures: an informative message need not be valuable, a relevance screen can become unsafe if treated as a send rule, a shortlist can miss value-bearing semantic bundles, and wording can leak control content even when it parses to the same record. We test these failures in controlled synthetic boundary environments with observable latent state, finite semantic inventory, legal recipients, costs, and counterfactual continuations. The logged boundary bundles pair contexts and legal act inventories with forced or randomized branch outcomes, costs, downstream deltas, and wording-policy metadata, matching the identification contracts in Appendix D. The suite is diagnostic rather than benchmark-competitive.

The results support the mechanism-level predictions. Decision-focused usefulness separates from generic informativeness, confirming that a Stage I screen should target recipient-side decision relevance rather than generic likelihood change. The full screen–control–realize actor im-

Diagnostic	Theory prediction	Controlled result	Safety meaning
Decision-focused relevance	Generic informativeness is not the right Stage I screen.	AUPRC 1.000 versus 0.861 for generic likelihood; Stage I AUROC 0.994 versus 0.707.	Avoids messages that are predictive but irrelevant to downstream control.
Stage I versus Stage II	Relevance is not a send rule; value and cost decide the act.	Full actor strictly improves over $23/26$ required variants, with three non-strict ties.	Prevents high-relevance messages from bypassing value and cost checks.
Proposal and bundle loss	Top-1 proposals can omit valuable semantic combinations.	Width-2 bundles improve return by $+1.775$ and reduce miss rate from 0.500 to 0.000.	Shows omitted semantic bundles are a real control approximation gap.
Wording validity	Parser equivalence is not control equivalence.	Support leakage is 0.000 in the safe channel-side regime and 1.000 in the leaky regime.	Detects raw text as an unsanctioned control channel.

Table 1. **Controlled diagnostics test safety-relevant communication failure modes.** Results are reported on the controlled synthetic boundary suite. Higher AUPRC, AUROC, and return are better; lower miss rate and support leakage are better. The table supports diagnostic predictions under a finite semantic interface and does not assert broad deployed-agent benchmark superiority.

proves over fixed communication policies, pooled single-agent behavior, and component ablations in the controlled sweep; this indicates that separating Stage I relevance from Stage II value is operationally useful in the diagnostic setting. Bundle diagnostics show that top-1 semantic proposals can miss value-bearing combinations, matching the shortlist-loss term in Theorem 4.5. The wording stress test shows that a system can have zero leakage when wording variation is channel-side only, and complete leakage when raw text crosses the semantic boundary. The medium and large reduced suites each beat $6/6$ required baselines, which checks stability within the same finite-interface diagnostic family.

These experiments should be read narrowly. The controlled environment exposes the hidden boundary state, finite semantic inventory, legal recipients, communication cost, and counterfactual continuations needed to evaluate Stage I relevance, Stage II semantic value, shortlist loss, bundle loss, and wording drift. The experiments do not establish broad real-world agent performance, robustness to adversarial agents, or global monotonic improvement. Appendix J gives the full protocol, baselines, additional ablations, scale checks, and reproducibility details.

6. Scope, Limitations, Responsible Use, and Conclusion

Communication in LLM-agent systems is a costly boundary action, not just generated text. Boundary control exposes the ready-agent choice among local action, yield, or a recipient-targeted semantic record; Stage I prunes only certified local negatives; Stage II chooses by net semantic value and cost; and wording follows record selection.

The exact claims are local and roundwise: semantic interface, act inventory, context map, costs, continuation and wording policies, proposer, shortlist rule, prior, temperature, and improvement distribution are fixed for the round. The framework does not claim exact RL over arbitrary strings, a final semantic interface, parser-consistent wording safety, passive-log identification of untried acts, simultaneous-ready decentralized control, global monotonic improvement, broad benchmark superiority, or a full adversarial security guarantee.

The main responsible-use risk is false assurance. A miscalibrated boundary controller could suppress warnings, over-escalate weak evidence, route sensitive content to the wrong recipient, or hide control-relevant information in raw wording. The framework mitigates these risks only when its diagnostics are measured: legal act inventories must be logged, pruned acts need calibrated low-gain certificates before hard pruning, value estimates must be checked for relative action error, context aliasing must be monitored, and wording policies must be tested for support leakage and within-class control drift.

The central empirical evidence uses synthetic controlled environments rather than human-subject data, scraped datasets, or LLM-generated judgments as primary metrics. The manuscript is intended as a primarily human-authored submission; any AI assistance for editing or formatting must be verified by the authors and must not supply unverified results, analysis, or claims.

The payoff of the finite boundary view is auditability. Systems can learn whether to communicate, whom to target, and what content to expose while separating boundary-decision error, relevance/value confusion, proposal loss, relative value error, context aliasing, distribution shift, and wording drift. For agents operating in open-ended environments, this separation turns communication from an opaque transcript event into a measurable control-plane decision.

References

Amato, C., Konidaris, G., Kaelbling, L. P., and How, J. P. Modeling and Planning with Macro-Actions in Decentralized POMDPs. *Journal of Artificial Intelligence Research*, 64:817–859, March 2019. ISSN 1076-9757. doi: 10.1613/jair.1.11418.

Bi, Z., Lu, M., Li, Y., Roy, S., Guan, W., Ziyadi, M., and Wang, X. OPTAGENT: Optimizing Multi-Agent LLM Interactions Through Verbal Reinforcement Learning for Enhanced Reasoning, October 2025.

Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Chan, C.-M., Yu, H., Lu, Y., Hung, Y.-H., Qian, C., Qin, Y., Cong, X., Xie, R., Liu, Z., Sun, M., and Zhou, J. AgentVerse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors, October 2023.

Chen, Y., Sun, Y., Wang, H., Zhang, X., Shen, X., Li, W., and Zhang, W. Contextual Counterfactual Credit Assignment for Multi-Agent Reinforcement Learning in LLM Collaboration. https://arxiv.org/abs/2603.06859v1, March 2026.

Cohen, P. R. and Levesque, H. J. Performatives in a Rationally Based Speech Act Theory. In *28th Annual Meeting of the Association for Computational Linguistics*, pp. 79–88, Pittsburgh, Pennsylvania, USA, June 1990. Association for Computational Linguistics. doi: 10.3115/981823.981834.

Das, A., Gervet, T., Romoff, J., Batra, D., Parikh, D., Rabbat, M., and Pineau, J. TarMAC: Targeted Multi-Agent Communication. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 1538–1546. PMLR, May 2019.

Du, Y., Li, S., Torralba, A., Tenenbaum, J. B., and Mordatch, I. Improving Factuality and Reasoning in Language Models through Multiagent Debate. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 11733–11763. PMLR, July 2024.

Eo, S., Moon, H., Zi, E. H., Park, C., and Lim, H. Debate Only When Necessary: Adaptive Multiagent Collaboration for Efficient LLM Reasoning, May 2025.

Fan, Y., Zhang, J., Cai, K., Yang, J., Tang, C., Wang, J., and Wang, K. Cost-Effective Communication: An Auction-based Method for Language Agent Interaction. https://arxiv.org/abs/2511.13193v2, November 2025.

Foerster, J., Assael, I. A., de Freitas, N., and Whiteson, S. Learning to Communicate with Deep Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Foundation for Intelligent Physical Agents. FIPA communicative act library specification. FIPA Specification SC00037J, 2002. URL http://www.fipa.org/specs/fipa00037/.

Fourney, A., Bansal, G., Mozannar, H., Tan, C., Salinas, E., Erkang, Zhu, Niedtner, F., Proebsting, G., Bassman, G., Gerrits, J., Alber, J., Chang, P., Loynd, R., West, R., Dibia, V., Awadallah, A., Kamar, E., Hosn, R., and Amershi, S. Magentic-One: A Generalist Multi-Agent System for Solving Complex Tasks, November 2024.

Gan, B., Zhao, Y., Zhang, T., Huang, J., Yusu, L., Teo, S. X., Zhang, C., and Shi, W. MASTER: A Multi-Agent System with LLM Specialized MCTS. In Chiruzzo, L., Ritter, A., and Wang, L. (eds.), *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 9409–9426, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-189-6. doi: 10.18653/v1/2025.naacl-long.476.

Goldman, C. V. and Zilberstein, S. Optimizing information exchange in cooperative multi-agent systems. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '03, pp. 137–144, New York, NY, USA, July 2003. Association for Computing Machinery. ISBN 978-1-58113-683-8. doi: 10.1145/860575.860598.

Goldman, C. V. and Zilberstein, S. Communication-Based Decomposition Mechanisms for Decentralized MDPs. *Journal of Artificial Intelligence Research*, 32:169–202, May 2008. ISSN 1076-9757. doi: 10.1613/jair.2466.

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Wang, J., Zhang, C., Wang, Z., Yau, S. K. S., Lin, Z., Zhou, L., Ran, C., Xiao, L., Wu, C., and Schmidhuber, J. MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework. In *The Twelfth International Conference on Learning Representations*, October 2023.

Hu, M., Zhou, Y., Fan, W., Nie, Y., Ye, Z., Xia, B., Sun, T., Jin, Z., Li, Y., Zhang, Z., Wang, Y., Ye, Q., Ghanem, B., Luo, P., and Li, G. OWL: Optimized Workforce Learning for General Multi-Agent Assistance in Real-World Task Automation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, October 2025.

Jiang, E. H., Li, L., Sun, R., Liang, X., Li, Y., Wu, Y., Luo, H., Li, H., Zhang, Z., Kang, Z., Chang, K.-W., and Wu, Y. N. Agent Q-Mix: Selecting the Right Action for LLM Multi-Agent Systems through Reinforcement Learning, April 2026.

Jiang, J. and Lu, Z. Learning Attentional Communication for Multi-Agent Cooperation. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Kim, D., Moon, S., Hostallero, D., Kang, W. J., Lee, T., Son, K., and Yi, Y. Learning to Schedule Communication in Multi-agent Reinforcement Learning. In *International Conference on Learning Representations*, September 2018.

Labrou, Y. and Finin, T. A semantics approach for KQML—a general purpose communication language for software agents. In *Proceedings of the Third International Conference on Information and Knowledge Management*, CIKM '94, pp. 447–455, New York, NY, USA, November 1994. Association for Computing Machinery. ISBN 978-0-89791-674-5. doi: 10.1145/191246.191320.

Li, G., Hammoud, H. A. A. K., Itani, H., Khizbullin, D., and Ghanem, B. CAMEL: Communicative Agents for "Mind" Exploration of Large Language Model Society. In *Thirty-Seventh Conference on Neural Information Processing Systems*, November 2023.

Liao, J., Wen, M., Wang, J., and Zhang, W. MARFT: Multi-Agent Reinforcement Fine-Tuning, November 2025.

Liu, S., Chen, T., Liang, Z., Lyu, X., and Amato, C. LLM Collaboration With Multi-Agent Reinforcement Learning, December 2025.

Liu, S., Chen, T., Amiri, R., and Amato, C. Learning Decentralized LLM Collaboration with Multi-Agent Actor Critic. In *Workshop on Multi-Agent Learning and Its Opportunities in the Era of Generative AI*, March 2026.

Lu, Y., Hu, Y., Zhao, X., and Cao, J. DyTopo: Dynamic Topology Routing for Multi-Agent Reasoning via Semantic Matching, February 2026.

Park, C., Han, S., Guo, X., Ozdaglar, A. E., Zhang, K., and Kim, J.-K. MAPoRL: Multi-Agent Post-Co-Training for Collaborative Large Language Models with Reinforcement Learning. In Che, W., Nabende, J., Shutova, E., and Pilehvar, M. T. (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 30215–30248, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.1459.

Pynadath, D. V. and Tambe, M. The Communicative Multiagent Team Decision Problem: Analyzing Teamwork Theories and Models. *Journal of Artificial Intelligence Research*, 16:389–423, June 2002. ISSN 1076-9757. doi: 10.1613/jair.1024.

Qian, C., Liu, W., Liu, H., Chen, N., Dang, Y., Li, J., Yang, C., Chen, W., Su, Y., Cong, X., Xu, J., Li, D., Liu, Z., and Sun, M. ChatDev: Communicative Agents for Software Development. In Ku, L.-W., Martins, A., and Srikumar, V. (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15174–15186, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.810.

Roth, M., Simmons, R., and Veloso, M. What to Communicate? Execution-Time Decision in Multi-agent POMDPs. In Gini, M. and Voyles, R. (eds.), *Distributed Autonomous Robotic Systems 7*, pp. 177–186. Springer Japan, Tokyo, 2006. ISBN 978-4-431-35878-7. doi: 10.1007/4-431-35881-1_18.

Schick, T., Dwivedi-Yu, J., Dessi, R., Raileanu, R., Lomeli, M., Hambro, E., Zettlemoyer, L., Cancedda, N., and Scialom, T. Toolformer: Language Models Can Teach Themselves to Use Tools. In *Thirty-Seventh Conference on Neural Information Processing Systems*, November 2023.

Singh, A., Jain, T., and Sukhbaatar, S. Learning when to Communicate at Scale in Multiagent Cooperative and Competitive Tasks. In *International Conference on Learning Representations*, September 2018.

Sukhbaatar, S., szlam, a., and Fergus, R. Learning Multiagent Communication with Backpropagation. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Wang, D., Yin, D., Desai, R., Li, L., Celikyilmaz, A., and Ni, A. Learning to Interrupt in Language-based Multiagent Communication, April 2026.

Wang, Z., Wang, Y., Liu, X., Ding, L., Zhang, M., Liu, J., and Zhang, M. AgentDropout: Dynamic Agent Elimination for Token-Efficient and High-Performance LLM-Based Multi-Agent Collaboration. In Che, W., Nabende, J., Shutova, E., and Pilehvar, M. T. (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 24013–24035, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-251-0. doi: 10.18653/v1/2025.acl-long.1170.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Burger, D., and Wang, C. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversations. In *First Conference on Language Modeling*, August 2024.

Xuan, P., Lesser, V., and Zilberstein, S. Communication decisions in multi-agent cooperation: Model and experiments. In *Proceedings of the Fifth International Conference on Autonomous Agents*, AGENTS '01, pp. 616–623, New York, NY, USA, May 2001. Association for Computing Machinery. ISBN 978-1-58113-326-4. doi: 10.1145/375735.376469.

Yao, H., Da, L., Liu, X., Fleming, C., Chen, T., and Wei, H. LangMARL: Natural Language Multi-Agent Reinforcement Learning, April 2026.

Yu, Y., Liu, H., Jin, H., Yuan, X., Kuang, P., and Wang, H. Learning to Communicate: Toward End-to-End Optimization of Multi-Agent Language Systems, April 2026.

Zhang, G., Yue, Y., Li, Z., Yun, S., Wan, G., Wang, K., Cheng, D., Yu, J. X., and Chen, T. Cut the Crap: An Economical Communication Pipeline for LLM-based Multi-Agent Systems, October 2024.

Zhang, G., Yue, Y., Sun, X., Wan, G., Yu, M., Fang, J., Wang, K., Chen, T., and Cheng, D. G-Designer: Architecting Multi-agent Communication Topologies via Graph Neural Networks. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 76678–76692. PMLR, October 2025a.

Zhang, Y., Liu, X., and Xiao, C. MetaAgent: Automatically Constructing Multi-Agent Systems Based on Finite State Machines. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 75667–75694. PMLR, October 2025b.

A. Roadmap, Notation, and Exact Round Contract

The main text states the core idea: communication is a boundary-level control decision over local action, yield, and finite recipient-targeted semantic records, with natural-language wording generated only after a record is selected. The appendices make the exactness boundary explicit. They are organized around the questions needed to evaluate the theory: what is the boundary process, when is a semantic record a valid control abstraction, what data identify the relevant targets, what Stage I can certify, what Stage II optimizes, and which extensions preserve the finite one-chooser geometry.

This appendix fixes notation and gives the frozen-round contract used by all exact statements. The contract is intentionally narrow. It is not a limitation hidden in small print: it is the reason exact claims are possible for language-agent communication at all.

A.1. Appendix guide

Question	Where it is answered
What is the boundary-level process?	Appendix B defines the semi-Markov boundary process, the control–audit split, completed events, canonical records, and cost convention.
When is a semantic record an exact control abstraction?	Appendix C defines semantic records as control-equivalence quotients and proves wording-invariance, raw-wording impossibility, and interface-refinement results.
What data identify the Stage I and Stage II targets?	Appendix D states replay, randomization, simulator, and observational identification conditions, plus non-identification results for zero-propensity actions.
What does Stage I certify?	Appendix E proves the send-versus-yield KL identity, conditioning/coarsening results, local pruning certificates, and decision-value bounds.
What does Stage II optimize?	Appendix F proves average-then-softmax, the aliasing-gap identity, shortlist restriction loss, and local approximation-regret bounds.
Which extensions preserve the exact geometry?	Appendix G records one-effective-chooser, structured-action, budget, and simultaneous-ready results.
Which standard inequalities are used?	Appendix H collects total-variation, KL, softmax, concentration, and Blackwell-garbling lemmas used in the proofs.
What is and is not claimed?	Appendix I lists exact claims, approximation claims, controlled empirical diagnostics, extension status, and explicit non-claims.
What controlled evidence supports the mechanism?	Appendix J gives the controlled suite, baselines, mechanism diagnostics, scale checks, and empirical scope.
How can the experiments be reproduced?	Appendix K records scripts, configurations, environment assumptions, randomness, compute, and responsible-use details.
Where are workshop submission notes recorded?	Appendix L records AIWILD-specific anonymity, packaging, and LLM-use notes. Reproducibility and compute details are in Appendix K.

Table A.1. **The appendix roadmap answers the theory questions in dependency order.** Proofs are placed near the definitions and theorems they support, rather than collected in a separate proof appendix.

A.2. Controller pseudocode

Algorithm A.1 gives the full controller pseudocode, including logging fields for proposal-recall, shortlist-loss, and boundary-bundle diagnostics.

Algorithm A.1 Communication boundary controller at boundary b

Require: local context c, high-level acts $\mathcal{U}_i^{\mathrm{hi}}$, semantic interface Σ_{sem}, wording policy $\bar{\pi}_{\mathrm{text}}$
Require: predictive-usefulness scorer I_{pred}, net value estimator $\bar{Q}^{\mathrm{net}}$, prior μ, temperature τ

$U_{\mathrm{full}}(c) \leftarrow \mathcal{U}_i^{\mathrm{hi}}(c)$
$P_{\mathrm{send}}(c) \leftarrow \textsc{ProposeSendActs}(c, \Sigma_{\mathrm{sem}})$
for all $u \in P_{\mathrm{send}}(c)$ **do**
  $I_{\mathrm{pred}}(u \mid c) \leftarrow \textsc{PredictiveUsefulness}(c, u)$
end for
$R_{\mathrm{send}}(c) \leftarrow \textsc{RecallBiasedShortlist}(P_{\mathrm{send}}(c), I_{\mathrm{pred}})$

Remove only certified local negatives, if such certificates are available.
$F_{\mathrm{impl}}(c) \leftarrow \mathcal{A}_i^{\mathrm{phys,hi}}(c) \cup \{\varnothing\} \cup R_{\mathrm{send}}(c)$
for all $u \in F_{\mathrm{impl}}(c)$ **do**
  $V(u) \leftarrow \bar{Q}^{\mathrm{net}}(c, u)$
end for
Sample $u^\star \sim \pi(\cdot \mid c)$, where $\pi(u \mid c) \propto \mu(u \mid c) \exp(V(u)/\tau)$ on $F_{\mathrm{impl}}(c)$.
if $u^\star = (j, a)$ is a send act **then**
  Sample $m \sim \bar{\pi}_{\mathrm{text}}(\cdot \mid c, u^\star)$.
  Deliver $E^\star = (u^\star, m)$ and canonical record $\Gamma(E^\star)$.
else
  Execute local action or yield $u^\star$.
end if

A.3. Notation

The notation below separates four objects that are easily conflated: the hidden boundary snapshot, the local context seen by the actor, the finite semantic control record, and the surface wording used for human-readable communication.

Symbol	Meaning
b	Boundary index. A boundary occurs after one high-level event has completed and before the runtime commits to the next high-level event.
t_b	Physical time of boundary b. Boundary intervals may have random duration.
B^b	Control-relevant hidden boundary snapshot. It contains all persistent variables that can affect future controlled evolution.
L^b	Audit or transcript state. It may contain raw text, logs, debugging traces, or human-readable material not consumed by future control.
i_b	Ready agent or effective chooser at boundary b.
$C(B^b)$, c^b	Deployed local context observed by the ready agent. The actor must use one distribution for all hidden states inducing the same context.
$\mathcal{A}_i^{\mathrm{phys,hi}}$	Finite roundwise set of noncommunicative high-level macro-actions for agent i.
$\varnothing$	The round's explicit yield act. Its transition meaning is fixed by the runtime and continuation policy.
Σ_{sem}, $\Sigma_{\mathrm{sem}}(c)$	Finite semantic interface for the frozen round, and the subset of records legal at context c. Elements may be typed schema instances, finite bins, or finite pointers to exposed objects.
$u = (j, a)$	Semantic send act: send finite record $a \in \Sigma_{\mathrm{sem}}$ to recipient j.
$\mathcal{C}_i^{\mathrm{hi}}$	Set of communicative high-level acts available to agent i.
m	Natural-language wording or realization emitted after a send act is selected.
$\mathcal{D}_a$	Delivered-object class canonicalized as semantic record a: $\{D \in \mathcal{D} : g^\star(D) = a\}$.
$\mathcal{M}_a(c, u)$	Admissible wording set for semantic record a after fixing context c and send act $u = (j, a)$. We write $\mathcal{M}_a$ only as a local shorthand once c, u are fixed.
E^b	Completed event. If the act is a local action or yield, $E^b = U^b$; if communicative, $E^b = (U^b, M^b)$.
D^b	Delivered runtime object, including the semantic envelope and any human-readable wording transported by the runtime.
$\Gamma(D^b)$	Canonical control record consumed by future control. Raw wording may remain in L^b for audit.
G^b	Net boundary-interval return, equal to gross interval reward minus communication cost under the external-pricing convention.
Δ_b	Duration of the interval between boundaries b and $b + 1$.
$Q^\pi(B, E)$	Boundary-level action value of completed event E at hidden boundary snapshot B, followed by continuation policy π.
$Q_{\mathrm{hi}}^{\pi_0}(B, u)$	Semantic-layer value of high-level act u under fixed continuation policy π_0 and fixed wording policy. For send acts, it averages over wording.
$\bar{Q}_d(c, u)$	Conditional mean semantic value, $\mathbb{E}_d[Q_{\mathrm{hi}}^{\pi_0}(B, u) \mid C(B) = c]$, under improvement distribution d.
P_x^u, $P_x^\varnothing$	Recipient-side target laws after intervention u and after yield, conditioned on recipient abstraction x.
$\mu(\cdot \mid c)$	Full-support prior over feasible acts at context c.
τ	Temperature in the local KL-regularized objective.

Table A.2. **Core notation separates hidden state, local context, semantic records, and wording.** Symbols are grouped by the boundary process and frozen-round actor used in the theory.

A.4. Frozen-round exactness contract

All exact results are relative to a single frozen improvement round. Practical systems may update the semantic interface, context extractor, cost convention, proposer, shortlist rule, critic, or wording policy between rounds. Such updates are not forbidden; they are approximation layers outside the theorem being invoked.

Exactness Contract A.1: one frozen semantic-control round

During round k, exact statements are relative to a fixed tuple

$$\Theta_k = \big(\phi_k, \Gamma_k, g_k^\star, \hat{g}_k, \Sigma_{\mathrm{sem},k}, \mathcal{U}_k, \bar{\pi}_{\mathrm{text},k}, \pi_{0,k}, C_k, c_{\mathrm{comm},k}, \mathcal{T}_k, d_k, \mu_k, \tau_k, \mathcal{R}_k\big),$$

where:

1. ϕ_k is the control-relevant quotient of the boundary state used by semantic-invariance statements.
2. Γ_k is the canonical control-record map consumed by future control.
3. $g_k^\star$ is the ideal semantic membership map and $\hat{g}_k$ the operational parser/checker.
4. $\Sigma_{\mathrm{sem},k}$ is the finite semantic interface for the round.
5. $\mathcal{U}_k$ is the finite high-level act set exposed to the local actor.
6. $\bar{\pi}_{\mathrm{text},k}$ is the fixed wording policy, and $\pi_{0,k}$ is the fixed continuation policy.
7. C_k is the deployed local context extractor.
8. $c_{\mathrm{comm},k}$ is the explicit communication-cost convention.
9. $\mathcal{T}_k$ is the Stage I target package: recipient abstraction, decision-focused target, horizon, and any decision-sufficient coarsening.
10. d_k is the improvement distribution used by Stage II.
11. μ_k and τ_k are the local prior and temperature.
12. $\mathcal{R}_k$ contains fixed proposal, shortlist, and pruning machinery that affects the feasible set seen by Stage II.

If any component is updated, the system has entered a new round. Exact claims from round k do not automatically transfer to round $k+1$.

The contract has three practical consequences.

First, the finite semantic interface is not a claim that all useful communication can be represented forever by a fixed small schema. It is a claim that, during one improvement round, the controller optimizes over a finite exposed control layer.

Second, the local actor theorem is not a centralized full-state theorem. The deployed actor sees only $c = C(B)$. Therefore it must use one act distribution for every hidden boundary snapshot consistent with that context.

Third, wording is fixed when values are defined. If a wording-policy update changes downstream behavior, then Stage I target laws and Stage II semantic values should be refreshed before using the exact actor again.

A.5. Running example map

The C7 example from the main text is useful because it makes all layers visible without needing a full benchmark.

Formal object	C7 interpretation
Boundary b	The simulator has finished a local check and the planner is about to allocate the next batch of budget.
Hidden snapshot B^b	True candidate state, planner state, simulator history, safety constraints, tool outputs, scheduler state, and private memories.
Local context c^b	What the simulator actually sees: its instability result, relevant public plan state, and exposed runtime metadata.
Local macro-action	Run another simulation, continue local analysis, or perform a local verification step.
Yield $\varnothing$	Send no record and take no new local macro-action; the fixed continuation policy decides what event occurs next.
Send act $u = (j, a)$	Send record a to planner, literature agent, or safety checker.
Semantic record a	Example: `warn(candidate=C7, issue=instability, threshold_bin=above-450K, evidence=sim17)` Fields such as confidence, urgency, provenance, or evidence pointers are semantic if they affect downstream control.
Wording m	A natural-language realization such as "C7 appears unstable above 450K in simulation 17."
Canonical record $\Gamma(D)$	The machine-readable envelope and sanctioned fields future policies may read. Raw wording may be retained only for audit.
Stage I target Y_H	A decision-focused recipient variable, e.g., the planner's next branch choice, plan revision, or verifier state over the next H steps.
Stage II value	Net downstream return of warning, asking literature, handing off to safety, acting locally, or yielding, including communication cost.

Table A.3. **The C7 example instantiates each formal layer of the boundary-control model.** The table maps the running scenario to hidden state, local context, semantic records, wording, and Stage I/II targets.

A.6. How to read exact and approximate statements

The appendices use three kinds of statements.

Exact roundwise statements. These hold under the frozen-round contract in Appendix A.4. Examples are the average-then-softmax local actor, semantic control-equivalence, and quotient-state wording invariance.

Approximation analyses. These state how errors enter when practical systems use learned predictors, critics, shortlists, coarsened recipient abstractions, stochastic retrieval, or approximate wording support. Approximation analyses are not exactness assumptions; they are diagnostics for gaps from the exact target.

Architectural consequences and non-claims. Some statements are design consequences: future control should consume canonical records; raw wording can be a covert channel; passive logs do not identify untried acts. These statements protect the scope of the theory. They are not empirical claims about any particular runtime.

Scope caution

The paper does not claim exact control over arbitrary natural-language strings, identification from passive logs in general, or monotonic end-to-end return improvement from one local actor update. Exactness stops at the frozen finite semantic layer; practical systems must measure the gap from that layer.

B. Boundary Process, Canonical Records, and Cost

This appendix defines the boundary-level object used throughout the paper. The goal is not to model every token prefix or every low-level runtime operation. The goal is to identify the boundary after a high-level event has completed, where a ready agent's choice can change future controlled evolution. Once that boundary object is well-defined, local action, yield, and semantic send acts can be compared by ordinary boundary-level values.

B.1. Boundary-level semi-Markov process

All random variables are defined on a common probability space, and all measurable spaces may be taken to be standard Borel. At physical time t, the hidden environment and runtime state may include an environment state S_t, agent-local observations and memories, scheduler state, transport state, tool outputs, and any persistent variables used by deployed policies. The exact results below do not require a complete low-level executor. They require only the induced boundary-level kernel.

Let

$$0 = t_0 < t_1 < t_2 < \cdots$$

be the sequence of completed-event boundaries. At each boundary b, exactly one effective chooser is responsible for the high-level decision in the exact core. In the main text this chooser is the ready agent i_b, though Appendix G records one-effective-chooser variants where the chooser is a scheduler, router, or public coordinator.

The control-relevant boundary snapshot is

$$B^b = \big(S_{t_b}, Q^b, i_b, H_1^b, \ldots, H_N^b\big),$$

where Q^b is public runtime, scheduler, or transport state, and H_i^b is agent i's complete local control state at boundary b. For exactness, H_i^b contains every persistent agent-local variable that can affect future controlled evolution: observations, private tool outputs, internal scratchpad or memory, prior sent control records, prior received control records, and any other private state used by the deployed policy.

The deployed actor does not observe B^b. The ready agent observes a local execution context

$$c^b = C(B^b) = \zeta_{i_b}(H_{i_b}^b, Q^b, i_b),$$

where ζ_{i_b} is the context extractor actually used by the runtime. It may serialize memory, retrieve documents, expose tool outputs, or select a bounded prompt view. The exact local actor is optimized over policies measurable with respect to $C(B)$, not over full-state policies.

Proposition B.1 (Boundary-kernel realization and exact state sufficiency)**.** *Assume:*

1. *B^b contains all persistent variables that can affect future control-relevant evolution.*
2. *There exists a standard-Borel noise space $(\Xi, \mathcal{X})$, a stochastic kernel*

$$\Lambda(d\xi \mid B^b, L^b, E^b),$$

and measurable maps F_B, F_L, F_G, F_Δ such that, with fresh noise $\Xi_b \sim \Lambda(\cdot \mid B^b, L^b, E^b)$,

$$\begin{aligned} B^{b+1} &= F_B(B^b, L^b, E^b, \Xi_b), \\ L^{b+1} &= F_L(B^b, L^b, E^b, \Xi_b), \\ G^b &= F_G(B^b, L^b, E^b, \Xi_b), \\ \Delta_b &= F_\Delta(B^b, L^b, E^b, \Xi_b). \end{aligned}$$

3. *For control-level values stated only as functions of B, the marginal law of (B^{b+1}, G^b, Δ_b) factors through (B^b, E^b): for every L, L' compatible with B, E,*

$$\begin{aligned} &\mathsf{Law}\Big(B^{b+1}, G^b, \Delta_b \,\Big|\, {}^{B^b=B,\ L^b=L,}_{E^b=E}\Big) \\ &\quad = \mathsf{Law}\Big(B^{b+1}, G^b, \Delta_b \,\Big|\, {}^{B^b=B,\ L^b=L',}_{E^b=E}\Big). \end{aligned}$$

4. *If full local histories are replaced by summaries, the induced control-relevant boundary kernel factors through those summaries.*

Then there exists a well-defined semi-Markov kernel

$$\mathcal{K}_G(dB', dL', dG, d\Delta \mid B, L, E)$$

such that, for every measurable set $A \subseteq \mathcal{B} \times \mathcal{L} \times \mathbb{R} \times \mathbb{N}_+$,

$$\Pr\big((B^{b+1}, L^{b+1}, G^b, \Delta_b) \in A \mid B^0, L^0, E^0, \ldots, B^b, L^b, E^b\big) = \Pr\big((B^{b+1}, L^{b+1}, G^b, \Delta_b) \in A \mid B^b, L^b, E^b\big).$$

Moreover, item 3 induces a control-level marginal kernel

$$\mathcal{K}_G^{\mathrm{ctrl}}(dB', dG, d\Delta \mid B, E)$$

for the variables entering $Q^\pi(B, E)$. If item 3 fails, the exact control state is (B, L), or B must be enlarged to include the control-relevant part of L. If a summary is used without the factorization property in item 4, the summarized boundary process is not exact in general.

Proof. Define the combined map

$$F = (F_B, F_L, F_G, F_\Delta).$$

For fixed (B, L, E), the law of

$$F(B, L, E, \Xi), \qquad \Xi \sim \Lambda(\cdot \mid B, L, E),$$

is a probability measure on $\mathcal{B} \times \mathcal{L} \times \mathbb{R} \times \mathbb{N}_+$. Define

$$\mathcal{K}_G(A \mid B, L, E) := \int \mathbf{1}[F(B, L, E, \xi) \in A]\, \Lambda(d\xi \mid B, L, E).$$

Because Ξ_b is fresh and its conditional law depends on the earlier past only through (B^b, L^b, E^b), the conditional law of the next boundary variables depends only on (B^b, L^b, E^b). This proves the semi-Markov property.

Item 3 says that the projection of this kernel onto (B', G, Δ) is the same for all audit states L compatible with (B, E). This common projected law defines $\mathcal{K}_G^{\mathrm{ctrl}}(dB', dG, d\Delta \mid B, E)$. If the projected law changes with L, then L is control-relevant and cannot be omitted from the exact value state.

If full local histories are replaced by summaries and the kernel does not factor through those summaries, then two histories with the same summary can induce different next-boundary laws. The summarized state is then not Markov for the controlled process. □

B.2. Control state versus audit state

The paper separates the exact control state from the audit transcript. This distinction is required for finite semantic records to be exact control objects while still permitting rich human-readable logs.

Control plane versus audit plane

The exact control theorems are stated on B^b, or on a quotient $\phi(B^b)$ of B^b. The audit state L^b may contain raw wording, full transcripts, debugging traces, and human-readable logs. Future control should consume the canonical record $\Gamma(D^b)$, or a specified quotient of the delivered event, rather than unrestricted raw wording. Otherwise nominally harmless rephrasings can become hidden control channels.

In the strongest exact-core architecture, raw wording remains in L^b for audit, user display, or replay inspection, but it is not an input to future policies. Future policies condition on control-side objects: semantic envelopes, typed fields, sanctioned auxiliary fields, and explicit pointers. If a system instead lets future policies read arbitrary retained wording, then the semantic abstraction must account for that wording, or the exact record/wording separation fails; Proposition C.5 formalizes this point.

B.3. High-level acts, completed events, and canonical records

For the ready agent $i = i_b$, the roundwise high-level act set is

$$\mathcal{U}_i^{\mathrm{hi}} = \mathcal{A}_i^{\mathrm{phys,hi}} \cup \{\varnothing\} \cup \mathcal{C}_i^{\mathrm{hi}}, \qquad \mathcal{C}_i^{\mathrm{hi}} = \{(j, a) : j \neq i,\ a \in \Sigma_{\mathrm{sem}}\}.$$

Here $\mathcal{A}_i^{\mathrm{phys,hi}}$ is the finite set of noncommunicative macro-actions exposed during the round. The act $\varnothing$ is the round's explicit yield act. It is not an undefined null action; it is a specified boundary act whose transition behavior is defined by the runtime and fixed continuation policy. In runtimes where every nonsend means local continuation, $\varnothing$ can be folded into $\mathcal{A}_i^{\mathrm{phys,hi}}$. We keep it explicit because Stage I compares candidate sends to a yield baseline.

A chosen high-level act becomes a completed event. If the act is noncommunicative,

$$E^b = U^b \qquad \text{for } U^b \in \mathcal{A}_i^{\mathrm{phys,hi}} \cup \{\varnothing\}.$$

If the act is communicative, $U^b = (j, a)$, the runtime constructs a control envelope and a surface realization. A convenient singleton-recipient representation is

$$A^b = (i_b, j, a, \kappa^b), \qquad M^b \in \Sigma_{\mathrm{nat}}^*, \qquad E^b = (A^b, M^b),$$

where κ^b is an optional finite bundle of sanctioned control-side fields, such as confidence bins, urgency bins, provenance flags, or pointers to objects exposed during the round. The main text often suppresses κ^b and writes $u = (j, a)$.

The delivered runtime object is

$$D^b = (A^b, M^b).$$

The canonical control record is

$$\Gamma(D^b).$$

In the strongest exact architecture, $\Gamma(D^b)$ contains the semantic envelope A^b and only the auxiliary fields intentionally allowed to affect future control. The natural-language

text M^b can be retained in the audit state, but future control is not allowed to extract additional unsanctioned task-relevant bits from it.

B.4. Finite semantic interfaces, schema fields, and pointers

The exact Stage II theorem requires the high-level act set exposed during the round to be finite. This includes both the physical macro-action set and the communicative semantic records. The round exposes a finite inventory Σ_{sem}; when legality depends on context, $\Sigma_{\text{sem}}(c) \subseteq \Sigma_{\text{sem}}$ denotes the records legal at context c.

A semantic record $a \in \Sigma_{\text{sem}}$ may be more structured than a coarse intent label. It may include:

- an act type, such as `warn`, `ask`, `handoff`, or `verify`;
- a finite recipient or role field;
- finite slots or bins, such as threshold bins, confidence bins, urgency bins, or provenance categories;
- finite pointers to objects exposed during the round.

Pointers require care. A pointer index is finite only relative to a fixed exposed inventory. If selecting a pointer reveals previously private unbounded content, then the referenced content must be treated as part of the communication interface, part of the communication cost, or an explicit approximation gap. Otherwise a finite-looking record can transmit unbounded information through the pointer target.

The ideal semantic membership map can be written as

$$g^\star : \mathcal{D} \to \Sigma_{\text{sem}}^{\perp}, \qquad \Sigma_{\text{sem}}^{\perp} = \Sigma_{\text{sem}} \cup \{\perp\},$$

where $\mathcal{D}$ is the delivered-object space and $\perp$ denotes invalid, ambiguous, or out-of-schema realizations. For a record a, its canonical delivered-object class is

$$\mathcal{D}_a = \{D \in \mathcal{D} : g^\star(D) = a\}.$$

After a context c and semantic send act $u = (j, a)$ fix the control envelope A_u, the admissible wording set is

$$\mathcal{M}_a(c, u) = \{m : (A_u, m) \in \mathcal{D}_a\}.$$

When c, u are fixed, later wording results write $\mathcal{M}_a$ as shorthand for $\mathcal{M}_a(c, u)$. If the runtime carries an explicit semantic envelope, $g^\star$ may be trivial on the envelope and used only to certify that the accompanying text is an admissible gloss. If the runtime transports only raw text, then $g^\star$ becomes a parser, and parser inconsistency, support leakage, and within-class covert channels become central approximation risks.

If a phrasing changes control, it is not harmless wording

Relative to a fixed interface, a feature of a delivered message counts as wording only if changing that feature while staying in the same semantic class does not materially change control-relevant future dynamics, except through explicitly sanctioned channel-side terms such as latency or compression cost. If confidence, urgency, threshold precision, provenance, or evidence strength changes downstream control, it belongs in the semantic record or in a measured approximation term.

B.5. Communication cost convention

Let R^b denote the discounted gross interval reward accumulated between boundaries b and $b+1$:

$$R^b = \sum_{k=0}^{\Delta_b - 1} \gamma^k r_{t_b + k}.$$

The paper uses an external-pricing convention for communication burden:

$$G^b = R^b - c_{\text{comm}}(E^b),$$

where G^b is the net boundary-interval return used in boundary-level values. Equivalently, communication cost may be absorbed into the environment reward if that convention is used consistently. The important point is to avoid double counting.

A typed communication cost can include more than token count:

$$c_{\text{comm}}(E^b) = \mathbf{1}[E^b \text{ communicative}]\Big(c_{\text{call}}(A^b, Q^b) + \lambda_{\text{rate}} \ell_{\text{ref}}(D^b \mid C^b_{\text{code}}) + \lambda_{\text{ctx}} \operatorname{occ}(D^b) + \lambda_{\text{lat}} \Delta_b^{\text{comm}}\Big) + \lambda_{\text{act}} \mathbf{1}[\text{wake}].$$

Here c_{call} charges invocation or routing overhead, ℓ_{ref} is a reference code length conditioned on a shared or public pre-delivery coder context, occ measures future context occupancy, Δ_b^{comm} measures communication-induced latency, and the final term charges waking dormant agents. The indicator matters: noncommunicative macro-actions should not be charged as communication unless the cost model says so explicitly.

Cost accounting

Latency, context occupancy, and activation cost should be charged either through the environment reward or through c_{comm}, not both. Stage II values in the paper are net values unless explicitly called gross values. When pruning compares a local value-gain upper bound to $\underline{c}_{\text{comm}}(u)$, that lower bound must use the same cost convention as G to avoid double counting.

Any code-length proxy should condition only on information shared before delivery. A sender-private code can make an informative message appear artificially cheap: if

the sender privately knows bit S and sends $D = S$, then $H(D \mid S) = 0$, even though the recipient learns one full bit. Thus sender-private conditioning is not a valid measure of recipient-side communication burden unless that private context is itself shared or separately transmitted.

B.6. Boundary values and semantic-layer values

The following values use the control-level marginal sufficiency condition in Proposition B.1. If the audit state changes the law of (B', G, Δ) after conditioning on (B, E), then these definitions should be read with (B, L) in place of B, or with B enlarged accordingly.

Given a continuation policy π, define the boundary value

$$V^\pi(B) = \mathbb{E}_\pi\left[\sum_{r\geq 0}\gamma^{t_{b+r}-t_b}G^{b+r}\,\middle|\, B^b = B\right],$$

assuming the usual boundedness or termination conditions needed for the expectation to exist. The completed-event action value is

$$Q^\pi(B,E) = \mathbb{E}\big[G^b + \gamma^{\Delta_b}V^\pi(B^{b+1})\,\big|\, B^b = B,\ E^b = E\big].$$

The high-level semantic actor chooses u, not the exact wording. Therefore, under a fixed wording policy $\bar{\pi}_{\text{text}}$, the semantic-layer value of a high-level act is

$$Q^{\pi_0}_{\text{hi}}(B,u) = \begin{cases} Q^{\pi_0}(B,u), & u \in \mathcal{A}^{\text{phys,hi}}_i \cup \{\varnothing\}, \\ \mathbb{E}_{M\sim\bar{\pi}_{\text{text}}(\cdot\mid C(B),u)}\left[Q^{\pi_0}(B,(u,M))\right], & u \in \mathcal{C}^{\text{hi}}_i \end{cases}$$

This definition is the bridge between the completed-event boundary process and the finite semantic actor. It also explains why wording-policy updates change the semantic values: if $\bar{\pi}_{\text{text}}$ changes, the expectation defining $Q^{\pi_0}_{\text{hi}}$ changes.

For a local context c, Stage II uses the conditional mean

$$\bar{Q}_d(c,u) = \mathbb{E}_{B\sim d}\left[Q^{\pi_0}_{\text{hi}}(B,u)\,|\,C(B) = c\right],$$

where d is the improvement distribution for the round. Different choices of d define different local improvement targets; distribution-shift sensitivity is analyzed in Appendix F.

C. Semantic Abstraction, Wording Validity, and Interface Refinement

The finite semantic interface is not merely a human-designed label set. It is an action abstraction. A semantic record is exact only when the delivered events it groups together are interchangeable for the future control problem being solved. This appendix makes that statement formal.

The key idea is to distinguish raw delivered events from the control-relevant quotient of their consequences. If future policies consume canonical records or quotient states, harmless wording can be ignored. If future policies may read unrestricted raw text, nominally equivalent rephrasings can become covert control channels.

C.1. Semantic records as control-equivalence quotients

Let e denote a completed event. For a local action or yield act, $e = u$. For a communicative act, $e = (u, m)$, where $u = (j, a)$ is the semantic send act and m is the wording.

Let $\phi : \mathcal{B} \to \Phi$ be a control-relevant quotient of the boundary state. For example, $\phi(B)$ may retain all state variables future policies can read while discarding raw audit text. For a boundary snapshot B and completed event e, define the quotient next-boundary kernel

$$K^\phi_{B,e} = \mathsf{Law}\big(\phi(B'), G, \Delta \mid B, e\big),$$

a probability measure on

$$\mathcal{Y}_\phi := \Phi \times \mathbb{R} \times \mathbb{N}_+.$$

Definition C.1 (Control equivalence)**.** *Two completed events e and e' are ϕ-control-equivalent on a set of boundary states $\mathcal{B}_0$ if*

$$K^\phi_{B,e} = K^\phi_{B,e'} \qquad \textit{for every } B \in \mathcal{B}_0.$$

When $\mathcal{B}_0$ is clear from context, we write $e \sim_\phi e'$.

Theorem C.2 (Semantic records as control-equivalence quotients)**.** *Fix a policy class Π_ϕ such that for every $\pi \in \Pi_\phi$ there exists a measurable $\widetilde{V}^\pi : \Phi \to \mathbb{R}$ with*

$$V^\pi(B') = \widetilde{V}^\pi(\phi(B')) \qquad \textit{for all } B'.$$

If $e \sim_\phi e'$ at boundary state B, then

$$Q^\pi(B,e) = Q^\pi(B,e') \qquad \textit{for every } \pi \in \Pi_\phi.$$

If instead

$$\left\|K^\phi_{B,e} - K^\phi_{B,e'}\right\|_{\text{TV}} \leq \varepsilon$$

and the Bellman integrand

$$h_\pi(\varphi, g, \Delta) = g + \gamma^\Delta \widetilde{V}^\pi(\varphi)$$

has span at most S_π, then

$$|Q^\pi(B,e) - Q^\pi(B,e')| \leq S_\pi\varepsilon.$$

Proof. Because $V^\pi(B') = \widetilde{V}^\pi(\phi(B'))$,

$$Q^\pi(B,e) = \int\Big(g + \gamma^\Delta\widetilde{V}^\pi(\varphi)\Big)\,K^\phi_{B,e}(d\varphi, dg, d\Delta).$$

If $K^{\phi}_{B,e} = K^{\phi}_{B,e'}$, the two integrals are identical.

For the approximate statement, let

$$h_\pi(\varphi, g, \Delta) = g + \gamma^{\Delta} \widetilde{V}^{\pi}(\varphi).$$

Since h_π has span at most S_π, shifting it by a constant gives a function in $[0, S_\pi]$ without changing the difference of expectations. By total-variation duality for bounded functions,

$$\left| \int h_\pi \, dK^{\phi}_{B,e} - \int h_\pi \, dK^{\phi}_{B,e'} \right| \le S_\pi \left\| K^{\phi}_{B,e} - K^{\phi}_{B,e'} \right\|_{\mathrm{TV}} \le S_\pi.$$

□

Proposition C.3 (Kernel equality is the strongest event-level abstraction test)**.** *For fixed B, e, e', the following are equivalent:*

$$K^{\phi}_{B,e} = K^{\phi}_{B,e'}$$

and

$$\int h \, dK^{\phi}_{B,e} = \int h \, dK^{\phi}_{B,e'} \qquad \text{for every bounded measurable } h : \mathcal{Y}_\phi \to \mathbb{R}.$$

Proof. If the kernels are equal, the equality of integrals is immediate. Conversely, take $h = \mathbf{1}_A$ for each measurable set $A \subseteq \mathcal{Y}_\phi$. Then

$$K^{\phi}_{B,e}(A) = K^{\phi}_{B,e'}(A)$$

for every measurable A, so the probability measures are equal. □

Corollary C.4 (Maximal mergeable semantic classes)**.** *For a fixed quotient ϕ and boundary-state set $\mathcal{B}_0$, the relation $e \sim_\phi e'$ is an equivalence relation. Its equivalence classes are the maximal event classes that can be merged while preserving exact interchangeability for all continuation policies whose values factor through ϕ, provided equality of quotient kernels is required for every $B \in \mathcal{B}_0$.*

Proof. Reflexivity, symmetry, and transitivity follow from equality of probability measures. Theorem C.2 gives interchangeability for all policies whose values factor through ϕ. Proposition C.3 shows that equality of quotient kernels is the strongest event-level abstraction test based on all bounded Bellman test functions on $\mathcal{Y}_\phi$. □

Finite semantic interfaces are finite quotients or finite approximations

The exact control-equivalence quotient need not be finite. If raw wording or arbitrary hidden transcript features affect future control, the exact quotient may be as large as the delivered-event space. The paper's exact finite-interface claims apply when the round exposes a finite quotient or a finite value-equivalent abstraction. Otherwise, the finite interface is an approximation whose within-class drift must be measured.

C.2. Why unrestricted raw wording breaks exact semantic classes

The quotient view explains why future control should consume canonical records rather than unrestricted raw wording. If raw wording is available to future policies, then wording can be used as a hidden channel even when all wordings parse to the same semantic record.

Proposition C.5 (Pure-text exact semantics is impossible if future control can read raw wording)**.** *Suppose:*

1. *For a fixed context c and send act $u = (j, a)$, the admissible wording set $\mathcal{M}_a(c, u)$ contains two distinct realizations $m_0 \neq m_1$.*
2. *After taking (u, m_r) for $r \in \{0, 1\}$, future control can observe the retained raw wording m_r.*
3. *There is a later decision information state, reached with positive probability after both (u, m_0) and (u, m_1), identical in all non-wording control variables but retaining m_r, at which two feasible continuation actions a_0, a_1 have different continuation returns with nonzero expected contribution.*

Then there exists a continuation policy π such that

$$Q^{\pi}(B, (u, m_0)) \neq Q^{\pi}(B, (u, m_1)).$$

Proof. Construct a continuation policy that, at the later decision state, chooses a_0 if the retained wording is m_0 and chooses a_1 if the retained wording is m_1. By assumption, the non-wording control variables are the same at that later decision state, the state is reached with positive probability, and the two continuation actions have different returns with nonzero expected contribution. Therefore the continuation value after (u, m_0) differs from the continuation value after (u, m_1). Hence

$$Q^{\pi}(B, (u, m_0)) \neq Q^{\pi}(B, (u, m_1)).$$

□

The proposition is not saying that raw text can never be stored. It is saying that if future policies may condition on raw wording, then raw wording is part of the control state. A semantic class that ignores raw wording is exact only relative to a quotient or policy class that does not use those raw wording differences.

C.3. Quotient-state wording invariance

We now specialize the control-equivalence theorem to the record/wording split. Fix semantic send act $u = (j, a)$. For any fixed boundary state B, write $\mathcal{M}_a$ as local shorthand for $\mathcal{M}_a(C(B), u)$. Let $m \in \mathcal{M}_a$ be an admissible wording. Define

$$K^{\phi}_{B,u,m} = \mathsf{Law}\big(\phi(B'), G, \Delta \mid B, (u, m)\big).$$

Theorem C.6 (Quotient-state wording invariance)**.** *Fix a semantic send act $u = (j, a)$, a control-relevant quotient ϕ, and a policy class Π_ϕ whose values factor through ϕ. If, for every B and all $m, m' \in \mathcal{M}_a$,*

$$K^{\phi}_{B,u,m} = K^{\phi}_{B,u,m'},$$

then

$$Q^{\pi}(B, (u, m)) = Q^{\pi}(B, (u, m')) \qquad \text{for every } \pi \in \Pi_\phi.$$

If instead

$$\left\| K^{\phi}_{B,u,m} - K^{\phi}_{B,u,m'} \right\|_{\mathrm{TV}} \le \varepsilon_\phi$$

and

$$g + \gamma^{\Delta} \widetilde{V}^{\pi}(\varphi) \in [L_\Pi, U_\Pi]$$

for all relevant (φ, g, Δ) and $\pi \in \Pi_\phi$, then

$$\sup_{\pi \in \Pi_\phi} |Q^{\pi}(B, (u, m)) - Q^{\pi}(B, (u, m'))| \le (U_\Pi - L_\Pi)\varepsilon_\phi.$$

Proof. This is Theorem C.2 applied to the completed events $e = (u, m)$ and $e' = (u, m')$. The approximate statement uses the span bound $S_\pi \le U_\Pi - L_\Pi$. □

The full-state kernel theorem in the main text is the special case $\phi(B) = B$. The quotient version is often the practically relevant statement. In a raw transcript state, two different wordings usually produce different raw states even when they have identical future control meaning. The exact semantic architecture therefore stores or exposes the canonical control record $\Gamma(D)$ to future policies and places raw text in the audit state.

C.4. Semantic drift versus channel-side variation

A wording change can be useful without changing semantic control content. For example, a shorter wording may reduce context occupancy, or a clearer wording may reduce latency or parsing failure, while leaving the recipient's control-relevant state unchanged. The paper therefore separates semantic drift from channel-side variation.

Proposition C.7 (Control-state invariance with sanctioned channel-side variation)**.** *Fix boundary snapshot B, send act $u = (j, a)$, and policy class Π_ϕ. Suppose that for every realization $m \in \mathcal{M}_a$ and policy $\pi \in \Pi_\phi$,*

$$Q^{\pi}(B, (u, m)) = \kappa(B, u, m) + \int \Big(g + \gamma^{\Delta} \widetilde{V}^{\pi}(\varphi') \Big) K^{\phi}_{B,u}(d\varphi', dg, d\Delta),$$

where the quotient kernel $K^{\phi}_{B,u}$ does not depend on m, while $\kappa(B, u, m)$ is an explicitly sanctioned channel-side term. Here g excludes the channel-side term represented by κ, or equivalently κ is the difference between two stated cost conventions. Then for all $m, m' \in \mathcal{M}_a$,

$$Q^{\pi}(B, (u, m)) - Q^{\pi}(B, (u, m')) = \kappa(B, u, m) - \kappa(B, u, m').$$

Thus within-class value variation is purely channel-side under this decomposition.

Proof. The integral term is identical for m and m', because $K^{\phi}_{B,u}$ does not depend on the wording. Subtracting the two value expressions leaves exactly the difference of the channel-side terms. □

This proposition is useful for wording optimization. It says that a realization model may improve compression, latency, or context occupancy inside a semantic class without inducing semantic drift, provided those effects are represented as sanctioned channel-side terms and the control-relevant quotient dynamics remain fixed.

C.5. Within-class value spread and support leakage

Exact quotient invariance is strong. In practical systems, the relevant question is often whether within-class variation is small enough to treat as approximation error.

Proposition C.8 (Within-class realization drift is controlled by value spread)**.** *Fix boundary snapshot B, send act $u = (j, a)$, and continuation policy π_0. Suppose*

$$\sup_{m, m' \in \mathcal{M}_a} |Q^{\pi_0}(B, (u, m)) - Q^{\pi_0}(B, (u, m'))| \le \varepsilon_{\mathrm{sem}}.$$

Then for any two realization policies ρ, ρ' supported on $\mathcal{M}_a$,

$$|\mathbb{E}_{m \sim \rho} Q^{\pi_0}(B, (u, m)) - \mathbb{E}_{m \sim \rho'} Q^{\pi_0}(B, (u, m))| \le \varepsilon_{\mathrm{sem}}.$$

Proof. The values

$$v(m) = Q^{\pi_0}(B, (u, m)), \qquad m \in \mathcal{M}_a,$$

lie in an interval of width at most $\varepsilon_{\mathrm{sem}}$. The expectation of v under any distribution supported on $\mathcal{M}_a$ lies in the same interval. Therefore the difference between two such expectations is at most $\varepsilon_{\mathrm{sem}}$. □

Corollary C.9 (Support leakage adds an outside-mass penalty)**.** *Fix B and $u = (j, a)$. Assume the within-class spread bound in Proposition C.8 and suppose the full realization-support value span is at most S_Q:*

$$\mathrm{sp}\big(Q^{\pi_0}(B, (u, \cdot))\big) \le S_Q.$$

Let $\eta \in [0, 1]$, and let ρ, ρ' be two realization policies satisfying

$$\rho(\mathcal{M}_a) \ge 1 - \eta, \qquad \rho'(\mathcal{M}_a) \ge 1 - \eta.$$

Then

$$|\mathbb{E}_\rho Q^{\pi_0}(B, (u, M)) - \mathbb{E}_{\rho'} Q^{\pi_0}(B, (u, M))| \le \varepsilon_{\mathrm{sem}} + 2\eta S_Q.$$

Proof. Let $v(m) = Q^{\pi_0}(B, (u, m))$. Let ρ_a and ρ'_a be the normalized restrictions of ρ and ρ' to $\mathcal{M}_a$, whenever the corresponding mass is positive. If a mass is zero, the condition $\rho(\mathcal{M}_a) \ge 1 - \eta$ implies $\eta = 1$, and the bound is trivial.

Because v has span at most S_Q, replacing ρ by its normalized restriction ρ_a changes the expectation by at most

$$S_Q\, \rho(\mathcal{M}_a^c) \le \eta S_Q.$$

The same holds for ρ'. By Proposition C.8,

$$\big|\mathbb{E}_{\rho_a} v - \mathbb{E}_{\rho'_a} v\big| \le \varepsilon_{\mathrm{sem}}.$$

The triangle inequality gives

$$|\mathbb{E}_\rho v - \mathbb{E}_{\rho'} v| \le \eta S_Q + \varepsilon_{\mathrm{sem}} + \eta S_Q = \varepsilon_{\mathrm{sem}} + 2\eta S_Q.$$

□

Wording diagnostics

A wording update should report three quantities when possible:

1. support leakage, $\Pr[M \notin \mathcal{M}_a(c, u)]$;
2. within-class value spread, $\sup_{m, m' \in \mathcal{M}_a(c,u)} |Q(B, (u, m)) - Q(B, (u, m'))|$;
3. whether observed value changes are semantic drift or sanctioned channel-side variation.

C.6. Soft option value of interface refinement

When within-class wording or pragmatic variation changes downstream control, the right response is to refine the semantic interface. The next result quantifies the value of exposing a hidden within-class distinction as an explicit finite semantic choice. A semantic field is not chosen because it is linguistically salient; it is chosen because changing it changes downstream boundary value.

Let a coarse semantic record a hide a finite set of refined records

$$G_a = \{v_1, \dots, v_K\}.$$

Let $\rho_a(\cdot \mid c)$ be a full-support reference law over G_a, and let $Q(c, v)$ be the fixed refined value of variant v at context c. If the coarse interface exposes only a, with reference realization over the variants, its coarse value is

$$\bar{Q}_a(c) = \sum_{v \in G_a} \rho_a(v \mid c) Q(c, v).$$

The comparison is therefore against the coarse interface whose act a realizes refined variants according to ρ_a. If a deployed coarse act has an independently estimated value not equal to this reference average, the refinement gain below compares to the reference-realization coarse act, not to that independently estimated value.

If the refined interface lets the controller choose $v \in G_a$, regularized to ρ_a, its inner group value is

$$V_a(c) = \max_{r \in \Delta(G_a)} \left[\sum_{v \in G_a} r(v) Q(c, v) - \tau \mathrm{KL}(r \| \rho_a(\cdot \mid c)) \right].$$

Theorem C.10 (Soft option value of semantic-interface refinement)**.** *Under the setup above,*

$$V_a(c) = \tau \log \sum_{v \in G_a} \rho_a(v \mid c) e^{Q(c,v)/\tau}.$$

The inner refinement gain is

$$\Delta_{\mathrm{refine}}(c, a) = V_a(c) - \bar{Q}_a(c) = \tau \log \mathbb{E}_{v \sim \rho_a(\cdot \mid c)} \left[e^{(Q(c,v) - \bar{Q}_a(c))/\tau} \right] \ge 0.$$

It is zero if and only if $Q(c, v)$ is constant on the support of $\rho_a(\cdot \mid c)$.

If the within-group value span is

$$S_a(c) = \max_{v \in G_a} Q(c, v) - \min_{v \in G_a} Q(c, v),$$

then

$$0 \le \Delta_{\mathrm{refine}}(c, a) \le \min \left\{ S_a(c),\ \frac{S_a(c)^2}{8\tau} \right\}.$$

For small within-group variation or large τ,

$$\Delta_{\mathrm{refine}}(c, a) = \frac{\mathrm{Var}_{\rho_a}(Q(c, V))}{2\tau} + O\left(\frac{\mathbb{E}_{\rho_a} |Q(c, V) - \bar{Q}_a(c)|^3}{\tau^2} \right).$$

Proof. The expression for $V_a(c)$ is the finite KL-soft-value identity applied to action set G_a, prior $\rho_a(\cdot \mid c)$, and value $Q(c, \cdot)$.

Let

$$X = Q(c, V), \qquad V \sim \rho_a(\cdot \mid c).$$

Then $\bar{Q}_a(c) = \mathbb{E}[X]$ and

$$V_a(c) = \tau \log \mathbb{E}[e^{X/\tau}].$$

Thus

$$\Delta_{\text{refine}}(c, a) = \tau \log \mathbb{E}[e^{X/\tau}] - \mathbb{E}[X].$$

Let $X_0 = X - \mathbb{E}[X]$. Then

$$\Delta_{\text{refine}}(c, a) = \tau \log \mathbb{E}[e^{X_0/\tau}].$$

By Jensen's inequality,

$$\mathbb{E}[e^{X_0/\tau}] \geq e^{\mathbb{E}[X_0]/\tau} = 1,$$

so $\Delta_{\text{refine}}(c, a) \geq 0$. Equality in Jensen holds if and only if X_0 is constant almost surely, equivalently $Q(c, v)$ is constant on the support of ρ_a.

For the $S_a(c)^2/(8\tau)$ bound, X_0 is mean-zero and lies in an interval of width $S_a(c)$. Hoeffding's lemma gives

$$\log \mathbb{E}[e^{\lambda X_0}] \leq \frac{\lambda^2 S_a(c)^2}{8}.$$

Taking $\lambda = 1/\tau$ yields

$$\Delta_{\text{refine}}(c, a) \leq \tau \cdot \frac{S_a(c)^2}{8\tau^2} = \frac{S_a(c)^2}{8\tau}.$$

Also, $V_a(c) \leq \max_v Q(c, v)$, while $\bar{Q}_a(c) \geq \min_v Q(c, v)$, so

$$\Delta_{\text{refine}}(c, a) \leq S_a(c).$$

The expansion follows from the cumulant expansion

$$\log \mathbb{E}[e^{\lambda X_0}] = \frac{\lambda^2}{2} \operatorname{Var}(X) + O(\lambda^3 \mathbb{E}|X_0|^3)$$

with $\lambda = 1/\tau$, multiplied by τ. □

Corollary C.11 (Global soft-value gain from splitting one coarse record)**.** *Suppose the top-level coarse act set contains record a and other acts $\mathcal{O}$, with a's coarse value equal to the ρ_a-averaged refined value $\bar{Q}_a(c)$. Let the old partition function be*

$$Z_{\text{old}} = \sum_{w \in \mathcal{O}} \mu(w \mid c) e^{Q(c,w)/\tau} + \mu(a \mid c) e^{\bar{Q}_a(c)/\tau}.$$

After replacing a by refined acts $v \in G_a$ with prior

$$\widetilde{\mu}(v \mid c) = \mu(a \mid c) \rho_a(v \mid c),$$

the new partition function is

$$Z_{\text{new}} = \sum_{w \in \mathcal{O}} \mu(w \mid c) e^{Q(c,w)/\tau} + \mu(a \mid c) e^{V_a(c)/\tau}.$$

The global local-soft-value gain is

$$\Delta_{\text{global}}(c, a) = \tau \log \frac{Z_{\text{new}}}{Z_{\text{old}}} = \tau \log \left(1 + p_{\text{old}}(a \mid c) \left[e^{\Delta_{\text{refine}}(c,a)/\tau} - 1\right]\right)$$

where

$$p_{\text{old}}(a \mid c) = \frac{\mu(a \mid c) e^{\bar{Q}_a(c)/\tau}}{Z_{\text{old}}}.$$

In particular,

$$0 \leq \Delta_{\text{global}}(c, a) \leq \Delta_{\text{refine}}(c, a).$$

Proof. The expression for Z_{new} follows from

$$\sum_{v \in G_a} \rho_a(v \mid c) e^{Q(c,v)/\tau} = e^{V_a(c)/\tau}.$$

Then

$$\frac{Z_{\text{new}}}{Z_{\text{old}}} = 1 + \frac{\mu(a \mid c) e^{\bar{Q}_a(c)/\tau}}{Z_{\text{old}}} \left(e^{(V_a(c) - \bar{Q}_a(c))/\tau} - 1\right),$$

which is the displayed formula. Nonnegativity follows from $\Delta_{\text{refine}} \geq 0$. The upper bound $\Delta_{\text{global}} \leq \Delta_{\text{refine}}$ follows because $0 \leq p_{\text{old}}(a \mid c) \leq 1$, so

$$1 + p_{\text{old}}(a \mid c)(e^x - 1) \leq e^x \qquad \text{for } x \geq 0.$$

Multiplying the logarithm by τ gives the claim. □

Interface-refinement rule

A field should be promoted into the semantic record when the value of exposing it exceeds the added interface, estimation, parser, and communication burden. The refinement gain is driven by within-class relative value variation:

$$\Delta_{\text{refine}}(c, a) \approx \frac{\operatorname{Var}_{\rho_a}(Q(c, V))}{2\tau}.$$

If all variants inside a class have the same value up to negligible error, splitting the class adds little control value.

C.7. Semantic-interface diagnostics

The semantic abstraction can fail in several ways. The following table summarizes the most important diagnostics.

D. Interventional Identification and Boundary-Data Contracts

Stage I and Stage II are not ordinary supervised labels. They are interventional boundary quantities. Stage I asks what a recipient-side target would do if a candidate semantic act were sent rather than yield. Stage II asks what downstream net value would follow if the ready agent took a candidate high-level act at the boundary. Passive logs identify these quantities only under additional causal assumptions. This appendix states the estimands, gives exact identification regimes, corrects the observational identification statement, and records a non-identification result that explains why candidate-complete boundary bundles are needed.

Question	Failure mode	Diagnostic or response
Does future control read raw text?	Within-class wording becomes a covert channel.	Move raw text to audit state, consume $\Gamma(D)$, or use a quotient-state theorem.
Do two wordings induce different recipient actions?	The interface is too coarse.	Promote the differing feature into a semantic field, or measure within-class value spread.
Does a wording update leave the admissible class?	Support leakage.	Track $\Pr[M \notin \mathcal{M}_a(c,u)]$ and apply leakage bounds such as Corollary C.9.
Does wording only change latency or compression?	Channel-side variation, not semantic drift.	Represent the effect through sanctioned $\kappa(B,u,m)$ or the communication-cost model.
Does a pointer reveal private unbounded content?	Finite-looking record transmits unbounded information.	Treat target content as part of the interface, cost, or approximation gap.
Should a class be split?	Within-class variants have different downstream values.	Estimate the refinement gain in Theorem C.10.

Table C.1. **Semantic-interface failures are diagnosed by checking whether wording changes future control.** Each row pairs a failure mode with the corresponding response or measurement.

Unproposed acts are not identified

If a legal act u is never proposed, forced, replayed, randomized, or structurally modeled at context c, its interventional value at c is not identified from passive runtime logs. A runtime candidate pool is therefore not the same object as the legal semantic inventory.

D.1. The two interventional estimands

There are two distinct targets.

Stage I target laws. For a recipient abstraction x, candidate send act $u = (j, a)$, yield act $\varnothing$, fixed continuation policy π_0, and fixed wording policy $\bar{\pi}_{\text{text}}$, Stage I uses the interventional laws

$$P_x^u = \mathsf{Law}^{\pi_0,\bar{\pi}_{\text{text}}}(Y_H \mid X_j = x, \mathrm{do}(U = u)),$$

and

$$P_x^{\varnothing} = \mathsf{Law}^{\pi_0,\bar{\pi}_{\text{text}}}(Y_H \mid X_j = x, \mathrm{do}(U = \varnothing)).$$

The target Y_H should be decision-focused: the recipient's next high-level action, plan revision, artifact state, verifier state, or another quantity future control actually uses.

Stage II conditional mean semantic values. For a fixed continuation policy π_0, wording policy $\bar{\pi}_{\text{text}}$, and improvement distribution d over boundary snapshots, Stage II uses

$$\bar{Q}_d(c,u) = \mathbb{E}_{B\sim d}\left[Q_{\text{hi}}^{\pi_0}(B,u) \,|\, C(B) = c\right].$$

This is not generally a pointwise full-state value. It is the exact target for a deployed actor that observes $c = C(B)$ and must use one action distribution for all hidden states sharing c.

These two targets are related but not interchangeable. Stage I measures decision-relevant predictive change relative to yield. Stage II measures downstream net value under cost.

D.2. Exact identification regimes

The interventional quantities above are identified exactly in at least three clean regimes.

Resettable branch replay. If the system can reset to an exact boundary snapshot B and force each candidate act u, then the branch outcome under $\mathrm{do}(U = u)$ directly samples the interventional kernel. For Stage I, replay estimates target laws over Y_H. For Stage II, replay estimates the one-step frozen-continuation return

$$Y_{\text{ret}}^u = G^u + \gamma^{\Delta^u} V^{\pi_0}(B'^u)$$

or a Monte Carlo continuation return. For send acts, Y_{ret}^u includes the fixed wording-policy realization distribution unless the branch intervention explicitly fixes the wording.

Randomized boundary interventions. If, at boundary contexts where act u is queried, the data-collection policy randomizes over legal acts with known positive propensities, then standard randomized-experiment estimators identify both Stage I laws and Stage II values, conditional on the variables randomized over.

Known simulator or structural boundary model. If the boundary kernel

$$\mathcal{K}_G(dB', dG, d\Delta \mid B, E)$$

is known, or if a simulator can sample it under forced interventions, then both Stage I and Stage II targets can be computed or simulated directly.

Passive logs can also identify the targets under adjustment assumptions. That case is more delicate and is treated next.

D.3. Boundary-level observational identification under ignorability

The adjustment theorem must condition on the local context being queried. A common but incorrect shortcut is to

write

$$\mathbb{E}[Y^u_{\text{ret}} \mid C = c] = \mathbb{E}[\mathbb{E}[Y_{\text{ret}} \mid S, U = u] \mid C = c]$$

for an arbitrary adjustment variable S. This is not valid unless S includes or refines C. The safe theorem conditions on (S, C).

Theorem D.1 (Boundary-level observational identification under context-aware ignorability)**.** *Fix a frozen round. Let S be a pre-intervention adjustment variable measurable at the boundary, and let $C = C(B)$ be the deployed local context. For act u, assume:*

1. ***Consistency:*** *if the realized act is u, then observed post-boundary variables equal their potential variables under $\text{do}(U = u)$.*
2. ***Positivity:***
$$\Pr(U = u \mid S = s, C = c) > 0$$
whenever the conditional quantity is queried.
3. ***Boundary ignorability:*** *for each act u, the potential variables of interest are conditionally independent of U given (S, C).*

Then:

1. *For any Stage I target Y_H,*
$$\mathsf{Law}(Y^u_H \mid S = s, C = c) = \mathsf{Law}(Y_H \mid S = s, C = c, U = u).$$
2. *Let*
$$Y^u_{\text{ret}} := G^u + \gamma^{\Delta^u} V^{\pi_0}(B'^u)$$
be the one-step frozen-continuation return, including the fixed wording-policy realization distribution for send acts unless wording is explicitly fixed by the intervention. Define
$$m_u(s, c) := \mathbb{E}[Y_{\text{ret}} \mid S = s, C = c, U = u],$$
where $Y_{\text{ret}} = G + \gamma^{\Delta} V^{\pi_0}(B')$ is the observed one-step frozen-continuation return on realized branches. For any improvement distribution d whose conditional law $d(S \mid C = c)$ is specified and for which the conditional potential-outcome regression
$$\mathbb{E}[Y^u_{\text{ret}} \mid S = s, C = c] = m_u(s, c)$$
is invariant between the logged data-generating distribution and d,
$$\bar{Q}_d(c, u) = \mathbb{E}_{S \sim d(\cdot \mid C = c)} \left[m_u(S, c) \right].$$
In particular, if d is the logged boundary distribution, then
$$\bar{Q}_d(c, u) = \mathbb{E}\left[\mathbb{E}[Y_{\text{ret}} \mid S, C, U = u] \mid C = c \right].$$

Proof. Fix act u.

For item 1, consistency gives $Y_H = Y^u_H$ on the event $\{U = u\}$. Boundary ignorability gives

$$Y^u_H \perp U \mid S, C.$$

By positivity, conditioning on $U = u$ is valid, and therefore

$$\mathsf{Law}(Y^u_H \mid S = s, C = c) = \mathsf{Law}(Y^u_H \mid S = s, C = c, U = u) = \mathsf{Law}(Y_H \mid S$$

For item 2, define

$$Y^u_{\text{ret}} = G^u + \gamma^{\Delta^u} V^{\pi_0}(B'^u).$$

By consistency and ignorability,

$$\mathbb{E}[Y^u_{\text{ret}} \mid S = s, C = c] = \mathbb{E}[Y_{\text{ret}} \mid S = s, C = c, U = u] = m_u(s, c).$$

Now take expectation under the conditional improvement distribution $d(S \mid C = c)$:

$$\bar{Q}_d(c, u) = \mathbb{E}_d[Y^u_{\text{ret}} \mid C = c] = \mathbb{E}_{S \sim d(\cdot \mid C = c)} \left[\mathbb{E}[Y^u_{\text{ret}} \mid S, C = c] \right] = \mathbb{E}_{S \sim d(\cdot \mid C}$$

The logged-distribution case is the special case where d is the observational boundary distribution. □

Remark D.2 (Why conditioning on C matters)**.** *The context C cannot be dropped in general. Let $C \sim \text{Bernoulli}(1/2)$, let S be independent of C, and suppose the potential return under u is $Y^u_{\text{ret}} = C$. Even if treatment is randomized and ignorability holds given S, we have*

$$\mathbb{E}[Y^u_{\text{ret}} \mid C = c] = c,$$

but

$$\mathbb{E}[\mathbb{E}[Y^u_{\text{ret}} \mid S] \mid C = c] = 1/2.$$

Thus $\mathbb{E}[\mathbb{E}[Y_{\text{ret}} \mid S, U = u] \mid C = c]$ is not a valid target unless S carries the relevant context information or the inner regression conditions on C.

D.4. Zero-propensity non-identification

The next theorem formalizes a hard limit. If an act is never taken at a context, passive logs do not identify what would have happened had that act been forced there.

Theorem D.3 (Zero-propensity non-identification)**.** *Fix local context c and legal act u. Suppose passive logs are generated by a behavior policy β satisfying*

$$\Pr_\beta(U = u \mid C = c) = 0.$$

In a nonparametric class of boundary kernels, the interventional value

$$\bar{Q}(c, u) = \mathbb{E}[Q^{\pi_0}_{\text{hi}}(B, u) \mid C(B) = c]$$

is not identifiable from those logs. More precisely, there exist two boundary-kernel models M_+ and M_- that induce the same distribution over all logged trajectories under β, but satisfy

$$\bar{Q}_+(c, u) \neq \bar{Q}_-(c, u).$$

The analogous Stage I law

$$\mathsf{Law}(Y_H \mid C = c, \mathrm{do}(U = u))$$

is also not identifiable in general.

Proof. Choose a measurable set

$$A \subseteq \{B : C(B) = c\}$$

with positive conditional mass under the target boundary distribution. Construct two models M_+ and M_- as follows.

For every state-action pair other than (B, u) with $B \in A$, let the two models have identical transition, reward, duration, and observation kernels. At (B, u) for $B \in A$, let M_+ give net return $+1$ and transition to a terminal absorbing state, while M_- gives net return -1 and transitions to the same absorbing state.

Under behavior policy β,

$$\Pr_\beta(U = u \mid C = c) = 0.$$

Thus the state-action set on which the models differ is never sampled in logged trajectories. Since M_+ and M_- agree on all state-action pairs that can occur under β, they induce exactly the same logged-data distribution.

However,

$$Q_+(B, u) = +1, \qquad Q_-(B, u) = -1 \qquad \text{for } B \in A.$$

Because A has positive conditional mass within $C = c$, the conditional mean values differ by

$$2\ \Pr(B \in A \mid C(B) = c)$$

relative to the common contribution outside A:

$$\bar{Q}_+(c, u) \neq \bar{Q}_-(c, u).$$

Therefore $\bar{Q}(c, u)$ is not identifiable.

For Stage I, modify only the target law under $\mathrm{do}(U = u)$ on A: in M_+, set $Y_H = 1$ almost surely; in M_-, set $Y_H = 0$ almost surely. The logged distributions remain identical because u is never taken at $C = c$, but the interventional target laws differ on a positive-mass set. Hence the Stage I law is also not identifiable. □

Corollary D.4 (Runtime-pool-only non-identification)**.** *Suppose logs record only a runtime candidate pool $P_b \subsetneq \mathcal{U}_{\mathrm{legal}}(b)$, and a legal act $u \in \mathcal{U}_{\mathrm{legal}}(b) \setminus P_b$ is never forced, randomized, replayed, or structurally modeled at boundaries with context c. Then the value and recipient-side eventfulness of u at c are not identifiable from those logs in a nonparametric boundary-kernel class.*

Proof. Apply Theorem D.3 to the event that u is outside the runtime candidate pool. If the data-generating process never executes or otherwise observes u at c, the passive logged distribution can be held fixed while the unobserved interventional branch is changed arbitrarily. □

D.5. Boundary bundles and full legal inventories

A training-ready boundary dataset should expose enough information to estimate both Stage I eventfulness and Stage II within-boundary value/ranking targets. A convenient unit is the boundary bundle.

Definition D.5 (Boundary bundle)**.** *A boundary bundle for boundary b consists of:*

1. *a stable boundary identity;*
2. *the local context $c^b = C(B^b)$;*
3. *the full legal high-level inventory $\mathcal{U}_{\mathrm{legal}}(b)$;*
4. *the runtime proposal or shortlist P_b, if any;*
5. *branch specifications for forced local, chosen, and selected alternate acts;*
6. *branch outcomes for those forced acts, including reward, duration, cost, and downstream task state;*
7. *replay, force-rule, and paired-seed integrity metadata;*
8. *recipient-side next-H deltas or artifact-state changes sufficient to derive Stage I labels such as* $\mathrm{eventful}(b, u)$*;*
9. *the semantic record and wording policy used for any send act.*

A boundary is candidate-complete *for a subset $S_b \subseteq \mathcal{U}_{\mathrm{legal}}(b)$ if branch outcomes and recipient-side next-H deltas are available for every $u \in S_b$.*

The full legal inventory matters even if only a few branches are replayed. It defines the denominator for proposal recall, shortlist loss, and missed soft-optimal mass.

Proposition D.6 (Candidate-complete replay supports estimation of finite local branch targets)**.** *Fix boundary context c, a finite candidate set $S \subseteq \mathcal{U}_{\mathrm{legal}}(b)$, fixed continuation policy π_0, and fixed wording policy. Suppose a boundary bundle contains replay-valid branch outcomes for every $u \in S$ from the same boundary state B^b. Then:*

1. *replayed branch outcomes are samples of the interventional branch-return distributions for $u \in S$, conditional on B^b;*

2. *recipient-side next-H deltas identify the corresponding eventfulness labels for $u \in S$, relative to the specified baseline branch;*

3. *repeated independent replay or repeated bundles over boundaries with $C(B) = c$ estimate $\mathbb{E}[Q_{\text{hi}}^{\pi_0}(B, u) \mid C(B) = c]$ for $u \in S$, under the boundary distribution sampled by the bundles.*

Proof. Replay validity means that forcing branch u from boundary B^b samples the same interventional kernel as $\text{do}(U = u)$ at B^b. Thus branch returns are samples of the interventional branch-return distribution, proving item 1.

The eventfulness label is defined as a function of the recipient-side next-H delta relative to a specified baseline branch. If that delta is logged, the label is identified by construction, proving item 2.

For item 3, average the branch-return samples over independent replays or bundles whose contexts equal c. The resulting conditional average estimates the conditional expectation under the distribution from which those bundles were sampled. □

D.6. Operational versus content-only predictive usefulness

The main Stage I score compares a candidate send act against yield:

$$D_{\text{KL}}(P_x^u \| P_x^{\varnothing}).$$

This is the correct operational quantity when the question is whether communication should occur at all. In some runtimes, however, send-versus-yield mixes semantic content effects with timing, scheduling, activation, or transport effects.

When the runtime permits it, define a matched-null act 0_u that preserves the transport and timing footprint of u but carries no recipient-relevant content. The content-only diagnostic is

$$D_{\text{KL}}(P_x^u \| P_x^{0_u}).$$

A large send-versus-yield KL but small matched-null KL suggests that the act matters operationally because of timing, scheduling, or activation rather than because of semantic content.

Target	Identified by	Common failure mode
Stage I law P_x^u	Replay, randomization, simulator, or valid adjustment.	Passive logs with confounding or zero propensity for u.
Stage I eventfulness label	Recipient-side next-H delta under forced branch.	Logs contain only final task score or no recipient-local delta.
Stage II value $\bar{Q}_d(c, u)$	Branch returns under u, reweighted or averaged under $d(B \mid C = c)$.	Runtime pool omits u; no branch outcome exists.
Shortlist recall	Full legal inventory and approximate values or bounds for omitted acts.	Logs record only the proposed runtime pool.
Wording drift	Within-class branch comparison or quotient-state audit.	Future control reads raw wording or parser support leaks.
Matched-null content effect	A null branch with same transport footprint as u.	Yield comparison can confound content with scheduling or activation.

Table D.1. **Boundary-level data must identify interventional Stage I and Stage II targets.** The checklist highlights the support and logging failures that make passive runtime data insufficient.

D.7. Identification diagnostics

E. Stage I: Predictive Usefulness, Decision Value, and Pruning

Stage I is a screening layer. It asks whether a candidate send act changes a decision-focused recipient target relative to yield. This appendix proves the KL identity, shows why coarse conditioning can create false negatives, gives the conservative pruning certificate, analyzes learned predictor error, and connects predictive usefulness to decision-theoretic value of information.

Stage I is a no-go certificate, not a go certificate

A calibrated small send-versus-yield usefulness score can justify pruning under a specified recipient-local proxy and cost bound. A large true usefulness score means the target distribution changes, and a large estimated score only suggests such a change. Final act choice still belongs to Stage II value optimization.

E.1. Semantic send-versus-yield KL identity

Proposition E.1 (Semantic send-versus-yield KL identity)**.** *Fix recipient state x, send act $u = (j, a)$, continuation policy π_0, wording policy $\bar{\pi}_{\text{text}}$, and decision-focused target Y_H. Let*

$$P_x^u = \mathsf{Law}^{\pi_0, \bar{\pi}_{\text{text}}}(Y_H \mid X_j = x, \text{do}(U = u))$$

and

$$P_x^{\varnothing} = \mathsf{Law}^{\pi_0,\bar{\pi}_{\text{text}}}(Y_H \mid X_j = x, \mathrm{do}(U = \varnothing)).$$

If $p_u(\cdot \mid x)$ and $p_{\varnothing}(\cdot \mid x)$ are the Bayes-optimal predictors under those interventions, then

$$\mathbb{E}_{Y_H \sim P_x^u}\left[\log p_u(Y_H \mid x) - \log p_{\varnothing}(Y_H \mid x)\right] = D_{\mathrm{KL}}(P_x^u \| P_x^{\varnothing}).$$

Proof. Under Bayes-optimal log loss, the predictive distribution for the send intervention is $p_u(\cdot \mid x) = P_x^u$, and the predictive distribution for the yield intervention is $p_{\varnothing}(\cdot \mid x) = P_x^{\varnothing}$. Therefore

$$\mathbb{E}_{Y_H \sim P_x^u}\left[\log p_u(Y_H \mid x) - \log p_{\varnothing}(Y_H \mid x)\right] = \int \log \frac{dP_x^u}{dP_x^{\varnothing}}\, dP_x^u = D_{\mathrm{KL}}(P_x^u \| P_x^{\varnothing}),$$

with both sides interpreted as $+\infty$ when $P_x^u \not\ll P_x^{\varnothing}$. □

The identity is useful because it gives a principled screening target. It is not a value theorem. A message can change a target distribution without changing the recipient's optimal downstream decision, and it can be too costly even when it is predictive.

E.2. Conditioning granularity and false negatives

The notation P_x^u suppresses additional pre-intervention variables. If omitted variables matter, coarse conditioning can wash out useful communication.

Let X and Z be pre-intervention variables, with X a recipient abstraction and Z an additional variable. For $\xi \in \{u, \varnothing\}$, define

$$P_{x,z}^{\xi} = \mathsf{Law}(Y_H \mid X = x, Z = z, \mathrm{do}(U = \xi)),$$

and

$$P_x^{\xi} = \mathsf{Law}(Y_H \mid X = x, \mathrm{do}(U = \xi)).$$

Proposition E.2 (Coarsening pre-intervention conditioning reduces Stage I KL)**.** *Because Z is pre-intervention, its conditional law*

$$\lambda_x(dz) = \mathsf{Law}(Z \mid X = x)$$

is the same under both interventions, and

$$P_x^{\xi} = \int P_{x,z}^{\xi} \lambda_x(dz).$$

Then

$$D_{\mathrm{KL}}(P_x^u \| P_x^{\varnothing}) \le \int D_{\mathrm{KL}}(P_{x,z}^u \| P_{x,z}^{\varnothing})\, \lambda_x(dz).$$

Proof. The two coarse laws are common mixtures:

$$P_x^u = \int P_{x,z}^u \lambda_x(dz), \qquad P_x^{\varnothing} = \int P_{x,z}^{\varnothing} \lambda_x(dz).$$

By KL contraction under common mixing,

$$D_{\mathrm{KL}}\left(\int P_{x,z}^u \lambda_x(dz) \middle\| \int P_{x,z}^{\varnothing} \lambda_x(dz)\right) \le \int D_{\mathrm{KL}}(P_{x,z}^u \| P_{x,z}^{\varnothing}) \lambda_x(dz).$$

This is the claim. □

Remark E.3 (False negatives)**.** *Coarsening can create false negatives. A warning may be highly useful only when the planner lacks a specific threshold or provenance field. If Stage I conditions only on a coarser abstraction, the useful cases may be averaged together with cases where the recipient already knows enough, reducing the apparent KL.*

E.3. Local pruning certificate

Stage I can safely hard-prune only under a one-sided certificate. The certificate is local to a specified recipient-side proxy problem and communication-cost lower bound.

Theorem E.4 (Predictive usefulness can certify local pruning)**.** *Let $Z_H = f(Y_H)$ be decision-sufficient for a recipient-local H-step problem under π_0. Let $\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x, u)$ be the corresponding recipient-local proxy gain from sending u rather than yielding at recipient state x, before charging communication cost. Let $P_{Z_H,x}^u$ and $P_{Z_H,x}^{\varnothing}$ be the push-forward laws of Z_H. If the recipient-local utility span is at most V_H, then*

$$\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x, u) \le V_H \left\| P_{Z_H,x}^u - P_{Z_H,x}^{\varnothing} \right\|_{\mathrm{TV}} \le V_H \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P_x^u \| P_x^{\varnothing})}.$$

The KL term is interpreted as $+\infty$ when $P_x^u \not\ll P_x^{\varnothing}$. Consequently, if a calibrated upper bound $U_u(x)$ satisfies

$$V_H \sqrt{\frac{1}{2} U_u(x)} \le \underline{c}_{\mathrm{comm}}(u),$$

then pruning u is locally safe at x for the specified proxy problem and cost lower bound.

Proof. Let

$$p = P_{Z_H,x}^u, \qquad q = P_{Z_H,x}^{\varnothing}.$$

By decision sufficiency, there is a class of recipient-local decision rules $\mathcal{A}_{\mathrm{loc}}^{(H)}$ and a utility representation $r_H(a, z)$ with span at most V_H such that

$$\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x, u) = \sup_{a \in \mathcal{A}_{\mathrm{loc}}^{(H)}} \mathbb{E}_{Z \sim p}[r_H(a, Z)] - \sup_{a \in \mathcal{A}_{\mathrm{loc}}^{(H)}} \mathbb{E}_{Z \sim q}[r_H(a, Z)].$$

Shift utilities so that $r_H(a, z) \in [0, V_H]$. Then

$$\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x, u) \le \sup_{a \in \mathcal{A}_{\mathrm{loc}}^{(H)}} \left(\mathbb{E}_p[r_H(a, Z)] - \mathbb{E}_q[r_H(a, Z)]\right).$$

For each fixed a, total-variation duality gives

$$\mathbb{E}_p[r_H(a, Z)] - \mathbb{E}_q[r_H(a, Z)] \le V_H \|p - q\|_{\mathrm{TV}}.$$

Taking the supremum gives

$$\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x,u) \le V_H \|p-q\|_{\mathrm{TV}}.$$

Pinsker's inequality gives

$$\|p-q\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(p\|q)}.$$

Since $Z_H = f(Y_H)$, KL data processing gives

$$D_{\mathrm{KL}}(p\|q) = D_{\mathrm{KL}}(P^u_{Z_H,x}\|P^{\varnothing}_{Z_H,x}) \le D_{\mathrm{KL}}(P^u_x\|P^{\varnothing}_x).$$

Combining the inequalities proves the gross proxy-gain bound. The pruning statement follows by applying the calibrated upper bound $U_u(x)$ and comparing the resulting gross local-gain upper bound to $\underline{c}_{\mathrm{comm}}(u)$. □

Corollary E.5 (Sender-expected local pruning)**.** *Let $b^j(\cdot \mid I_i)$ be the sender-side posterior over recipient abstract state. Define*

$$\bar{\Delta}_{\mathrm{loc}}^{\mathrm{gross}}(u \mid I_i) = \mathbb{E}_{x\sim b^j(\cdot|I_i)}[\Delta_{\mathrm{loc}}^{\mathrm{gross}}(x,u)].$$

Then

$$\bar{\Delta}_{\mathrm{loc}}^{\mathrm{gross}}(u \mid I_i) \le V_H \sqrt{\frac{1}{2}\mathbb{E}_{x\sim b^j(\cdot|I_i)}\left[D_{\mathrm{KL}}(P^u_x\|P^{\varnothing}_x)\right]}.$$

Consequently, if U_u is a calibrated upper bound on the sender-expected KL and

$$V_H\sqrt{\frac{1}{2}U_u} \le \underline{c}_{\mathrm{comm}}(u),$$

then pruning u is sender-expected recipient-locally safe for the chosen proxy problem.

Proof. Take the expectation of Theorem E.4's pointwise bound:

$$\bar{\Delta}_{\mathrm{loc}}^{\mathrm{gross}}(u \mid I_i) \le V_H \mathbb{E}_x\left[\sqrt{\frac{1}{2}D_{\mathrm{KL}}(P^u_x\|P^{\varnothing}_x)}\right].$$

Apply Jensen's inequality to the concave square-root function. □

E.4. Learned Stage I predictor errors

In deployment, Stage I may use learned predictive models. The exact KL identity then becomes a target rather than an automatic guarantee.

Proposition E.6 (Plug-in decomposition for learned Stage I log-score gaps)**.** *Fix x and u. Let P^u_x and $P^{\varnothing}_x$ be true interventional laws. Assume $P^u_x \ll P^{\varnothing}_x$, and let $q_\eta(\cdot \mid x,u)$ and $q_\eta(\cdot \mid x,\varnothing)$ be learned predictors that are strictly positive on the support of P^u_x. Define*

$$\delta_\eta(Y;x,u) = \log q_\eta(Y \mid x,u) - \log q_\eta(Y \mid x,\varnothing), \qquad Y \sim P^u_x.$$

Then

$$\mathbb{E}[\delta_\eta(Y;x,u)] = D_{\mathrm{KL}}(P^u_x\|P^{\varnothing}_x) - \mathcal{E}_u(x) + \mathcal{B}_u(x),$$

where

$$\mathcal{E}_u(x) = D_{\mathrm{KL}}(P^u_x\|q_\eta(\cdot \mid x,u))$$

is the send-predictor calibration error, and

$$\mathcal{B}_u(x) = \mathbb{E}_{Y\sim P^u_x}\left[\log\frac{P^{\varnothing}_x(Y)}{q_\eta(Y \mid x,\varnothing)}\right]$$

is the transported-baseline error. Consequently,

$$\left|\mathbb{E}[\delta_\eta(Y;x,u)] - D_{\mathrm{KL}}(P^u_x\|P^{\varnothing}_x)\right| \le \mathcal{E}_u(x) + \mathbb{E}_{Y\sim P^u_x}\left[\left|\log\frac{P^{\varnothing}_x(Y)}{q_\eta(Y \mid x,\varnothing)}\right|\right].$$

If $P^u_x \not\ll P^{\varnothing}_x$, the true KL is $+\infty$, and the decomposition should be interpreted in the extended-real sense whenever the displayed terms are well-defined.

Proof. Add and subtract the true likelihood ratio:

$$\begin{aligned}\mathbb{E}[\delta_\eta(Y;x,u)] &= \mathbb{E}_{P^u_x}\left[\log\frac{P^u_x(Y)}{P^{\varnothing}_x(Y)}\right] + \mathbb{E}_{P^u_x}\left[\log\frac{q_\eta(Y \mid x,u)}{P^u_x(Y)}\right] \\ &\quad + \mathbb{E}_{P^u_x}\left[\log\frac{P^{\varnothing}_x(Y)}{q_\eta(Y \mid x,\varnothing)}\right] \\ &= D_{\mathrm{KL}}(P^u_x\|P^{\varnothing}_x) - D_{\mathrm{KL}}(P^u_x\|q_\eta(\cdot \mid x,u)) + \mathcal{B}_u(x).\end{aligned}$$

The deviation bound follows by the triangle inequality. □

Proposition E.7 (Sender-side recipient-model error for bounded screeners)**.** *Let $g_u(x) \in [0, G_{\max}]$ be any bounded usefulness functional. Let $b^j(\cdot \mid I_i)$ be the true sender-side posterior over recipient abstract state and let $\hat{b}^j_\xi(\cdot \mid I_i)$ be an approximation. Then*

$$\left|\mathbb{E}_{x\sim\hat{b}^j_\xi}[g_u(x)] - \mathbb{E}_{x\sim b^j}[g_u(x)]\right| \le G_{\max}\left\|\hat{b}^j_\xi(\cdot \mid I_i) - b^j(\cdot \mid I_i)\right\|_{\mathrm{TV}}.$$

Proof. Apply total-variation duality to the bounded function g_u. □

Raw KL scores are unbounded. If robust finite-sample pruning is the goal, clipped KL, bounded proper scoring-rule divergences, or calibrated upper-confidence bounds are often preferable to unbounded point estimates.

E.5. Decision value and Blackwell dominance

The KL identity explains a log-loss screening score. The fundamental recipient-side quantity, however, is the value of information for a downstream decision problem.

Let Θ be a finite decision-sufficient hidden variable for a recipient-local decision problem at state x. A send act u induces a signal S_u with kernel

$$K_u(ds \mid \theta, x).$$

Let $\mathcal{A}$ be finite, and let the recipient choose $a \in \mathcal{A}$, with utility $r(a, \theta) \in [0, V]$. Define the value of observing S_u:

$$\mathrm{Val}(S_u \mid x) = \mathbb{E}_s \left[\max_a \mathbb{E}[r(a, \Theta) \mid x, s] \right],$$

and the no-signal value

$$\mathrm{Val}_0(x) = \max_a \mathbb{E}[r(a, \Theta) \mid x].$$

The value of information is

$$\mathrm{VoI}_u(x) = \mathrm{Val}(S_u \mid x) - \mathrm{Val}_0(x).$$

Theorem E.8 (Recipient-local value of information bound)**.** *Under the setup above,*

$$0 \le \mathrm{VoI}_u(x) \le V \mathbb{E}_s \left[\| p(\Theta \mid x, s) - p(\Theta \mid x) \|_{\mathrm{TV}} \right].$$

Moreover,

$$\mathrm{VoI}_u(x) \le V \sqrt{\frac{1}{2} I(\Theta; S_u \mid x)}.$$

Proof. For any distribution p over Θ, define

$$v(p) = \max_a \sum_\theta p(\theta) r(a, \theta).$$

This is the maximum of linear functions of p, hence convex.

Let

$$p_s(\theta) = p(\theta \mid x, s), \qquad p_0(\theta) = p(\theta \mid x).$$

Since $\mathbb{E}_s[p_s] = p_0$, Jensen's inequality gives

$$\mathbb{E}_s[v(p_s)] \ge v(p_0),$$

so $\mathrm{VoI}_u(x) \ge 0$.

For the upper bound, for any p, q,

$$|v(p) - v(q)| \le \sup_a \left| \sum_\theta (p(\theta) - q(\theta)) r(a, \theta) \right|.$$

Since $r(a, \theta) \in [0, V]$, total-variation duality gives

$$|v(p) - v(q)| \le V \|p - q\|_{\mathrm{TV}}.$$

Therefore

$$\mathrm{VoI}_u(x) = \mathbb{E}_s[v(p_s)] - v(p_0) \le \mathbb{E}_s |v(p_s) - v(p_0)| \le V \mathbb{E}_s \|p_s - p_0\|_{\mathrm{TV}}.$$

Pinsker's inequality gives

$$\|p_s - p_0\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2} D_{\mathrm{KL}}(p_s \| p_0)}.$$

By Jensen's inequality,

$$\mathbb{E}_s \sqrt{\frac{1}{2} D_{\mathrm{KL}}(p_s \| p_0)} \le \sqrt{\frac{1}{2} \mathbb{E}_s D_{\mathrm{KL}}(p_s \| p_0)}.$$

Finally,

$$\mathbb{E}_s D_{\mathrm{KL}}(p(\Theta \mid x, s) \| p(\Theta \mid x)) = I(\Theta; S_u \mid x).$$

Combining the inequalities proves the result. □

Values above use maxima because the recipient-local action set is finite. For general compact or measurable action spaces, replace maxima by suprema and assume the relevant suprema are attained, or use ε-optimal rules.

Theorem E.9 (Blackwell dominance implies larger decision value)**.** *Let two communication acts induce signals S_v and S_u. Suppose S_u is a garbling of S_v: there exists a Markov kernel $T(ds_u \mid s_v)$ such that*

$$P(ds_u \mid \theta, x) = \int T(ds_u \mid s_v) P(ds_v \mid \theta, x).$$

Then for every bounded recipient-local decision problem,

$$\mathrm{Val}(S_v \mid x) \ge \mathrm{Val}(S_u \mid x).$$

Proof. Take any decision rule $\delta_u(a \mid s_u)$ for the signal S_u. Construct a decision rule for S_v by simulating the garbling:

$$\delta_v(a \mid s_v) = \int \delta_u(a \mid s_u) T(ds_u \mid s_v).$$

The joint law of (Θ, S_u, A) induced by first drawing S_v, then garbling it through T, then applying δ_u, is the same as the law induced by observing S_u directly and applying δ_u. Thus every S_u-based decision rule can be simulated by an S_v-based decision rule with the same expected utility. Optimizing over all S_v-based rules can only improve value, so

$$\mathrm{Val}(S_v \mid x) \ge \mathrm{Val}(S_u \mid x).$$

□

Corollary E.10 (Belief change without decision change has zero decision value)**.** *Let a_0 be optimal without observing S_u. If a_0 remains optimal for every posterior $p(\Theta \mid x, s)$, then*

$$\mathrm{VoI}_u(x) = 0.$$

More generally, if posterior-optimal actions differ from a_0 only on an event A with probability p, then

$$\mathrm{VoI}_u(x) \le V p.$$

Proof. If a_0 remains optimal for every signal realization s, then

$$v(p_s) = \mathbb{E}[r(a_0, \Theta) \mid x, s].$$

Therefore

$$\mathbb{E}_s[v(p_s)] = \mathbb{E}_s \mathbb{E}[r(a_0, \Theta) \mid x, s] = \mathbb{E}[r(a_0, \Theta) \mid x] = v(p_0),$$

so $\mathrm{VoI}_u(x) = 0$.

For the second statement, the gain is zero on A^c and at most V on A. Hence the expected gain is at most Vp. □

Remark E.11 (Why this does not replace Stage II)**.** *The value-of-information bound is recipient-local and proxy-dependent. Stage II remains necessary because the final boundary act must account for global downstream return, communication cost, other agents, and environment dynamics. The decision-value view explains why generic informativeness is insufficient; it does not replace the semantic control objective.*

E.6. Stage I diagnostics

Symptom	Likely cause	Diagnostic or response
High KL, low task value	Target changes but downstream decision does not.	Compare to recipient-local decision value or Stage II value.
Low KL, missed useful act	Recipient abstraction x is too coarse or omits pre-intervention covariates.	Refine x or include sender-visible z; use Proposition E.2.
Hard pruning harms return	Gross-gain upper bound uncalibrated or proxy not decision-sufficient.	Replay pruned acts; check V_H, target sufficiency, and cost bound.
Send-vs-yield confounds timing	Scheduling or activation effect, not semantic content.	Use matched-null diagnostic $D_{\mathrm{KL}}(P_x^u \| P_x^{0_u})$.
Model-model KL unstable	Raw KL is unbounded and predictor ratios are poorly calibrated.	Clip, use bounded scores, or calibrate with interventional data.

Table E.1. **Stage I diagnostics distinguish predictive usefulness from downstream value.** High or low KL is actionable only after checking calibration, conditioning, pruning, and timing confounds.

F. Stage II: Local Actor, Aliasing, and Approximation Budget

Stage II is the control layer. The deployed actor sees local context, not the hidden boundary snapshot. It must therefore choose one distribution for all hidden states compatible with that context. This appendix proves the average-then-softmax rule, quantifies the exact value lost to local-context aliasing, explains why averaging hidden-state teacher policies solves the wrong problem, and gives a local approximation budget for screen-then-control implementations.

Average values, not policies

The exact local actor softmaxes

$$\bar{Q}(c,u) = \mathbb{E}[Q(B,u) \mid C(B) = c].$$

Averaging hidden-state softmax teachers solves a different forward-KL distillation problem. The correct local-control aggregation is average values first, then softmax.

F.1. Average-then-softmax local actor

Theorem F.1 (Average-then-softmax local actor)**.** *Fix a continuation policy π_0, fixed wording policy $\bar{\pi}_{\mathrm{text}}$, improvement distribution d, finite feasible act set $F(c)$, full-support prior $\mu(\cdot \mid c)$, and temperature $\tau > 0$. Assume $F(c)$ is context-measurable and every $u \in F(c)$ is legal for every positive-mass hidden boundary state B with $C(B) = c$. If legality varies within a context class, the exact theorem applies after refining the context or after excluding acts that are illegal on any hidden state with positive conditional mass. Illegal actions may be assigned value $-\infty$, but such actions are outside the finite-value feasible set used by the theorem. Assume the values below are finite and integrable. Define*

$$\bar{Q}_d(c,u) = \mathbb{E}_{B\sim d}\left[Q_{\mathrm{hi}}^{\pi_0}(B,u) \,|\, C(B) = c\right].$$

The unique optimizer of

$$\max_{\pi(\cdot|c)\in\Delta(F(c))} \sum_{u\in F(c)} \pi(u \mid c)\bar{Q}_d(c,u) - \tau \mathrm{KL}(\pi(\cdot \mid c) \| \mu(\cdot \mid c))$$

is

$$\pi_{\mathrm{hi}}^{\star}(u \mid c) = \frac{\mu(u \mid c)\exp(\bar{Q}_d(c,u)/\tau)}{\sum_{v\in F(c)} \mu(v \mid c)\exp(\bar{Q}_d(c,v)/\tau)}.$$

Proof. For fixed c, abbreviate

$$Q(u) = \bar{Q}_d(c,u), \qquad \mu(u) = \mu(u \mid c), \qquad \pi(u) = \pi(u \mid c).$$

The objective is

$$\sum_u \pi(u)Q(u) - \tau \sum_u \pi(u) \log \frac{\pi(u)}{\mu(u)}.$$

Because $\tau > 0$ and μ has full support, the objective is strictly concave on the simplex. Introduce a Lagrange multiplier λ for $\sum_u \pi(u) = 1$:

$$\mathcal{L}(\pi,\lambda) = \sum_u \pi(u)Q(u) - \tau \sum_u \pi(u) \log \frac{\pi(u)}{\mu(u)} + \lambda \left(\sum_u \pi(u) - 1\right).$$

Stationarity gives

$$0 = Q(u) - \tau \left(\log \frac{\pi(u)}{\mu(u)} + 1\right) + \lambda.$$

Thus

$$\pi(u) = \mu(u)\exp(Q(u)/\tau) \cdot Z^{-1},$$

where Z normalizes. Substituting $Q(u) = \bar{Q}_d(c,u)$ gives the stated actor. Strict concavity gives uniqueness. □

The corresponding optimal local soft value is

$$V^{\star}(c) = \tau \log \sum_{u\in F(c)} \mu(u \mid c)\exp(\bar{Q}_d(c,u)/\tau).$$

F.2. Exact local-context aliasing gap

The local actor cannot choose a different policy for each hidden boundary snapshot B with the same context c. The next theorem quantifies the exact value gap between a hidden-state soft controller and the best context-measurable soft controller.

Fix context c. Let $B \sim d(\cdot \mid C = c)$, and write $Q(B, u)$ for $Q^{\pi_0}_{\text{hi}}(B, u)$. Assume $Q(B, u)$ is finite and integrable for all $u \in F(c)$, and that the displayed log-sum-exp expectations are finite. Define

$$\bar{Q}(c, u) = \mathbb{E}[Q(B, u) \mid C(B) = c].$$

The hidden-state soft teacher is

$$\nu_B(u) = \frac{\mu(u \mid c)e^{Q(B,u)/\tau}}{\sum_v \mu(v \mid c)e^{Q(B,v)/\tau}}.$$

The exact local actor is

$$\pi^\star_{\text{loc}}(u \mid c) = \frac{\mu(u \mid c)e^{\bar{Q}(c,u)/\tau}}{\sum_v \mu(v \mid c)e^{\bar{Q}(c,v)/\tau}}.$$

Theorem F.2 (Exact local-context aliasing gap)**.** *Define the full-information soft value*

$$V^\star_{\text{full}}(c) = \mathbb{E}\left[\tau \log \sum_u \mu(u \mid c)e^{Q(B,u)/\tau} \,\middle|\, C(B) = c\right],$$

and the local-context soft value

$$V^\star_{\text{loc}}(c) = \tau \log \sum_u \mu(u \mid c)e^{\bar{Q}(c,u)/\tau}.$$

Then

$$V^\star_{\text{full}}(c) - V^\star_{\text{loc}}(c) = \tau\mathbb{E}\left[\mathrm{KL}\left(\pi^\star_{\text{loc}}(\cdot \mid c) \| \nu_B(\cdot)\right) \mid C(B) = c\right].$$

Moreover,

$$\pi^\star_{\text{loc}}(\cdot \mid c) = \arg\min_{\pi \in \Delta(F(c))} \mathbb{E}\left[\mathrm{KL}(\pi \| \nu_B) \mid C(B) = c\right].$$

The aliasing gap is zero iff the hidden-state teachers agree almost surely, equivalently iff there exists a scalar $\alpha(B)$ such that

$$Q(B, u) = \bar{Q}(c, u) + \alpha(B) \qquad \textit{for all } u$$

almost surely under $B \mid C = c$.

Proof. For fixed B, define

$$Z_B = \sum_u \mu(u \mid c)e^{Q(B,u)/\tau}.$$

Then

$$\nu_B(u) = \frac{\mu(u \mid c)e^{Q(B,u)/\tau}}{Z_B}.$$

For any distribution π over $F(c)$,

$$\log \frac{\pi(u)}{\nu_B(u)} = \log \frac{\pi(u)}{\mu(u \mid c)} - \frac{Q(B, u)}{\tau} + \log Z_B.$$

Therefore

$$\mathrm{KL}(\pi \| \nu_B) = \mathrm{KL}(\pi \| \mu(\cdot \mid c)) - \frac{1}{\tau}\sum_u \pi(u)Q(B, u) + \log Z_B.$$

Taking conditional expectation over $B \mid C = c$,

$$\mathbb{E}[\mathrm{KL}(\pi \| \nu_B) \mid C = c] = \mathrm{KL}(\pi \| \mu) - \frac{1}{\tau}\sum_u \pi(u)\bar{Q}(c, u) + \mathbb{E}[\log Z_B \mid C = c$$

The last term does not depend on π. Minimizing the expected reverse KL is therefore equivalent to maximizing

$$\sum_u \pi(u)\bar{Q}(c, u) - \tau\mathrm{KL}(\pi \| \mu),$$

whose unique optimizer is $\pi^\star_{\text{loc}}$.

Now set $\pi = \pi^\star_{\text{loc}}$ and multiply the displayed equality by τ:

$$\tau\mathbb{E}[\mathrm{KL}(\pi^\star_{\text{loc}} \| \nu_B) \mid C = c] = \tau\mathbb{E}[\log Z_B \mid C = c] - \left[\sum_u \pi^\star_{\text{loc}}(u)\bar{Q}(c, u) - \tau\mathrm{KL}(\pi^\star_{\text{loc}} \| \mu)\right].$$

The first term is $V^\star_{\text{full}}(c)$. The bracketed term is the optimal local soft value $V^\star_{\text{loc}}(c)$. This proves the identity.

The gap is zero iff the expected KL is zero, which holds iff $\pi^\star_{\text{loc}} = \nu_B$ almost surely. Since μ has full support, two prior-softmax distributions are equal iff their logits differ by an additive constant. Therefore $\nu_B = \pi^\star_{\text{loc}}$ iff there exists $\alpha(B)$ such that

$$Q(B, u) = \bar{Q}(c, u) + \alpha(B)$$

for all u. □

F.3. Forward-KL teacher averaging solves a different problem

The aliasing theorem gives the reverse-KL characterization. A common distillation approximation averages hidden-state teachers:

$$\pi_{\text{fwd}}(\cdot \mid c) = \mathbb{E}[\nu_B(\cdot) \mid C(B) = c].$$

This is not the exact local-control actor in general.

Proposition F.3 (Reverse-KL local actor versus forward-KL mean teacher)**.** *With ν_B as above,*

$$\pi^\star_{\text{loc}}(\cdot \mid c) = \arg\min_{\pi \in \Delta(F(c))} \mathbb{E}\left[\mathrm{KL}(\pi \| \nu_B) \mid C(B) = c\right].$$

By contrast,

$$\pi_{\mathrm{fwd}}(\cdot \mid c) = \mathbb{E}[\nu_B(\cdot) \mid C(B) = c]$$

is the unique minimizer of

$$\arg\min_{\pi \in \Delta(F(c))} \mathbb{E}\left[\mathrm{KL}(\nu_B \| \pi) \,|\, C(B) = c\right].$$

Proof. The reverse-KL statement is proved in Theorem F.2. For the forward-KL statement,

$$\mathbb{E}[\mathrm{KL}(\nu_B \| \pi) \mid C = c] = \text{const} - \sum_u \mathbb{E}[\nu_B(u) \mid C = c] \log \pi(u).$$

Minimizing this cross-entropy over the simplex gives

$$\pi(u) = \mathbb{E}[\nu_B(u) \mid C = c],$$

with uniqueness on the support because ν_B has full support whenever μ does. □

Proposition F.4 (Teacher-averaging mismatch bound)**.** *Let π_{fwd} and $\pi^\star_{\mathrm{loc}}$ be defined as above. Then*

$$\|\pi_{\mathrm{fwd}}(\cdot \mid c) - \pi^\star_{\mathrm{loc}}(\cdot \mid c)\|_1 \le \frac{1}{\tau}\mathbb{E}\left[\mathrm{sp}\big(Q(B,\cdot) - \bar{Q}(c,\cdot)\big) \,\middle|\, C(B) = c\right].$$

In particular, the two coincide whenever

$$Q(B,u) = \bar{Q}(c,u) + \alpha(B)$$

for all u almost surely within the context class.

Proof. Let

$$S_c(z) = \mathrm{Soft}_{\mu(\cdot|c)}(z).$$

Then

$$\pi_{\mathrm{fwd}} - \pi^\star_{\mathrm{loc}} = \mathbb{E}\left[S_c(Q(B,\cdot)/\tau) - S_c(\bar{Q}(c,\cdot)/\tau) \,\middle|\, C = c\right].$$

Taking ℓ_1-norms and applying Jensen,

$$\|\pi_{\mathrm{fwd}} - \pi^\star_{\mathrm{loc}}\|_1 \le \mathbb{E}\left[\left\|S_c(Q(B,\cdot)/\tau) - S_c(\bar{Q}(c,\cdot)/\tau)\right\|_1 \,\middle|\, C = c\right].$$

The span Lipschitz bound for softmax gives

$$\left\|S_c(Q/\tau) - S_c(\bar{Q}/\tau)\right\|_1 \le \frac{1}{\tau}\,\mathrm{sp}(Q - \bar{Q}).$$

Combining proves the claim. If $Q(B,\cdot) = \bar{Q}(c,\cdot) + \alpha(B)\mathbf{1}$, the additive shift cancels in the softmax, so the two policies coincide. □

F.4. Shortlist restriction loss

If Stage I passes only a subset of acts to Stage II, the exact local actor is restricted. The loss is exactly the missed soft-optimal mass.

Proposition F.5 (Exact restriction loss for shortlists)**.** *Fix context c, full finite act set $\mathcal{U}$, full-support prior $\mu(\cdot \mid c)$, value vector Q, and temperature $\tau > 0$. Let*

$$V^\star(c) = \max_{\pi \in \Delta(\mathcal{U})} \left[\sum_u \pi(u)Q(u) - \tau \mathrm{KL}(\pi \| \mu)\right],$$

and let $V^\star_S(c)$ be the same optimum restricted to policies supported on a nonempty subset $S \subseteq \mathcal{U}$. If $\pi^\star_Q$ is the full-set soft actor, then

$$V^\star(c) - V^\star_S(c) = -\tau \log \pi^\star_Q(S).$$

Proof. Let

$$Z = \sum_{u \in \mathcal{U}} \mu(u \mid c) e^{Q(u)/\tau}, \qquad Z_S = \sum_{u \in S} \mu(u \mid c) e^{Q(u)/\tau}.$$

The soft value identity gives

$$V^\star(c) = \tau \log Z, \qquad V^\star_S(c) = \tau \log Z_S.$$

Also,

$$\pi^\star_Q(S) = \frac{Z_S}{Z}.$$

Therefore

$$V^\star(c) - V^\star_S(c) = -\tau \log \frac{Z_S}{Z} = -\tau \log \pi^\star_Q(S).$$

□

Corollary F.6 (Restriction-loss upper bound from value bounds)**.** *Suppose for every u,*

$$L(u) \le Q(u) \le U(u).$$

Then

$$V^\star(c) - V^\star_S(c) \le \tau \log\left(1 + \frac{\sum_{u \notin S} \mu(u \mid c) e^{U(u)/\tau}}{\sum_{u \in S} \mu(u \mid c) e^{L(u)/\tau}}\right).$$

Proof. From Proposition F.5,

$$V^\star(c) - V^\star_S(c) = \tau \log\left(1 + \frac{\sum_{u \notin S} \mu(u \mid c) e^{Q(u)/\tau}}{\sum_{u \in S} \mu(u \mid c) e^{Q(u)/\tau}}\right).$$

Upper-bound the numerator by replacing $Q(u)$ with $U(u)$ outside S, and lower-bound the denominator by replacing $Q(u)$ with $L(u)$ inside S. □

F.5. Unified local approximation regret

The implemented controller may use both a shortlist and approximate values. The next result gives a compact local approximation budget.

Let

$$F_Q^\mu(\pi) = \sum_{u \in \mathcal{U}} \pi(u)Q(u) - \tau \mathrm{KL}(\pi \| \mu).$$

Theorem F.7 (Shortlist plus critic-error local regret)**.** *Fix context c, full feasible set $\mathcal{U}$, true value vector Q, full-support prior μ, and temperature $\tau > 0$. Let*

$$\pi_Q^\star = \mathrm{Soft}_\mu(Q/\tau)$$

be the full exact soft actor. Let $S \subseteq \mathcal{U}$ be a nonempty shortlist, and let

$$\mu_S(u) = \frac{\mu(u)}{\mu(S)} \qquad (u \in S)$$

be the prior restricted to S. Let $\widehat{Q}$ be an estimated value vector on S, and let

$$\widehat{\pi}_S = \mathrm{Soft}_{\mu_S}(\widehat{Q}/\tau)$$

be the deployed shortlist actor, viewed as a policy on $\mathcal{U}$ by assigning probability zero to $\mathcal{U} \setminus S$. Then

$$F_Q^\mu(\pi_Q^\star) - F_Q^\mu(\widehat{\pi}_S) \le -\tau \log \pi_Q^\star(S) + \mathrm{sp}_S(\widehat{Q} - Q),$$

where

$$\mathrm{sp}_S(\widehat{Q} - Q) = \max_{u \in S}(\widehat{Q}(u) - Q(u)) - \min_{u \in S}(\widehat{Q}(u) - Q(u)).$$

If

$$\|\widehat{Q} - Q\|_{\infty,S} \le \varepsilon,$$

then

$$F_Q^\mu(\pi_Q^\star) - F_Q^\mu(\widehat{\pi}_S) \le -\tau \log \pi_Q^\star(S) + 2\varepsilon.$$

Proof. Let

$$\pi_{Q,S}^\star = \arg\max_{\pi \in \Delta(S)} F_Q^\mu(\pi)$$

be the exact restricted soft actor on S. Decompose:

$$F_Q^\mu(\pi_Q^\star) - F_Q^\mu(\widehat{\pi}_S) = \Big[F_Q^\mu(\pi_Q^\star) - F_Q^\mu(\pi_{Q,S}^\star)\Big] + \Big[F_Q^\mu(\pi_{Q,S}^\star) - F_Q^\mu(\widehat{\pi}_S)\Big].$$

The first term is the restriction loss:

$$F_Q^\mu(\pi_Q^\star) - F_Q^\mu(\pi_{Q,S}^\star) = -\tau \log \pi_Q^\star(S).$$

For policies supported on S,

$$\mathrm{KL}(\pi \| \mu) = \mathrm{KL}(\pi \| \mu_S) - \log \mu(S).$$

Therefore

$$F_Q^\mu(\pi) = F_Q^{\mu_S}(\pi) + \tau \log \mu(S),$$

and the additive constant cancels in restricted regret:

$$F_Q^\mu(\pi_{Q,S}^\star) - F_Q^\mu(\widehat{\pi}_S) = F_Q^{\mu_S}(\pi_{Q,S}^\star) - F_Q^{\mu_S}(\widehat{\pi}_S).$$

By the critic-gap result in Proposition F.8, proved below, applied on the finite set S,

$$F_Q^{\mu_S}(\pi_{Q,S}^\star) - F_Q^{\mu_S}(\widehat{\pi}_S) \le \mathrm{sp}_S(\widehat{Q} - Q).$$

Combining proves the first bound. The second follows from

$$\mathrm{sp}_S(\widehat{Q} - Q) \le 2\|\widehat{Q} - Q\|_{\infty,S}.$$

□

Local approximation budget

The implemented screen-then-control actor is locally good when:

1. the shortlist retains high full-soft-optimal mass $\pi_Q^\star(S)$;
2. value-estimation error has small relative-action span on S;
3. wording drift, distribution shift, and model error are represented as value-error terms with controlled span.

$$\text{local regret} \;\le\; \underbrace{-\tau \log \pi_Q^\star(S)}_{\text{shortlist loss}} + \underbrace{\mathrm{sp}_S(\widehat{Q} - Q)}_{\text{relative value error}}.$$

F.6. Critic error, actor sensitivity, and distribution shift

Proposition F.8 (Local objective gap under value-estimation error)**.** *Let $\pi_Q^\star$ and $\pi_{\widehat{Q}}^\star$ be exact soft actors under Q and $\widehat{Q}$, respectively, on the same finite act set and prior. Then*

$$F_Q(\pi_Q^\star) - F_Q(\pi_{\widehat{Q}}^\star) \le \mathrm{sp}(\widehat{Q} - Q).$$

In particular, if $\|\widehat{Q} - Q\|_\infty \le \varepsilon$, then

$$F_Q(\pi_Q^\star) - F_Q(\pi_{\widehat{Q}}^\star) \le 2\varepsilon.$$

Proof. For any π,

$$F_Q(\pi) - F_{\widehat{Q}}(\pi) = \sum_u \pi(u)(Q(u) - \widehat{Q}(u)).$$

Let $e = Q - \widehat{Q}$. Since $\pi_{\widehat{Q}}^\star$ maximizes $F_{\widehat{Q}}$,

$$\begin{aligned} F_Q(\pi_Q^\star) - F_Q(\pi_{\widehat{Q}}^\star) &= F_{\widehat{Q}}(\pi_Q^\star) - F_{\widehat{Q}}(\pi_{\widehat{Q}}^\star) + \sum_u \pi_Q^\star(u)e(u) - \sum_u \pi_{\widehat{Q}}^\star(u)e(\\ &\le \max_u e(u) - \min_u e(u) = \mathrm{sp}(e). \end{aligned}$$

Since $\mathrm{sp}(e) = \mathrm{sp}(\widehat{Q} - Q)$, the first claim follows. The sup-norm bound follows from $\mathrm{sp}(e) \le 2\|e\|_\infty$. □

Proposition F.9 (Span-based actor sensitivity)**.** *Let*

$$\pi^\star_Q = \mathrm{Soft}_\mu(Q/\tau), \qquad \pi^\star_{\widehat{Q}} = \mathrm{Soft}_\mu(\widehat{Q}/\tau).$$

Then

$$\|\pi^\star_{\widehat{Q}} - \pi^\star_Q\|_1 \le \frac{\mathrm{sp}(\widehat{Q} - Q)}{\tau} \le \frac{2\|\widehat{Q} - Q\|_\infty}{\tau}.$$

Proof. Apply the shift-invariant softmax Lipschitz bound to logits Q/τ and $\widehat{Q}/\tau$:

$$\|\,\mathrm{Soft}_\mu(\widehat{Q}/\tau) - \mathrm{Soft}_\mu(Q/\tau)\|_1 \le \mathrm{sp}((\widehat{Q} - Q)/\tau) = \frac{\mathrm{sp}(\widehat{Q} - Q)}{\tau}.$$

□

Proposition F.10 (Distribution-shift sensitivity of $\bar{Q}_d$)**.** *Let d_1 and d_2 be two improvement distributions. Suppose for fixed u,*

$$Q^{\pi_0}_{\mathrm{hi}}(B, u) \in [L_Q, U_Q]$$

for all B. For context c,

$$\left|\bar{Q}_{d_1}(c, u) - \bar{Q}_{d_2}(c, u)\right| \le (U_Q - L_Q)\,\|d_1(\cdot \mid C = c) - d_2(\cdot \mid C = c)\|_{\mathrm{TV}}.$$

Proof. Apply total-variation duality to the bounded function

$$B \mapsto Q^{\pi_0}_{\mathrm{hi}}(B, u) - L_Q \in [0, U_Q - L_Q]$$

under the two conditional hidden-state laws. □

If the total value error decomposes as

$$\widehat{Q} - Q = e_{\mathrm{critic}} + e_{\mathrm{wording}} + e_{\mathrm{shift}} + e_{\mathrm{model}},$$

then

$$\mathrm{sp}(\widehat{Q} - Q) \le \mathrm{sp}(e_{\mathrm{critic}}) + \mathrm{sp}(e_{\mathrm{wording}}) + \mathrm{sp}(e_{\mathrm{shift}}) + \mathrm{sp}(e_{\mathrm{model}}),$$

because span is subadditive. This is the local way to combine critic error, wording drift, and distribution shift in the approximation budget.

F.7. Boundary-level performance-difference identity

The Stage II theorem optimizes a local regularized objective under the improvement distribution d. It is not, by itself, a monotonic-improvement theorem for the full return J. The next identity shows what additional term would have to be controlled.

Theorem F.11 (Boundary-level performance-difference lemma)**.** *Fix a wording policy and the induced boundary-level semi-Markov process. Let π and π_0 be two high-level policies on the same local-context class. Define*

$$V^{\pi_0}(B) = \mathbb{E}_{\pi_0}\left[\sum_{b \ge 0} \gamma^{t_b} G^b \,\middle|\, B^0 = B\right],$$

$$Q^{\pi_0}(B, u) = \mathbb{E}\left[G^0 + \gamma^{\Delta_0} V^{\pi_0}(B^1) \,\middle|\, B^0 = B, U^0 = u\right],$$

and

$$A^{\pi_0}(B, u) = Q^{\pi_0}(B, u) - V^{\pi_0}(B).$$

Let the discounted boundary occupancy measure of π be

$$\rho_\pi(A) = \mathbb{E}_\pi\left[\sum_{b \ge 0} \gamma^{t_b} \mathbf{1}\{B^b \in A\}\right].$$

Assume ρ_π is finite, the two policies have the same initial boundary distribution, and the telescoping tail term vanishes. For example, this holds under bounded values with $\gamma < 1$ and $t_b \to \infty$, or in finite episodic problems. Then

$$J(\pi) - J(\pi_0) = \int \rho_\pi(dB) \sum_u \pi(u \mid C(B)) A^{\pi_0}(B, u).$$

Proof. Define

$$\delta_b = G^b + \gamma^{\Delta_b} V^{\pi_0}(B^{b+1}) - V^{\pi_0}(B^b).$$

By definition of Q^{π_0} and A^{π_0},

$$\mathbb{E}[\delta_b \mid B^b, U^b] = A^{\pi_0}(B^b, U^b).$$

For finite n,

$$\sum_{b=0}^{n} \gamma^{t_b} \delta_b = \sum_{b=0}^{n} \gamma^{t_b} G^b + \sum_{b=0}^{n} \gamma^{t_b + \Delta_b} V^{\pi_0}(B^{b+1}) - \sum_{b=0}^{n} \gamma^{t_b} V^{\pi_0}(B^b).$$

Since $t_{b+1} = t_b + \Delta_b$, this telescopes:

$$\sum_{b=0}^{n} \gamma^{t_b} \delta_b = \sum_{b=0}^{n} \gamma^{t_b} G^b - V^{\pi_0}(B^0) + \gamma^{t_{n+1}} V^{\pi_0}(B^{n+1}).$$

Taking expectations under π and using the vanishing tail term gives

$$\mathbb{E}_\pi\left[\sum_{b \ge 0} \gamma^{t_b} \delta_b\right] = J(\pi) - \mathbb{E}[V^{\pi_0}(B^0)].$$

With the same initial distribution,

$$\mathbb{E}[V^{\pi_0}(B^0)] = J(\pi_0).$$

Now condition on B^b:

$$\mathbb{E}_\pi[\delta_b \mid B^b] = \sum_u \pi(u \mid C(B^b)) A^{\pi_0}(B^b, u).$$

Substituting and using the definition of ρ_π proves the identity. □

Corollary F.12 (Occupancy-shift decomposition)**.** *Assume both ρ_π and ρ_{π_0} are finite boundary occupancy measures. Define*

$$\bar{A}_{\pi_0,\pi}(B) = \sum_u \pi(u \mid C(B)) A^{\pi_0}(B, u).$$

Then

$$J(\pi) - J(\pi_0) = \int \rho_{\pi_0}(dB)\bar{A}_{\pi_0,\pi}(B) + \int (\rho_\pi - \rho_{\pi_0})(dB)\bar{A}_{\pi_0,\pi}(B).$$

If $|\bar{A}_{\pi_0,\pi}(B)| \le A_{\max}$, then

$$J(\pi) - J(\pi_0) \ge \int \rho_{\pi_0}(dB)\bar{A}_{\pi_0,\pi}(B) - A_{\max}\|\rho_\pi - \rho_{\pi_0}\|_{\mathrm{TV}}.$$

Here $\|\rho_\pi - \rho_{\pi_0}\|_{\mathrm{TV}}$ is the total variation norm of the signed finite measure $\rho_\pi - \rho_{\pi_0}$.

Proof. Add and subtract the ρ_{π_0} term in Theorem F.11. The bound follows from total-variation duality applied to the signed measure $\rho_\pi - \rho_{\pi_0}$ and bounded test function $\bar{A}_{\pi_0,\pi}$. □

No global monotonic-improvement theorem

The local Stage II actor is the exact optimizer of a local regularized objective under frozen objects. It is not a general guarantee that replacing π_0 by the local actor improves J. A full monotonic theorem would require controlling the occupancy-shift term in Corollary F.12.

F.8. Stage II diagnostics

Gap	Mathematical quantity	Diagnostic
Local-context aliasing	$V^\star_{\mathrm{full}}(c) - V^\star_{\mathrm{loc}}(c)$.	Reverse-KL teacher dispersion in Theorem F.2.
Wrong teacher aggregation	$\|\pi_{\mathrm{fwd}} - \pi^\star_{\mathrm{loc}}\|_1$.	Compare average-then-softmax with softmax-then-average.
Shortlist loss	$-\tau \log \pi^\star_Q(S)$.	Estimate retained soft-optimal mass.
Critic error	$\mathrm{sp}(\widehat{Q} - Q)$.	Branch replay, calibration, confidence bounds.
Actor sensitivity	$\mathrm{sp}(\widehat{Q} - Q)/\tau$.	Lower τ requires sharper relative values.
Distribution shift	$\|d_1(\cdot \mid C = c) - d_2(\cdot \mid C = c)\|_{\mathrm{TV}}$.	Compare hidden-state posteriors or context-conditioned replay distributions.
Global improvement	Occupancy-shift term in Corollary F.12.	Trust regions or explicit occupancy-shift analysis.

Table F.1. **Stage II diagnostics track the approximation gaps that perturb the local actor.** Each row names the measurable gap and the relevant theorem diagnostic.

G. Structured Extensions and Hard Boundaries

The exact core of the paper uses the simplest geometry: one effective chooser, one selected high-level act, and a finite feasible act set during a frozen round. This appendix records which extensions preserve that geometry and which do not. The guiding principle is simple: any extension that can still be represented as a finite act choice by one effective chooser inherits the same local softmax structure. Extensions that require simultaneous decentralized product policies are a different optimization problem.

What generalizes and what does not

The average-then-softmax theorem generalizes to any *one-effective-chooser finite-action* boundary problem. It does not automatically solve simultaneous-ready decentralized product-policy control. Serialization, public coordination, or coordinate updates add structure; they do not erase the underlying coupling.

G.1. One effective chooser per boundary

The main text describes the ready agent as the chooser. The theorem, however, does not depend on the chooser being a literal agent. It depends on there being one effective decision module at the boundary.

Theorem G.1 (Single effective chooser extension)**.** *Suppose that at each boundary there is exactly one effective chooser R, where R may be the ready agent, a scheduler, a public coordinator, a router, or another designated control module. Let the chooser observe local context $C_R(B)$ and choose from a finite context-indexed act set $\mathcal{U}_R(C_R(B))$. Let $\mu_R(\cdot \mid c)$ be a full-support prior on that feasible act set, and define*

$$\bar{Q}_R(c, u) = \mathbb{E}[Q_R(B, u) \mid C_R(B) = c]$$

under the frozen continuation law and improvement distribution. For $\tau > 0$, the unique optimizer of

$$\max_{\pi_R(\cdot|c) \in \Delta(\mathcal{U}_R(c))} \sum_{u \in \mathcal{U}_R(c)} \pi_R(u \mid c)\bar{Q}_R(c, u) - \tau \mathrm{KL}(\pi_R(\cdot \mid c) \| \mu_R(\cdot \mid c))$$

is

$$\pi^\star_R(u \mid c) = \frac{\mu_R(u \mid c)\exp(\bar{Q}_R(c, u)/\tau)}{\sum_{v \in \mathcal{U}_R(c)} \mu_R(v \mid c)\exp(\bar{Q}_R(c, v)/\tau)}.$$

Proof. Condition on $C_R(B) = c$. The objective decomposes pointwise in c. For fixed c, the problem is a fi-

nite KL-regularized linear optimization over the simplex $\Delta(\mathcal{U}_R(c))$. The same Lagrange multiplier calculation as in Theorem F.1 gives the displayed softmax. Strict concavity follows from $\tau > 0$ and full support of μ_R, so the optimizer is unique. □

This extension is useful for orchestration-heavy runtimes. A scheduler or public coordinator can choose among finite boundary acts while preserving the same exact local geometry. What changes is the information pattern: the chooser's context $C_R(B)$ may differ from any individual agent's local context.

G.2. Finite structured actions

The semantic acts used in the paper are structured records. The theory does not require those records to be atomic labels.

Proposition G.2 (Finite structured-action closure)**.** *Let the chooser's feasible act set at context c be any finite structured set*

$$\mathcal{U}(c) \subseteq \mathcal{T}_1 \times \cdots \times \mathcal{T}_K,$$

where each factor $\mathcal{T}_k$ is finite. Then every exact local-control result that depends only on the one-chooser finite-action geometry applies unchanged.

Proof. The local actor theorem requires only a finite feasible set, a full-support prior on that set, a context-measurable policy, and well-defined act values. It is indifferent to whether an act is represented as an atomic label or as a tuple of finite fields. □

Corollary G.3 (Typed schemas, bounded bundles, and single-ready broadcast)**.** *During a frozen round, each of the following preserves the exact finite-action geometry whenever the resulting high-level act set remains finite:*

1. *typed semantic schemas with finite fields and bins;*
2. *bounded bundles of at most K semantic atoms from a finite atom inventory;*
3. *single-ready broadcast acts of the form (J, a), where*

$$\varnothing \neq J \subseteq \mathcal{N} \setminus \{i_b\}, \qquad a \in \Sigma_{\mathrm{sem}},$$

and the recipient set inventory is finite.

Proof. Each construction produces a finite structured action set. Apply Proposition G.2. □

This corollary is intentionally modest. It says that these objects remain finite actions for the local actor theorem. It does not say that broadcast value is additive across recipients, that bundles lack synergy, or that a particular schema is a good interface. Those are modeling and approximation questions.

G.3. Hierarchical grouped-action decomposition

Grouped actions are useful when a coarse act, such as `warn`, is refined into a set of more specific records. The next theorem is the finite-action algebra behind the interface-refinement result in Appendix C.6.

Theorem G.4 (Exact hierarchical decomposition for grouped actions)**.** *Let the refined feasible act set $\widetilde{\mathcal{U}}(c)$ be partitioned into disjoint groups*

$$\widetilde{\mathcal{U}}(c) = \bigsqcup_{a \in \mathcal{A}(c)} G_a(c).$$

Assume the refined prior factorizes as

$$\widetilde{\mu}(v \mid c) = \mu(a \mid c)\rho_a(v \mid c), \qquad v \in G_a(c),$$

where $\mu(\cdot \mid c)$ is a full-support group prior and $\rho_a(\cdot \mid c)$ is a full-support within-group reference law. Let $\widetilde{Q}(c, v)$ be the refined value. Define the group soft value

$$V_a(c) = \tau \log \sum_{v \in G_a(c)} \rho_a(v \mid c) \exp(\widetilde{Q}(c, v)/\tau).$$

Then the exact refined KL-regularized optimizer factorizes as

$$\widetilde{\pi}^\star(v \mid c) = p^\star(a \mid c) r_a^\star(v \mid c), \qquad v \in G_a(c),$$

where

$$r_a^\star(v \mid c) = \frac{\rho_a(v \mid c) \exp(\widetilde{Q}(c, v)/\tau)}{\sum_{w \in G_a(c)} \rho_a(w \mid c) \exp(\widetilde{Q}(c, w)/\tau)},$$

and

$$p^\star(a \mid c) = \frac{\mu(a \mid c) \exp(V_a(c)/\tau)}{\sum_{a' \in \mathcal{A}(c)} \mu(a' \mid c) \exp(V_{a'}(c)/\tau)}.$$

Proof. Any refined policy can be written as

$$\widetilde{\pi}(v) = p(a) r_a(v), \qquad v \in G_a(c),$$

where p is a distribution over groups and r_a is a conditional distribution within group a. The prior factorization gives

$$\log \frac{\widetilde{\pi}(v)}{\widetilde{\mu}(v)} = \log \frac{p(a)}{\mu(a)} + \log \frac{r_a(v)}{\rho_a(v)}.$$

Therefore

$$\mathrm{KL}(\widetilde{\pi} \| \widetilde{\mu}) = \mathrm{KL}(p \| \mu) + \sum_a p(a) \mathrm{KL}(r_a \| \rho_a).$$

The value term also decomposes:

$$\sum_v \widetilde{\pi}(v)\widetilde{Q}(c,v) = \sum_a p(a) \sum_{v\in G_a(c)} r_a(v)\widetilde{Q}(c,v).$$

Thus the refined objective is

$$\sum_a p(a)\left[\sum_{v\in G_a(c)} r_a(v)\widetilde{Q}(c,v) - \tau \mathrm{KL}(r_a\|\rho_a)\right] - \tau\mathrm{KL}(p\|\mu).$$

For each fixed group a, the inner bracket is a finite KL-regularized problem whose optimum is $V_a(c)$, attained by $r_a^\star$. Substituting these optima leaves

$$\max_p \left[\sum_a p(a)V_a(c) - \tau\mathrm{KL}(p\|\mu)\right],$$

whose optimizer is $p^\star$. Combining $p^\star$ and $r_a^\star$ gives the claimed refined policy. □

Corollary G.5 (Refined interfaces weakly enlarge the exact policy class)**.** *Under the setup of Theorem G.4, any coarse policy $\pi(a \mid c)$ has the refined embedding*

$$\widetilde{\pi}(v \mid c) = \pi(a \mid c)\rho_a(v \mid c), \qquad v \in G_a(c).$$

Therefore optimizing over the refined policy class cannot reduce the exact optimum, before adding any interface-complexity, estimation, or communication burden.

Proof. The embedded policy fixes the within-group conditional to ρ_a. Thus the comparison is relative to the coarse interface whose group value is the ρ_a-averaged refined value. Since the refined policy class contains every such embedded coarse policy and may also choose non-reference within-group conditionals, the refined optimum is at least the optimum over the embedded coarse subclass. □

G.4. Budget-constrained semantic actors

The main text uses scalarized communication costs. Some deployments instead impose explicit budgets, such as maximum expected context occupancy, message count, activation cost, or latency. In finite contexts and finite action sets, budget constraints still preserve the same softmax form through shadow prices.

Theorem G.6 (Budget-constrained actor under Slater feasibility)**.** *Assume:*

1. *a finite context set $\mathcal{C}$ with probabilities $p(c) > 0$;*
2. *finite feasible act sets $\mathcal{U}(c)$;*
3. *full-support priors $\mu(\cdot \mid c)$;*
4. *bounded values $Q(c,u)$;*
5. *resource vectors $g(c,u) \in \mathbb{R}^R$;*
6. *budget vector $\beta \in \mathbb{R}^R$;*
7. *Slater feasibility: there exists a policy $\widetilde{\pi}$ such that, for every resource r,*

$$\sum_c p(c)\sum_u \widetilde{\pi}(u \mid c) g_r(c,u) < \beta_r.$$

Consider

$$\max_\pi \sum_c p(c)\left[\sum_u \pi(u \mid c)Q(c,u) - \tau\mathrm{KL}(\pi(\cdot \mid c)\|\mu(\cdot \mid c))\right]$$

subject to

$$\sum_c p(c)\sum_u \pi(u \mid c) g_r(c,u) \le \beta_r, \qquad r = 1,\dots,R.$$

Then there exists a finite multiplier vector $\lambda^\star \in \mathbb{R}_+^R$ such that the unique optimal policy has the form

$$\pi^\star(u \mid c) = \frac{\mu(u \mid c)\exp\left(\frac{Q(c,u)-(\lambda^\star)^\top g(c,u)}{\tau}\right)}{\sum_v \mu(v \mid c)\exp\left(\frac{Q(c,v)-(\lambda^\star)^\top g(c,v)}{\tau}\right)}.$$

The multipliers satisfy complementary slackness:

$$\lambda_r^\star\left(\sum_c p(c)\sum_u \pi^\star(u \mid c) g_r(c,u) - \beta_r\right) = 0 \qquad \textit{for all } r.$$

Proof. The feasible set is a finite-dimensional product of simplexes intersected with linear inequalities, hence convex and compact. The objective is continuous and strictly concave in π, because it is linear minus τ times a KL term to a full-support prior. Slater feasibility gives strong duality and the KKT conditions for the equivalent convex minimization problem obtained by negating the objective.

For $\lambda \ge 0$, the Lagrangian is

$$\mathcal{L}(\pi,\lambda) = \sum_c p(c)\left[\sum_u \pi(u \mid c)Q(c,u) - \tau\mathrm{KL}(\pi(\cdot \mid c)\|\mu(\cdot \mid c))\right] - \sum_{r=1}^R \lambda_r\left(\sum_c p(c)\sum_u \pi(u \mid c) g_r(c,u) - \beta_r\right).$$

Rearranging,

$$\mathcal{L}(\pi,\lambda) = \lambda^\top\beta + \sum_c p(c)\left[\sum_u \pi(u \mid c)\big(Q(c,u) - \lambda^\top g(c,u)\big) - \tau\mathrm{KL}(\pi(\cdot \mid$$

For fixed λ, maximization over π decomposes by context. Each context-wise maximizer is the KL-softmax actor with adjusted value

$$Q^\lambda(c,u) = Q(c,u) - \lambda^\top g(c,u).$$

By strong duality and KKT, some finite $\lambda^\star \geq 0$ attains the dual optimum and the corresponding $\pi_{\lambda^\star}$ is primal optimal. Complementary slackness is one of the KKT conditions. □

Proposition G.7 (Finite shadow prices can fail without a constraint qualification)**.** *Without a condition such as Slater feasibility, the constrained optimum may require a limiting multiplier rather than a finite $\lambda^\star$. Consider one context, two actions u_0, u_1, full-support prior μ, resource costs*

$$g(u_0) = 0, \qquad g(u_1) = 1,$$

and budget $\beta = 0$. Then every feasible policy satisfies $\pi(u_1) = 0$, but for every finite λ,

$$\frac{\mu(u_1)\exp((Q(u_1)-\lambda)/\tau)}{\mu(u_0)\exp(Q(u_0)/\tau) + \mu(u_1)\exp((Q(u_1)-\lambda)/\tau)} > 0.$$

Thus no finite adjusted-value softmax satisfies the constraint exactly.

Proof. The constraint is

$$\pi(u_1) \leq 0,$$

so feasibility requires $\pi(u_1) = 0$. The displayed softmax probability of u_1 is strictly positive for every finite λ, because the prior has full support and exponentials are positive. The feasible optimum can be approached only as $\lambda \to \infty$. □

G.5. Simultaneous-ready product policies

A centralized coordinator over finite joint acts preserves the softmax geometry. Decentralized product policies do not.

Proposition G.8 (Centralized finite joint coordinator)**.** *Suppose that at a joint boundary a single coordinator chooses a finite joint act*

$$u_R = (u_i)_{i \in R}$$

from a finite joint feasible set $\mathcal{U}_R(c_R)$, using joint context c_R, full-support joint prior $\mu_R(\cdot \mid c_R)$, and conditional mean value $\bar{Q}_R(c_R, u_R)$. Then the unique local KL-regularized optimizer is

$$\pi_R^\star(u_R \mid c_R) = \frac{\mu_R(u_R \mid c_R)\exp(\bar{Q}_R(c_R,u_R)/\tau)}{\sum_{v_R \in \mathcal{U}_R(c_R)} \mu_R(v_R \mid c_R)\exp(\bar{Q}_R(c_R,v_R)/\tau)}.$$

Proof. This is Theorem G.1 applied to the finite joint action space. □

The difficulty appears when the joint policy is required to factor:

$$\pi_R(u_R \mid c_R) = \prod_{i \in R} \pi_i(u_i \mid c_i).$$

The joint value generally couples the agents' actions, so the objective need not be jointly concave in the product-policy parameters.

Theorem G.9 (Coordinate softmax best response for product policies)**.** *Let R be a set of simultaneously ready agents. Each agent $i \in R$ observes local context C_i, chooses $u_i \in \mathcal{U}_i(C_i)$, and uses a policy $\pi_i(u_i \mid C_i)$. The joint policy is the product*

$$\pi(u_R \mid C_R) = \prod_{i \in R} \pi_i(u_i \mid C_i).$$

Fix an improvement distribution over boundary states and contexts, and consider

$$\mathcal{I}(\pi_1, \ldots, \pi_{|R|}) = \mathbb{E}_{B,U_R}\left[Q(B,U_R)\right] - \tau \sum_{i \in R} \mathbb{E}_{C_i}\left[\mathrm{KL}(\pi_i(\cdot \mid C_i) \| \mu_i(\cdot \mid C_i)\right]$$

Hold π_{-i} fixed. Define agent i's coordinate value

$$\bar{Q}_i(c_i, u_i) = \mathbb{E}\left[Q(B, u_i, U_{-i}) \,\middle|\, C_i = c_i,\ U_{-i} \sim \prod_{k \neq i} \pi_k(\cdot \mid C_k)\right].$$

Equivalently, the expectation first draws $B \sim d(\cdot \mid C_i = c_i)$ and then draws U_{-i} conditionally on B from $\prod_{k \neq i} \pi_k(\cdot \mid C_k(B))$. Then the exact coordinate maximizer is

$$\pi_i^\star(u_i \mid c_i) = \frac{\mu_i(u_i \mid c_i)\exp(\bar{Q}_i(c_i,u_i)/\tau)}{\sum_v \mu_i(v \mid c_i)\exp(\bar{Q}_i(c_i,v)/\tau)}.$$

Proof. Holding π_{-i} fixed, all regularization terms for agents $k \neq i$ are constant with respect to π_i. The expected value term can be written as

$$\mathbb{E}_{C_i}\left[\sum_{u_i} \pi_i(u_i \mid C_i)\bar{Q}_i(C_i, u_i)\right].$$

Thus, up to constants independent of π_i,

$$\mathcal{I}(\pi_i; \pi_{-i}) = \mathbb{E}_{C_i}\left[\sum_{u_i} \pi_i(u_i \mid C_i)\bar{Q}_i(C_i,u_i) - \tau \mathrm{KL}(\pi_i(\cdot \mid C_i) \| \mu_i(\cdot \mid C_i))\right].$$

This decomposes pointwise in c_i. Each pointwise problem is a finite KL-regularized local actor problem, whose unique maximizer is the displayed softmax. □

Proposition G.10 (Product-policy objective can be nonconcave)**.** *Consider two simultaneously acting agents with binary actions $\{0,1\}$, uniform priors, and independent policies. Let*

$$p = \pi_1(1), \qquad q = \pi_2(1),$$

and consider

$$\mathcal{I}(p,q) = \mathbb{E}[\mathbf{1}\{U_1 = U_2\}] - \tau \mathrm{KL}(\pi_1 \| \mu_1) - \tau \mathrm{KL}(\pi_2 \| \mu_2).$$

For $\tau < 1/2$, $\mathcal{I}$ is not jointly concave in (p,q).

Proof. We have

$$\mathbb{E}[\mathbf{1}\{U_1 = U_2\}] = pq + (1-p)(1-q) = 1 - p - q + 2pq.$$

The KL-to-uniform terms are

$$\mathrm{KL}(\pi_1 \| \mu_1) = p\log(2p) + (1-p)\log(2(1-p)),$$

and similarly for π_2. At $p = q = 1/2$, the Hessian is

$$H = \begin{pmatrix} -4\tau & 2 \\ 2 & -4\tau \end{pmatrix}.$$

Its eigenvalues are $-4\tau + 2$ and $-4\tau - 2$. If $\tau < 1/2$, one eigenvalue is positive, so the objective is not jointly concave. □

Thus coordinate updates preserve a local softmax best-response form, but they do not provide the global one-simplex theorem. Coordinate ascent may converge only to a stationary point or local optimum; no global optimality is claimed. Simultaneous decentralized product-policy control remains a separate parent class.

G.6. Deliberately omitted or compressed extensions

Several mathematically valid extensions are deliberately not emphasized because they are not central to the paper's claim.

General convex regularizers. Replacing KL by a general convex regularizer yields the usual convex-conjugate actor

$$\pi^\star \in \partial\Omega^*(\bar{Q}).$$

This is standard convex analysis. The paper focuses on KL because it gives the transparent softmax form used by the controller.

Proper-loss Stage I. The log-loss/KL identity extends to any proper scoring rule through the induced proper-loss divergence. This is correct, but KL is the operating object in the main method because it connects directly to Pinsker-style pruning bounds.

Activation submodularity. Information-surrogate activation objectives such as $I(Y; H_S)$ can be submodular under strong conditional-independence assumptions, and can fail sharply under synergy. These facts are useful for proposal heuristics but are orthogonal to the paper's task-value semantic control objective.

Control-as-inference. A literal reward-weighted posterior over semantic acts yields a log-mean-exp score, not the conditional mean semantic value used by the local actor theorem. This is a useful interpretive caveat but not part of the exact derivation.

Broadcast interaction decompositions. Broadcast value may decompose additively only under strong no-interaction assumptions. In general, redundancy, common knowledge, and synergy create higher-order set-function terms. Since broadcast remains a finite structured act in the single-ready setting, the main theory does not need a separate broadcast-value decomposition.

H. Technical Lemmas

This appendix collects the standard tools used in the proofs. It can be skipped on first reading. The lemmas are stated for finite or standard-Borel spaces at the level needed by the paper.

H.1. Total variation and bounded test functions

For probability measures P, Q on a common measurable space, define

$$\|P - Q\|_{\mathrm{TV}} = \sup_A |P(A) - Q(A)| = \frac{1}{2}\int |dP - dQ|.$$

Lemma H.1 (Total variation duality for bounded functions)**.** *If f is measurable and $0 \le f \le V$, then*

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \le V\|P - Q\|_{\mathrm{TV}}.$$

If instead $|f| \le V$, then

$$|\mathbb{E}_P[f] - \mathbb{E}_Q[f]| \le 2V\|P - Q\|_{\mathrm{TV}}.$$

Proof. For the first claim, write $g = f/V \in [0,1]$. The variational characterization of total variation gives

$$\sup_{0 \le g \le 1} |\mathbb{E}_P[g] - \mathbb{E}_Q[g]| = \|P - Q\|_{\mathrm{TV}}.$$

Multiplying by V proves the first bound.

For the second claim, define

$$g = \frac{f + V}{2V} \in [0,1].$$

Then

$$\mathbb{E}_P[f] - \mathbb{E}_Q[f] = 2V\left(\mathbb{E}_P[g] - \mathbb{E}_Q[g]\right),$$

and the first argument applies. □

H.2. Pinsker and KL data processing

Lemma H.2 (Pinsker's inequality)**.** *If $P \ll Q$, then*

$$\|P - Q\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2}D_{\mathrm{KL}}(P\|Q)}.$$

Proof. Let $\delta = \|P - Q\|_{\mathrm{TV}}$, and let A be a measurable set attaining the total variation supremum up to arbitrary approximation. By data processing under the binary map $x \mapsto \mathbf{1}_A(x)$,

$$D_{\mathrm{KL}}(P\|Q) \geq D_{\mathrm{KL}}(\mathrm{Bern}(p)\|\mathrm{Bern}(q)),$$

where $p = P(A)$, $q = Q(A)$, and $|p - q|$ can be taken arbitrarily close to δ. The elementary binary inequality

$$D_{\mathrm{KL}}(\mathrm{Bern}(p)\|\mathrm{Bern}(q)) \geq 2(p-q)^2$$

then gives

$$D_{\mathrm{KL}}(P\|Q) \geq 2\delta^2.$$

Taking square roots proves the result. □

Lemma H.3 (KL data processing)**.** *Let f be measurable. Let $P_f = P \circ f^{-1}$ and $Q_f = Q \circ f^{-1}$ be the pushforward laws. Then*

$$D_{\mathrm{KL}}(P_f\|Q_f) \leq D_{\mathrm{KL}}(P\|Q).$$

Proof. If $P \not\ll Q$, the right-hand side is $+\infty$ and the claim is trivial. Assume $P \ll Q$, and let

$$L = \frac{dP}{dQ}.$$

The Radon–Nikodym derivative of the pushforward law satisfies

$$\frac{dP_f}{dQ_f}(y) = \mathbb{E}_Q[L \mid f = y].$$

Therefore

$$D_{\mathrm{KL}}(P_f\|Q_f) = \mathbb{E}_Q\left[\mathbb{E}_Q[L \mid f] \log \mathbb{E}_Q[L \mid f]\right].$$

Since $x \mapsto x \log x$ is convex, Jensen's inequality gives

$$\mathbb{E}_Q[L \mid f] \log \mathbb{E}_Q[L \mid f] \leq \mathbb{E}_Q[L \log L \mid f].$$

Taking expectations yields

$$D_{\mathrm{KL}}(P_f\|Q_f) \leq \mathbb{E}_Q[L \log L] = D_{\mathrm{KL}}(P\|Q).$$

□

H.3. KL contraction under common mixing

Lemma H.4 (KL under common mixing)**.** *Let λ be a probability measure on an index space $\mathcal{Z}$. Let $z \mapsto P_z$ and $z \mapsto Q_z$ be measurable families of probability measures on a common measurable space. Then*

$$D_{\mathrm{KL}}\left(\int P_z\,\lambda(dz)\middle\|\int Q_z\,\lambda(dz)\right) \leq \int D_{\mathrm{KL}}(P_z\|Q_z)\,\lambda(dz).$$

Proof. Define joint measures on $\mathcal{Z} \times \mathcal{Y}$:

$$\widetilde{P}(dz, dy) = \lambda(dz)P_z(dy), \qquad \widetilde{Q}(dz, dy) = \lambda(dz)Q_z(dy).$$

Projecting $(z, y) \mapsto y$ gives the two mixture laws. By KL data processing,

$$D_{\mathrm{KL}}\left(\int P_z\,\lambda(dz)\middle\|\int Q_z\,\lambda(dz)\right) \leq D_{\mathrm{KL}}(\widetilde{P}\|\widetilde{Q}).$$

The joint KL decomposes as

$$D_{\mathrm{KL}}(\widetilde{P}\|\widetilde{Q}) = \int D_{\mathrm{KL}}(P_z\|Q_z)\lambda(dz),$$

because the λ-terms cancel. This proves the claim. □

H.4. Soft value identity

Let $\mathcal{U}$ be finite, let μ be a full-support prior on $\mathcal{U}$, let $Q : \mathcal{U} \to \mathbb{R}$, and let $\tau > 0$. Define

$$Z_Q = \sum_{u \in \mathcal{U}} \mu(u) e^{Q(u)/\tau},$$

and

$$\pi_Q^\star(u) = \frac{\mu(u)e^{Q(u)/\tau}}{Z_Q}.$$

Lemma H.5 (Soft value identity)**.** *For every $\pi \in \Delta(\mathcal{U})$,*

$$\sum_u \pi(u)Q(u) - \tau\mathrm{KL}(\pi\|\mu) = \tau \log Z_Q - \tau\mathrm{KL}(\pi\|\pi_Q^\star).$$

Consequently,

$$\max_{\pi \in \Delta(\mathcal{U})}\left[\sum_u \pi(u)Q(u) - \tau\mathrm{KL}(\pi\|\mu)\right] = \tau \log Z_Q,$$

with unique maximizer $\pi_Q^\star$.

Proof. Since

$$\pi_Q^\star(u) = \frac{\mu(u)e^{Q(u)/\tau}}{Z_Q},$$

we have

$$\log\frac{\pi(u)}{\mu(u)} = \log\frac{\pi(u)}{\pi_Q^\star(u)} + \frac{Q(u)}{\tau} - \log Z_Q.$$

Therefore

$$\begin{aligned}\sum_u \pi(u)Q(u) - \tau\sum_u \pi(u)\log\frac{\pi(u)}{\mu(u)} &= \sum_u \pi(u)Q(u) \\ &\quad - \tau\sum_u \pi(u)\left[\log\frac{\pi(u)}{\pi_Q^\star(u)} + \frac{Q(u)}{\tau}\right. \\ &= \tau\log Z_Q - \tau\mathrm{KL}(\pi\|\pi_Q^\star).\end{aligned}$$

The maximum is attained uniquely when the KL term is zero, i.e., at $\pi = \pi_Q^\star$. □

H.5. Softmax Lipschitz bounds

For a finite act set and full-support prior μ, define

$$\mathrm{Soft}_\mu(z)(u) = \frac{\mu(u)e^{z_u}}{\sum_v \mu(v)e^{z_v}}.$$

For $v \in \mathbb{R}^{|\mathcal{U}|}$, define the span seminorm

$$\mathrm{sp}(v) = \max_u v_u - \min_u v_u.$$

Lemma H.6 (ℓ_∞-to-ℓ_1 softmax Lipschitz bound)**.** *For any* $z, \hat{z} \in \mathbb{R}^{|\mathcal{U}|}$,

$$\|\mathrm{Soft}_\mu(z) - \mathrm{Soft}_\mu(\hat{z})\|_1 \le 2\|z - \hat{z}\|_\infty.$$

Proof. Let $\nu = \mathrm{Soft}_\mu(z)$. For a direction h,

$$(J(z)h)_u = \nu(u)\left(h_u - \sum_v \nu(v)h_v\right).$$

Therefore

$$\|J(z)h\|_1 \le \sum_u \nu(u)\,|h_u - \mathbb{E}_\nu[h]|\,.$$

If $\|h\|_\infty \le 1$, then both h_u and $\mathbb{E}_\nu[h]$ lie in $[-1, 1]$, so the last display is at most 2. Thus the Jacobian operator norm from ℓ_∞ to ℓ_1 is at most 2, uniformly in z. Integrating the Jacobian along the line segment from $\hat{z}$ to z proves the bound. □

Lemma H.7 (Shift-invariant span softmax bound)**.** *For any* $z, \hat{z} \in \mathbb{R}^{|\mathcal{U}|}$,

$$\|\mathrm{Soft}_\mu(z) - \mathrm{Soft}_\mu(\hat{z})\|_1 \le \mathrm{sp}(z - \hat{z}).$$

Proof. Softmax is invariant to common additive shifts:

$$\mathrm{Soft}_\mu(z + a\mathbf{1}) = \mathrm{Soft}_\mu(z).$$

Therefore, for every $a \in \mathbb{R}$,

$$\|\mathrm{Soft}_\mu(z) - \mathrm{Soft}_\mu(\hat{z})\|_1 = \|\mathrm{Soft}_\mu(z - a\mathbf{1}) - \mathrm{Soft}_\mu(\hat{z})\|_1\,.$$

Using Lemma H.6,

$$\|\mathrm{Soft}_\mu(z) - \mathrm{Soft}_\mu(\hat{z})\|_1 \le 2\inf_a \|(z - \hat{z}) - a\mathbf{1}\|_\infty.$$

The minimizing a is the midpoint of the maximum and minimum entries of $z - \hat{z}$, and

$$2\inf_a \|(z - \hat{z}) - a\mathbf{1}\|_\infty = \mathrm{sp}(z - \hat{z}).$$

□

H.6. Hoeffding and small-variation expansion

Lemma H.8 (Hoeffding's lemma)**.** *If X is a real random variable with $\mathbb{E}[X] = 0$ and $X \in [a, b]$ almost surely, then for every $\lambda \in \mathbb{R}$,*

$$\log \mathbb{E}[e^{\lambda X}] \le \frac{\lambda^2(b-a)^2}{8}.$$

Proof. By convexity of the exponential, for $x \in [a, b]$,

$$e^{\lambda x} \le \frac{b-x}{b-a}e^{\lambda a} + \frac{x-a}{b-a}e^{\lambda b}.$$

Taking expectations and using $\mathbb{E}[X] = 0$ gives

$$\mathbb{E}[e^{\lambda X}] \le \frac{b}{b-a}e^{\lambda a} - \frac{a}{b-a}e^{\lambda b}.$$

The scalar function on the right is bounded by

$$\exp\left(\frac{\lambda^2(b-a)^2}{8}\right),$$

which is Hoeffding's elementary two-point bound. Taking logarithms proves the result. □

Lemma H.9 (Bounded cumulant expansion)**.** *Let X be bounded, mean zero, and let $|\lambda|$ be sufficiently small. Then*

$$\log \mathbb{E}[e^{\lambda X}] = \frac{\lambda^2}{2}\,\mathrm{Var}(X) + O\Big(|\lambda|^3\mathbb{E}|X|^3 e^{|\lambda|\|X\|_\infty}\Big).$$

In particular, if X is uniformly bounded and $\lambda \to 0$,

$$\log \mathbb{E}[e^{\lambda X}] = \frac{\lambda^2}{2}\,\mathrm{Var}(X) + O(\lambda^3).$$

Proof. Using Taylor's theorem,

$$e^{\lambda X} = 1 + \lambda X + \frac{\lambda^2 X^2}{2} + O(|\lambda|^3|X|^3 e^{|\lambda|\|X\|_\infty}).$$

Taking expectations and using $\mathbb{E}[X] = 0$,

$$\mathbb{E}[e^{\lambda X}] = 1 + \frac{\lambda^2}{2}\mathbb{E}[X^2] + O(|\lambda|^3\mathbb{E}|X|^3 e^{|\lambda|\|X\|_\infty}).$$

Since $\log(1+y) = y + O(y^2)$ and $y = O(\lambda^2)$, the displayed expansion follows. □

H.7. Blackwell garbling

The Stage I decision-value appendix uses only the easy direction of Blackwell's theorem: a more informative experiment can simulate any decision rule available to its garbling.

Lemma H.10 (Garbling cannot increase Bayes value)**.** *Let S_v and S_u be signals about Θ. Suppose S_u is a garbling of S_v: there exists a Markov kernel $T(ds_u \mid s_v)$ such that*

$$P(ds_u \mid \theta) = \int T(ds_u \mid s_v) P(ds_v \mid \theta).$$

Then for every bounded decision problem with action a and utility $r(a, \theta)$, the Bayes value of observing S_v is at least the Bayes value of observing S_u.

Proof. Take any decision rule $\delta_u(a \mid s_u)$ using S_u. Construct a rule using S_v by simulating the garbling:

$$\delta_v(a \mid s_v) = \int \delta_u(a \mid s_u) T(ds_u \mid s_v).$$

The joint distribution of (Θ, A) induced by observing S_v, garbling to S_u, and applying δ_u, is exactly the distribution induced by observing S_u directly and applying δ_u. Therefore every S_u-based rule can be matched by an S_v-based rule with the same expected utility. Optimizing over all S_v-based rules can only do better. □

I. Claim Matrix, Non-Claims, and Diagnostics

This appendix is a checklist for interpreting the paper. It separates exact theorem claims, controlled empirical diagnostics, approximation analyses, extension status, architectural recommendations, and explicit non-claims. The separation is important because the paper's core thesis is not that arbitrary communication always helps; it is that communication boundary control exposes a finite, tractable, and auditable layer for deciding when, whom, and what to communicate.

Nonclaim summary

The paper does not claim exact control over arbitrary natural-language strings, identification from passive logs in general, or global monotonic improvement from a local actor update. Exactness is roundwise: one effective chooser, finite exposed high-level acts, fixed semantic interface, fixed local context, fixed cost convention, fixed continuation policy, and fixed wording policy.

I.1. Core exact claims

Claim	Exact statement	Where
Boundary process	If the boundary snapshot contains all persistent control-relevant state and the control-relevant marginal is audit-state invariant, completed events induce a well-defined semi-Markov boundary kernel for control values.	Proposition B.1
Semantic abstraction	Semantic records are exact when their delivered events induce the same control-relevant quotient kernel.	Theorem C.2
Raw wording boundary	If future policies can read unrestricted raw wording, within-class wording can become a hidden control channel.	Proposition C.5
Stage I KL identity	Under exact interventional evaluation and Bayes-optimal log loss, send-versus-yield predictive improvement equals $D_{\mathrm{KL}}(P_x^u \| P_x^{\varnothing})$.	Proposition E.1
Local pruning	Small calibrated decision-relevant KL can upper-bound gross recipient-local proxy gain and justify local pruning under a specified proxy and cost lower bound.	Theorem E.4
Stage II local actor	For a finite context-measurable frozen act set, fixed wording policy, fixed continuation policy, and fixed improvement distribution, the local KL-regularized actor softmaxes conditional mean semantic values.	Theorem F.1
Aliasing gap	The cost of local-context aliasing is an exact reverse-KL disagreement between the deployed local actor and hidden-state soft teachers.	Theorem F.2
Wording invariance	Within-class wording is interchangeable when it preserves the control-relevant quotient boundary kernel. Approximate drift is bounded by total variation times value span.	Theorem C.6
Interface refinement	Splitting a coarse semantic class has an exact soft option value driven by within-class value dispersion, relative to the stated reference-realization coarse baseline.	Theorem C.10
Approximation regret	For a shortlist actor with approximate values, local regret is bounded by missed soft-optimal mass plus relative value-error span.	Theorem F.7

Table I.1. **Exact claims are scoped to frozen-round communication boundary control.** Each row states the theorem-level claim and where it is established.

I.2. Approximation analyses

The following statements are not exactness assumptions. They quantify what changes when a practical system departs from the frozen finite-interface ideal.

Approximation layer	Analysis	Where
Passive logged data	Stage I and Stage II targets are interventional. Observational identification requires consistency, positivity, and context-aware ignorability.	Theorem D.1
Zero propensity	If a legal act is never taken, replayed, randomized, or modeled at context c, its interventional value is not identified in a nonparametric model class.	Theorem D.3
Runtime-pool-only logs	A runtime candidate pool is not a full legal inventory; omitted legal acts are non-identifiable without additional structure.	Corollary D.4
Coarse Stage I conditioning	Suppressing relevant pre-intervention variables can only reduce exact send-versus-yield KL and can create false negatives.	Proposition E.2
Learned Stage I predictors	Plug-in log-score gaps decompose into true KL minus send-predictor calibration error plus transported-baseline error.	Proposition E.6
Sender-side recipient model	Bounded usefulness scores inherit a TV error from an imperfect sender belief over recipient state.	Proposition E.7
Shortlists	Restricting Stage II to a subset incurs exact local loss $-\tau \log \pi^{\star}_{Q}(S)$.	Proposition F.5
Critic error	Local objective error is controlled by the span of relative value-estimation error.	Proposition F.8
Actor sensitivity	Softmax policy perturbation is bounded by value-error span divided by temperature.	Proposition F.9
Distribution shift	Changing the hidden-state distribution within a context class changes $\bar{Q}_d(c, u)$ by a TV-sensitive amount.	Proposition F.10
Wording drift	Within-class realization updates are safe only up to measured value spread and support leakage.	Proposition C.8, Corollary C.9
Global return	The local actor theorem is not a monotonic-improvement theorem; a full J-improvement statement requires occupancy-shift control.	Theorem F.11, Corollary F.12

Table I.2. **Approximation analyses measure departures from the frozen finite-interface ideal.** Each row names the practical gap, summarizes the analysis, and points to the supporting result.

I.3. Controlled empirical diagnostics

The empirical claims in the paper are diagnostic claims about the controlled synthetic boundary suite. They are not claims of broad deployed-agent superiority.

Empirical claim	Evidence and scope	Where
Decision-focused relevance is a better Stage I target than generic informativeness.	Controlled diagnostics show AUPRC 1.000 for decision-focused usefulness versus 0.861 for generic likelihood, and Stage I AUROC 0.994 versus 0.707 for the strongest generic baseline.	Tables J.1 and J.3
The finite screen–control–realize actor is useful in the controlled suite.	The main actor strictly improves over $23/26$ required controlled baselines and achieves mean return 4.191 in the diagnostic campaign; the remaining three are non-strict ties under the pre-specified win criterion, and the full baseline sweep reports mean return 4.369.	Tables J.1 and J.2
Proposal and shortlist restrictions can hide semantic value.	Width-2 semantic bundles improve mean return by $+1.775$ over top-1 and reduce miss rate from 0.500 to 0.000.	Tables J.1 and J.3
Wording validity must be measured as control drift.	The controlled wording stress test has support leakage 0.000 in the safe channel-side regime and 1.000 in the leaky regime.	Tables J.1 and J.3
Reduced scale suites preserve the controlled mechanism.	The medium and large controlled scale suites each beat $6/6$ required baselines.	Table J.4

Table I.3. **Empirical claims are scoped to controlled diagnostics.** The paper uses these experiments to test diagnostic predictions of the theory, not to claim broad benchmark performance.

I.4. Extension status

Extension	Status	Reason
Ready agent replaced by scheduler or coordinator	Exact	One effective chooser over finite acts preserves the local softmax geometry.
Finite typed semantic schemas	Exact	A finite schema inventory is a finite structured action set.
Bounded semantic bundles	Exact if finite	Fixed maximum bundle width over finite atoms remains finite.
Single-ready broadcast	Exact if finite	Broadcast recipient sets are finite structured acts in the single-chooser regime.
Grouped/refined actions	Exact	Prior factorization gives exact hierarchical decomposition.
Explicit resource budgets	Exact under finite support and Slater	Strong duality yields adjusted-value softmax with finite shadow prices.
Budget constraints without Slater	Boundary/limiting case	Finite shadow prices can fail; limiting multipliers may be required.
Centralized joint coordinator	Exact	A finite joint act is still one finite act chosen by one effective chooser.
Simultaneous-ready decentralized product policies	Not solved globally	Joint product-policy objective can be nonconcave. Coordinate best responses remain softmax.
General convex regularizers	Standard extension	Actor is given by the conjugate-gradient/subgradient map; not central to this paper.
Proper-loss Stage I	Standard extension	Proper scoring rules induce divergences, but KL/log loss is the paper's operating target.
Activation submodularity	Omitted surrogate	Concerns an information surrogate, not task-value semantic control.
Control-as-inference	Interpretive only	Literal reward-weighted inference gives log-mean-exp scores, not the mean values in the theorem.

Table I.4. **Extensions preserve exactness only when they keep a finite one-chooser geometry.** The status column separates exact variants from boundary cases, standard extensions, and unsupported regimes.

I.5. Architectural consequences

The theory also implies several architectural lessons. These are not separate performance theorems, but they are direct consequences of the exactness conditions.

Line of work	Primary object optimized	Communication event	Boundary-control distinction
Fixed and adaptive LLM scaffolds	Roles, routes, turns, tools, topology, interruptions, retained content	Usually specified or exposed by the scaffold	Boundary act competes with local action and yield.
MARL communication	Costly communication action or channel under partial observability	Defined by the decision model	Language-facing finite semantic record plus wording realization.
LLM textual credit and verbal RL	Produced messages, feedback, traces, collaborative behavior	Interaction has occurred	Decision before the communication event is realized.
Structured outputs and tool calls	Schema instance, parser target, tool arguments	Usually invoked by a task protocol	Schema-like record is the finite control act.
This paper	**Boundary event: local action, yield, or semantic send**	**Chosen at ready-agent boundary**	**Finite event-level choice exposes relevance/value separation, context aliasing, and wording drift.**

Table I.5. **Positioning by optimized communication object.** Communication boundary control isolates the communication event itself as the learnable boundary decision, rather than optimizing only wording, routing, topology, pruning, or textual credit inside an already instantiated communication process.

Canonical records are the robust control-plane object. Future policies should consume $\Gamma(D)$, not unrestricted raw wording. Raw text can be retained in an audit state for human inspection, but if future policies can read arbitrary wording, the wording is part of the control state.

A semantic field is anything that changes downstream control. Confidence, urgency, threshold precision, evidence strength, provenance, and pointer choice are not harmless phrasing if they affect future decisions. They should be represented as semantic fields, sanctioned channel-side terms, or approximation errors.

Interface refinement should target value dispersion. A class should be split when its within-class variants have materially different downstream values. Theorem C.10 gives the soft option value of such a split relative to the stated reference-realization coarse baseline.

Stage I should be recall-biased. Predictive usefulness is a screen, not the final communication objective. Hard pruning is justified only by calibrated upper bounds on gross local proxy gain under a specified local proxy and cost lower bound.

Stage II needs relative action values. The local actor is sensitive to relative value differences, not common additive shifts. Critic diagnostics should therefore emphasize span error, not only absolute value error.

Boundary-bundle data are a data contract, not an implementation detail. Training a theory-aligned Stage I/Stage II controller requires boundary-level interventional support: legal inventories, forced or replayed branch outcomes, recipient-side next-H deltas, and integrity metadata. Schema-valid runtime logs are insufficient if they are runtime-pool-only or lack alternate branch outcomes.

I.6. Explicit non-claims

For clarity, the paper does *not* claim any of the following.

1. **No exact RL over arbitrary natural-language strings.** Exactness stops at a finite semantic control layer under a fixed wording policy.
2. **No claim that the first semantic interface is final.** The interface may be refined between rounds. Exact claims are relative to the interface fixed during the round.
3. **No claim that parser-consistent wording is automatically harmless.** Parser consistency is not enough. Within-class wording must preserve control-relevant quotient dynamics or have measured value drift.
4. **No guarantee that token-space trust regions enforce semantic invariance.** A text-side KL penalty can stabilize wording, but it does not prove that wording preserves control meaning.
5. **No identification from passive logs in general.** Stage I and Stage II targets are interventional. Passive logs require adjustment assumptions and positivity; zero-propensity acts are not identifiable nonparametrically.
6. **No guarantee that proposal shortlists are globally safe.** Shortlisting is an approximation layer. Its local

loss is missed soft-optimal mass, and omitted useful acts create restriction loss.

7. **No global monotonic-improvement theorem for J.** The local actor optimizes a frozen local regularized objective. Full return improvement additionally depends on occupancy shift.

8. **No solution to simultaneous-ready decentralized product control.** The one-chooser theorem does not extend verbatim to decentralized product policies. Coordinate best responses are softmax, but the joint problem can be nonconcave.

9. **No claim that act/wording separation is permanent.** The split is a useful control abstraction under a fixed interface. Future architectures may jointly optimize intent and realization, but then their exactness conditions differ.

10. **No broad practical superiority claim from incomplete boundary data.** The controlled diagnostics support diagnostic claims under the synthetic finite-interface setting; they are not evidence of broad deployed-agent superiority.

I.7. Diagnostic checklist

The table below summarizes what should be measured when implementing communication boundary control.

Gap	Diagnostic	Relevant result
Boundary state insufficiency	Audit whether two histories with the same state summary induce different next-boundary laws.	Proposition B.1
Raw-text covert channel	Compare policies with access to raw wording versus canonical records only.	Proposition C.5
Semantic class too coarse	Estimate within-class value spread or refinement gain.	Proposition C.8, Theorem C.10
Support leakage	Estimate $\Pr[M \notin \mathcal{M}_a(c, u)]$.	Corollary C.9
Stage I calibration	Compare predicted usefulness to replayed or randomized target laws.	Proposition E.6
Stage I false negatives	Refine recipient abstraction or include omitted pre-intervention covariates.	Proposition E.2
Unsafe pruning	Check calibrated gross-gain upper bound, local utility span, and communication-cost lower bound.	Theorem E.4
Runtime-pool-only data	Count legal acts with no branch support.	Theorem D.3
Shortlist loss	Estimate retained full-soft-optimal mass $\pi^\star_Q(S)$.	Proposition F.5
Local-context aliasing	Estimate reverse-KL teacher dispersion.	Theorem F.2
Wrong distillation target	Compare average-then-softmax with softmax-then-average.	Proposition F.3
Critic error	Measure span of relative value error across feasible acts.	Proposition F.8
Distribution shift	Compare $d(B \mid C = c)$ across training and deployment.	Proposition F.10
Global improvement risk	Bound or monitor occupancy shift after policy updates.	Corollary F.12
Budget feasibility	Check Slater-like slack before interpreting finite shadow prices.	Theorem G.6, Proposition G.7

Table I.6. **The diagnostic checklist maps observed failures to the approximation layer responsible.** The relevant result column identifies which theorem or proposition explains each gap.

How to use the checklist

The checklist maps observed failures to the responsible approximation layer rather than requiring every implementation to prove every bound. A harmful shortlist indicates restriction loss; wording that changes the planner's next action indicates a coarse semantic interface or an unsanctioned channel.

J. Controlled Experiment Details

This appendix describes the controlled experiments used in Section 5. The experiments are diagnostic rather than benchmark-competitive: each family isolates one approximation layer predicted by the theory. The controlled environment exposes the hidden boundary state, finite semantic inventory, legal recipients, communication cost, and counterfactual continuations needed to evaluate Stage I relevance, Stage II semantic value, shortlist loss, bundle loss, and wording drift.

J.1. Controlled boundary suite

Each episode contains a ready-agent boundary at which the actor chooses among local action, yield, and recipient-targeted semantic sends. A send is a finite canonical record; wording is generated or perturbed only after the record is selected. The suite includes families in which generic information is predictive but not decision-relevant, low-shift messages are valuable despite small generic likelihood changes, recipient targeting matters, paraphrases should be control-equivalent, and bundles of semantic records carry value not available to a top-1 candidate.

The principal metric for policy comparison is mean net return, including the communication cost convention used by the round. Mechanism diagnostics use AUPRC and AUROC for Stage I relevance, shortlist recall and miss rate for proposal/shortlist diagnostics, support leakage for wording validity, and oracle gaps for remaining value-estimation headroom.

Diagnostic question	Controlled result	Interpretation
Does decision-focused relevance differ from generic informativeness?	Decision-focused AUPRC 1.000 versus generic-likelihood AUPRC 0.861; Stage I AUROC 0.994 versus 0.707 for the strongest generic baseline.	Stage I should target recipient-side decision relevance, not generic predictive informativeness.
Does the full boundary actor beat simple communication policies and component ablations?	Main actor mean return 4.191, communication cost 1.44, send rate 0.80; it strictly improves over $23/26$ required controlled baselines, with the remaining three non-strict ties under the pre-specified win criterion.	The finite screen–control–realize decomposition gives useful structure in the controlled boundary problem.
Can shortlist or proposal width hide semantic value?	Width-2 semantic bundles improve mean return by $+1.775$ over top-1 and reduce miss rate from 0.500 to 0.000.	Proposal and shortlist loss are real approximation terms, not bookkeeping.
Does wording validity need a control-drift test?	The controlled wording stress test has support leakage 0.000 in the safe channel-side regime and 1.000 in the leaky regime.	Parser-equivalent wording is safe only when it preserves induced control-relevant dynamics.
Do reduced scale suites preserve the controlled mechanism?	The medium and large reduced suites each beat $6/6$ required baselines, with send rate 0.80.	The mechanism is stable across the tested controlled scale variants, within the same finite-interface setting.

Table J.1. **Controlled diagnostics test the theory-scoped failure modes.** Metrics are reported on the controlled synthetic boundary suite; higher AUPRC, AUROC, mean return, and baseline-win counts are better, while lower leakage and miss rate are better. The table supports diagnostic predictions under a finite semantic interface and does not assert broad benchmark superiority.

J.2. Baselines and ablations

The controlled baseline sweep compares the main screen–control–realize actor against fixed communication policies, pooled single-agent behavior, and component ablations that remove or alter the Stage I target, Stage II value objective, recipient conditioning, proposal width, bundle width, and wording interface. These baselines test diagnostic predictions, not broad leaderboard rank.

Table J.1 reports the claim-driven diagnostic campaign, where the main actor return is 4.191. Table J.2 reports the separate full baseline sweep, where the same actor achieves 4.369. The difference is the evaluation campaign, not a conflicting metric. Table J.2 shows selected rows from the full 26-variant controlled baseline sweep; the anonymous supplementary aggregate files contain all required variants, metrics, pairwise deltas, and the pre-specified win indicators.

Variant	Return ↑	Success ↑	Cost ↓	Send/boundary
Main boundary actor	**4.369**	**0.600**	1.440	0.800
No communication	1.739	0.000	**0.000**	0.000
Always communicate	2.452	0.433	1.800	1.000
Budgeted heuristic	2.391	0.433	1.800	1.000
Fixed star	1.917	0.300	1.800	1.000
Fixed chain	1.968	0.133	0.360	0.200
Single-agent pooled	4.153	**0.600**	1.800	1.000
Stage I global threshold	3.746	**0.600**	1.440	0.800
Top-1 shortlist	2.953	0.433	1.440	0.800
Stage II reward only	4.173	**0.600**	1.440	0.800
Recipient model removed	3.103	0.500	1.440	0.800
Recipient-model oracle	4.593	**0.600**	1.440	0.800

Table J.2. **Selected rows from the controlled baseline sweep.** Results are reported on the synthetic controlled boundary suite. Return includes the round's communication cost. Bold marks the highest non-oracle return and all tied best success values. The recipient-model oracle is an upper comparator and is not a required baseline.

The main actor strictly improves over $23/26$ required controlled baselines; the remaining three are non-strict ties under the pre-specified win criterion. The strongest non-oracle comparator in this sweep is the single-agent pooled baseline, which reaches the same success rate but lower return and higher communication cost. The recipient-model oracle exceeds the main actor; this is expected and useful because it quantifies headroom from imperfect recipient state. The three tied required variants are Stage II Distillation, Stage II Prior-Aware Softmax-Then-Average, and Stage II Global Temp. The recipient-model oracle is an upper comparator, not a required baseline.

J.3. Mechanism diagnostics

Mechanism	Measured diagnostic	Conclusion
Decision-focused Stage I	Decision-focused AUPRC 1.000; generic-likelihood AUPRC 0.861; Stage I AUROC 0.994 versus 0.707.	Generic predictive information is not the right Stage I target.
Pruning calibration	Low-shift shortlist recall 1.000; warn-pair recall 1.000.	Low usefulness supports conservative pruning only when calibrated and scoped.
Bundle reasoning	Width-2 bundles improve mean return by $+1.775$ over top-1 and reduce miss rate from 0.500 to 0.000.	Restricting to a single semantic candidate can lose value.
Wording validity	Controlled stress-test support leakage 0.000 in the safe channel-side regime and 1.000 in the leaky regime.	Parser consistency is insufficient; control drift must be measured.
Oracle headroom	Oracle value gap 0.71875.	Remaining performance loss is concentrated in value estimation rather than in the exact finite-action formulation.

Table J.3. **Controlled mechanism diagnostics.** Each diagnostic corresponds to one approximation term in the theory: relevance, pruning, proposal or bundle loss, wording drift, and value-estimation headroom.

J.4. Reduced scale suites

We also evaluate reduced medium and large synthetic scale suites that preserve the same semantic-control mechanism while changing the controlled environment size. These are not broad scaling claims; they are robustness checks within the same diagnostic family.

Suite	Return ↑	Cost ↓	Send/boundary	Required wins
Medium controlled scale	4.085	1.440	0.800	$6/6$
Large controlled scale	4.081	1.440	0.800	$6/6$

Table J.4. **Reduced scale-suite robustness.** The main actor preserves the controlled mechanism advantage in both reduced scale suites.

J.5. Empirical scope

The experiments in this appendix are controlled diagnostics. They are designed to test the theory's mechanism-level predictions under a finite semantic interface, observable counterfactual branches, and a fixed communication-cost convention. They are not presented as a broad deployed-agent benchmark.

K. Reproducibility, Compute, and Responsible Use

This appendix records the information needed to reproduce the controlled experiments reported in the paper. It is written for an anonymous submission: paths are relative to the project root or to the anonymized supplementary package, and absolute machine-local paths are represented by the `PROJECT_DATA_ROOT` environment variable used in the commands below.

K.1. Code entry points and configurations

The controlled synthetic results are generated by three orchestration scripts:

- `scripts/run_synthetic_foundation_campaign.py`: runs the claim-driven controlled diagnostic campaign and reduced scale suites.
- `scripts/run_claim_suite.py`: runs the individual mechanism diagnostics, including Stage I relevance, pruning calibration, Stage II objective diagnostics, wording regimes, bundle reasoning, oracle gaps, and cost sweeps.
- `scripts/run_synthetic_baseline_suite.py`: runs the controlled baseline and ablation sweep.

The corresponding configuration surface is:

- `configs/experiment/baseline_suite_manifest.yaml`;
- `configs/experiment/baseline_suite_scale_medium_manifest.yaml`;
- `configs/experiment/baseline_suite_scale_large_manifest.yaml`;
- `configs/experiment/synthetic_baseline_suite.yaml`;
- `configs/experiment/claim_suite/`;
- `configs/benchmark/synthetic_claim_grade.yaml`.

The reported controlled results use evaluation mode and seed 7. The controlled suite is generated from the repository's synthetic benchmark configuration rather than from a collected dataset. All central numbers in Tables J.1 to J.4 are computed from structured run summaries and aggregate JSON/CSV reports emitted by these scripts.

K.2. Reproduction commands

The following commands reproduce the controlled results when run from the project root with the controlled-suite dependencies installed:

```
PYTHONPATH=src
PROJECT_DATA_ROOT=/path/to/output \
  python scripts/run_synthetic_foundation_campaign.py
  \
  --claim-grade-baseline-manifest \
  configs/experiment/baseline_suite_manifest.yaml
  \
  --scale-medium-baseline-manifest \
  configs/experiment/baseline_suite_scale_medium_manifest.yaml
  \
  --scale-large-baseline-manifest \
  configs/experiment/baseline_suite_scale_large_manifest.yaml
  \
  --strict-preflight

PYTHONPATH=src
PROJECT_DATA_ROOT=/path/to/output \
  python scripts/run_claim_suite.py \
  --config-dir
  configs/experiment/claim_suite \
  --mode eval \
  --strict-preflight

PYTHONPATH=src
PROJECT_DATA_ROOT=/path/to/output \
  python
  scripts/run_synthetic_baseline_suite.py \
  --suite-manifest \
  configs/experiment/baseline_suite_manifest.yaml
  \
  --mode eval \
  --strict-preflight
```

The `PROJECT_DATA_ROOT` directory stores generated runs, aggregate metrics, pairwise deltas, communication-frontier CSVs, and Markdown tables. The paper tables should be regenerated from the structured aggregate files, not by copying console output.

K.3. Expected result artifacts

The reproduction commands emit the complete result artifacts used by the paper: structured run summaries for each configured suite; aggregate JSON/CSV metric tables for the claim-driven diagnostics, full 26-variant baseline sweep, mechanism diagnostics, and medium/large reduced scale suites; pairwise deltas and pre-specified win indicators; communication-frontier CSVs; and Markdown tables used to transcribe the LaTeX tables. The main PDF prints compact result tables for readability, while the anony-

mous supplementary aggregate files retain the complete controlled-suite outputs behind every reported number.

K.4. Environment

The controlled experiments were run with Python 3.13 and the repository package installed as `native-multi-agent`. The run summaries record the following key package versions for the controlled suite: `PyYAML==6.0.2`, `pytest==8.4.2`, and `wandb==0.25.1`. The controlled synthetic suite runs on a single CPU-capable machine and does not use a GPU, TPU, or specialized accelerator path.

K.5. Randomness and statistical treatment

The main controlled results use a fixed evaluation seed, seed 7, and are reported as deterministic diagnostic readouts of the configured controlled suite. We therefore do not use these numbers to make population-level claims over arbitrary task distributions. The paper's empirical claims are scoped to diagnostic testing in the controlled suite. Broader deployment claims would require repeated runs, stronger comparator sets, and uncertainty reporting over task sampling, model stochasticity, and runtime variation.

K.6. Compute and storage

The reported controlled experiments require single-machine CPU execution, local disk storage for run summaries, and enough memory to hold the synthetic benchmark state and aggregate tables. The stored artifacts are structured JSON summaries, CSV metric tables, pairwise deltas, communication-frontier CSVs, and Markdown tables. No GPU hours are incurred for the central controlled experiments because no GPU-specific computation path is used.

K.7. Data, licenses, and human subjects

The central experiments use synthetic controlled environments generated by the code and configs above. They do not introduce a new dataset, do not scrape human-generated data, and do not involve human subjects or crowdsourcing. The implementation is released under the license included in the anonymous supplementary package.

K.8. Use of LLMs and agents

The reported controlled synthetic diagnostics do not use LLM-generated judgments as primary metrics. The paper studies LLM-agent communication as an object of analysis, but the central empirical evidence is generated by the controlled synthetic suite rather than by hosted language-model judgments.

L. AIWILD Submission Notes

This appendix records workshop-specific packaging notes for the anonymous AIWILD submission version. The main paper uses the ICML 2026 AIWILD workshop style file in review mode and omits venue-specific checklist material not requested by the workshop. References and supplementary material are outside the workshop's regular-track page limit.

Anonymity. All submitted material should remain double-blind: the main PDF, supplement, code package, generated metadata, and any linked material must omit author names, affiliations, personal or lab URLs, non-anonymous repository owners, acknowledgments, machine-local paths, and identifiable logs. The submitted code package should exclude `.git`, caches, local run logs, archives not needed for reproduction, and generated files containing user-specific paths.

LLM-use posture. The controlled diagnostics reported in the paper do not use LLM-generated judgments as primary metrics. Any AI assistance used for writing, editing, or formatting must be verified by the authors; the submitted paper should remain primarily human-authored and should not include AI-generated results, citations, experiments, or analysis that the authors have not checked.