INCENTIVE ALIGNMENT RESHAPES COORDINATION OVER UNRELIABLE COMMUNICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

When communication is unreliable, agents with the same objective may need different participation rules. We study Bayesian coordination games in which utilities interpolate between private and group-average payoffs. For full-quorum relay lines with heterogeneous links and an informed root that vetoes unsafe operation, we characterize all coordinating observation-monotone pure equilibria: each consists of a message-responsive prefix and an unconditionally acting suffix, determined by three posterior-odds inequalities. Permuting the same link reliabilities can change the equilibrium assignment while preserving end-to-end reliability, so no scalar summary of the channel determines the local policies. Under a uniform veto condition, equilibrium existence and the best certified welfare are monotone in alignment, although a canonical equilibrium selection need not be. The characterization yields an exact compiler that separates equilibrium certification, welfare-based selection, and local execution. Experiments with open-weight language models distinguish the formal certificate used for audit from the representation given to the executor. On fresh relay games, executing the full audit certificate directly loses welfare relative to not coordinating at all, whereas the same certificate compiled into a local decision-table execution card is executed almost exactly and recovers the welfare of the selected policy; violations of the declared fallback remain when the certificate's assumptions fail. The results connect an exact coordination structure to a local execution interface while separating policy fidelity from validity enforcement.

1 INTRODUCTION

Multi-agent systems increasingly consist of autonomous components, among them language-model agents, that must commit to a joint action while information about the environment reaches them through unreliable channels: a message may be delayed, dropped, or never confirmed. Classical results rule out protocols that *guarantee* simultaneous coordination over such channels, and even a shared objective removes strategic disagreement without removing the information constraint (Halpern & Moses, 1990; Rubinstein, 1989). In practice the question is quantitative rather than binary: given the incentives and the communication structure, which agents should condition their participation on a message they may never receive, and how can the resulting policy be handed to an agent so that it is executed as designed, including when its preconditions fail? These are the questions of this paper.

Consider an example. Three agents must act together on an operation. Agent 1 observes whether it is SAFE and, if so, sends a message through agents 2 and 3. Each relay agent sees only whether its own message arrived. Should an agent that received nothing participate? The answer can differ across agents, even when their payoffs are identical.

Take prior $\pi = 0.4$, reward $U = 1$ for a safe full quorum, loss $L = 1$ for acting without a quorum, and loss $H = 10$ for an unsafe quorum. Let utilities place weight $\alpha = 0.8$ on the group-average payoff. With link reliabilities $(q_1, q_2) = (0.8, 0.1)$, the only coordinating pure equilibrium has agent 2 act on receipt and agent 3 act regardless. Swapping the links to $(0.1, 0.8)$ preserves their product and multiset, but the only coordinating cutoff equilibrium makes both relay agents act regardless. Thus neither end-to-end reliability nor any permutation-invariant summary determines the local policy assignment.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Lossy communication creates uncertainty about other agents' information and participation. We ask how that uncertainty interacts with incentives in a fixed protocol. Utilities take the standard form $u_i^\alpha = (1-\alpha)u_i + \alpha\bar{u}$ (Collard, 1978; Levine, 1998). Here *alignment* is the scalar incentive weight α, not a property of a trained language model. Increasing it changes best responses without changing the communication process.

Our main result is a complete characterization within a specified equilibrium class. On a full-quorum relay line, every coordinating observation-monotone pure equilibrium with an informed root veto consists of a responsive prefix and a blind suffix. Three inequalities determine whether a cutoff is an equilibrium. They also yield exact validity intervals and monotone comparative statics for equilibrium existence and the welfare envelope. A solver census shows that replacing heterogeneous links by a uniform channel with the same end-to-end reliability recommends a non-equilibrium in 11.7% of coordinating rows.

Using this structure requires two further choices: an equilibrium must be selected, and its local rules must be executed by an agent that reads its environment through language. Both choices turn out to matter as much as certification itself. Our contributions are the following.

1. **An exact characterization of relay coordination.** For full-quorum relay lines with heterogeneous links and an informed root veto, we prove that every coordinating observation-monotone pure equilibrium is a responsive-prefix/blind-suffix cutoff policy and that three posterior-odds inequalities decide each cutoff exactly (Theorem 1). The result gives exact validity intervals in the alignment weight, shows that the assignment depends on the order of the links and not only on their reliabilities, and yields monotone comparative statics for equilibrium existence and the certified welfare envelope (Theorem 2), while a canonical selection from the family need not be monotone.
2. **A compiler that separates certification, selection, and execution.** The characterization compiles into an exact procedure that enumerates the certified family, selects a deployment policy by welfare with all-WAIT as an outside option, and emits two objects per agent: an *audit certificate* carrying the witness, validity interval, assumptions, and fallback, and an *execution card* that projects it onto the agent's local rule with guard precedence (§4). A solver census quantifies the cost of ignoring channel order and shows that the same machinery can choose the channel ordering.
3. **An empirical account of the language-model interface.** In a sandbox that implements the game, we measure realized joint outcomes rather than local fidelity alone and show that equilibrium selection, not only execution, determines whether coordination pays (§5.2). In a preregistered format ablation on fresh games, direct execution of the full audit certificate loses welfare, whereas the same content compiled into a decision-table card is executed almost exactly, matching a deterministic validator (§5.3); reconstruction of the equilibrium from the game description remains unreliable unless the extracted game is handed to the exact solver (§5.4). The remaining failures are localized to fallback compliance when assumptions are violated.

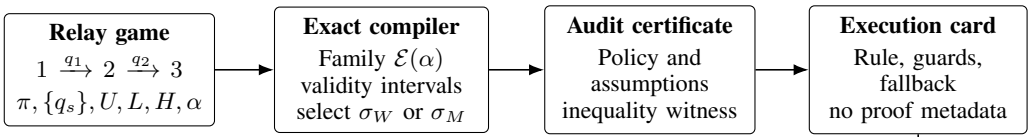
2 Model and objective

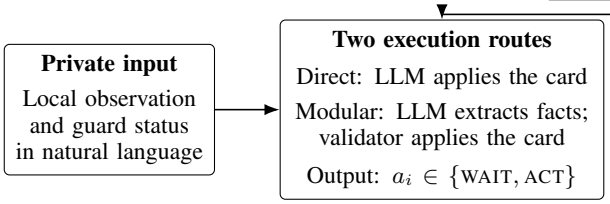
A finite Bayesian game has agents $N = \{1, \dots, n\}$, worlds Ω with prior P, information partitions O_i, and actions $a_i \in \{\text{WAIT}, \text{ACT}\}$. An information cell is an agent's *type*. Throughout the main text, $n \geq 2$ and success requires a full quorum. The base payoff is zero for WAIT, $U > 0$ for ACT in a safe full quorum, $-H < 0$ in an unsafe full quorum, and $-L < 0$ if the quorum fails. Utilities are

$$u_i^\alpha(w, a) = (1-\alpha)u_i(w, a) + \frac{\alpha}{n}\sum_{j=1}^{n} u_j(w, a), \qquad \alpha \in [0, 1]. \tag{1}$$

Thus $\alpha = 0$ gives private payoffs and $\alpha = 1$ gives a shared objective. The evaluation criterion is base welfare $W(\sigma) = \mathbb{E}_{w\sim P}[\sum_i u_i(w, \sigma(w))]$. Since $\sum_i u_i^\alpha = \sum_i u_i$, changing α does not change a fixed policy's welfare.

Relay protocol. The root, agent 1, observes the state, which is SAFE with probability π. It sends a message only in SAFE. Hop s delivers with probability q_s, independently of other hops; the first failure stops forwarding. Define $Q_0 = 1$ and $Q_k = \prod_{h=1}^{k} q_h$. World SAFE_t, in which the message





Blind-agent card: example

Veto	Parameters	Action
not YES	any	WAIT
YES	not YES	WAIT
YES	YES	ACT

UNKNOWN counts as not YES.

Figure 1: From a relay game to local execution. The compiler first identifies the equilibrium family and selects a policy. Its audit certificate records the mathematical witness; the execution card retains only the local rule and guards. At execution time, the agent receives its own observation as text, not the global state. The table illustrates a blind agent: either failed or unconfirmed guard requires WAIT, while confirmed guards permit ACT regardless of message receipt. A language model may apply the card directly, or a validator may apply it to extracted facts.

reaches agents $1, \ldots, t$, has probability $\pi Q_{t-1}(1 - q_t)$ for $t < n$, and πQ_{n-1} for $t = n$. The remaining mass $1 - \pi$ is UNSAFE. Forwarding is automatic, not an agent decision. Non-root agents observe only receipt or non-receipt. We assume $0 < \pi < 1$ and $0 < q_s < 1$; Appendix B treats boundary probabilities.

Equilibrium and coordination. We use pure Bayesian Nash equilibrium (BNE), imposing weak best-response inequalities at each positive-probability type. A profile *coordinates* if all agents act on some positive-probability safe history. All-WAIT is a strict BNE at every α: a unilateral ACT fails the quorum and costs $(1 - \beta\alpha)L > 0$, where $\beta = (n - 1)/n$. Coordination is therefore an additional selection objective, not a consequence of equilibrium alone. Our characterization concerns observation-monotone equilibria with a *root veto*: the root waits in UNSAFE.

3 Structure of Relay Coordination

Lemma 1 (Marginal payoffs)**.** *If an agent's action completes a safe full quorum, the utility gain of* ACT *over* WAIT *is* $G_n(\alpha) = U + \beta\alpha L$*. If the quorum fails either way, the gain is* $-C_n(\alpha) = -(1 - \beta\alpha)L$*. For a type facing only these two events,* ACT *is a best response exactly when its posterior success odds are at least*

$$R_n(\alpha) = \frac{C_n(\alpha)}{G_n(\alpha)} = \frac{(1 - \beta\alpha)L}{U + \beta\alpha L}. \tag{2}$$

The positive $\beta\alpha L$ term is the share of the other agents' failed-quorum losses that participation avoids. Alignment lowers the threshold from L/U to $L/(nU + (n - 1)L)$, but the latter remains positive. A shared objective does not remove the information constraint. The same local mechanism drives the two-agent electronic mail game under iterated dominance (Rubinstein, 1989; Dimitri, 2004); we do not develop that case here.

Cutoff policies. For $r \in \{1, \ldots, n\}$, let σ^r prescribe: the root acts iff SAFE; agents $2, \ldots, r$ act iff their message arrives; agents $r + 1, \ldots, n$ act regardless. We call the first group *responsive* and the second *blind*. A non-root pure strategy is observation-monotone if its action on receipt is at least its action on non-receipt, ordering WAIT $<$ ACT. This includes always-WAIT, responsive, and blind policies, but excludes the anti-responsive policy.

Define

$$A_r = \frac{Q_{r-1}}{1 - Q_{r-1}} \quad (A_1 = +\infty), \qquad \eta_j = \frac{\pi Q_{j-2}(1 - q_{j-1})}{1 - \pi Q_{j-2}}, \quad j \geq 2. \tag{3}$$

A_r is the root's success odds under σ^r. For an agent receiving nothing, η_j compares failure at its own incoming edge with an unsafe state or a failure earlier in the line.

Theorem 1 (Exact cutoff characterization)**.** *Assume full support. Then σ^r is a pure BNE if and only if*

$$A_r \geq R_n(\alpha), \qquad \textit{root participates in } \textsc{safe}, \tag{4}$$

$$\eta_r \leq R_n(\alpha) \qquad \textit{if } r \geq 2, \quad \textit{last responsive agent waits on non-receipt}, \tag{5}$$

$$\eta_{r+1} \geq R_n(\alpha) \qquad \textit{if } r < n, \quad \textit{first blind agent acts on non-receipt}, \tag{6}$$

and, when $r = 1$, $H \geq \beta\alpha L$. Moreover, every coordinating observation-monotone root-veto pure BNE is a cutoff policy.

Why only three types matter. The root acts in SAFE when the responsive prefix succeeds with probability Q_{r-1}. A responsive agent i that receives a message knows that earlier hops succeeded, so its success probability is $Q_{r-1}/Q_{i-1} \geq Q_{r-1}$. The root's inequality therefore covers every responsive receipt type. On non-receipt, an earlier responsive agent knows that another responsive agent downstream will wait. Only the last responsive agent can be pivotal: its action succeeds exactly when its own incoming edge failed after all earlier hops succeeded, giving odds η_r.

A blind agent $j > r$ receiving nothing is pivotal when the responsive prefix succeeded but delivery stopped before j. Its success odds are

$$B_{r,j} = \frac{\pi(Q_{r-1} - Q_{j-1})}{1 - \pi Q_{r-1}}. \tag{7}$$

These increase with j and equal η_{r+1} at the first blind agent. Receipt instead guarantees a safe quorum if that agent acts. Finally, the unsafe root can complete a quorum only when $r = 1$, giving the separate veto condition.

For completeness, coordination rules out always-WAIT agents. If a blind agent precedes a responsive one, non-receipt at the former implies non-receipt at the latter. The latter waits, making the former's blind action strictly costly. Blind agents must therefore form a suffix. Appendix A gives the full type-by-type proof. Anti-responsive coordinating equilibria lie outside this argument and are handled by the general solver.

Corollary 1 (Homogeneous links)**.** *If $q_s = q$ for every hop, no interior cutoff $1 < r < n$ is an equilibrium. Under $H \geq \beta\alpha L$, a coordinating cutoff exists iff*

$$\max\left\{\frac{q^{n-1}}{1-q^{n-1}}, \frac{\pi(1-q)}{1-\pi}\right\} \geq R_n(\alpha). \tag{8}$$

Here η_j strictly decreases with j, so the last-responsive and first-blind inequalities cannot both hold at an interior cutoff. Heterogeneity removes this obstruction. In the introductory example, $R_3 = 7/23$, $A_3 = 2/23$, and $\eta_2 = 2/15$: neither uniform policy coordinates in equilibrium. Yet $A_2 = 4$ and $\eta_2 \leq R_3 \leq \eta_3 = 36/85$ certify the hybrid σ^2. It is valid on $[147/242, 1]$. After swapping the links, only σ^1 is a certified cutoff, valid on $[3/8, 1]$. Exhaustive enumeration also finds an anti-responsive equilibrium in the swapped game, so uniqueness there is only within the cutoff class.

3.1 Equilibrium families, welfare, and selection

Let $\mathcal{E}(\alpha) = \{r : \sigma^r \text{ is a BNE at } \alpha\}$. For a fixed cutoff,

$$W_r = n(U+L)\pi Q_{r-1} - L\left[\pi \sum_{k=0}^{r-1} Q_k + (n-r)\right]. \tag{9}$$

The first term rewards safe full quorums; the second charges the expected number of acting agents. For $D = nU + (n-1)L$,

$$W_{j-1} - W_j = (1 - \pi Q_{j-2})(D\eta_j - L). \tag{10}$$

A shallower cutoff increases welfare exactly when $\eta_j > R_n(1)$. This identity yields a comparative-statics result for the equilibrium family, not for every choice from it.

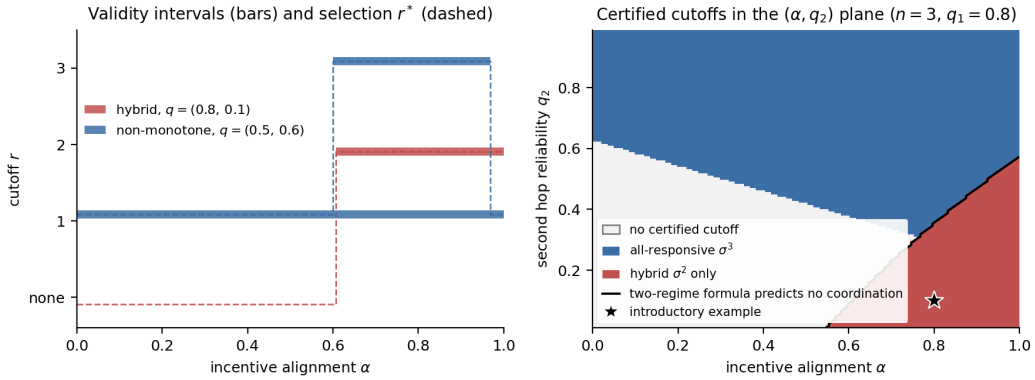


Figure 2: Cutoff validity and channel heterogeneity. Left: validity intervals and the maximal-cutoff selection on two three-agent games. The selection can move $1 \to 3 \to 1$ even though existence and the welfare envelope are monotone. Right: varying (α, q_2) with $q_1 = 0.8$ produces a hybrid-only region missed by restricting attention to the two uniform policies. The star marks the introductory example.

Theorem 2 (Uniform-veto comparative statics)**.** *Under full support and $H \geq \beta L$, as α increases: (i) once $\mathcal{E}(\alpha)$ is nonempty, it remains nonempty; (ii) $\min \mathcal{E}(\alpha)$ weakly decreases where defined; and (iii) $\max_{r \in \mathcal{E}(\alpha)} W_r$ weakly increases where defined.*

For $r \in \mathcal{E}(\alpha)$ and $\alpha' \geq \alpha$, set $\eta_1 = 0$ and $k = \max\{j \leq r : \eta_j \leq R_n(\alpha')\}$. The cutoff k remains certified, and every intervening step crosses $\eta_j > R_n(1)$, so $W_k \geq W_r$. The full proof is in Appendix A. Without the uniform veto, an all-blind equilibrium can disappear when the root becomes willing to act in UNSAFE.

We distinguish two selections. The *canonical* selection is $\sigma_M = \sigma^{r^*}$ with $r^* = \max \mathcal{E}(\alpha)$, using all-WAIT when the set is empty. It minimizes the number of blind agents within the family, but need not maximize welfare or provide additional safety. For deployment we instead use

$$\sigma_W \in \arg \max_{\sigma \in \{\text{all-WAIT}\} \cup \{\sigma^r : r \in \mathcal{E}(\alpha)\}} W(\sigma). \tag{11}$$

Its welfare is nonnegative by construction and inherits the monotonic welfare envelope of Theorem 2, with a zero floor. This is optimality within the certified cutoff family plus WAIT, not among all equilibria or all decentralized policies. In contrast, at $\pi = 7/10$, $(q_1, q_2) = (1/2, 3/5)$, and $U = L = 1, H = 10$, r^* follows $1 \to 3 \to 1$ as α grows (Figure 2). Certification alone does not specify which equilibrium to deploy.

4 EXACT COMPILATION AND LOCAL EXECUTION

Computing validity intervals. Every cutoff condition compares a fixed information-structure quantity with the decreasing threshold $R_n(\alpha)$. For finite $c \geq 0$, the crossing $R_n(\alpha) = c$ occurs at

$$\alpha(c) = \frac{L - cU}{\beta L(1 + c)}. \tag{12}$$

Retaining only crossings in $[0, 1]$ gives at most $2n-1$ interior candidate boundaries: $n-1$ root odds, $n-1$ edge-failure odds, and one veto boundary. Each cutoff's validity set is a closed interval, a singleton, or empty. Prefix products and interval endpoints require $O(n)$ rational operations; sorting costs $O(n \log n)$. Explicitly listing the whole family in every cell can require quadratic output size.

For arbitrary finite games with utilities affine in α, a general solver partitions the axis at the roots of all primitive interim comparisons and checks every open cell and boundary separately. The pure-BNE and pure-action-dominance survivor correspondences are constant between these roots (Proposition 1, Appendix C). This finite partition does not itself imply monotonicity or efficient enumeration. The relay characterization supplies the additional structure. On the reported implementation,

general enumeration takes 111 seconds at $n = 6$, compared with 7 milliseconds for the structured compiler; this is a small-scale implementation comparison (Appendix E).

Direct interim-payoff enumeration agrees with the cutoff conditions on all $71{,}640$ cases in the reported verification grid. Exact equality tests check boundary inclusion separately; the general solver also searches for equilibria outside the cutoff class (Appendix E).

Audit certificate. For a selected policy, the certificate records its local rules, solution concept, exact validity interval, assumptions, and a witness of the cutoff inequalities. The validity interval is the set where that policy is a BNE, not the region where a selection rule chooses it. For the introductory hybrid at $\alpha = 4/5$, the witness is $A_2 = 4$, $\eta_2 = 2/15$, and $\eta_3 = 36/85$, compared with $R_3 = 7/23$. Its validity interval is $[147/242, 1]$ and its alignment margin is $233/1210$. Agent 3 receives a blind rule, subject to the veto and parameter guards. The margin measures distance to a validity-interval endpoint, not payoff slack or an error probability. Boundary inclusion matters: at $\pi = 3/4$, $(q_1, q_2) = (4/5, 3/5)$, $U = L = 1$, and $H = 10$, the hybrid is a BNE only at $\alpha = 3/8$ (Appendix D).

Uncertain parameters. Under interval-valued parameters, the compiler must certify *one fixed policy* throughout the parameter box. It is not enough that every parameter point admits some, possibly different, equilibrium. Each cutoff inequality is checked at its adverse coordinatewise extreme: A_r increases with prefix reliability, while η_j increases with πQ_{j-2} and decreases with q_{j-1}. Failure to certify a common policy leads to deferral, not a claim that coordination is impossible everywhere in the box (Appendix D).

Execution card. The audit object is not the instruction we need an agent to interpret at every decision. We compile a local card that contains only the selected rule, the required guards, and their precedence. Let v indicate that the root veto is in force, p that the parameters are confirmed, and o_i the relevant local observation. Both guards take values YES, NO, or UNKNOWN. The card implements

$$K_i(v, p, o_i) = \begin{cases} \text{WAIT}, & \sigma_W = \text{all-WAIT or } v \neq \text{YES or } p \neq \text{YES}, \\ (\sigma_W)_i(o_i), & \text{otherwise.} \end{cases} \tag{13}$$

Unknown observations required by a responsive or root rule also produce WAIT; a blind rule does not require message receipt. A decision table presents these cases without asking the executor to re-derive the equilibrium. The table and audit policy have the same local semantics on confirmed inputs. This equivalence concerns the specified rules, not an LLM's compliance. We test direct card execution separately from a deterministic validator applied to model-extracted facts.

5 Experiments

We evaluate three distinct questions: whether channel structure changes the recommended policy, whether a selected policy has useful welfare, and whether a language interface preserves its local rules. The LLM studies use Qwen3-30B-A3B-Instruct-2507 and Mistral-Small-24B-Instruct-2501. The environment implements the one-shot relay protocol; agents do not choose messages or communicate during action selection. Setups and full tables are in Appendices F–I.

Joint outcomes and metrics. For an estimated behavioral profile $\hat{\sigma}$, let $p_i(w)$ be agent i's ACT probability at its local type in world w. Set $x(w) = U$ in SAFE and $x(w) = -H$ in UNSAFE. Independent action draws conditional on types give

$$W(\hat{\sigma}) = \sum_{w \in \Omega} P(w) \left[n\{x(w) + L\} \prod_i p_i(w) - L \sum_i p_i(w) \right]. \tag{14}$$

This charges each actor $-L$ and refunds $x(w) + L$ to all agents on a full quorum. Quorum probabilities similarly sum $P(w) \prod_i p_i(w)$ over safe or unsafe worlds. These expectations are exact conditional on the estimated action frequencies, which remain subject to finite-sample error. We report means of $W/(nU)$ across games; all-WAIT has value zero. Fidelity measures valid-arm agreement

with the selected action. Invalid/uncertain-arm compliance measures adherence to the WAIT fallback. These stress-arm rates are separate from valid-arm welfare, not components of a combined safety score.

5.1 CHANNEL REDUCTION AND ORDERING

For each $n \in \{3, 4, 5\}$, we draw 150 random reliability multisets and enumerate all distinct orderings at 21 alignment values. Among 5,749 multiset–alignment cells with coordination in some ordering, the certified family changes across orderings in 58.6%. Replacing the links by their geometric mean preserves end-to-end reliability, but its selected policy is not a BNE in 11.7% of 43,724 coordinating ordering–alignment rows. Conditional on a non-BNE recommendation, mean worst-type regret is $0.41U$ (maximum $4.1U$). A boundary-enriched stratum gives 62.7% family changes and 13.7% non-BNE recommendations.

The same solver can choose an ordering to maximize the best certified welfare, including WAIT. This objective varies across orderings in 36% of coordinating cells. Within those cells, the best ordering gains $0.11\,nU$ over the worst and $0.03\,nU$ over the median. A strong simple heuristic places the least reliable link first: it is optimal in 81% of these cells and loses only $0.009\,nU$ on average to the optimum. Appendix E reports denominators and the boundary-enriched results.

5.2 SELECTION CHANGES THE DEPLOYMENT OBJECTIVE

The initial sandbox study uses 24 games: 12 random games with a certified cutoff and 12 near a hybrid validity boundary. Ten language families render local observations and guard status. A post-hoc extension compares σ_M with σ_W on exactly these games, without removing unfavorable instances. The canonical selection has mean welfare $-0.021\,nU$. Welfare selection raises it to $0.207\,nU$ by choosing WAIT in seven games and another cutoff in two (Table 1). The remaining fifteen policies are unchanged.

For a target $\sigma_T \in \{\sigma_M, \sigma_W\}$ and an executed behavioral profile $\hat{\sigma}$, the welfare difference separates into

$$W(\sigma_W) - W(\hat{\sigma}) = \underbrace{W(\sigma_W) - W(\sigma_T)}_{\text{selection difference}} + \underbrace{W(\sigma_T) - W(\hat{\sigma})}_{\text{execution difference}}. \tag{15}$$

A faithful executor cannot fix an unprofitable target. Conversely, choosing σ_W removes the selection difference but does not ensure faithful execution. The execution difference need not be nonnegative for every behavioral profile, since σ_W optimizes only over the certified family plus WAIT.

Table 1: Selection and execution on the same 24 sandbox games. Entries are mean normalized welfare; all-WAIT earns 0. The welfare-selection extension is post hoc. "Certificate + validator" applies a deterministic rule to facts from the same response used by the direct certificate condition, not to gold facts. Full conditions and strata are in Appendix G.

	Qwen3-30B		Mistral-24B	
Execution	Canonical σ_M	Welfare σ_W	Canonical σ_M	Welfare σ_W
Exact selected policy	-0.021	0.207	-0.021	0.207
Local rule, direct	-0.186	0.086	-0.066	0.082
Full certificate, direct	-0.410	-0.223	-0.382	-0.117
Certificate + validator	-0.057	0.177	-0.098	0.170
Extraction + validator	-0.272	0.064	-0.037	0.200

Language-mediated execution preserves much of this gain when the final action is validated: the certificate-response validator achieves 0.177 on Qwen and 0.170 on Mistral, while extraction-only plus validation reaches 0.200 on Mistral but only 0.064 on Qwen. Direct full-certificate execution remains below WAIT on both models. The positive result therefore depends on both policy selection and execution, not simply on finding a coordinating BNE. A separate solver census over 400 random instances per n finds negative canonical welfare in 22.8%, 34.4%, and 38.4% of coordinating cells for $n = 3, 4, 5$ (Appendix G). These rates describe the sampled generator, not an arbitrary population of games.

Most of the gain on the sandbox comes from declining unprofitable coordination. In the two multiple-cutoff games, however, choosing cutoff 1 rather than cutoff 4 or 3 raises exact welfare from 0.012 to 0.200 and from 0.127 to 0.538, respectively. Welfare selection can thus require more blind agents, not fewer; it is not a robustness criterion.

5.3 Compiling the audit object into an execution card

A preregistered study, frozen after a separate pilot, evaluates 24 fresh games and six language families, with $15{,}840$ prompts per model. Nine games select all-WAIT, so the evaluation includes both participation and stand-down decisions. All formats carry the same selected σ_W, guards, and fallback. We compare the full audit object, a compact prose card, an action-first ordering of the full content, and a decision-table card. A two-stage condition first extracts facts and then asks an LLM to apply the table. A deterministic condition applies the table to extracted facts. The latter is not an oracle because its inputs can be wrong.

Table 2: Format ablation on 24 fresh games. W: mean valid-arm welfare in units of nU; Fid.: valid-arm action agreement; V/U: WAIT compliance when the veto is violated / parameters are uncertain. Higher is better in each column, but the metrics concern different arms. Exact selected welfare is 0.164; all-WAIT earns 0. All entries are point estimates. Full false-refusal and conditional-error results appear in Appendix H.

	Qwen3-30B				Mistral-24B			
Representation / pipeline	W	Fid.	V	U	W	Fid.	V	U
Full audit certificate	-0.206	.811	.843	.721	-0.072	.941	.689	.730
Compact prose card	.014	.964	.921	.980	.066	.965	.692	.739
Action-first audit	.161	.999	.888	.810	.003	.961	.974	.988
Decision-table card	**.164**	**1.000**	.924	.992	**.147**	**.997**	.997	.968
Extract + LLM table	-0.007	.957	.951	.860	.140	.996	1.000	1.000
Extract + validator	-0.007	.957	.953	1.000	.144	.997	1.000	1.000

The decision table gives the best direct-execution welfare on both models (Table 2). Qwen matches the exact target at the reported precision, with fidelity 1.000 and welfare 0.164. Mistral reaches fidelity 0.997 and welfare 0.147, a reported loss of $0.017\,nU$. The full audit object instead yields -0.206 and -0.072. Error conditional on correctly extracted facts falls from $0.262/0.185$ under the full certificate to $0.004/0.017$ under the table (Qwen/Mistral). The compact prose card and content reordering are not uniformly sufficient: action-first nearly matches the target on Qwen but has welfare 0.003 on Mistral.

These gains are not a guarantee of fallback compliance. The Qwen table still acts in 7.6% of veto-violated samples, and Mistral acts in 3.2% of uncertain samples. Nor does the result show that an LLM outperforms a correct deterministic rule on identical facts. Qwen's extraction-only validator has welfare -0.007 because its fact-extraction context differs from direct card execution. The ablation supports separating audit information from local action rules; it does not establish general grounding robustness or end-to-end safety.

5.4 Reconstructing equilibria

A paired study on 48 held-out games asks for any BNE cutoff, the maximal cutoff, the welfare-maximizing coordinating cutoff, or a full equilibrium profile. Unlike the later deployment objective, the welfare-cutoff task does not include WAIT as an outside option. The six strata cover unique and multiple cutoffs, no in-class coordination, boundary equality, singleton validity, and non-monotone selection. Conditions share a $7{,}000$-token budget, eight samples, and an output schema.

From the game description alone, maximal-cutoff accuracy is 0.177 on Qwen and 0.299 on Mistral (Table 3). Supplying the theorem raises Qwen to 0.688, but Mistral's response depends on the task: full-profile BNE validity improves from 0.373 to 0.549, while maximal-cutoff accuracy does not improve. Generated code executes in only 69% and 24% of snippets, respectively. Handing the extracted game to the exact solver achieves 0.992–1.000 across the reported tasks. The successful

Table 3: Paired reconstruction on 48 held-out games. Any: a valid BNE cutoff or a correct NONE answer; Max: the exact maximal-cutoff answer. All conditions use the same games and budget. Generated code is written by the model; the exact-tool condition instead delegates an extracted game to the compiler. Full tasks and strata appear in Appendix I.

	Qwen3-30B		Mistral-24B	
Condition	Any	Max	Any	Max
Game only	0.271	0.177	0.445	0.299
Theorem supplied	0.635	0.688	0.445	0.263
Generated code	0.591	0.469	0.159	0.055
Exact tool	1.000	1.000	0.995	0.992

interface does not require the executor to reconstruct an equilibrium: the compiler computes it before local execution.

6 Related work

Communication and incentives. The electronic mail game and common-knowledge analysis establish limits of coordination under unreliable communication (Rubinstein, 1989; Halpern & Moses, 1990). Dimitri's heterogeneous-failure analysis studies payoff-dependent cutoff coordination in email games (Dimitri, 2004). Awaya and Krishna analyze diffusion among payoff-bearing adopters on lines and trees, including conditional and unconditional adoption (Awaya & Krishna, 2024). We add a participating veto source and edge-specific reliability, and characterize the coordinating observation-monotone root-veto class along an incentive axis. Linear altruism and welfare blending are established models (Levine, 1998; Apt & Schäfer, 2014; Chen et al., 2014); our results concern how their equilibrium correspondence interacts with this information structure.

Decentralized policies and formal interfaces. Multi-agent RL methods such as QMIX and MAPPO learn decentralized policies from centralized training information (Rashid et al., 2018; Yu et al., 2022). Our setting instead assumes a known one-shot game and computes its policies exactly. Policy certificates in RL bound policy quality (Dann et al., 2019), and shielding enforces formal constraints during execution (Alshiekh et al., 2018). Our certificates establish equilibrium validity under stated game assumptions, a different guarantee. ReAct and PAL combine language processing with actions or program execution (Yao et al., 2023; Gao et al., 2023). We study a narrower interface question: how the same formally computed local rule behaves when presented as an audit object, prose instruction, or decision table. Exact solution of a finite game is standard (Avis et al., 2010). The relay characterization reduces cutoff certification to three interim tests.

7 Discussion and limitations

Heterogeneous relay coordination admits an exact policy structure, but equilibrium certification, selection, and execution solve different problems. The compiler identifies admissible cutoffs; welfare selection can reject an unprofitable coordinating equilibrium; a local decision table can preserve much of the selected benefit through language. The full audit certificate remains useful for verification, even when showing it directly to an executor degrades behavior.

The theorem is restricted to full-quorum lines, observation-monotone pure strategies, and a root veto. Partial quorums need not have cutoff structure (Appendix J). Welfare optimality is relative to the certified family under known parameters. The language studies use two models, small one-shot games, and constructed narratives; forwarding and model parameters are not learned. High valid-arm fidelity does not imply reliable assumption checking, and low catastrophe rates in the original sandbox do not establish safety outside its protocol. Decision-table execution reduces the observed representation error without removing these limitations.

AI USE STATEMENT

Generative AI tools were used for literature discovery, text editing, and code assistance. The authors are responsible for all claims, proofs, implementations, and reported results, and for the final content of the submission.

REPRODUCIBILITY STATEMENT

Appendices A–D give proofs, boundary conventions, and certificate semantics. Appendices E–I describe the sampled games, study comparisons, scoring, and additional results. The welfare-selection extension is identified as post hoc; the subsequent fresh-game card ablation has a separate preregistration. The experimental artifact records prompts, model settings, study-level registration and deviation records, and per-query outputs. Exact-payoff checks and table-generation scripts accompany the computational results.

REFERENCES

Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, pp. 2669–2678, 2018. doi: 10.1609/aaai.v32i1.11797.

Krzysztof R. Apt and Guido Schäfer. Selfishness level of strategic games. *Journal of Artificial Intelligence Research*, 49:207–240, 2014. doi: 10.1613/jair.4164.

David Avis, Gabriel D. Rosenberg, Rahul Savani, and Bernhard von Stengel. Enumeration of Nash equilibria for two-player games. *Economic Theory*, 42(1):9–37, 2010.

Yu Awaya and Vijay Krishna. Spreading information via social networks: An irrelevance result, 2024. arXiv preprint arXiv:2402.05276.

Po-An Chen, Bart de Keijzer, David Kempe, and Guido Schäfer. Altruism and its impact on the price of anarchy. *ACM Transactions on Economics and Computation*, 2(4):17:1–17:45, 2014. doi: 10.1145/2597893.

David Collard. *Altruism and Economy: A Study in Non-Selfish Economics*. Martin Robertson, Oxford, 1978.

Christoph Dann, Lihong Li, Wei Wei, and Emma Brunskill. Policy certificates: Towards accountable reinforcement learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 1507–1516. PMLR, 2019.

Nicola Dimitri. Efficiency and equilibrium in the electronic mail game; the general case. *Theoretical Computer Science*, 314(3):335–349, 2004. doi: 10.1016/j.tcs.2003.12.016.

Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Jamie Callan, and Graham Neubig. PAL: Program-aided language models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 10764–10799. PMLR, 2023. URL `https://proceedings.mlr.press/v202/gao23f.html`.

Joseph Y. Halpern and Yoram Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990.

David K. Levine. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3):593–622, 1998.

Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 4295–4304. PMLR, 2018. URL `https://proceedings.mlr.press/v80/rashid18a.html`.

Ariel Rubinstein. The electronic mail game: Strategic behavior under "almost common knowledge". *American Economic Review*, 79(3):385–391, 1989.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2210.03629.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of PPO in cooperative multi-agent games. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24611–24624, 2022. doi: 10.52202/068431-1787.

A PROOFS OF THE RELAY RESULTS

All relay proofs use $n \geq 2, U, L, H > 0$, independent hop delivery, and a fixed forwarding protocol. Full support is assumed except in Appendix B. Strategies are pure maps from private observations to actions. Inequalities are weak for BNE, so a tied type can follow either best response.

A.1 STRATEGIC EQUIVALENCE AND MARGINAL PAYOFFS

Writing $\beta = (n-1)/n$, equation equation 1 becomes

$$u_i^\alpha = (1-\beta\alpha)u_i + \frac{\alpha}{n}\sum_{j\neq i} u_j = (1-\beta\alpha)\left(u_i + \lambda \sum_{j\neq i} u_j\right), \qquad \lambda = \frac{\alpha}{n-(n-1)\alpha}.$$

The factor $1-\beta\alpha \geq 1/n$ is positive. The two forms therefore have the same best-response comparisons, pure equilibria, and pure-action dominance relations.

To prove Lemma 1, consider a pivotal safe action. The agent's base payoff changes from 0 to U; each of the other $n-1$ agents changes from $-L$ to U. Hence the blended gain is

$$(1-\alpha)U + \frac{\alpha}{n}\{U + (n-1)(U+L)\} = U + \beta\alpha L.$$

If the quorum fails either way, only the agent's base payoff changes, from 0 to $-L$. The blended gain is $-(1-\alpha+\alpha/n)L = -(1-\beta\alpha)L$. Weighting these two gains by conditional event probabilities proves equation equation 2. The unsafe pivotal gain at a root that would complete an all-blind quorum is $-H + \beta\alpha L$.

A.2 THEOREM 1: NECESSITY AND SUFFICIENCY

Fix a cutoff r. The following cases exhaust the positive-probability types.

Root in SAFE. All other agents act iff every responsive agent receives a message, which has probability Q_{r-1}. Acting is optimal iff $Q_{r-1}G_n \geq (1-Q_{r-1})C_n$, or $A_r \geq R_n(\alpha)$.

Responsive agent $i \leq r$ after receipt. The state is safe and all earlier hops succeeded. The conditional chance that the remaining responsive agents receive is $Q_{r-1}/Q_{i-1} \geq Q_{r-1}$. The root inequality therefore implies that acting is a best response.

Responsive agent $i < r$ after non-receipt. Agent $i+1$ also receives nothing and waits. A full quorum is impossible regardless of i's action. Acting loses $C_n > 0$, so waiting is strictly optimal.

Last responsive agent $r \geq 2$ after non-receipt. Acting is pivotal only when hops through $r-2$ succeeded and hop $r-1$ failed. This event has prior mass $\pi Q_{r-2}(1-q_{r-1})$. All remaining non-receipt worlds, of mass $1-\pi Q_{r-2}$, fail the quorum either way. Thus waiting is optimal iff $\eta_r \leq R_n(\alpha)$.

Blind agent $j > r$ after receipt. The state is safe and all responsive agents received. Every other agent acts. Acting is strictly optimal.

Blind agent $j > r$ after non-receipt. Acting completes a safe quorum when the responsive prefix succeeded but delivery stopped before j. Its success odds are

$$B_{r,j} = \frac{\pi(Q_{r-1} - Q_{j-1})}{1-\pi Q_{r-1}}.$$

These odds weakly increase with j and are minimized at $j = r+1$, where $B_{r,r+1} = \eta_{r+1}$. Every blind non-receipt type acts optimally iff $\eta_{r+1} \geq R_n(\alpha)$.

Root in UNSAFE. For $r \geq 2$, agent 2 waits and the quorum is impossible; the root strictly prefers waiting. For $r = 1$, all other agents act blindly, so the root prefers waiting iff $H \geq \beta\alpha L$.

The stated inequalities are therefore necessary and sufficient.

Completeness in the observation-monotone class. A full quorum on a positive-probability safe history requires every agent to participate. The root's safe action must be ACT, and each non-root observation-monotone strategy must be responsive or blind, since always-WAIT prevents coordination. Suppose a blind agent j precedes a responsive agent $k > j$. At j's non-receipt type, the relay has stopped before j, so k also receives nothing and waits. The blind action then loses $C_n > 0$ in every possible world. This contradicts equilibrium. All blind agents form a suffix, proving completeness. An anti-responsive downstream agent need not wait on non-receipt, which is why this argument does not extend to all pure equilibria. □

A.3 COROLLARY 1: HOMOGENEOUS LINKS

If $q_s = q$, then $\eta_j = \pi q^{j-2}(1-q)/(1-\pi q^{j-2})$ strictly decreases with j. No interior cutoff can satisfy $\eta_r \le R_n(\alpha) \le \eta_{r+1}$. The all-blind cutoff exists iff $\eta_2 \ge R_n(\alpha)$ and the veto condition holds. The all-responsive cutoff requires both $A_n \ge R_n(\alpha)$ and $\eta_n \le R_n(\alpha)$. If $A_n \ge R_n(\alpha)$ but $\eta_n > R_n(\alpha)$, then $\eta_2 > R_n(\alpha)$ and the all-blind cutoff exists. If both odds in equation equation 8 fall below $R_n(\alpha)$, neither cutoff exists. The corollary is an existence test, not a sufficient test for the all-responsive cutoff alone. □

A.4 WELFARE AND ADJACENT CUTOFFS

Under the root veto, the probability of a full safe quorum is πQ_{r-1}. Charge every actor $-L$, then refund $U + L$ to each of the n agents on success. The expected number of actors is $\pi \sum_{k=0}^{r-1} Q_k + (n-r)$: the root and responsive agents act only after the relevant safe prefix, while the $n-r$ blind agents act in every state. This gives equation equation 9.

For $j \ge 2$, subtracting adjacent welfare values gives

$$\begin{aligned} W_{j-1} - W_j &= n(U+L)\pi Q_{j-2}(1-q_{j-1}) - L(1-\pi Q_{j-1}) \\ &= D\pi Q_{j-2}(1-q_{j-1}) - L(1-\pi Q_{j-2}) \\ &= (1-\pi Q_{j-2})(D\eta_j - L), \end{aligned}$$

where $D = nU + (n-1)L$ and $n(U+L) = D + L$. Since $R_n(1) = L/D$, the welfare comparison follows. □

A.5 THEOREM 2: COMPARATIVE STATICS

Let $t = R_n(\alpha)$ and $t' = R_n(\alpha') \le t$ for $\alpha' \ge \alpha$. Given $r \in \mathcal{E}(\alpha)$, define $\eta_1 = 0$ and $k = \max\{j \in \{1, \dots, r\} : \eta_j \le t'\}$. This set is nonempty. We check the cutoff k:

$$\begin{aligned} A_k \ge A_r \ge t \ge t', & \\ \eta_k \le t' & \qquad \text{if } k \ge 2, \\ \eta_{k+1} > t' & \qquad \text{if } k < r. \end{aligned}$$

If $k = r < n$, the blind inequality is inherited from $\eta_{r+1} \ge t \ge t'$; if $k = n$, none is required. The uniform veto gives $H \ge \beta L \ge \beta\alpha' L$ for $k = 1$. Hence $k \in \mathcal{E}(\alpha')$, proving persistence. Choosing $r = \min \mathcal{E}(\alpha)$ gives $k \le r$, proving the minimum-cutoff claim.

For the welfare claim, start from a maximizer r at α. Every $j \in \{k+1, \dots, r\}$ has $\eta_j > t' \ge R_n(1)$, so each step to a shallower cutoff increases welfare by equation equation 10. Thus $W_k \ge W_r$, and the envelope at α' is at least W_k. Adding all-WAIT extends the envelope by a zero floor. Before any coordinating equilibrium exists this value is zero; after existence it cannot decrease. □

Selection example. At $\pi = 7/10$, $(q_1, q_2) = (1/2, 3/5)$, $U = L = 1$, and $H = 10$, cutoff 1 is certified throughout $[0, 1]$, while cutoff 3 is certified exactly on $[3/5, 153/158]$. The maximal selection is therefore $1 \to 3 \to 1$. This does not contradict welfare-envelope monotonicity. At $\pi = 3/5$ and $(q_1, q_2) = (3/10, 4/5)$, the same selection can switch from cutoff 1 to 3 while base welfare falls from 1 to $-3/50$.

B BOUNDARY PROBABILITIES

The odds formulation assumes full support. For $\pi, q_s \in [0,1]$, some types are unreachable. The following conditions test a *specified cutoff*; we do not claim a complete characterization of all equilibrium strategies at these degenerate probabilities.

Write $G = U + \beta\alpha L$ and $C = (1-\beta\alpha)L$, and define

$$\begin{aligned} S_r &= Q_{r-1}G - (1-Q_{r-1})C, \\ N_r &= \pi Q_{r-2}(1-q_{r-1})G - (1-\pi Q_{r-2})C, && r \geq 2, \\ B_{r+1} &= \pi Q_{r-1}(1-q_r)G - (1-\pi Q_{r-1})C, && r < n. \end{aligned} \tag{16}$$

Then σ^r is a BNE iff

$$\begin{gathered} S_r \geq 0 \quad \text{when } \pi > 0, \qquad N_r \leq 0 \quad (r \geq 2), \qquad B_{r+1} \geq 0 \quad (r < n), \\ H \geq \beta\alpha L \quad \text{when } r = 1 \text{ and } \pi < 1. \end{gathered} \tag{17}$$

To certify *coordinating* behavior rather than a cutoff BNE alone, additionally require $\pi Q_{r-1} > 0$, so that a safe full quorum has positive probability. Importantly, N_r and B_{r+1} need no additional reachability mask. Their relevant non-receipt type probabilities are $1-\pi Q_{r-1}$ and $1-\pi Q_r$, respectively. If one of these type probabilities is zero, its entire unnormalized expression is zero. Otherwise its sign is exactly the conditional best-response comparison multiplied by the positive type probability. Factors such as πQ_{r-2} are *not* non-receipt type probabilities.

The other types follow the proof of Theorem 1: impossible receipt types impose no condition; positive-probability receipt types inherit the root comparison; earlier non-receipt responsive types face a failed quorum. For deeper blind agents the unnormalized advantage is $\pi(Q_{r-1} - Q_{j-1})G - (1-\pi Q_{r-1})C$, which weakly increases with j.

As a boundary check, at $\pi = 0$ and $r < n$, $B_{r+1} = -C < 0$: blind agents cannot profitably act when the root always waits. At $\pi = 1$, the root's unsafe comparison is correctly omitted. These are different forms of unreachable-type handling.

C FINITE AFFINE PARTITION FOR GENERAL GAMES

Proposition 1 (Finite affine partition)**.** *Consider a finite Bayesian game with fixed prior and information partitions and utilities $u_i^\alpha(w,a) = c_i(w,a) + \alpha d_i(w,a)$. Let T_i and A_i be its type and action sets. Form all interim differences between ordered own-action pairs at a positive-probability type against each pure opponent type-contingent profile. Let $C_0 \subseteq [0,1]$ contain the roots of the nonconstant differences. The pure-BNE correspondence is constant on every connected component of $[0,1] \setminus C_0$. The same holds for the survivor correspondence under simultaneous iterated deletion of actions strictly dominated by another* pure *action at a type. Moreover,*

$$|C_0| \leq \sum_i |T_i|\,|A_i|(|A_i|-1)\prod_{j\neq i}|A_j|^{|T_j|}.$$

Proof. Each primitive difference is affine and, if nonconstant, changes sign at most once. Between roots, every primitive sign is constant, including identically zero differences. A pure-BNE condition is a finite conjunction of weak sign constraints. First-round pure-action dominance is also determined by these signs. Inductively, if the surviving action sets are fixed through a round, the next round uses comparisons from the same finite set, so its survivors are fixed. Finite deletion terminates, proving the result. The bound counts all primitive comparisons; duplicate roots only reduce the count. □

At roots, equality can create separate behavior because equilibrium is weak and dominance is strict. The compiler must solve root points as well as open cells. The proposition gives neither a polynomial-time general algorithm nor constancy of mixed-equilibrium probabilities. It is supporting machinery for exact enumeration, rather than the structural contribution of the relay result.

D CERTIFICATE SPECIFICATION AND INTERVAL UNCERTAINTY

Audit fields. An audit record identifies the game and policy class, the selected cutoff or all-WAIT, local rules, the equilibrium solution concept, and the validity assumptions. It includes the exact inequalities, validity-interval endpoints with inclusion flags, a model identifier, and an oracle re-check where available. Distinct statuses indicate certified validity, no coordinating policy in the specified class, parameter uncertainty, out-of-scope models, or invalid assumptions. "No coordinating policy in class" does not mean no equilibrium exists.

Three sets must remain distinct: (i) the alignments where a named policy is a BNE, (ii) the alignments where a named selection chooses it, and (iii) cells where the whole equilibrium family is constant. For instance, at $\pi = 3/4$, $(q_1, q_2) = (4/5, 3/5)$, $U = L = 1$, and $H = 10$, the hybrid is a BNE exactly at $\alpha = 3/8$. Its validity set is a singleton, not an open cell.

For a nonempty policy-validity interval $[a, b]$ and query $\alpha \in [a, b]$, the alignment margin is $\min\{\alpha - a, b - \alpha\}$. It is zero at an endpoint and for a singleton. For the introductory hybrid at $\alpha = 4/5$, it is $233/1210$. This is a geometric summary, not a measure of payoff robustness or LLM reliability.

Interval-valued parameters. For a full-support box of possible $(\pi, q_1, \ldots, q_{n-1})$, A_r increases in each link of its prefix. Writing $x = \pi Q_{j-2}$ gives $\eta_j = x(1-q_{j-1})/(1-x)$, which increases in x and decreases in q_{j-1}. Each cutoff inequality can therefore be bounded over the box by coordinatewise extrema. A fixed cutoff is certified over the box only when every required inequality holds at its adverse extreme. This test keeps separate

$$\exists r\ \forall \theta \text{ in the box} : r \in \mathcal{E}_\theta(\alpha) \quad \text{and} \quad \forall \theta \text{ in the box } \exists r : r \in \mathcal{E}_\theta(\alpha).$$

The latter does not provide one executable policy for an unknown parameter. Failure to certify a common policy is also not proof that every point lacks a coordinating cutoff. Deferral requests better parameter estimates or uses the declared WAIT fallback. It does not constitute an adaptation algorithm.

Local semantic projection. For a blind agent, the execution card in Figure 1 ignores message receipt only after both guards are affirmed. Responsive agents add a receipt check; the root adds a safe-state check. All-WAIT has priority over every other rule. These are mutually exclusive and exhaustive cases over the three-valued guard fields, with required unknown observations resolved to WAIT. The rule is a finite lookup table, so its semantic agreement with the selected policy can be checked without querying an LLM. The figure is a semantic illustration, not a verbatim experimental prompt. Its mathematical agreement with the audit object does not imply that a model will execute it correctly.

E SOLVER VERIFICATION AND ORDERING CENSUS

E.1 VERIFICATION AND COST

The verification artifact uses rational arithmetic for priors, payoffs, and alignment values. The direct solver constructs worlds and compares interim utilities, rather than using the cutoff inequalities as its equilibrium test. A separate floating-point implementation checks away from boundaries. Relabeling agents and worlds and positively rescaling payoffs provide additional consistency checks.

The reported relay grid uses $n \in \{2, 3, 4, 5\}$, hop schedules over $\{0.1, 0.5, 0.8, 0.95\}$, priors $\{0.1, 0.5, 0.9\}$, payoff triples $(1, 1, 10)$, $(10, 20, 100)$, and $(10, 20, 5)$, five alignments, and every cutoff. The last payoff triple exercises a failing veto. All $71{,}640$ predicted-versus-enumerated cutoff checks agree. Exact equality tests perturb odds above and below a boundary; weak equilibrium includes equality.

In a separate 300-instance structural search, 46 of 229 coordinating pure BNE lie outside the cutoff family: 15 lack the root veto and 35 contain anti-responsive agents, with four in both categories. These counts are not all root-veto anti-responsive equilibria. A targeted search also finds root-veto anti-responsive examples. General-compiler cross-checks reproduce cutoff validity and contain each structured boundary among their candidate roots.

Table 4: Compiler runtimes in seconds on a laptop CPU, means over five timed runs of one relay instance per size; the table illustrates scaling rather than a distributional benchmark.

n	Primitive comparisons	General	Structured
3	112	0.11	< 0.001
4	480	1.1	< 0.001
5	1984	17	< 0.001
6	8064	111	0.007

E.2 Scalar reduction and ordering

For each $n \in \{3, 4, 5\}$, each stratum contains 150 base instances. Priors lie in $\{4/20, \ldots, 16/20\}$; reliabilities in $\{1/20, \ldots, 19/20\}$, with at least two distinct links; and integer payoffs satisfy $U \in [1, 5]$, $L \in [1, 10]$, $H \in [20, 60]$. We enumerate all distinct link orderings on $\alpha \in \{k/20 : k = 0, \ldots, 20\}$. The random stratum is unconditioned; the boundary stratum requires an interior validity boundary within 0.03 of a grid point for some ordering.

A correspondence-change rate uses multiset–alignment cells with a coordinating cutoff in at least one ordering. The scalar baseline substitutes $\bar{q} = (\prod_s q_s)^{1/(n-1)}$, computes its selected cutoff, and evaluates that policy in each true ordering. Non-equilibrium rates and welfare gaps use ordering–alignment rows that admit a coordinating cutoff. They are not rates over every possible game.

For ordering p, define $F(p) = \max(\{0\} \cup \{W_r(p) : r \in \mathcal{E}_p(\alpha)\})$. Ordering gains below compare the best with the worst ordering, conditional on F varying. The ascending heuristic is optimal in 81% of these random-stratum cells (75% boundary), with mean shortfall $0.009\, nU$ (0.015 boundary).

Table 5: Ordering census. Scalar gap is relative to the best true cutoff with WAIT allowed; it combines the channel reduction with its selection convention and is not an isolated causal cost of losing channel order.

Stratum	n	Cells	$\mathcal{E}$ changes	Scalar non-BNE	Scalar gap	Order gain	Opt. hybrid
Random	3	1929	.382	.085	.110	.083	.116
	4	1909	.578	.098	.140	.115	.276
	5	1911	.800	.126	.236	.144	.177
	All	5749	.586	.117	.205	.112	.185
Boundary	3	2014	.439	.114	.081	.094	.232
	4	1926	.685	.166	.155	.136	.315
	5	1906	.769	.132	.204	.165	.209
	All	5846	.627	.137	.184	.126	.255

The scalar policy's mean gap to the true *canonical* selection is instead $-0.04\, nU$ in both strata. An invalid recommendation and a low-welfare selection are distinct failures; a higher welfare value does not make a non-BNE recommendation an equilibrium. Hybrid-only families occur in 11–14% of coordinating rows.

F Language experiments: protocol and evaluation

Environment. The sandbox samples the state and fixed-protocol delivery failures, provides each agent only its own observation, and applies simultaneous actions. Since actions are independently sampled conditional on local types, per-type action frequencies define a behavioral profile. We evaluate that profile over the full world distribution. A separate simulation draws one response from each type's pool for 200 episodes per game and condition; the reported mean absolute discrepancy is $0.05\, nU$. This is a consistency check on sampled rollouts, not a second independent set of LLM queries.

Equation equation 14 gives the exact world-distribution evaluation used here. It is obtained by charging each expected actor $-L$ and replacing that charge by $x(w)$ if all act. The calculation

eliminates rollout sampling error conditional on the fitted action probabilities, not estimation error in those probabilities.

Metrics. Fidelity is agreement with the designated local action in the valid arm. False refusal is a WAIT response where that action is ACT; it is undefined for all-WAIT games and is not a complete error metric. In particular, acting against a stand-down instruction is not a false refusal. Violated-arm and uncertain-arm compliance are the fractions choosing the specified WAIT fallback. Worst-type regret is the largest expected gain from a unilateral action deviation at a positive-probability type under the stated game and behavioral opponents.

For target σ_T and achieved profile $\hat{\sigma}$, we report $\{W(\sigma_T) - W(\hat{\sigma})\}/(nU)$. Equation equation 15 separates the selection and execution differences. A difference need not be nonnegative for every arbitrary behavioral profile: σ_W is optimal only within the specified certified family plus WAIT. Welfare in the tables evaluates the valid game, whereas compliance evaluates the separate stress arms. We do not aggregate these into an expected deployment return under an assumed rate of assumption failures.

Initial sandbox design. The 24 seeded games have $n \in \{3, 4, 5\}$, eight per size, $\pi \in \{4/20, \ldots, 16/20\}$, $q_s \in \{1/20, \ldots, 19/20\}$, and integer $U \in [1, 5]$, $L \in [1, 10]$, $H \in [20, 60]$. Twelve random games have an alignment sampled on the $1/100$ grid conditional on a coordinating cutoff. Twelve boundary games have an interior hybrid cutoff and an alignment within 0.03 of its validity boundary on the $1/1000$ grid. Seed: 20260924.

Six development language families are direct statements, event logs, distractor prose, status tables, dialogue, and checklists. Four families held out until the confirmatory run are revocation notices, conflicting entries with an explicit precedence rule, quoted email, and negation. Frames differ syntactically; paired texts within a frame change only the relevant fact slot. Labels are unambiguous by construction. The pilot fixed an undefined veto term and inconsistent missing-message labels before prompts were frozen.

The valid arm affirms the veto and parameter status. The violated arm announces a broken root veto and requires WAIT; root types are excluded from this arm's local compliance score. The uncertain arm states that parameters are unverified and requires WAIT. Original conditions use four completions per prompt, temperature 0.7, and a 512-token cap, totaling $23{,}232$ prompts per executor. Game-only uses four language families; it is not identical in coverage to the other conditions.

Separating extraction and action. A response can contain both extracted fields and the model's chosen action. A1 scores that action. A2 applies the deterministic rule to the fields from the *same response*, without reading the action or the model's certificate-validity judgment. A3 applies the rule to gold facts. Only A3 is an oracle. The extraction-only condition uses a different prompt that requests facts but no decision. The evidence condition additionally requires a verbatim supporting quote, treating an unverifiable affirmative field as unknown.

Literal field accuracy and operational accuracy differ: recording NO for a gold UNKNOWN guard is literally wrong but yields the correct defer action. Conditional action-error rates restrict to responses with the relevant facts correctly extracted. These diagnostics do not reveal the model's internal reasoning.

The original study uses game–family cluster bootstrapping for its held-out-family intervals. The added welfare and format-ablation tables report point estimates; no confidence interval or claim of statistical equivalence is inferred from a rounded score of 1.000. Independent games, not the total number of prompt completions, determine the breadth of task coverage.

G WELFARE-SELECTION EXTENSION

This extension is post hoc. It uses the same games, language families, arm structure, and prompt templates as the initial sandbox. Nine games change the selected policy: seven to all-WAIT and two to a different cutoff. Policy-bearing rule and certificate prompts are queried again for those games; game-only is queried with the welfare convention on all 24 games. The extension uses $6{,}312$ prompts per model with four samples. Extraction-only prompts do not mention the selected policy, so their

original outputs are reused and re-scored with σ_W. The reported comparisons therefore combine unchanged queries with newly queried changed-policy conditions, not two fully independent runs.

Table 6: Exact selection outcomes on the original 24 games. Welfare is normalized by nU. All seven games with negative canonical welfare are retained and select WAIT under σ_W.

Stratum	Games	$W(\sigma_M)$	$W(\sigma_W)$	Select WAIT	Other cutoff
Random	12	.067	.233	2	0
Boundary	12	$-.108$	.182	5	2
All	24	$-.021$	.207	7	2

Table 7: Exact welfare and selected policies for all 24 original sandbox games. The seven stand-down games are retained in every reported overall mean.

Game	Stratum	n	$\mathcal{E}(\alpha)$	r^*	$W(\sigma_M)/(nU)$	σ_W	$W(\sigma_W)/(nU)$
00	random	3	$\{3\}$	3	0.068	σ^3	0.068
01	random	3	$\{3\}$	3	0.041	σ^3	0.041
02	random	3	$\{1\}$	1	0.070	σ^1	0.070
03	random	3	$\{3\}$	3	0.505	σ^3	0.505
04	random	4	$\{1\}$	1	0.562	σ^1	0.562
05	random	4	$\{4\}$	4	0.144	σ^4	0.144
06	random	4	$\{1\}$	1	0.575	σ^1	0.575
07	random	4	$\{1\}$	1	-1.125	WAIT	0.000
08	random	5	$\{1\}$	1	0.440	σ^1	0.440
09	random	5	$\{2\}$	2	0.231	σ^2	0.231
10	random	5	$\{1\}$	1	-0.870	WAIT	0.000
11	random	5	$\{1\}$	1	0.160	σ^1	0.160
12	boundary	3	$\{2\}$	2	-0.179	WAIT	0.000
13	boundary	3	$\{2\}$	2	0.355	σ^2	0.355
14	boundary	3	$\{2\}$	2	0.124	σ^2	0.124
15	boundary	3	$\{2\}$	2	-1.762	WAIT	0.000
16	boundary	4	$\{2\}$	2	0.318	σ^2	0.318
17	boundary	4	$\{2\}$	2	-0.187	WAIT	0.000
18	boundary	4	$\{2\}$	2	0.323	σ^2	0.323
19	boundary	4	$\{3\}$	3	-0.140	WAIT	0.000
20	boundary	5	$\{2\}$	2	0.322	σ^2	0.322
21	boundary	5	$\{2\}$	2	-0.612	WAIT	0.000
22	boundary	5	$\{1,4\}$	4	0.012	σ^1	0.200
23	boundary	5	$\{1,3\}$	3	0.127	σ^1	0.538

Exact selection difference $W(\sigma_W) - W(\sigma_M)$ over the 24 games: 0.228 [0.082, 0.403] nU (game-clustered bootstrap, 2000 resamples). Ties in σ_W resolve to the larger cutoff, as in the compiler's welfare selection; no tie occurred on these games.

Separate selection census. This census uses 400 random instances for each $n \in \{3,4,5\}$ and 21 alignment grid points. It is distinct from the 150-instance-per-size *ordering* census. Table 11 conditions on a nonempty certified family. The share with negative canonical welfare increases across these sizes, but this is a descriptive result for the generator, not a theorem about increasing n.

Most of the change on the 24 sandbox games comes from declining unprofitable coordination, not from a novel rule for selecting among many equilibria. In games 22 and 23, the certified sets are $\{1,4\}$ and $\{1,3\}$; the welfare selection chooses cutoff 1. Their normalized exact welfare increases from .012 to .200 and from .127 to .538, respectively. Welfare selection can therefore increase unconditional participation and is not a robustness guarantee.

The stand-down games expose a distinct execution failure. On game 07, a plain WAIT instruction gives welfare $-.225$ on Qwen and $-.808$ on Mistral, while the validator yields zero. The deterministic result follows from the constant WAIT policy and does not demonstrate improved fact extraction. All seven stand-down games have zero welfare under deterministic execution.

Table 8: Qwen3-30B: normalized welfare by selection and stratum. All rows use the same games; the welfare extension is post hoc.

		All (24)		Random (12)		Boundary (12)	
Pipeline	Selection	W	Loss	W	Loss	W	Loss
Local rule	σ_M	-0.186	0.166	-0.202	0.269	-0.171	0.063
	σ_W	0.086	0.121	0.081	0.152	0.091	0.091
Full certificate	σ_M	-0.410	0.390	-0.380	0.447	-0.441	0.333
	σ_W	-0.223	0.431	-0.188	0.421	-0.259	0.440
Certificate + validator	σ_M	-0.057	0.037	0.003	0.063	-0.118	0.010
	σ_W	0.177	0.030	0.180	0.053	0.174	0.008
Extraction + validator	σ_M	-0.272	0.251	-0.344	0.411	-0.199	0.091
	σ_W	0.064	0.143	0.026	0.207	0.102	0.080
Evidence + validator	σ_M	-0.576	0.555	-0.358	0.425	-0.793	0.685
	σ_W	-0.084	0.291	-0.078	0.311	-0.090	0.272
Game description	σ_M	-0.330	0.309	-0.240	0.306	-0.421	0.312
	σ_W	-0.330	0.538	-0.240	0.473	-0.421	0.603

Table 9: Mistral-24B: normalized welfare by selection and stratum. All rows use the same games; the welfare extension is post hoc.

		All (24)		Random (12)		Boundary (12)	
Pipeline	Selection	W	Loss	W	Loss	W	Loss
Local rule	σ_M	-0.066	0.045	-0.003	0.070	-0.128	0.020
	σ_W	0.082	0.125	0.117	0.116	0.046	0.135
Full certificate	σ_M	-0.382	0.362	-0.387	0.453	-0.378	0.270
	σ_W	-0.117	0.324	-0.082	0.315	-0.152	0.333
Certificate + validator	σ_M	-0.098	0.077	-0.002	0.068	-0.194	0.086
	σ_W	0.170	0.038	0.189	0.044	0.150	0.031
Extraction + validator	σ_M	-0.037	0.016	0.043	0.024	-0.117	0.008
	σ_W	0.200	0.008	0.219	0.014	0.180	0.002
Evidence + validator	σ_M	-0.543	0.523	-0.571	0.638	-0.516	0.407
	σ_W	-0.094	0.301	-0.135	0.368	-0.052	0.234
Game description	σ_M	-0.357	0.336	-0.275	0.342	-0.438	0.330
	σ_W	-0.355	0.562	-0.267	0.500	-0.442	0.624

Table 10: Welfare track: realized welfare under σ_W (units of nU) with game-clustered 95% bootstrap intervals and the paired difference to exact σ_W execution (2,000 resamples of the 24 games).

Executor	Pipeline	Realized W under σ_W [95% CI]	Δ vs exact σ_W
Qwen3-30B	Local rule, direct	0.086 [0.012, 0.164]	−0.121 [-0.172, -0.074]
Qwen3-30B	Full certificate, direct	−0.223 [-0.391, -0.074]	−0.431 [-0.592, -0.279]
Qwen3-30B	Certificate + validator	0.177 [0.102, 0.255]	−0.030 [-0.056, -0.010]
Qwen3-30B	Extraction + validator	0.064 [-0.025, 0.142]	−0.143 [-0.233, -0.069]
Qwen3-30B	Evidence + validator	−0.084 [-0.190, 0.019]	−0.291 [-0.424, -0.176]
Qwen3-30B	Game description	−0.330 [-0.663, -0.074]	−0.538 [-0.850, -0.296]
Mistral-24B	Local rule, direct	0.082 [-0.069, 0.215]	−0.125 [-0.227, -0.051]
Mistral-24B	Full certificate, direct	−0.117 [-0.266, 0.006]	−0.324 [-0.451, -0.212]
Mistral-24B	Certificate + validator	0.170 [0.100, 0.248]	−0.038 [-0.060, -0.020]
Mistral-24B	Extraction + validator	0.200 [0.123, 0.279]	−0.008 [-0.015, -0.002]
Mistral-24B	Evidence + validator	−0.094 [-0.160, -0.035]	−0.301 [-0.401, -0.200]
Mistral-24B	Game description	−0.355 [-0.704, -0.094]	−0.562 [-0.882, -0.316]

Table 11: Separate selection census. W denotes mean normalized welfare over coordinating cells.

n	Cells	$W(\sigma_M)$	$W(\sigma_W)$	$W(\sigma_M) < 0$	Select WAIT	Other cutoff	$\lvert\mathcal{E}\rvert > 1$
3	4724	.149	.265	.228	.228	.110	.131
4	4450	.009	.228	.344	.324	.096	.116
5	4293	−.019	.214	.384	.363	.082	.102

H FRESH-GAME EXECUTION-CARD ABLATION

The card study has a separate preregistration (`card_ablation_prereg.md`, with recorded outcomes), frozen after a pilot with seed 999 on six disjoint games and three families, used only to check parsing and the card texts. Its fresh game set uses seed 20260929 and the sandbox generator and strata (12 random, 12 boundary-enriched; $n \in \{3, 4, 5\}$), contains 24 games, and selects all-WAIT in nine. It evaluates six language families (direct, table, checklist, revocation, conflict, negation; the two hardest development families and three of the four held-out families of the initial study), the three arms of the initial study, every agent type (8 per game on average; root types excluded from the violated arm), and the five prompted conditions, giving 15,840 prompts per model with 4 completions each at temperature 0.7 and a 512-token cap (63,360 samples). The two-stage condition issues one extraction call per prompt (4 completions) and one second call per completion (64-token cap) that receives only the extracted facts and the card. Fields are parsed by the same regular-expression parser as the initial study, failing closed. Ties in σ_W resolve to the larger cutoff, as in the compiler. Exact mean welfare is $0.164\,nU$ for σ_W and $-0.094\,nU$ for σ_M. The data are stored separately from the original sandbox and its post-hoc welfare extension.

The four single-call representations carry the same selected policy, guards, and fallback but vary how much audit content is visible and how the operational rule is presented. The full audit certificate retains all metadata. The compact prose card removes audit metadata. Action-first reorders the full content. The decision-table card presents the local cases directly. The two-stage LLM condition separates extraction from application of the table. The deterministic comparator uses extracted facts, not ground truth. As the extraction contexts differ, comparison with direct card execution is a pipeline comparison, not a controlled comparison of final decision rules on identical facts.

Table 12: Card-ablation false refusals and welfare loss. Refusal rates are conditional on a prescribed ACT; they do not count unwanted action under an all-WAIT policy. Losses are reported before rounding the component means, so subtraction of the displayed values need not agree exactly.

	Qwen3-30B		Mistral-24B	
Representation	False refusal	Welfare loss	False refusal	Welfare loss
Full audit certificate	.301	.370	.082	.237
Compact prose card	.056	.150	.006	.098
Action-first audit	.001	.003	.032	.161
Decision-table card	.000	.000	.001	.017
Extract + LLM table	.097	.171	.008	.024
Extract + validator	.097	.171	.006	.020

Table 13: Probability of an incorrect action conditional on correctly extracted relevant facts, as reported for the four direct formats. This diagnostic includes cases outside the valid-arm fidelity subset.

Representation	Qwen3-30B	Mistral-24B
Full audit certificate	.262	.185
Compact prose card	.030	.147
Action-first audit	.117	.031
Decision-table card	.004	.017

The action-first effect is model- and metric-dependent. It gives Qwen valid-arm welfare .161, yet uncertain-arm compliance is only .810. Mistral's compliance improves to .974/.988, while valid-arm welfare is .003. A monotone performance ordering from audit to prose to action-first to table is therefore not supported.

For Qwen, the table's valid-arm fidelity is 1.000 at reported precision, but conditional action error over the evaluated fact-correct cases is .004 and veto-violation compliance is .924. These quantities concern different subsets of samples and must not be collapsed into a zero-error claim. For Mistral, the corresponding conditional error is .017, and uncertain-arm compliance is .968. Residual failures concern guard enforcement as well as grounding. No new aggregate catastrophe probability is inferred from these local compliance rates.

Table 14: Card ablation: realized welfare (units of nU) with game-clustered 95% bootstrap intervals and paired differences against the full audit certificate and against the extract-plus-validator pipeline (2,000 resamples of the 24 games, computed from the raw logs).

Executor	Representation	Realized W [95% CI]	Δ vs full audit	Δ vs validator
Qwen3-30B	Full audit certificate	−0.206 [-0.326, -0.090]	0.000 [0.000, 0.000]	−0.199 [-0.313, -0.084]
Qwen3-30B	Compact prose card	0.014 [-0.077, 0.100]	0.221 [0.131, 0.316]	0.022 [-0.025, 0.072]
Qwen3-30B	Action-first audit	0.161 [0.090, 0.241]	0.368 [0.248, 0.490]	0.169 [0.076, 0.278]
Qwen3-30B	Decision-table card	0.164 [0.093, 0.242]	0.370 [0.252, 0.494]	0.171 [0.077, 0.284]
Qwen3-30B	Extract + LLM table	−0.007 [-0.115, 0.084]	0.199 [0.086, 0.315]	0.000 [0.000, 0.000]
Qwen3-30B	Extract + validator	−0.007 [-0.115, 0.084]	0.199 [0.086, 0.315]	0.000 [0.000, 0.000]
Mistral-24B	Full audit certificate	−0.072 [-0.171, 0.021]	0.000 [0.000, 0.000]	−0.217 [-0.301, -0.137]
Mistral-24B	Compact prose card	0.066 [-0.044, 0.176]	0.139 [0.058, 0.230]	−0.078 [-0.135, -0.025]
Mistral-24B	Action-first audit	0.003 [-0.105, 0.115]	0.076 [0.021, 0.142]	−0.141 [-0.216, -0.073]
Mistral-24B	Decision-table card	0.147 [0.069, 0.230]	0.220 [0.141, 0.302]	0.003 [-0.023, 0.028]
Mistral-24B	Extract + LLM table	0.140 [0.068, 0.219]	0.212 [0.135, 0.294]	−0.004 [-0.010, 0.000]
Mistral-24B	Extract + validator	0.144 [0.072, 0.224]	0.217 [0.138, 0.302]	0.000 [0.000, 0.000]

I PAIRED EQUILIBRIUM RECONSTRUCTION

The paired reconstruction suite contains 48 seeded games with $n \in \{3, 4\}$, eight in each stratum: unique cutoff, multiple cutoffs, no in-class coordination, boundary equality, singleton validity, and a non-monotone selection transition. Each prompt describes the game and cutoff class and asks one of four questions: T_{any} requests any BNE cutoff or NONE; $T_{\max}$ the largest BNE cutoff; T_W the highest-base-welfare *coordinating cutoff*; and T_{prof} the canonical full type-to-action profile. Thus T_W in this study is not the later deployment selection with WAIT as an outside option. Its tie-break chooses larger r.

Conditions share a 7,000-completion-token budget, temperature .7, eight samples, and an output schema. Game-only receives the task description. Formula adds the cutoff inequalities. Code adds the inequalities and requests a Python function, executed for up to ten seconds with the standard library. Tool asks the model to extract the game as JSON and delegates computation to the exact compiler. Pilot changes to the budget and extraction contract are recorded before the main run.

Table 15: Paired reconstruction. T_{any} and T_{prof} report BNE validity and agreement with the canonical answer separately. $T_{\max}$ and T_W report exact selection accuracy.

Executor	Condition	T_{any}		$T_{\max}$	T_W	T_{prof}	
		BNE	=conv.	exact	exact	BNE	=conv.
Mistral-24B	game only	0.445	0.240	0.299	0.279	0.373	0.250
Mistral-24B	formula	0.445	0.240	0.263	0.362	0.549	0.276
Mistral-24B	code (runs 0.24)	0.159	0.089	0.055	0.029	–	–
Mistral-24B	tool	0.995	0.995	0.992	0.992	–	–
Qwen3-30B	game only	0.271	0.188	0.177	0.307	0.418	0.271
Qwen3-30B	formula	0.635	0.497	0.688	0.505	0.909	0.768
Qwen3-30B	code (runs 0.69)	0.591	0.401	0.469	0.328	–	–
Qwen3-30B	tool	1.000	1.000	1.000	1.000	–	–

The formula helps Qwen on several tasks, but there is no uniform effect across models. Mistral's full-profile validity rises from .373 to .549 and welfare-cutoff accuracy from .279 to .362, despite unchanged T_{any} validity. Non-monotone games are also not uniformly near-zero: game-only Mistral achieves .422 on T_{any} and .250 on $T_{\max}$. Exact selection is not consistently harder than finding any equilibrium, and the table does not establish that welfare selection is harder on both models. The reliable comparison is that the exact-tool pipeline is near ceiling while unaided and code-generating pipelines retain substantial errors.

J SCOPE BEYOND A FULL QUORUM

For a quorum of $m < n$, a root veto no longer prevents an unsafe quorum from forming without the root, and best responses depend on whether the other agents already form a quorum. In the

Table 16: Exact reconstruction accuracy by stratum. Each cell has eight games and 64 samples.

Executor	Task	Condition	unique	multiple	none	boundary eq	singleton	nonmono
Mistral-24B	T any	game only	0.375	0.641	0.031	0.453	0.750	0.422
Mistral-24B	T any	formula	0.375	0.672	0.125	0.328	0.766	0.406
Mistral-24B	T any	code	0.188	0.188	0.031	0.125	0.266	0.156
Mistral-24B	T any	tool	0.984	1.000	1.000	1.000	0.984	1.000
Mistral-24B	T max	game only	0.219	0.453	0.156	0.328	0.391	0.250
Mistral-24B	T max	formula	0.328	0.203	0.078	0.359	0.219	0.391
Mistral-24B	T max	code	0.078	0.016	0.062	0.062	0.000	0.109
Mistral-24B	T max	tool	1.000	0.984	1.000	0.984	1.000	0.984
Mistral-24B	T wel	game only	0.281	0.203	0.016	0.516	0.375	0.281
Mistral-24B	T wel	formula	0.297	0.422	0.188	0.453	0.406	0.406
Mistral-24B	T wel	code	0.016	0.047	0.031	0.000	0.047	0.031
Mistral-24B	T wel	tool	1.000	1.000	1.000	0.969	1.000	0.984
Qwen3-30B	T any	game only	0.250	0.641	0.031	0.219	0.406	0.078
Qwen3-30B	T any	formula	0.719	0.641	0.406	0.672	0.688	0.688
Qwen3-30B	T any	code	0.406	0.422	0.734	0.578	0.719	0.688
Qwen3-30B	T any	tool	1.000	1.000	1.000	1.000	1.000	1.000
Qwen3-30B	T max	game only	0.125	0.453	0.000	0.219	0.234	0.031
Qwen3-30B	T max	formula	0.750	0.672	0.578	0.625	0.688	0.812
Qwen3-30B	T max	code	0.422	0.578	0.625	0.469	0.297	0.422
Qwen3-30B	T max	tool	1.000	1.000	1.000	1.000	1.000	1.000
Qwen3-30B	T wel	game only	0.422	0.406	0.000	0.234	0.516	0.266
Qwen3-30B	T wel	formula	0.547	0.328	0.703	0.500	0.375	0.578
Qwen3-30B	T wel	code	0.234	0.188	0.516	0.359	0.250	0.422
Qwen3-30B	T wel	tool	1.000	1.000	1.000	1.000	1.000	1.000

strategically equivalent utility $u_i + \lambda \sum_{j \neq i} u_j$, with k other actors, the marginal utility of acting is $-L$ if $k \leq m-2$; $U+\lambda(m-1)(U+L)$ if $k = m-1$ in SAFE; $-H+\lambda(m-1)(L-H)$ if $k = m-1$ in UNSAFE; and U or $-H$ if $k \geq m$. When $H > L$ the alignment terms of the two pivotal cases have opposite signs, so the single-odds reduction of Lemma 1 does not carry over. An exhaustive pure-BNE scan of 760 sampled instances ($n \in \{3, 4, 5\}$, $2 \leq m < n$, 21 alignments each) found no violation of upward-closed coordination support, but only 20–54% of the coordinating equilibria were cutoff profiles, and at $(n, m) = (4, 2)$ and $(5, 4)$ a small share (0.4–0.5%) reached the quorum in UNSAFE. Partial quorums are therefore outside the scope of Theorem 1; this scan is a scope check, not a theory of the partial-quorum case.

K Initial sandbox: grounding versus enforcement by pipeline and language family

The initial sandbox study (Appendix F) deployed the canonical selection σ_M. Its per-pipeline results separate reading the environment from applying the rule. Fidelity is valid-arm agreement with the certified action; false refusal is WAIT where the certified action is ACT; compliance is the WAIT rate in the violated and uncertain arms; the gap is to the best certified cutoff with all-WAIT allowed, i.e. to $W(\sigma_W)$; "danger" is P(UNSAFE quorum) with the veto broken and the root modeled as always acting. Operational fact accuracy of the three pivotal facts is 0.88–0.96 (Qwen3-30B) and 0.97–0.99 (Mistral-24B) across pipelines and does not degrade on the four held-out families. Given correctly extracted facts, the full-certificate executors still choose the wrong action in 0.33 and 0.28 of samples (local rule: 0.03 and 0.02), act in 0.48 and 0.42 of uncertain-arm samples although they record the parameters as unconfirmed in 0.99 of them, and recognize the broken veto yet act in 0.37 and 0.45 of violated-arm samples. Realized danger is near zero because after a broken veto the models almost never act without a message (10 of 1,840 blind-type samples on Qwen3-30B), and acting on a received message is outcome-safe since a message implies SAFE. Pre-registered predictions: P1 (validator compliance $\geq$ own-action compliance), P2 (evidence gate raises refusals) and P4 (game-only lowest fidelity) were supported; P3 (rule and certificate within 0.05 fidelity) was rejected, the certificate wrapper lowering own-action fidelity by 0.13–0.18; P5 (held-out degradation) was not supported.

Table 17: Initial sandbox, σ_M deployed, all arms and families (92,928 samples per executor). Code rows are exact policies evaluated in the true game.

Executor	Pipeline	Fidelity	False ref.	Compl. viol.	Compl. unc.	Gap (nU)	Max regret	Danger
code	exact compiler	–	–	–	–	0.23	0.00	–
code	always-WAIT	–	–	–	–	0.21	0.00	–
code	all-responsive	–	–	–	–	0.54	4.31	–
code	all-blind	–	–	–	–	0.25	4.52	–
code	scalar reliability	–	–	–	–	0.25	2.30	–
code	gold-facts validator (oracle)	1.000	0.000	1.000	1.000	0.23	0.00	0.000
Mistral-24B	rule only, own action	0.990	0.009	0.962	0.982	0.27	0.31	0.000
Mistral-24B	certificate, own action	0.916	0.109	0.498	0.585	0.59	1.30	0.003
Mistral-24B	certificate → validator	0.990	0.014	0.995	1.000	0.31	0.55	0.000
Mistral-24B	extract → validator	0.998	0.003	1.000	1.000	0.24	0.18	0.000
Mistral-24B	extract + evidence → validator	0.909	0.122	0.999	1.000	0.75	2.21	0.000
Mistral-24B	game only, convention stated	0.754	0.320	0.517	0.518	0.56	4.31	0.000
Qwen3-30B	rule only, own action	0.974	0.035	0.906	0.959	0.39	0.34	0.000
Qwen3-30B	certificate, own action	0.785	0.291	0.730	0.520	0.62	4.37	0.000
Qwen3-30B	certificate → validator	0.994	0.009	0.947	0.990	0.26	0.36	0.000
Qwen3-30B	extract → validator	0.954	0.062	0.951	1.000	0.48	1.22	0.000
Qwen3-30B	extract + evidence → validator	0.890	0.149	0.980	1.000	0.78	7.18	0.000
Qwen3-30B	game only, convention stated	0.760	0.325	0.530	0.500	0.54	4.31	0.000

Table 18: Per-family operational fact accuracy and gold-action agreement (all arms). †: family held out until the confirmatory run.

Executor	Family	rule only, A1		certificate, A1		extract → validator	
		op. facts	gold agr.	op. facts	gold agr.	op. facts	gold agr.
Mistral-24B	direct	0.994	0.987	0.995	0.638	0.999	1.000
Mistral-24B	log	0.996	0.974	0.996	0.679	0.995	1.000
Mistral-24B	distractor	0.998	0.994	0.998	0.625	0.998	1.000
Mistral-24B	table	0.809	0.972	0.762	0.641	0.998	1.000
Mistral-24B	dialogue	0.999	0.978	0.996	0.706	0.998	1.000
Mistral-24B	checklist	0.988	0.971	0.982	0.675	0.996	1.000
Mistral-24B	revocation†	0.996	0.974	0.984	0.765	0.990	0.997
Mistral-24B	conflict†	0.994	0.978	0.980	0.678	0.978	0.997
Mistral-24B	email†	0.995	0.976	0.990	0.769	0.994	0.999
Mistral-24B	negation†	0.992	0.990	0.998	0.640	0.995	0.999
Qwen3-30B	direct	1.000	0.998	0.902	0.680	0.988	1.000
Qwen3-30B	log	0.987	0.974	0.977	0.661	0.925	0.991
Qwen3-30B	distractor	0.993	0.990	0.925	0.622	0.863	1.000
Qwen3-30B	table	0.932	1.000	0.773	0.705	0.672	0.933
Qwen3-30B	dialogue	1.000	0.991	1.000	0.718	0.986	1.000
Qwen3-30B	checklist	0.832	0.725	0.804	0.599	0.866	0.866
Qwen3-30B	revocation†	0.932	0.996	0.906	0.734	0.765	0.925
Qwen3-30B	conflict†	1.000	0.895	0.998	0.679	1.000	1.000
Qwen3-30B	email†	0.943	0.935	0.973	0.732	0.867	0.997
Qwen3-30B	negation†	0.870	0.997	0.909	0.605	0.847	0.988

Table 19: Held-out families only: estimates with two-way (game × family) cluster-bootstrap 95% intervals (500 resamples).

Executor	Pipeline	Fidelity	False refusal	Compl. violated	Compl. uncertain
Mistral-24B	rule only, own action	0.985 [0.97, 0.99]	0.018 [0.01, 0.03]	0.968 [0.95, 0.98]	0.983 [0.97, 0.99]
Mistral-24B	certificate, own action	0.860 [0.80, 0.92]	0.182 [0.11, 0.25]	0.541 [0.45, 0.62]	0.694 [0.55, 0.84]
Mistral-24B	certificate → validator	0.997 [0.99, 1.00]	0.004 [0.00, 0.01]	1.000 [1.00, 1.00]	1.000 [1.00, 1.00]
Mistral-24B	extract → validator	0.994 [0.99, 1.00]	0.007 [0.00, 0.01]	1.000 [1.00, 1.00]	1.000 [1.00, 1.00]
Mistral-24B	extract + evidence → validator	0.877 [0.84, 0.91]	0.166 [0.12, 0.22]	0.999 [1.00, 1.00]	1.000 [1.00, 1.00]
Mistral-24B	game only, convention stated	0.755 [0.71, 0.81]	0.325 [0.27, 0.37]	0.523 [0.50, 0.56]	0.531 [0.50, 0.57]
Qwen3-30B	rule only, own action	0.940 [0.81, 1.00]	0.081 [0.00, 0.25]	0.965 [0.92, 0.99]	0.965 [0.90, 1.00]
Qwen3-30B	certificate, own action	0.765 [0.71, 0.82]	0.318 [0.26, 0.36]	0.785 [0.60, 0.97]	0.537 [0.51, 0.57]
Qwen3-30B	certificate → validator	0.999 [0.99, 1.00]	0.002 [0.00, 0.01]	0.997 [0.99, 1.00]	1.000 [1.00, 1.00]
Qwen3-30B	extract → validator	0.938 [0.84, 1.00]	0.084 [0.00, 0.22]	1.000 [1.00, 1.00]	1.000 [1.00, 1.00]
Qwen3-30B	extract + evidence → validator	0.879 [0.78, 0.95]	0.164 [0.06, 0.30]	0.998 [0.99, 1.00]	1.000 [1.00, 1.00]
Qwen3-30B	game only, convention stated	0.760 [0.71, 0.81]	0.324 [0.27, 0.37]	0.500 [0.50, 0.50]	0.500 [0.50, 0.50]

L EARLIER LANGUAGE-INTERFACE STUDIES

Before the sandbox, smaller pre-registered studies on a wider executor panel (Qwen3-30B-A3B-Instruct-2507, Qwen3-8B with thinking disabled unless stated, Mistral-Small-24B-Instruct-2501, Phi-4; temperature 0.7; 8 samples per query unless stated) established the execution ceiling and motivated the sandbox and card designs. We keep only the results that bound or motivate the final claims; prompts, registration records and per-query logs are in the artifact.

Execution ceiling. On the introductory hybrid relay, a homogeneous $n = 4$ relay and a two-agent path swept densely in α (64 samples per type and point), certificate conditions agreed with the certified policy at 1.000 on Qwen3-30B ($3552/3552$) and 0.847 pooled on Qwen3-8B ($1097/1296$; 0.70–0.93 per condition), switched inside the predicted grid cell where the certified policy changes across a boundary, and cost ≈7 completion tokens. Fixed-depth belief scaffolds never realized the blind type's unconditional ACT above the hybrid boundary. Execution transferred to 120 held-out generated relay lines (501 game–alignment pairs, $n \leq 5$): agreement 1.000 (30B), 0.997 (8B), 0.994 (Mistral-24B); Phi-4 was below ceiling (0.76) with both truncation at a 256-token cap and disagreement among parsed responses (0.81).

Recommendation formats. Supplying the correct recommendation as an action label, a local policy rule, a machine-readable certificate or a full natural-language certificate was indistinguishable on Qwen3-30B (all at ceiling); on Mistral-24B and Phi-4 the bare label was the *worst* format and execution rose with structure, the machine-readable certificate reaching ≥ 0.996 on every checkpoint. This is the earlier evidence that a certificate whose local action is already resolved executes best, which the card ablation (Appendix H) tests directly.

Validity factorial. A $3\times3\times3$ factorial crossed certificate validity (correct; stale after an announced veto break; internally inconsistent) with assumption presentation (omitted; prose; machine-readable guard) and fallback (none; soft; explicit per-agent WAIT) on the all-blind relay. Without assumption text and fallback every model executed the stale certificate (uninformed ACT rate 0.95–1.000). Stating the assumption enabled detection but not protection: with no fallback Qwen3-30B flagged the stale certificate in $61/61$ parsed responses yet acted in 0.95 of samples. Only assumption text plus an explicit fallback cut exposure (0.42 / 0.20 / 0.02 on 30B / Mistral / Phi-4), soft fallbacks were rationalized away, and the metadata taxed valid certificates (execution 0.75–0.86 against 0.94–1.000 without metadata). Qwen3-8B followed the stale certificate in every cell with thinking disabled and reached exposure 0.000 in the strongest cell with thinking enabled.

Single-template narrative study. On 48 generated games with one narrative template per role, the veto violation stated only in a narrative bulletin was extracted correctly in 0.26 (30B) and 0.51 (Mistral) of samples, so a validator fed "veto intact" applied the certificate (compliance 0.37–0.63), and one ambiguous UNSAFE template was read as SAFE in $249/256$ samples. Grounding, not enforcement, bound there; the sandbox's defined terms and ten families were the response.

Superseded studies. Three further studies are archived in the artifact and are superseded by the designs reported in the main text: an answer-hidden reconstruction study without a stated selection convention (agreement with the compiler-selected policy 0.46–0.60; the only condition producing catastrophic quorums was free-form reasoning, p_{cat} up to 0.12) and its convention-stated replication at a larger budget (0.73), whose comparison does not isolate the convention because budget and format also changed, which is why the paired design of Appendix I was run; and a five-episode adaptation case study that showed exact execution and no adaptation, in which a certificate compiled from a wrong channel parameter was executed faithfully into the wrong policy (true-model welfare $0.20 \to -0.56$), while a pre-registered hypothesis that the alignment margin predicts agreement was null.