DECENTRALIZED LOCAL SEARCH IN NEAR-POTENTIAL BINARY GAMES

Anonymous authors
Paper under double-blind review

ABSTRACT

We study decentralized binary programming games in which agents make constrained binary decisions under coupled costs, limited computation, and unreliable communication. We propose Decentralized Local Search (DLS), which replaces exact best-response optimization with randomized local search and uses locally maintained empirical-frequency estimates over a random time-varying network. Communication may become increasingly sparse, provided successful information propagation persists over time. For near-potential games, we establish finite-time belief-tracking guarantees and almost-sure convergence to an approximate Nash equilibrium set. Under a best-response margin condition, DLS reaches the pure Nash equilibrium set in finite time almost surely; if equilibria are strict, the joint decisions eventually stabilize at a pure Nash equilibrium. Numerical experiments show that DLS preserves equilibrium-seeking performance while substantially reducing the cost of repeated exact optimization.

1 INTRODUCTION

Many multi-agent decision problems are combinatorial at the level of the individual agent: a sensor selects which measurements to use (Joshi & Boyd, 2009), an autonomous vehicle which target to engage (Arslan et al., 2007), a transmitter or access point which channels to use (Kamiya et al., 2017), and a network user which route to take (Roughgarden & Tardos, 2002). In such problems, each agent's decision is naturally represented by a binary or discrete vector subject to local constraints such as capacities, budgets, ranges, or exclusivity rules, while its cost depends on the decisions of other agents through congestion, interference, competition, or shared resources. This makes the interaction a constrained combinatorial game.

Two difficulties make decentralized equilibrium seeking particularly challenging in this setting. The first is computational. Even a single exact best response may require solving an NP-hard binary integer program (Karp, 1972), while computing a Nash equilibrium is PPAD-complete in general (Daskalakis et al., 2009). Repeatedly solving such problems to optimality can therefore be prohibitively expensive, particularly in online or large-scale settings. This motivates learning-based approaches in which equilibrium behavior emerges from repeated, comparatively simple local updates rather than from directly solving the equilibrium problem (Fudenberg & Levine, 1998; Young, 2004). The challenge considered here is sharper because even the local best-response step may itself be a difficult combinatorial optimization problem.

The second difficulty is informational and concerns resilience of the communication network. In autonomous and networked systems, communication may become intermittent because of mobility and limited communication ranges, environmental interference, bandwidth or energy constraints, or adversarial disruptions such as jamming and denial-of-service attacks (Fan et al., 2011; Tsao et al., 2022; Ishii et al., 2022). These effects are particularly relevant in multi-robot and UAV systems, where changing relative positions can break communication links and wireless channels are vulnerable to interference and cyberattacks. Requiring persistent connectivity or uniformly positive communication rates can therefore be restrictive. We instead seek a learning rule that tolerates long and increasingly sparse communication gaps, provided successful information propagation persists over time.

This paper develops such a learning rule for near-potential binary games Candogan et al. (2013b). Our main idea is to replace both expensive computation and demanding communication require-

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.

ments with inexpensive local mechanisms. Rather than computing an exact best response, each agent performs a randomized local search over single-bit perturbations of its current binary decision. Rather than observing the actions of all other agents, it maintains local estimates of their fading-memory empirical frequencies through communication over a random time-varying network. Inertia creates periods during which actions remain fixed, allowing these local estimates to become accurate even under sparse communication. Randomized exploration, in turn, allows the local search to move through constrained binary spaces without solving the underlying integer program exactly.

The near-potential structure provides the global mechanism connecting these local operations. Although the agents' objectives need not be generated by an exact common potential, their unilateral incentives remain within δ of those of a nearby potential game. Consequently, sufficiently large unilateral improvements induce descent of the nearby potential, while sufficiently accurate local beliefs allow the decentralized search to identify improving decisions. This coupling between inexpensive local search, intermittent information exchange, and near-potential structure is what permits convergence without either exact best responses or persistent network connectivity. Our contributions are as follows.

A computationally light decentralized learning algorithm. We introduce Decentralized Local Search (DLS, Algorithms 1–2), in which each agent replaces an exact binary-program best response with a randomized search over single-bit-flip neighbors of its current incumbent. The search may explore lightly infeasible intermediate decisions and uses random exploration to escape restrictive local-search paths. Thus, DLS requires neither a centralized coordinator nor the repeated solution of a mixed-integer program at every update.

Resilience to sparse and deteriorating communication. Agents maintain local estimates of the others' fading-memory empirical frequencies through consensus over a random time-varying communication network. We establish finite-time tracking bounds under a substantially weaker requirement than persistent connectivity. Over blocks of B steps, it is sufficient that information originating from agent j reaches every agent with probability at least $p_{j,m}$ in block m, where $\sum_{m=0}^{\infty} p_{j,m} = \infty$. In particular, the probabilities $p_{j,m}$ may vanish with time and the realized graph need not be connected at any individual time step. When an agent holds its decision sufficiently long, all local estimates approach that decision with arbitrarily high probability (Lemma 1). This provides a form of resilience to increasingly sparse communication without requiring persistent information exchange.

Convergence from local search and imperfect information. We connect randomized local search and belief tracking through the near-potential structure. Under the network persistence and admissible-path reachability assumptions, DLS admits finite positive-probability improvement paths even though agents act only on local estimates and do not compute exact best responses. We first show convergence to the $\delta \prod_i |\mathcal{X}_i|$-approximate Nash equilibrium set of a δ-near-potential game. Under an additional best-response margin condition, every non-equilibrium profile admits a finite realization that strictly decreases the nearby potential; finiteness of the action space then implies almost-sure finite-time reachability of the pure Nash equilibrium set. If pure Nash equilibria are strict, belief tracking makes a sufficiently accurate visit absorbing, so the joint actions stabilize almost surely at a pure Nash equilibrium while the local estimates converge to it (Theorem 1).

Computational and scalability gains in practice. We evaluate DLS on nine classes of binary games with up to 100 agents (Section 6), including vehicle–target assignment and problems with knapsack, cardinality, and range constraints. DLS reaches an exact Nash equilibrium in all 50 random seeds on 32 of the 36 instances for which exact best responses remain tractable, while achieving social costs comparable to solver-based best-response dynamics. More importantly, DLS remains effective in regimes where exact optimization becomes impractical: in the cubic interference game the centralized formulation cannot even be constructed for 64 or more agents, whereas DLS reaches an exact equilibrium in every seed with up to 100 agents. When individual best-response problems themselves become large, DLS attains 95–97% of the welfare of solver-based dynamics at roughly one tenth of their per-update computation time.

2 Related Work

Integer programming games. Integer programming games model strategic interactions with integer-constrained decisions (Carvalho et al., 2025). Computational approaches include sample

generation (Carvalho et al., 2022), equilibrium inequalities for computing and selecting pure Nash equilibria (Dragotto & Scatamacchia, 2023), and outer approximations for mixed-equilibrium computation (Carvalho et al., 2026). Payoff approximations enable approximate equilibrium computation in nonlinear games (Duguet et al., 2026), while random-restart best-response dynamics address larger instances (Lee et al., 2024). These methods provide solver-based benchmarks. DLS couples solver-free binary local search with distributed information tracking and admits approximate-equilibrium convergence guarantees under explicit structural assumptions.

Decentralized learning and communication. Distributed fictitious play estimates opponents' empirical frequencies over time-varying networks in potential and near-potential games (Arefizadeh & Eksin, 2019; Aydın et al., 2021). Log-linear learning variants accommodate restricted action availability and limited payoff information (Marden & Shamma, 2012). Complementary consensus results characterize persistent information flow and convergence under unreliable communication, including time-varying success probabilities and structured dependence (Touri & Nedic, 2010; Shi et al., 2012; 2015). These results provide background for our tracking analysis, although their network and dependence assumptions differ. We establish approximate-equilibrium convergence for constrained binary local search coupled to distributed estimation over random communication networks.

Combinatorial search. Exact methods for binary and mixed-integer optimization, such as branch-and-bound and branch-and-cut, can certify global optimality, but repeated best-response solves may incur substantial computational cost. Heuristic alternatives include GRASP (Festa & Resende, 2002), variable neighborhood search (Mladenović & Hansen, 1997), and binary-decision-diagram heuristics (Bergman et al., 2014). Lokketangen & Glover (1998) develop tabu search combining memory-guided simplex pivots with objective and integer-infeasibility criteria. For large-scale binary programs, Umetani (2017) combines variable associations, multi-flip neighborhoods, incremental evaluation, and adaptive constraint penalties. These methods focus on single-agent optimization. DLS addresses the additional challenges of strategic interaction and imperfect information by combining constrained binary local search with inertial updates and distributed empirical-frequency tracking over random communication networks. Rather than requiring globally optimal responses at each update, our framework establishes convergence to approximate Nash equilibria in near-potential games under the stated assumptions.

Coverage and resource allocation. Distributed coverage games incorporate energy and sensing constraints (Zhu & Martínez, 2013), while limited-information methods address coverage in non-convex discrete environments (Iwase et al., 2024). Utility-design methods characterize the efficiency of equilibria in covering and related resource-allocation games (Paccagnan & Marden, 2021). Our objective is complementary: learning equilibria for prescribed, individually constrained binary costs, rather than designing utilities or optimizing a particular geometric coverage model.

3 Near-Potential Binary Programming Games

We consider a non-cooperative binary programming game $\Gamma = \langle \mathcal{N}, \{\mathcal{X}_i\}_{i\in\mathcal{N}}, \{c_i\}_{i\in\mathcal{N}} \rangle$, where each agent $i \in \mathcal{N}$ selects a binary decision vector by solving

$$\begin{aligned} \min_{\mathbf{x}_i} \quad & c_i(\mathbf{x}_i, \mathbf{x}_{-i}) \\ \text{s.t.} \quad & \mathbf{A}_i \mathbf{x}_i \leq \mathbf{b}_i \\ & \mathbf{x}_i \in \{0,1\}^{K_i}, \end{aligned} \tag{1}$$

where $\mathbf{A}_i \in \mathbb{Z}^{M_i \times K_i}$ and $\mathbf{b}_i \in \mathbb{Z}^{M_i}$ define agent i's individual feasible set $\mathcal{X}_i := \{\mathbf{x}_i \in \{0,1\}^{K_i} : \mathbf{A}_i\mathbf{x}_i \leq \mathbf{b}_i\}$, and the cost $c_i : \{0,1\}^{\sum_{n=1}^{N} K_n} \to \mathbb{R}$ couples agent i's decision to the joint decisions $\mathbf{x}_{-i} \in \mathcal{X}_{-i} \subseteq \{0,1\}^{\sum_{j\neq i} K_j}$ of other agents. With slight abuse of notation, we extend the cost c_i to mixed strategies $\sigma_i \in \Delta\mathcal{X}_i$ by defining the expected cost $c_i(\sigma_i, \sigma_{-i})$ in the standard game-theoretic manner.

Definition 1 (Near-Potential Game Candogan et al. (2013a)). *Given two games* $\Gamma = (\mathcal{N}, \{\mathcal{X}_i, c_i\}_{i\in\mathcal{N}})$ *and* $\hat{\Gamma} = (\mathcal{N}, \{\mathcal{X}_i, \hat{c}_i\}_{i\in\mathcal{N}})$ *sharing the same agent and decision sets, their maxi-*

mum pairwise difference is

$$d(\Gamma, \hat{\Gamma}) := \max_{i \in \mathcal{N},\, x_i' \in \mathcal{X}_i,\, x \in \mathcal{X}} \big| \big(c_i(x_i', x_{-i}) - c_i(x_i, x_{-i})\big) - \big(\hat{c}_i(x_i', x_{-i}) - \hat{c}_i(x_i, x_{-i})\big) \big|, \tag{2}$$

which captures the worst-case discrepancy in unilateral cost differences across all agents and decision profiles between two games. A game Γ *is a* near-potential game *if there exists a potential game* $\hat{\Gamma}$ *Monderer & Shapley (1996) with* $d(\Gamma, \hat{\Gamma}) \leq \delta$ *for some* $\delta > 0$.

Agents observe others only through a random time-varying network $\mathcal{G}_t = (\mathcal{N}, \mathcal{E}_t)$. We allow communication to become increasingly sparse: Assumption 3 only requires persistent information propagation over finite blocks, with conditional success probabilities that may decay over time as long as their cumulative sum diverges. Thus, neither instantaneous connectivity nor uniformly positive communication rates are required. Under this network model, we develop a decentralized local-search algorithm for near-potential games and establish convergence to approximate Nash equilibria using low-cost decision updates.

4 Decentralized Learning in Binary Games

In the learning mechanism, each agent i maintains a fading-memory empirical frequency vector $\mathbf{f}_i(t) \in [0,1]^{K_i}$, as a discounted average of its own past binary decisions with fading parameter $\alpha \in (0,1)$,

$$\mathbf{f}_i(t) = (1-\alpha)\mathbf{f}_i(t-1) + \alpha \mathbf{x}_i(t), \tag{3}$$

where the initial empirical frequency is set to $\mathbf{f}_i(0) = \mathbf{x}_i(0)$, and $\mathbf{x}_i(t) \in \mathcal{X}_i$ denotes the binary decision vector chosen by agent i at time t. Similarly, each agent keeps the local copy for the empirical frequency of other agents $j \in \mathcal{N} \setminus \{i\}$ via consensus updates,

$$\hat{\mathbf{f}}_j^i(t+1) = \sum_{n \in \mathcal{N}_i(t)} w_{j,n}^i(t) \left(\hat{\mathbf{f}}_j^n(t) + (\mathbf{f}_j(t+1) - \mathbf{f}_j(t)) \chi_{\{n=j\}} \right), \tag{4}$$

where $\chi_{\{k=j\}} = 1$ if $n = j$ and 0 otherwise. The term $(\mathbf{f}_{j,t+1} - \mathbf{f}_{j,t})\chi_{\{k=j\}}$ injects the latest update into the consensus chain at the source node j, and propagates outward over time, where we define $\mathcal{N}_i(t) := \{n \in \mathcal{N} : (n,i) \in \mathcal{E}_t\} \cup \{i\}$.

4.1 Local Search with Inertia

DLS couples randomized constrained binary local search with inertial updates and distributed empirical-frequency estimation. At each time t, agent i retains its previous decision with probability $\epsilon_{\text{inertia}}$ and otherwise searches for an improving feasible response using its local estimates $\hat{\mathbf{f}}_{-i}^{\,i}(t)$. The inner search uses trace-based candidate selection introduced by Bertsimas et al. (2013), while randomized exploration supports the improvement-path reachability used in our equilibrium analysis. Candidates are evaluated using $c_i(\mathbf{x}_i, \hat{\mathbf{f}}_{-i}^{\,i}(t))$, without requiring an exact best-response solve. Our analysis establishes convergence of the resulting decentralized dynamics under the stated near-potential, communication, and reachability assumptions. Algorithm 1 explores the neighbors of solutions in a list SL, updating the incumbent on any feasible improvement, and otherwise admitting both feasible and infeasible neighbors based on the following principles:

Trace-based bookkeeping. For any $\mathbf{y}_i \in \{0,1\}^{K_i}$, define its violation and slack vectors as

$$\mathbf{y}_i^v := \max\{\mathbf{A}_i \mathbf{y}_i - \mathbf{b}_i, 0\}, \qquad \mathbf{y}_i^w := \mathbb{1}_{\{\mathbf{A}_i \mathbf{y}_i < \mathbf{b}_i\}},$$

and its constraint trace as $\tau(\mathbf{y}_i) := [\mathbf{y}_i^v; \mathbf{y}_i^w]$.e Thus, $\tau(\mathbf{y}_i)$ summarizes the relation of $\mathbf{y}_i$ to the constraints independently of the particular active variables.

Interesting candidates. Let $\mathbf{z}_i$ be the incumbent feasible action and let $\mathbf{y}_i$ be a one-bit-flip neighbor of the current expansion head. We call $\mathbf{y}_i$ *interesting* if

(A1) $\mathbf{y}_i^v \in \{0,1\}^{m_i}$;

Algorithm 1 Randomized Local-Search for Agent i

Require: Cost $\mathbf{c}_i(.)$, local empirical frequency estimates $\hat{\mathbf{f}}^i_{-i}(t)$, constraint data $(\mathbf{A}_i, \mathbf{b}_i)$, current incumbent $\mathbf{z}_i$, budget cap $\bar{Q}$, geometric parameter λ, exploration probability $\epsilon_{\text{exp}} \in (0, 1)$.

```
Draw Q_t ~ Geometric(λ) truncated at Q̄
Initialize solution list SL ← {z_i}, trace boxes TB[·] ← +∞
while SL ≠ ∅ do
    Pop expansion head x_i ← first element of SL
    for each y_i ∈ {0,1}^{K_i} with ‖y_i − x_i‖_1 = 1 do
        if y_i ∈ X_i and c_i(y_i, f̂^i_{-i}(t)) < c_i(z_i, f̂^i_{-i}(t)) then
            z_i ← y_i
            Reset SL ← {y_i}, TB[·] ← +∞
            break inner loop
        else if y_i satisfies (A1)-(A3) or Bernoulli(ε_exp) = 1 then
            TB[h(τ(y_i))] ← c_i(y_i, f̂^i_{-i}(t))
            Append y_i to SL
        end if
    end for
end while
return z_i
```

(A2) $\|\tau(\mathbf{y}_i) - \tau(\mathbf{z}_i)\|_1 \le Q_t$;

(A3) $\mathbf{y}_i$ has lower cost than every previously examined candidate assigned to the same trace box, i.e.,$c_i(\mathbf{y}_i, \hat{\mathbf{f}}^i_{-i}(t)) < c_i(\mathbf{x}_i, \hat{\mathbf{f}}^i_{-i}(t))$, for every previously examined $\mathbf{x}_i$ satisfying $h(\tau(\mathbf{x}_i)) = h(\tau(\mathbf{y}_i))$, where h maps traces to one of N_{TB} trace boxes.

Conditions (A1)–(A2) allow controlled exploration through lightly infeasible configurations, while (A3) prunes dominated candidates and uses hashing to limit memory requirements.

Unlike the deterministic local search of Bertsimas et al. (2013), we randomize the trace budget Q_t and introduce exploration probability $\epsilon_{\text{exp}} > 0$. Hence, candidates excluded by (A1)–(A3), including infeasible intermediate points, can still enter the search list with positive probability. The resulting decentralized procedure is summarized below.

Algorithm 2 DLS for Agent i

Input: Local estimates $\mathbf{f}^i_{-i}(0)$ and time-varying networks $\mathcal{G}_t = (\mathcal{N}, \mathcal{E}_t)$.

for $t = 1, 2, \cdots$ **do**

Each agent i follows the inertial local search,

$$\mathbf{x}_i(t) = \begin{cases} \mathbf{x}_i(t-1), & \text{w. p. } \epsilon_{\text{inertia}}, \\ \mathbf{z}_i \text{ (Algorithm 1)}, & \text{otherwise.} \end{cases}$$

Update $\mathbf{f}_i(t)$ (Eq. (3)) and local copies $\mathbf{f}^i_{-i}(t)$ (Eq. (4)).

end for

5 Convergence Analysis

We start by stating the structural assumptions: (i) connectivity condition ensuring that the time-varying communication network has persistent positive probability of connection, and (ii) a reachability condition ensuring that an improving response is attainable through paths of bounded infeasibility at worst.

Assumption 1 (Admissible-Path Reachability)**.** *For every $i \in \mathcal{N}$, $\mathbf{x}_{-i} \in \mathcal{X}_{-i}$, and $\mathbf{z}_i \in \mathcal{X}_i$ not being a best response, $\mathbf{z}_i \notin \operatorname{argmin}_{\mathbf{x}_i \in \mathcal{X}_i} c_i(\mathbf{x}_i, \mathbf{x}_{-i})$, there exist $\mathbf{y}^\star_i \in \mathcal{X}_i$ with $c_i(\mathbf{y}^\star_i, \mathbf{x}_{-i}) < c_i(\mathbf{z}_i, \mathbf{x}_{-i})$ and a finite sequence $\mathbf{z}_i, \mathbf{v}^1, \ldots, \mathbf{v}^{(L-1)}, \mathbf{y}^\star_i \in \{0,1\}^{K_i}$ of bounded length $L \le L_{\max}$ satisfying $\|\mathbf{v}^{(\ell)} - \mathbf{v}^{(\ell-1)}\|_1 = 1$ for each ℓ, $\|\mathbf{y}^\star_i - \mathbf{v}^{(L-1)}\|_1 = 1$, and $(\mathbf{A}_i \mathbf{v}^{(\ell)} - \mathbf{b}_i)_j \le 1$ for all $j \in \{1, \ldots, m_i\}$ and $\ell \in \{0, \ldots, L-1\}$.*

Assumption 2 (Consensus weights)**.** *For every $j \in \mathcal{N}$ and $t \geq 0$, let the communication matrix for agent j's beliefs be defined as $W_j(t) = \left[w^i_{j,n}(t)\right]_{i,n\in\mathcal{N}}$. The weights satisfy the following conditions, (i) $W_j(t)$ is row-stochastic, i.e.,$W_j(t)\mathbf{1} = \mathbf{1}$, (ii) The source agent preserves its own empirical frequency, and every available estimate receives a uniformly positive weight. Specifically, there exists $\eta \in (0,1)$ such that $e_j^\top W_j(t) = e_j^\top, \qquad w^i_{j,n}(t) \geq \eta$ for $n \in \mathcal{N}_i(t) \cup \{i\}$, and $w^i_{j,n}(t) = 0$ for $n \notin \mathcal{N}_i(t) \cup \{i\}$.*

Assumption 3 (Probabilistic information propagation)**.** *Fix an integer $B \geq 1$ and we define the block endpoints $t_m = t_0 + mB,\ m \geq 0$. For each $j \in \mathcal{N}$, we define the consensus matrix product over block m as, $\Psi_{j,m} = \overleftarrow{\prod_{\tau=t_m}^{t_m+B-1}} W_j(\tau)$. There exists a constant $\underline{\beta}_j \in (0,1]$ and a deterministic sequence $\{p_{j,m}\}_{m\geq 0} \subset (0,1]$ such that*

$$\mathbb{P}\left(\min_{i\in\mathcal{N}}[\Psi_{j,m}]_{ij} \geq \underline{\beta}_j \,\middle|\, \mathcal{F}_{t_m}\right) \geq p_{j,m} \qquad \text{a.s.} \tag{5}$$

Moreover, the block-success probabilities satisfy the persistence condition,$\sum_{m=0}^{\infty} p_{j,m} = \infty$.

In contrast to classical consensus and distributed-optimization analyses, which commonly impose uniformly positive mixing weights together with bounded intercommunication intervals or uniform joint connectivity (Tsitsiklis et al., 1986; Nedić & Ozdaglar, 2009; Nedić & Olshevsky, 2015; Moreau, 2005), our condition only requires conditional block-success probabilities $p_{j,m}$ satisfying $\sum_m p_{j,m} = \infty$. Hence, these probabilities may decay over time and the block-success events need not be independent or identically distributed, unlike common random-network models based on temporally i.i.d. link failures (Lobel & Ozdaglar, 2011). The row-stochasticity and positive-weight conditions follow standard consensus constructions (Touri & Nedić, 2014), while the source-preserving condition anchors the true empirical frequency at its originating agent (Eksin & Ribeiro, 2018; Swenson et al., 2018; Arefizadeh & Eksin, 2019; Aydın & Eksin, 2022).

Assumption 4 (Strict best-response margin within the approximate equilibrium set)**.** *For every $\mathbf{x} \in \mathcal{X}_{\delta h} \setminus \mathcal{X}_0$, there exists an agent $i \in \mathcal{N}$ and reachable decision $\mathbf{x}_i$ such that*

$$c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \min_{\mathbf{y}_i \in \mathcal{X}_i} c_i(\mathbf{y}_i, \mathbf{x}_{-i}) > \delta. \tag{6}$$

Theorem 1 (Almost-sure convergence to a pure Nash equilibrium)**.** *Suppose Assumptions 1–3 and 4 hold, and Γ is δ-near an exact potential game. Let $\mathcal{X}_0$ denote the set of pure Nash equilibria. Suppose further that every $\mathbf{x}^\star \in \mathcal{X}_0$ is strict, i.e., for every $i \in \mathcal{N}$ and every $\mathbf{y}_i \in \mathcal{X}_i \setminus \{\mathbf{x}^\star_i\}$,*

$$c_i(\mathbf{x}^\star_i, \mathbf{x}^\star_{-i}) < c_i(\mathbf{y}_i, \mathbf{x}^\star_{-i}). \tag{7}$$

Then there exist an almost surely finite random time τ and a pure Nash equilibrium $\mathbf{x}^\star \in \mathcal{X}_0$ such that

$$\mathbf{x}(t) = \mathbf{x}^\star, \qquad \text{for all } t \geq \tau, \quad \text{a.s.} \tag{8}$$

To our knowledge, this is the first decentralized better-response mechanism for near-potential games preserving the convergence guarantees of perfect-information scenarios Candogan et al. (2013a). Unlike the another local-search study Li et al. (2026), our method incorporates random communication networks, generalizes beyond exact potential games, and relaxes the feasible-connectivity assumption by allowing exploration through infeasible decisions.

6 Numerical Experiments

We evaluate DLS (Algorithm 2 with Algorithm 1) on eight binary programming games P1–P8 that instantiate (1) with increasingly demanding feasible sets and costs: a sensor-coverage game, its constrained and exact-cardinality variants, two games with asymmetric interactions, a game whose centralized problem is cubic, the autonomous vehicle–target assignment game of Arslan et al. (2007), and a game whose per-agent problem is a large quadratic program; a ninth, a routing game with player-specific beliefs (P9), is reported in Appendix E. Every configuration is run with 50 random seeds (initial decisions, inertia, exploration and link activations). The agents communicate over a directed Erdős–Rényi graph $G(N, 0.6)$ redrawn at every iteration and use uniform weights in the consensus update (4), i.e., each agent averages its copies with those of its active in-neighbors. Other

values $p \geq 0.05$ give essentially the same equilibria and social cost, only reached at a different speed (Appendix D). All other parameters are the same for all problems and instances: $\alpha = 0.2$, $\epsilon_{\text{inertia}} = 1 - 1/N$, $\lambda = 0.5$, $\bar{Q} = 5$, $\epsilon_{\text{exp}} = 0.15$, a cap of 60 expansions of SL per call of Algorithm 1, and horizon $T = 500$ for $N \leq 8$ and $T = 60N$ otherwise. Our goal is to test whether DLS reaches (approximate) Nash equilibria under decentralized information as Theorem 1 predicts, at what cost in welfare, and how it compares with simpler and with solver-based decentralized rules; centralized solutions are reported only as references.

Table 1: Results on P1–P7: social cost (mean $\pm$ std over 50 seeds) and number of seeds (out of 50) ending at an exact Nash equilibrium (DLS/FLS/EBR). On P7, WLU and ESU are the aligned and misaligned vehicle utilities. *: the centralized solver reached its time limit (600 s; 300 s for P5 at $N = 100$, where the optimality gap is 183%) and the best solution found is reported; –: no optimum computable. Mean Nash gaps for P4 are in Appendix C.

instance	greedy	optimum	DLS	FLS	EBR	exact NE
P1: sensor coverage						
N=4	-28.881	-28.920	-27.620 ± 0.903	-27.634 ± 0.891	-27.634 ± 0.891	50/50/50
N=6	-31.176	-31.241	-24.801 ± 0.364	-24.694 ± 0.418	-24.694 ± 0.418	50/50/50
N=8	-33.084	-33.265	-20.878 ± 0.941	-20.803 ± 0.962	-20.803 ± 0.962	50/50/50
N=32	-95.003	-95.003	-38.940 ± 2.383	-38.858 ± 2.645	-38.858 ± 2.645	50/50/50
N=100	-130.641	-130.641	-62.203 ± 1.340	-62.250 ± 1.164	-62.250 ± 1.164	50/50/50
P2: sensor coverage with energy budget and region caps						
N=4	-24.107	-26.516	-26.251 ± 0.000	-24.679 ± 1.501	-26.251 ± 0.000	50/22/50
N=6	-31.128	-31.194	-24.514 ± 0.733	-24.780 ± 0.755	-24.464 ± 0.748	50/46/50
N=8	-32.889	-33.018	-22.112 ± 1.564	-22.487 ± 1.510	-22.260 ± 1.446	50/50/50
N=32	-94.815	-94.860	-41.020 ± 2.361	-41.660 ± 1.797	-41.491 ± 2.298	50/50/50
N=100	-130.641	-130.641	-61.105 ± 1.151	-61.275 ± 1.187	-61.281 ± 1.185	50/50/50
P3: exact-cardinality assignment						
N=4	-28.881	-28.920	-27.850 ± 0.823	-27.620 ± 0.856	-27.620 ± 0.856	50/50/50
N=6	-27.436	-27.806	-24.786 ± 0.369	-24.748 ± 0.370	-24.748 ± 0.370	50/50/50
N=8	-16.371	-16.567	-11.515 ± 0.606	-11.483 ± 0.540	-11.483 ± 0.540	50/50/50
N=32	118.302	117.504	130.915 ± 1.331	130.750 ± 1.211	130.750 ± 1.211	50/50/50
N=100	3939.565	3938.403	3986.677 ± 2.603	3985.478 ± 2.683	3985.423 ± 2.703	11/21/20
P4: competing project selection, $\eta = 0.5$						
N=8	-182.381	-233.241	-223.772 ± 0.253	-214.979 ± 5.612	-223.704 ± 0.211	50/0/50
N=32	-504.482	-532.299	-460.595 ± 6.950	-479.662 ± 8.251	-464.656 ± 6.765	0/0/0
N=100	-688.207	-689.544	-230.826 ± 27.528	-222.970 ± 31.397	-229.014 ± 28.417	0/0/0
P5: wireless channel allocation						
N=8	-115.477	-139.656	-137.600 ± 1.104	-129.821 ± 4.619	-137.548 ± 0.862	50/0/50
N=32	-320.484	-343.165	-285.532 ± 5.966	-285.501 ± 6.807	-284.739 ± 6.025	50/0/50
N=100	-452.434	-474.050^*	-285.825 ± 9.215	-283.283 ± 8.705	-284.689 ± 8.570	50/13/50
P6: cubic interference						
N=8	-96.610	-100.228	-98.954 ± 0.801	-98.784 ± 0.768	-98.784 ± 0.768	50/50/50
N=16	-168.390	-172.431	-165.990 ± 1.616	-165.567 ± 2.214	-165.567 ± 2.214	50/50/50
N=32	-218.598	-221.832^*	-166.540 ± 5.562	-164.880 ± 7.268	-164.880 ± 7.268	50/50/50
N=64	-235.408	–	-146.999 ± 9.279	-149.319 ± 10.389	-149.319 ± 10.389	50/50/50
N=100	-256.977	–	-149.116 ± 6.892	-147.206 ± 8.056	-147.206 ± 8.056	50/50/50
P7: vehicle–target assignment (social cost $-U_g$)						
WLU N=20	-8.956	-8.961	-8.938 ± 0.017	-8.937 ± 0.015	-8.937 ± 0.015	50/50/50
WLU N=50	-26.457	-26.496	-26.393 ± 0.032	-26.396 ± 0.041	-26.397 ± 0.040	50/50/50
WLU N=100	-49.417	-49.499^*	-49.224 ± 0.048	-49.224 ± 0.049	-49.224 ± 0.047	50/49/50
ESU N=20	-8.956	-8.961	-8.334 ± 0.113	-8.332 ± 0.090	-8.332 ± 0.090	47/48/48
ESU N=50	-26.457	-26.496	-24.568 ± 0.193	-24.513 ± 0.204	-24.522 ± 0.203	50/50/50
ESU N=100	-49.417	-49.499^*	-46.061 ± 0.247	-46.048 ± 0.287	-46.028 ± 0.285	50/50/50

6.1 PROBLEMS

In all problems $x_{ik} \in \{0, 1\}$ indicates that agent i selects item k (a target, project, channel or vehicle target) and $n_k = \sum_i x_{ik}$ is the load of item k; Appendix B gives the full definitions, parameters and instance sizes. Each problem tests a different difficulty. **P1** (sensor coverage): sensors select up to B targets and pay for congestion; an exact potential game with a cardinality constraint, the base case. **P2**: P1 with an energy budget (a knapsack row) and a cap on the targets in each region. **P3**: P1 with exactly B targets, so that no single bit flip from a feasible decision is feasible and every improvement passes through an infeasible decision (Assumption 1). **P4** (competing project selection): knapsack budgets and an antisymmetric competition term that makes the game near-potential ($\eta = 0.5$). **P5** (wireless channel allocation): power budgets and asymmetric interference; a near-potential game

Table 2: Sensor coverage with own-target interactions, $K = 500$ (P8): no centralized optimum is computable and NE gaps are empirical (best improving deviation found). EBR is run with the time limit per update given in the column header.

		greedy	DLS	DLS	FLS	EBR, time limit per update		
		(centralized)	(cap 60)	(cap 300)		0.2 s	0.5 s	1 s
$N = 8$	social cost	-1707.4	-1747.5	-1748.5	-1742.0	-567.7	-1802.8	-1809.8
	std over seeds	–	14.0	16.8	20.5	42.7	10.4	9.2
	NE gap	–	13.7	13.2	13.8	179.9	3.9	2.5
$N = 32$	social cost	-5887.7	-6084.8	-6091.8	-6058.4	-1357.9	-6354.8	-6375.2
	std over seeds	–	35.0	31.6	29.2	117.3	26.1	22.3
	NE gap	–	13.6	13.5	14.7	201.1	4.2	2.9

whose centralized problem is a non-convex binary quadratic program. **P6** (cubic interference): a three-way interference term makes the centralized problem cubic, and its exact linearization cannot be built for $N \geq 64$. **P7** (vehicle–target assignment): the game of Arslan et al. (2007) with ranges, capacities and fuel budgets, under an aligned (WLU) and a misaligned (ESU) utility. **P8** (own-target interactions): P1 with $K = 500$ targets and a dense quadratic term between a sensor's own targets, so each best response is a binary quadratic program too large to solve exactly at every update.

6.2 Baselines

We compare DLS with two decentralized rules and two centralized references. The *decentralized* rules run inside Algorithm 2 with the same inertia, estimates and random network as DLS and differ only in how an agent updates its decision from $c_i(\cdot, \hat{\mathbf{f}}^i_{-i}(t))$: **FLS**, a feasible-only local search that performs a steepest descent over feasible single additions, removals and swaps of items and never evaluates an infeasible decision (no trace tracking or random exploration); and **EBR**, which computes an exact best response $\arg\min_{\mathbf{y}_i \in \mathcal{X}_i} c_i(\mathbf{y}_i, \hat{\mathbf{f}}^i_{-i})$ at each update with Gurobi 12 on one thread. EBR gives the lowest estimated cost of a single update but needs a solver on every agent, which resource-constrained devices such as small UAVs, sensors or radios may be unable to run. On P8, where each best response is a large quadratic program, we also report EBR-s, the same rule with a time limit of $s \in \{0.2, 0.5, 1\}$ seconds per call, warm-started from the current decision, and DLS-300, Algorithm 1 with an expansion cap of 300 instead of 60. The *centralized* references assume a planner that knows all costs and constraints and sets all decisions at once; they show the welfare such a planner can achieve, which Nash equilibria of near-potential games need not reach (the price of anarchy). **Greedy** starts from $\mathbf{x} = \mathbf{0}$ and repeatedly switches on the feasible (i, k) pair with the largest decrease of the social cost; **optimum** is the social-cost minimizer computed by Gurobi with the formulation of Appendix B (one thread, warm-started from greedy, 600 s limit, after which the best solution found is reported).

6.3 Evaluation

For every run we record the social cost $\sum_i c_i(\mathbf{x}(T))$ of the final decisions (mean and standard deviation over the 50 seeds) and the Nash gap $g(\mathbf{x}) = \max_{i \in \mathcal{N}} \Big[c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \min_{\mathbf{y}_i \in \mathcal{X}_i} c_i(\mathbf{y}_i, \mathbf{x}_{-i}) \Big]$, the largest cost reduction any agent can obtain by a unilateral feasible deviation, computed from exact best responses to the other agents' true final decisions. A profile $\mathbf{x}$ is an ϵ-Nash equilibrium if $g(\mathbf{x}) \leq \epsilon$, and we count a run as reaching an exact equilibrium if $g(\mathbf{x}(T)) < 10^{-9}$; entries such as $50/50/50$ in the tables give these counts for DLS/FLS/EBR. On P8 exact best responses are too expensive, so we estimate the gap by the largest improvement found by a 2 s solver run warm-started from the current decision or by a full feasible local descent; this estimate may be too low and cannot certify an equilibrium. Equilibrium attainment is our main criterion, since Theorem 1 concerns convergence to approximate equilibria; the social cost measures collective performance, and a low social cost does not imply that no agent can improve.

6.4 RESULTS

Tables 1 and 2 report the results. (The social-cost trajectories behind the tables are plotted in Appendix F.) On problems with tractable best responses (P1–P7), DLS reaches an exact Nash equilibrium in all 50 seeds at every tested size up to $N = 100$, except in the following three cases: on P3 at $N = 100$ none of the rules has fully settled at $T = 6000$ (gaps below 0.15 on costs of about 4000), on P4 with $\eta = 0.5$ and $N \geq 32$ all rules reach approximate equilibria only (mean gaps 1.5 and 8.6), which is the near-potential regime addressed by Theorem 1, and on P7 under the misaligned ESU three seeds at $N = 20$ end one deviation short. Wherever EBR is exact, DLS reaches the same equilibria at the same cost: the randomized local search finds the best response without a solver. FLS matches EBR on P1, P3 and P6, where each agent minimizes a linear cost under a cardinality constraint, so any decision that no single addition, removal or swap improves is already a best response. Under knapsack-type budgets this fails and FLS stops at decisions that are not best responses: it reaches an equilibrium in 22/50 and 46/50 seeds at $N = 4$ and $N = 6$ on P2, 0/50 on P4 at $N = 8$ (gap 7.9) and 0/50, 0/50, 13/50 on P5, against 50/50 for DLS. Its social cost can still be close to, or below, that of DLS because the agents do not minimize the social cost. The welfare of the equilibria is far from the centralized optimum on the larger instances (52% above on P1 at $N = 100$, 67% on P4 at $N = 100$, and on P5 at $N = 100$ 40% above the best solution the solver found, which is far from proved optimal). On P6, the centralized solver certifies globally optimal solutions at $N = 8$ and $N = 16$, but reaches the 600 s time limit at $N = 32$ with an 11.9% optimality gap. For $N \geq 64$, the centralized model is too large to construct using Gurobi within the available computational resources, while all decentralized rules reach exact Nash equilibria in all 50 runs even at $N = 100$. On P7 the aligned WLU yields equilibria within 0.3–0.6% of the optimum at every size, whereas the misaligned ESU still converges but loses 6.9–7.3% of the global utility. This reproduces the utility-design result of Arslan et al. (2007) with richer constraints and without any coordinator. On P8, where a DLS update takes about 0.1 s on one core, a solver limited to 0.2 s per update (EBR-0.2) fails outright, DLS reaches 95–97% of the welfare of EBR-1 (1 s limit per update) and improves on the centralized greedy by 2–3%, while from 0.5 s per update the solver is near-optimal; a five-fold larger local search (DLS-300) does not close that gap, and the empirical gaps (≈ 13 against ≈ 3 for EBR-1) show that local search ends a few percent short of a best response at this size. The largest configuration (P7, $N = 100$) took about 21 minutes for DLS and about 4 minutes each for FLS and EBR (50 seeds each); the consensus step costs $O(N^3K)$ per iteration because every agent keeps a copy of every other agent's frequency vector.

Communication connectivity. Theorem 1 gives no convergence rate, so we also vary the network on P1 and P5 at $N = 32$ (Appendix D). With any constant activation probability $p \geq 0.05$, 49–50 of 50 seeds reach an equilibrium with the same social cost as at $p = 0.6$, only more slowly (median 1,248 instead of 864 steps on P1 at $p = 0.05$). No seed reaches an equilibrium when the base graph is disconnected, and vanishing probabilities $p(t) \propto (1+t/100)^{-\gamma}$ behave like a constant p for $\gamma \leq 1$ but freeze the estimates for $\gamma = 2$ (1 and 29 of 50 seeds at an equilibrium on P1 and P5), consistent with the condition $\sum_m p_{j,m} = \infty$ of Assumption 3.

7 CONCLUSION

We proposed Decentralized Local Search (DLS), in which each agent of a binary programming game replaces the exact best response by a randomized local search that may pass through slightly infeasible decisions, and learns the other agents' decisions only through consensus over a random, possibly sparse network. We proved that the joint decisions reach the approximate Nash equilibrium set of a near-potential game in finite time almost surely, under a propagation condition that allows the communication probabilities to vanish. On nine games with up to 100 agents, DLS reaches, on the instances whose best response is tractable, the equilibria and social costs of exact best-response dynamics without a solver; it reaches equilibria where a feasible-only local search stalls under knapsack-type budgets, and it scales to games whose centralized model cannot be built.

AI USE STATEMENT

Claude Fable (Anthropic) and ChatGPT (OpenAI) were used to paraphrase, reformat and refine the language of this paper. We did not use generative AI tools for the research idea, the design of

the algorithm or the proofs. The authors reviewed and edited all AI-assisted content and take full responsibility for this article.

REFERENCES

Sina Arefizadeh and Ceyhun Eksin. Distributed fictitious play in potential games with time varying communication networks. In *2019 53rd Asilomar conference on signals, systems, and computers*, pp. 1755–1759. IEEE, 2019.

Gürdal Arslan, Jason R. Marden, and Jeff S. Shamma. Autonomous vehicle-target assignment: A game-theoretical formulation. *Journal of Dynamic Systems, Measurement, and Control*, 129(5): 584–596, 2007.

Sarper Aydın and Ceyhun Eksin. Decentralized inertial best-response with voluntary and limited communication in random communication networks. *Automatica*, 145:110566, 2022.

Sarper Aydın, Sina Arefizadeh, and Ceyhun Eksin. Decentralized fictitious play in near-potential games with time-varying communication networks. *IEEE Control Systems Letters*, 6:1226–1231, 2021.

David Bergman, Andre A Cire, Willem-Jan van Hoeve, and Tallys Yunes. Bdd-based heuristics for binary optimization. *Journal of Heuristics*, 20(2):211–234, 2014.

Dimitris Bertsimas, Dan A Iancu, and Dmitriy Katz. A new local search algorithm for binary optimization. *INFORMS Journal on Computing*, 25(2):208–221, 2013.

Ozan Candogan, Asuman Ozdaglar, and Pablo A Parrilo. Dynamics in near-potential games. *Games and Economic Behavior*, 82:66–90, 2013a.

Ozan Candogan, Asuman Ozdaglar, and Pablo A Parrilo. Near-potential games: Geometry and dynamics. *ACM Transactions on Economics and Computation (TEAC)*, 1(2):1–32, 2013b.

Margarida Carvalho, Andrea Lodi, and João Pedro Pedroso. Computing equilibria for integer programming games. *European Journal of Operational Research*, 303(3):1057–1070, 2022.

Margarida Carvalho, Gabriele Dragotto, Andrea Lodi, and Sriram Sankaranarayanan. Integer programming games. *Foundations and Trends in Optimization*, 7(4):264–391, 2025.

Margarida Carvalho, Gabriele Dragotto, Andrea Lodi, and Sriram Sankaranarayanan. The cut-and-play algorithm: computing nash equilibria via outer approximations. *Operations Research*, 74 (2):1070–1086, 2026.

Tapas Kumar Chandra. *The Borel-Cantelli Lemma*. Springer, 2012.

Constantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.

Gabriele Dragotto and Rosario Scatamacchia. The zero regrets algorithm: Optimizing over pure nash equilibria via integer programming. *INFORMS Journal on Computing*, 35(5):1143–1160, 2023.

Aloïs Duguet, Margarida Carvalho, Gabriele Dragotto, and Sandra Ulrich Ngueveu. Computing approximate nash equilibria for integer programming games. *Optimization Letters*, 20(1):231–255, 2026.

Ceyhun Eksin and Alejandro Ribeiro. Distributed fictitious play for multiagent systems in uncertain environments. *IEEE Transactions on Automatic Control*, 63(4):1177–1184, 2018. doi: 10.1109/TAC.2017.2747767.

Yuan Fan, Gang Feng, Yong Wang, and Jianbin Qiu. A novel approach to coordination of multiple robots with communication failures via proximity graph. *Automatica*, 47(8):1800–1805, 2011.

Paola Festa and Mauricio GC Resende. Grasp: An annotated bibliography. In *Essays and surveys in metaheuristics*, pp. 325–367. Springer, 2002.

Drew Fudenberg and David K. Levine. *The Theory of Learning in Games*. MIT Press, 1998.

Hideaki Ishii, Yuan Wang, and Shuai Feng. An overview on multi-agent consensus under adversarial attacks. *Annual Reviews in Control*, 53:252–272, 2022.

Tatsuya Iwase, Aurélie Beynier, Nicolas Bredeche, Nicolas Maudet, and Jason R Marden. Is limited information enough? an approximate multi-agent coverage control in non-convex discrete environments. *arXiv preprint arXiv:2401.03752*, 2024.

Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2):451–462, 2009.

Shotaro Kamiya, Koji Yamamoto, Takayuki Nishio, Masahiro Morikura, and Tomoyuki Sugihara. Spatial co-channel overlap mitigation through channel assignment in dense wlan: Potential game approach. *IEICE Transactions on Communications*, E100.B(7):1094–1104, 2017.

Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pp. 85–103. Plenum Press, 1972.

Hyunwoo Lee, Robert Hildebrand, Wenbo Cai, and İ Esra Büyüktahtakın. Random-restart best-response dynamics for large-scale integer programming games and their applications. *arXiv preprint arXiv:2409.04078*, 2024.

Zhe Li, Changdi Liu, Shaolin Tan, and Wei Wang. Lightweight log-linear learning with neighborhood search for equilibrium selection in finite potential games. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2026.

Ilan Lobel and Asuman Ozdaglar. Distributed subgradient methods for convex optimization over random networks. *IEEE Transactions on Automatic Control*, 56(6):1291–1306, 2011.

Arne Lokketangen and Fred Glover. Solving zero-one mixed integer programming problems using tabu search. *European journal of operational research*, 106(2-3):624–658, 1998.

Jason R Marden and Jeff S Shamma. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games and Economic Behavior*, 75(2):788–808, 2012.

Igal Milchtaich. Congestion games with player-specific payoff functions. *Games and Economic Behavior*, 13(1):111–124, 1996.

Nenad Mladenović and Pierre Hansen. Variable neighborhood search. *Computers & operations research*, 24(11):1097–1100, 1997.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.

Luc Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005.

Angelia Nedić and Alex Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 60(3):601–615, 2015.

Angelia Nedić and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

Dario Paccagnan and Jason R Marden. Utility design for distributed resource allocation—part ii: Applications to submodular, covering, and supermodular problems. *IEEE Transactions on Automatic Control*, 67(2):618–632, 2021.

Tim Roughgarden and Éva Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2):236–259, 2002.

Guodong Shi, Mikael Johansson, and Karl Henrik Johansson. Randomized gossiping with unreliable communication: Dependent or independent node updates. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pp. 4846–4851. IEEE, 2012.

Guodong Shi, Brian DO Anderson, and Karl Henrik Johansson. Consensus over random graph processes: Network borel–cantelli lemmas for almost sure convergence. *IEEE Transactions on Information Theory*, 61(10):5690–5707, 2015.

Brian Swenson, Ceyhun Eksin, Soummya Kar, and Alejandro Ribeiro. Distributed inertial best-response dynamics. *IEEE Transactions on Automatic Control*, 63(12):4294–4300, 2018.

Behrouz Touri and Angelia Nedic. On ergodicity, infinite flow, and consensus in random models. *IEEE Transactions on Automatic Control*, 56(7):1593–1605, 2010.

Behrouz Touri and Angelia Nedić. Product of random stochastic matrices. *IEEE Transactions on Automatic Control*, 59(2):437–448, 2014.

Kai-Yun Tsao, Thomas Girdler, and Vassilios G. Vassilakis. A survey of cyber security threats and solutions for uav communications and flying ad-hoc networks. *Ad Hoc Networks*, 133:102894, 2022.

John N. Tsitsiklis, Dimitri P. Bertsekas, and Michael Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9): 803–812, 1986.

Shunji Umetani. Exploiting variable associations to configure efficient local search algorithms in large-scale binary integer programs. *European Journal of Operational Research*, 263(1):72–81, 2017.

Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

H. Peyton Young. *Strategic Learning and Its Limits*. Oxford University Press, 2004.

Minghui Zhu and Sonia Martínez. Distributed coverage games for energy-aware mobile sensor networks. *SIAM Journal on Control and Optimization*, 51(1):1–27, 2013.

A Proofs and Analysis

A.1 Preliminaries

For any matrix A, let $[A]_{i,n}$ denote its (i,n)th entry, i.e., the entry in row i and column n. For any matrix-valued sequence $\{A(t)\}$ and integers $a \le b$, define the backward-ordered product by

$$\prod_{\tau=a}^{\overleftarrow{b}} A(\tau) := A(b)A(b-1)\cdots A(a),$$

with the convention $\prod_{\tau=a}^{\overleftarrow{b}} A(\tau) := I$ whenever $a > b$.

Let $\{\mathcal{F}_t\}_{t\ge 0}$ denote the natural filtration generated by the history of the process up to time t, i.e.,

$$\mathcal{F}_t := \sigma\Big(\Big\{\mathbf{x}(s), \mathbf{f}(s), \widehat{\mathbf{f}}(s), \mathcal{E}_s, \{W_j(s)\}_{j\in\mathcal{N}}, \{I_i(s)\}_{i\in\mathcal{N}}, \\ \{Q_i(s)\}_{i\in\mathcal{N}}, \{E_{i,\ell}(s)\}_{i\in\mathcal{N},\ell\ge 1} :\ 0 \le s \le t\Big\}\Big).$$

where $I_i(s) \in \{0,1\}$ denotes the inertia draw of agent i, $Q_i(s)$ denotes its randomized local-search budget, and $E_{i,\ell}(s) \in \{0,1\}$ denotes the ℓth Bernoulli exploration draw used by agent i during its local search at time s. Thus, $\mathcal{F}_{t_m}$ represents the complete history available at the beginning of block m. Thus, $\mathcal{F}_{t_m}$ contains the complete history available at the beginning of block m.

A.2 Intermediate Results

Lemma 1 (Finite-time tracking under vanishing communication probabilities)**.** *Suppose Assumptions 2–3 hold. Fix $j \in \mathcal{N}$, an initial block $m_0 \ge 0$, and an action $\mathbf{x}_j^\star \in \mathcal{X}_j$. Suppose that $\mathbf{x}_j(t) = \mathbf{x}_j^\star$ for time steps $t = t_{m_0}, \dots, t_{m_0+MB}$, and that the empirical frequencies and their local estimates evolve according to* (3) *and* (4)*, respectively. Then, for every $\varepsilon > 0$ and $\delta \in (0,1)$, there exists a finite integer $M_{\varepsilon,\delta} = M_{\varepsilon,\delta}(j, m_0)$ such that, for every $M \ge M_{\varepsilon,\delta}$ satisfying the following,*

$$\mathbb{P}\left(\max_{i\in\mathcal{N}} \left\|\widehat{\mathbf{f}}_j^i(t_{m_0} + MB) - \mathbf{x}_j^\star\right\| \ge \varepsilon \,\middle|\, \mathcal{F}_{t_{m_0}}\right) \le \delta. \tag{9}$$

Proof. For each source agent j, let $\widehat{F}_j(t) \in \mathbb{R}^{N\times K_j}$ denote the matrix of local beliefs across agents, with $[\widehat{F}_j(t)]_{i,:} = (\widehat{\mathbf{f}}_j^i(t))^\top$ for each $i \in \mathcal{N}$. Then, almost surely, we rewrite the consensus dynamics (4) as,

$$\widehat{F}_j(t+1) = W_j(t)\widehat{F}_j(t) + W_j(t)e_j\big(\mathbf{f}_j(t+1) - \mathbf{f}_j(t)\big)^\top. \tag{10}$$

Since $W_j(t)\mathbf{1} = \mathbf{1}$, subtracting $\mathbf{1}\mathbf{f}_j(t+1)^\top$ from both sides yields,

$$\widehat{F}_j(t+1) - \mathbf{1}\mathbf{f}_j(t+1)^\top = W_j(t)\big(\widehat{F}_j(t) - \mathbf{1}\mathbf{f}_j(t)^\top\big) \\ + \big(W_j(t)e_j - \mathbf{1}\big)\big(\mathbf{f}_j(t+1) - \mathbf{f}_j(t)\big)^\top. \tag{11}$$

The tracking error accordingly consists of a consensus term and a perturbation induced by the variation of $\mathbf{f}_j(t)$.

Let $t_m := t_0 + mB$. Thus, block m consists of the B consecutive iterations $t_m, \dots, t_m + B - 1$. Therefore, the perturbation term in (11) decays geometrically over the holding interval. From (11), we extend the difference by iterating the recursion from t_m to $t_{m+1} - 1$,

$$\widehat{F}_j(t_{m+B}) - \mathbf{1}\mathbf{f}_j(t_{m+B})^\top \\ = \Psi_{j,m}\big(\widehat{F}_j(t_m) - \mathbf{1}\mathbf{f}_j(t_m)^\top\big) \\ + \sum_{s=t_m}^{t_0+mB} \left(\prod_{\tau=s+1}^{\overleftarrow{t_0+mB-1}} W_j(\tau)\right)\big(W_j(s)e_j - \mathbf{1}\big)\big(\mathbf{f}_j(s+1) - \mathbf{f}_j(s)\big)^\top. \tag{12}$$

The first term represents the tracking error already present at the beginning of block m. For each $i \in \mathcal{N}$, its ith row is given by $e_i^\top \Psi_{j,m}\big(\widehat{F}_j(t_m) - \mathbf{1}\mathbf{f}_j(t_m)^\top\big) = \sum_{n\in\mathcal{N}} [\Psi_{j,m}]_{i,n}\big(\widehat{\mathbf{f}}_j^n(t_m) - \mathbf{f}_j(t_m)\big)^\top$. Since $e_j^\top W_j(t) = e_j^\top$ and $\widehat{\mathbf{f}}_j^j(t) = \mathbf{f}_j(t)$, the source agent has zero tracking error at every time. Hence, for each $i \in \mathcal{N}$, the ith row of the first term satisfies

$$\begin{aligned}\left\|\sum_{n\in\mathcal{N}} [\Psi_{j,m}]_{i,n}\big(\widehat{\mathbf{f}}_j^n(t_m) - \mathbf{f}_j(t_m)\big)\right\| &= \left\|\sum_{n\neq j} [\Psi_{j,m}]_{i,n}\big(\widehat{\mathbf{f}}_j^n(t_m) - \mathbf{f}_j(t_m)\big)\right\| \\ &\le \sum_{n\neq j} [\Psi_{j,m}]_{i,n}\left\|\widehat{\mathbf{f}}_j^n(t_m) - \mathbf{f}_j(t_m)\right\| \\ &= (1 - [\Psi_{j,m}]_{i,j}) \max_{r\in\mathcal{N}} \left\|\widehat{\mathbf{f}}_j^r(t_m) - \mathbf{f}_j(t_m)\right\|. \end{aligned} \tag{13}$$

Thus, defining the successful propagation event for source agent j over block m as $\mathcal{C}_{j,m} := \Big\{\min_{i\in\mathcal{N}} [\Psi_{j,m}]_{i,j} \ge \underline{\beta}_j\Big\}$, the inherited tracking error is contracted by the factor $1 - \underline{\beta}_j$ whenever $\mathcal{C}_{j,m}$ occurs, whereas on $\mathcal{C}_{j,m}^c$ row-stochasticity guarantees a contraction factor of at most one. Let $Z_{j,m} := \mathbf{1}_{\mathcal{C}_{j,m}}$. From the one-block bound, with $Z_{j,m} := \mathbf{1}_{\mathcal{C}_{j,m}}$, we have

$$\max_{i\in\mathcal{N}} \left\|\sum_{n\in\mathcal{N}} [\Psi_{j,m}]_{i,n}\big(\widehat{\mathbf{f}}_j^n(t_m) - \mathbf{f}_j(t_m)\big)\right\| \le \Big(1 - \underline{\beta}_j Z_{j,m}\Big) \max_{r\in\mathcal{N}} \left\|\widehat{\mathbf{f}}_j^r(t_m) - \mathbf{f}_j(t_m)\right\|. \tag{14}$$

During the holding interval, $\mathbf{x}_j(t) = \mathbf{x}_j^\star$. By (3), for any $t \ge t_m$ within block m, $\mathbf{f}_j(t) - \mathbf{x}_j^\star = (1-\alpha)^{t-t_m}\big(\mathbf{f}_j(t_m) - \mathbf{x}_j^\star\big)$. Consequently, $\|\mathbf{f}_j(t+1) - \mathbf{f}_j(t)\| = \alpha(1-\alpha)^{t-t_m}\|\mathbf{f}_j(t_m) - \mathbf{x}_j^\star\|$, showing that the variation of the empirical frequency decays geometrically within the block. This implies that the perturbation accumulated over successive blocks decays at rate $\mathcal{O}\big((1-\alpha)^B\big)$ and vanishes as the holding interval progresses.

$$\begin{aligned}&\max_{i\in\mathcal{N}} \left\|\sum_{s=t_m}^{t_m+B-1} e_i^\top \left(\prod_{\tau=s+1}^{\overleftarrow{t_m+B-1}} W_j(\tau)\right)\big(W_j(s)e_j - \mathbf{1}\big)\big(\mathbf{f}_j(s+1) - \mathbf{f}_j(s)\big)^\top\right\| \\ &\le \sum_{s=t_m}^{t_m+B-1} \|\mathbf{f}_j(s+1) - \mathbf{f}_j(s)\| = \alpha \sum_{s=t_m}^{t_m+B-1} (1-\alpha)^{s-t_m}\left\|\mathbf{f}_j(t_m) - \mathbf{x}_j^\star\right\| \\ &= \big(1 - (1-\alpha)^B\big)(1-\alpha)^{(m-m_0)B}\left\|\mathbf{f}_j(t_{m_0}) - \mathbf{x}_j^\star\right\|. \end{aligned} \tag{15}$$

Hence, the perturbation accumulated over block m is $\mathcal{O}\big((1-\alpha)^{(m-m_0)B}\big)$. By defining $E_j(t) := \max_{i\in\mathcal{N}}\left\|\widehat{\mathbf{f}}_j^i(t) - \mathbf{f}_j(t)\right\|$, we combine the contraction of the inherited tracking error (Eq. (14)) with the bound on the perturbation accumulated over block m (Eq. (15)), we obtain

$$E_j(t_m + B) \le \Big(1 - \underline{\beta}_j Z_{j,m}\Big) E_j(t_m) + \mathcal{O}\Big((1-\alpha)^{(m-m_0)B}\Big). \tag{16}$$

Iterating the block-level recursion over the M blocks $m = m_0, \dots, m_0 + M - 1$ yields

$$\begin{aligned}E_j(t_{m_0} + MB) \le &\left(\prod_{\ell=m_0}^{m_0+M-1}\Big(1 - \underline{\beta}_j Z_{j,\ell}\Big)\right) E_j(t_{m_0}) \\ &+ \sqrt{2}(1 - (1-\alpha)^B) \sum_{k=m_0}^{m_0+M-1}\left(\prod_{\ell=k+1}^{m_0+M-1}\Big(1 - \underline{\beta}_j Z_{j,\ell}\Big)\right)(1-\alpha)^{(k-m_0)B}. \end{aligned} \tag{17}$$

To characterize successful propagation probabilistically, let $\mathcal{G}_m := \mathcal{F}_{t_m}$. Since $\mathcal{C}_{j,m} \in \mathcal{G}_{m+1}$ and Assumption 3 implies $\sum_{m=0}^\infty \mathbb{P}(\mathcal{C}_{j,m} \mid \mathcal{G}_m) = \infty$ almost surely, the conditional Borel–Cantelli Lemma (Chandra, 2012, Theorem 5.1.2, p. 96) implies that $\mathcal{C}_{j,m}$ occurs infinitely often almost surely, equivalently, $\sum_{m=0}^\infty Z_{j,m} = \infty$ almost surely.

For any $\varepsilon > 0$ and $\delta_c > 0$, let $S_{j,R,M} := \sum_{\ell=m_0+R}^{m_0+M-1} Z_{j,\ell}$ for some $R < M$. For the initial-error term, $\prod_{\ell=m_0}^{m_0+M-1}(1-\underline{\beta}_j Z_{j,\ell}) \le \prod_{\ell=m_0+R}^{m_0+M-1}(1-\underline{\beta}_j Z_{j,\ell}) = (1-\underline{\beta}_j)^{S_{j,R,M}}$. Likewise, for every $k = m_0, \dots, m_0+R-1$, the product multiplying the kth perturbation contains all factors with indices $\ell = m_0+R, \dots, m_0+M-1$, and is therefore bounded by $(1-\underline{\beta}_j)^{S_{j,R,M}}$. Using $E_j(t_{m_0}) \le \sqrt{2}$ together with $(1-(1-\alpha)^B)\sum_{r=0}^{R-1}((1-\alpha)^B)^r \le 1$, the initial error and the first R perturbation terms contribute at most $2\sqrt{2}(1-\underline{\beta}_j)^{S_{j,R,M}}$.

For the remaining perturbations, corresponding to $k = m_0+R, \dots, m_0+M-1$, we use $\prod_{\ell=k+1}^{m_0+M-1}(1-\underline{\beta}_j Z_{j,\ell}) \le 1$. The corresponding geometric tail satisfies

$$\sqrt{2}\big(1-(1-\alpha)^B\big)\sum_{k=m_0+R}^{m_0+M-1}(1-\alpha)^{(k-m_0)B} \le \sqrt{2}(1-\alpha)^{RB}.$$

Combining the two parts gives

$$E_j(t_{m_0}+MB) \le 2\sqrt{2}(1-\underline{\beta}_j)^{S_{j,R,M}} + \sqrt{2}(1-\alpha)^{RB}. \tag{18}$$

Since $0 < 1-\alpha < 1$, choose R so that $\sqrt{2}(1-\alpha)^{RB} \le \varepsilon/2$, and select $K_\varepsilon \in \mathbb{N}$ satisfying $2\sqrt{2}(1-\underline{\beta}_j)^{K_\varepsilon} \le \varepsilon/2$. On the event $\{S_{j,R,M} \ge K_\varepsilon\}$, (18) gives $E_j(t_{m_0}+MB) \le \varepsilon$, so $\{S_{j,R,M} \ge K_\varepsilon\} \subseteq \{E_j(t_{m_0}+MB) \le \varepsilon\}$.

Because successful propagation occurs infinitely often almost surely, $S_{j,R,M} \to \infty$ almost surely as $M \to \infty$ for every finite R. Accordingly, for the prescribed $\delta_c > 0$, one can choose a sufficiently large finite M such that $\mathbb{P}(S_{j,R,M} \ge K_\varepsilon \mid \mathcal{F}_{t_{m_0}}) \ge 1-\delta_c$. The event inclusion above then gives $\mathbb{P}(E_j(t_{m_0}+MB) \le \varepsilon \mid \mathcal{F}_{t_{m_0}}) \ge 1-\delta_c$, or equivalently $\mathbb{P}(E_j(t_{m_0}+MB) > \varepsilon \mid \mathcal{F}_{t_{m_0}}) \le \delta_c$. Finally, during the holding interval, $\mathbf{x}_j(t) = \mathbf{x}_j^\star$, so $\|\mathbf{f}_j(t_{m_0}+MB) - \mathbf{x}_j^\star\| \le \sqrt{2}(1-\alpha)^{MB}$. Choosing M large enough so that this term is at most $\varepsilon/2$ and applying the preceding tracking bound with $\varepsilon/2$, the triangle inequality gives

$$\max_{i\in\mathcal{N}} \left\|\widehat{\mathbf{f}}_j^i(t_{m_0}+MB) - \mathbf{x}_j^\star\right\| \le E_j(t_{m_0}+MB) + \left\|\mathbf{f}_j(t_{m_0}+MB) - \mathbf{x}_j^\star\right\|.$$

Therefore, $\mathbb{P}\Big(\max_{i\in\mathcal{N}} \|\widehat{\mathbf{f}}_j^i(t_{m_0}+MB) - \mathbf{x}_j^\star\| \ge \varepsilon \mid \mathcal{F}_{t_{m_0}}\Big) \le \delta$, which proves (9).

□

Lemma 2 (Preservation of strict improving responses)**.** *There exists a sufficiently small $\xi > 0$ such that, if $\|\hat{\mathbf{f}}_j^i(t) - x_j(t)\| \le \xi$ for all $j \in \mathcal{N}\setminus\{i\}$, then any strict improvement identified by Algorithm 1 according to the local beliefs is also a strict improvement with respect to the true actions; that is, if $c_i(y_i, \hat{\mathbf{f}}_{-i}^{\,i}(t)) < c_i(x_i, \hat{\mathbf{f}}_{-i}^{\,i}(t))$, then it holds $c_i(y_i, x_{-i}(t)) < c_i(x_i, x_{-i}(t))$.*

Proof. We define the minimum nonzero unilateral cost difference as

$$\Delta_{\min} := \min_{\substack{i\in\mathcal{N},\, x_{-i},\, x_i, y_i \in \mathcal{X}_i: \\ c_i(y_i,x_{-i}) \ne c_i(x_i,x_{-i})}} \left|c_i(y_i,x_{-i}) - c_i(x_i,x_{-i})\right|. \tag{19}$$

Since the action sets are finite, the set over which the minimum is taken is finite, and therefore $\Delta_{\min} > 0$ is well defined and stricly greater than zero.

For each $i \in \mathcal{N}$ and $x_i, y_i \in \mathcal{X}_i$, define

$$g_{i,x_i,y_i}(f_{-i}) := c_i(y_i, f_{-i}) - c_i(x_i, f_{-i}).$$

For each $i \in \mathcal{N}$ and $x_i, y_i \in \mathcal{X}_i$, define the unilateral cost difference as $\Delta_i(x_i, y_i; f_{-i}) := c_i(y_i, f_{-i}) - c_i(x_i, f_{-i})$. Since the mixed extension of each cost function is multilinear ove the finite action sets, $\Delta_i(x_i, y_i; \cdot)$ is Lipschitz over a simplex of (finitely many) actions. Hence, there exists a finite constant $L > 0$, uniform over all i, x_i, y_i, such that

$$\left|\Delta_i(x_i, y_i; \hat{\mathbf{f}}_{-i}^{\,i}(t)) - \Delta_i(x_i, y_i; x_{-i}(t))\right| \le L\left\|\hat{\mathbf{f}}_{-i}^{\,i}(t) - x_{-i}(t)\right\|.$$

We choose $\xi > 0$ such that $L\xi < \Delta_{\min}/2$. If Algorithm 1 identifies y_i as a strict improvement over x_i according to the local beliefs, then we have $\Delta_i(x_i, y_i; \hat{\mathbf{f}}^{\,i}_{-i}(t)) < 0$ and it gives $\left|\Delta_i(x_i, y_i; \hat{\mathbf{f}}^{\,i}_{-i}(t)) - \Delta_i(x_i, y_i; x_{-i}(t))\right| < \frac{\Delta_{\min}}{2}$. If the corresponding true cost difference were positive, then $\Delta_i(x_i, y_i; x_{-i}(t)) \geq \Delta_{\min}$, which would imply $\Delta_i(x_i, y_i; \hat{\mathbf{f}}^{\,i}_{-i}(t)) > \Delta_{\min} - \frac{\Delta_{\min}}{2} = \frac{\Delta_{\min}}{2} > 0$, contradicting the strict improvement identified by Algorithm 1. Therefore, $\Delta_i(x_i, y_i; x_{-i}(t)) < 0$. The reverse sign-preservation argument is analogous. Thus, the sign of every nonzero unilateral cost difference is preserved. In particular, $\Delta_i(x_i, y_i; \hat{\mathbf{f}}^{\,i}_{-i}(t)) < 0$ implies $\Delta_i(x_i, y_i; x_{-i}(t)) < 0$, completing the proof.

□

Lemma 3 (Positive-probability improving local search)**.** *Suppose Assumptions 1–3 hold and $\epsilon_{\exp} \in (0, 1)$. For any $i \in \mathcal{N}$ such that $\mathbf{x}_i(t)$ is not a best response to $\mathbf{x}_{-i}(t)$, suppose $\|\hat{\mathbf{f}}^{\,i}_j(t) - \mathbf{x}_j(t)\| \leq \xi$ for all $j \neq i$, where $\xi > 0$ is as in Lemma 2. Let $R_i(Q_t)$ denote the number of intermediate points along an admissible improving path that require explicit exploration under search budget Q_t. Then Algorithm 1 selects a strict improvement with probability*

$$\mathbb{P}\big(c_i(\mathbf{x}_i^+, \mathbf{x}_{-i}(t)) < c_i(\mathbf{x}_i(t), \mathbf{x}_{-i}(t)) \,\big|\, \mathcal{F}_t\big) \geq \mathbb{E}_{Q_t}\left[\epsilon_{\exp}^{R_i(Q_t)}\right] \geq \epsilon_{\exp}^{L_{\max}-1} > 0, \tag{20}$$

where, under the truncated geometric distribution, $\mathbb{E}_{Q_t}[\epsilon_{\exp}^{r_i(Q_t)}] = \sum_{q=1}^{\bar{Q}} \frac{\lambda(1-\lambda)^{q-1}}{1-(1-\lambda)^{\bar{Q}}} \epsilon_{\exp}^{r_i(q)}$.

Proof. Suppose $\mathbf{x}_i(t)$ is not a best response to $\mathbf{x}_{-i}(t)$. By Assumption 1, there exists an admissible one-bit path $\mathbf{x}_i(t) = \mathbf{v}^{(0)}, \mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(L-1)}, \mathbf{y}_i^\star$, with $L \leq L_{\max}$, terminating at a feasible strict improvement $\mathbf{y}_i^\star$. Condition (A1) holds for every intermediate point along this path by Assumption 1.

For a realization $Q_t = q$, let $r_i(q)$ denote the number of intermediate points $\mathbf{v}^{(\ell)}$, $\ell = 1, \ldots, L-1$, that fail (A2) or (A3) under budget q and therefore require the exploration branch. Let $P_\ell(q)$ denote the probability of reaching $\mathbf{v}^{(\ell)}$ along the prescribed path, with $P_0(q) = 1$. Since an intermediate point satisfying (A2)–(A3) is appended to SL directly, whereas otherwise it is appended with probability $\epsilon_{\exp}$, independence of the exploration draws gives For a realization $Q_t = q$, let $r_i(q)$ denote the number of intermediate points $\mathbf{v}^{(\ell)}$, $\ell = 1, \ldots, L-1$, that fail (A2) or (A3) and therefore require the exploration branch. Each of the remaining intermediate points is appended to SL deterministically, whereas each of these $r_i(q)$ points is appended with probability $\epsilon_{\exp}$. Under independent exploration draws,

$$\mathbb{P}\Big(\mathbf{v}^{(1)}, \ldots, \mathbf{v}^{(L-1)} \text{ are successively appended to } SL \,\Big|\, Q_t = q, \mathcal{F}_t\Big) \geq \epsilon_{\exp}^{r_i(q)}.$$

Once $\mathbf{v}^{(L-1)}$ is reached, $\mathbf{y}_i^\star$ is examined as a feasible one-bit neighbor and, by Lemma 2, is accepted as a strict improvement. Therefore,

$$\mathbb{P}\big(c_i(\mathbf{x}_i^+, \mathbf{x}_{-i}(t)) < c_i(\mathbf{x}_i(t), \mathbf{x}_{-i}(t)) \,\big|\, Q_t = q, \mathcal{F}_t\big) \geq \epsilon_{\exp}^{r_i(q)}.$$

Once $\mathbf{v}^{(L-1)}$ is reached, $\mathbf{y}_i^\star$ is examined as its feasible one-bit neighbor. By Lemma 2, the belief-accuracy condition preserves its strict improvement, and hence Algorithm 1 accepts it; if another feasible strict improvement is found earlier, the desired event has already occurred. Consequently,

$$\mathbb{P}\big(c_i(\mathbf{x}_i^+, \mathbf{x}_{-i}(t)) < c_i(\mathbf{x}_i(t), \mathbf{x}_{-i}(t)) \,\big|\, Q_t = q, \mathcal{F}_t\big) \geq \epsilon_{\exp}^{r_i(q)}.$$

For the truncated geometric distribution on $q = 1, \ldots, \bar{Q}$, $\mathbb{P}(Q_t = q) = \lambda(1-\lambda)^{q-1}/(1-(1-\lambda)^{\bar{Q}})$. Averaging over Q_t therefore yields

$$\begin{aligned}\mathbb{P}\big(c_i(\mathbf{x}_i^+, \mathbf{x}_{-i}(t)) < c_i(\mathbf{x}_i(t), \mathbf{x}_{-i}(t)) \,\big|\, \mathcal{F}_t\big) \\ \geq \sum_{q=1}^{\bar{Q}} \frac{\lambda(1-\lambda)^{q-1}}{1-(1-\lambda)^{\bar{Q}}} \epsilon_{\exp}^{r_i(q)} = \mathbb{E}_{Q_t}\left[\epsilon_{\exp}^{r_i(Q_t)}\right].\end{aligned} \tag{21}$$

Since $r_i(q) \leq L - 1 \leq L_{\max} - 1$ and $\epsilon_{\exp} \in (0, 1)$,

$$\mathbb{E}_{Q_t}\left[\epsilon_{\exp}^{r_i(Q_t)}\right] \geq \epsilon_{\exp}^{L_{\max}-1} > 0,$$

which proves the claim. □

Corollary 1 (Finite-horizon unilateral improvement)**.** *Suppose Assumptions 1–3 hold and* $\epsilon_{\text{in}}, \epsilon_{\text{exp}} \in (0,1)$*. Let* $i \in \mathcal{N}$ *be such that* $\mathbf{x}_i(t)$ *is not a best response to* $\mathbf{x}_{-i}(t)$*, and let* $\xi > 0$ *be as in Lemma 2. For any* $\delta > 0$*, choose* $T = MB < \infty$ *sufficiently large according to the tracking result with accuracy* ξ *and failure probability* δ*. Then*

$$\begin{aligned}\mathbb{P}\left(\begin{array}{c}\mathbf{x}(s)=\mathbf{x}(t), \quad s=t,\dots,t+T,\\ c_i(\mathbf{x}_i(t+T+1),\mathbf{x}_{-i}(t)) < c_i(\mathbf{x}_i(t),\mathbf{x}_{-i}(t)),\\ \mathbf{x}_j(t+T+1)=\mathbf{x}_j(t), \quad \textit{for all } j\neq i\end{array}\;\middle|\;\mathcal{F}_t\right)\\ \geq (1-\delta)(1-\epsilon_{\text{in}})\epsilon_{\text{in}}^{N(T+1)-1}\mathbb{E}_{Q_t}\left[\epsilon_{\text{exp}}^{r_i(Q_t)}\right]\\ \geq (1-\delta)(1-\epsilon_{\text{in}})\epsilon_{\text{in}}^{N(T+1)-1}\epsilon_{\text{exp}}^{L_{\max}-1} > 0. \end{aligned} \tag{22}$$

Proof. All agents remain at $\mathbf{x}(t)$ for T consecutive iterations with probability $\epsilon_{\text{in}}^{NT}$. Conditional on this holding period, the tracking result with accuracy ξ guarantees sufficiently accurate beliefs at time $t+T$ with probability at least $1-\delta$. At the subsequent iteration, agent i invokes Algorithm 1 while all $j \neq i$ remain inertial with probability $(1-\epsilon_{\text{in}})\epsilon_{\text{in}}^{N-1}$. Lemma 3 then gives a true strict improvement with conditional probability at least $\mathbb{E}_{Q_t}[\epsilon_{\text{exp}}^{r_i(Q_t)}]$. Multiplying these conditional probabilities yields (22), while $r_i(Q_t) \leq L_{\max} - 1$ gives the uniform positive lower bound. □

Theorem 2 (Infinite visitation of the approximate equilibrium set)**.** *Suppose Assumptions 1–3 hold and* Γ *is* δ*-near an exact potential game. Let* $\{\mathbf{x}_t\}_{t\geq 0} \subseteq \mathcal{X}$ *denote the sequence of joint action profiles generated by DLS (Algorithm 2).*

We define the δh*-approximate equilibrium set as the set of action profiles at which no agent can decrease its cost by more than* δh *through a unilateral deviation,*

$$\mathcal{X}_{\delta h} := \left\{\mathbf{x}\in\mathcal{X} : c_i(\mathbf{x}_i,\mathbf{x}_{-i}) \leq \min_{\mathbf{y}_i\in\mathcal{X}_i} c_i(\mathbf{y}_i,\mathbf{x}_{-i}) + \delta h, \quad \textit{for all } i\in\mathcal{N}\right\},$$

where $h := |\mathcal{X}|$ *is the cardinality of the joint action set. Then the sequence of joint actions* $\{\mathbf{x}_t\}_{t\geq 0} \subseteq \mathcal{X}$ *visits* $\mathcal{X}_{\delta h}$ *infinitely often almost surely, that is,*

$$\mathbb{P}(\mathbf{x}_t \in \mathcal{X}_{\delta h} \textit{ infinitely often}) = 1.$$

Proof. Consider any time t such that $\mathbf{x}_t \notin \mathcal{X}_{\delta h}$. By the definition of $\mathcal{X}_{\delta h}$, there exists an agent $i \in \mathcal{N}$ whose best-response improvement is larger than δh. By the finite-horizon unilateral-improvement bound established above, with conditional probability at least

$$(1-\delta_c)(1-\epsilon_{\text{in}})\epsilon_{\text{in}}^{N(T+1)-1}\epsilon_{\text{exp}}^{L_{\max}-1} > 0,$$

the other agents remain fixed long enough for the local beliefs to become sufficiently accurate and agent i implements a strict improvement with respect to the true actions.

If the resulting action of agent i is not yet a best response, the same favorable event can occur again while the other agents remain fixed. Each such realization strictly decreases $c_i(\cdot,\mathbf{x}_{-i}(t))$. Since $\mathcal{X}_i$ is finite, there is a finite positive-probability realization under which agent i eventually reaches a best response $\mathbf{x}_i^\star$. The resulting profile $\mathbf{x}' = (\mathbf{x}_i^\star, \mathbf{x}_{-i}(t))$ satisfies

$$c_i(\mathbf{x}_i(t),\mathbf{x}_{-i}(t)) - c_i(\mathbf{x}_i^\star,\mathbf{x}_{-i}(t)) > \delta h.$$

Since Γ is δ-near the exact potential game $\widehat{\Gamma}$,

$$\begin{aligned}\widehat{\Phi}(\mathbf{x}') - \widehat{\Phi}(\mathbf{x}_t) &\leq c_i(\mathbf{x}_i^\star,\mathbf{x}_{-i}(t)) - c_i(\mathbf{x}_i(t),\mathbf{x}_{-i}(t)) + \delta\\ &< -\delta h + \delta < 0,\end{aligned}$$

where the case $\delta = 0$ follows directly from the exact potential property.

If $\mathbf{x}' \notin \mathcal{X}_{\delta h}$, the same favorable construction can be applied again from $\mathbf{x}'$. Along this particular realization, the nearby potential strictly decreases after each such macro-step, so no joint action can be repeated. Since $|\mathcal{X}| = h$, a favorable sequence reaches $\mathcal{X}_{\delta h}$ after finitely many, and at most $h-1$, such macro-steps.

This construction applies to every profile in $\mathcal{X}\backslash\mathcal{X}_{\delta h}$. Since $\mathcal{X}$ is finite and the finite-horizon improvement bound above is uniform, there exist finite H and $\rho > 0$ such that, from every profile outside $\mathcal{X}_{\delta h}$, the probability of visiting $\mathcal{X}_{\delta h}$ during the following H iterations is at least ρ. Other realizations may cycle, move to worse profiles, or remain outside the set; only this favorable realization is used in the argument.

If no visit occurs during the first H iterations, the process is again at some profile outside $\mathcal{X}_{\delta h}$, and the same bound applies to the following block. Successive conditioning therefore gives

$$\mathbb{P}(\mathbf{x}_s \notin \mathcal{X}_{\delta h}, \quad s = t+1, \dots, t+mH \mid \mathcal{F}_t) \le (1-\rho)^m.$$

Letting $m \to \infty$ shows that, whenever the process is outside $\mathcal{X}_{\delta h}$, the probability of never visiting this set again is zero.

If $\mathcal{X}_{\delta h}$ were visited only finitely often, there would be some finite time after which the process remains outside the set forever. The preceding argument shows that this event has probability zero. Therefore,

$$\mathbb{P}(\mathbf{x}_t \in \mathcal{X}_{\delta h} \text{ infinitely often}) = 1.$$

□

A.3 Proof of Theorem 1

Theorem 3 (Almost-sure convergence to a pure Nash equilibrium)**.** *Suppose Assumptions 1–3 and 4 hold, and Γ is δ-near an exact potential game. Let $\mathcal{X}_0$ denote the set of pure Nash equilibria. Suppose further that every $\mathbf{x}^\star \in \mathcal{X}_0$ is strict, i.e., for every $i \in \mathcal{N}$ and every $\mathbf{y}_i \in \mathcal{X}_i \setminus \{\mathbf{x}_i^\star\}$,*

$$c_i(\mathbf{x}_i^\star, \mathbf{x}_{-i}^\star) < c_i(\mathbf{y}_i, \mathbf{x}_{-i}^\star). \tag{23}$$

Then there exist an almost surely finite random time τ and a pure Nash equilibrium $\mathbf{x}^\star \in \mathcal{X}_0$ such that

$$\mathbf{x}(t) = \mathbf{x}^\star, \qquad \textit{for all } t \ge \tau, \quad \textit{a.s.} \tag{24}$$

Proof. Consider any profile $\mathbf{x} \notin \mathcal{X}_0$. First, suppose that $\mathbf{x} \notin \mathcal{X}_{\delta h}$. By the definition of $\mathcal{X}_{\delta h}$, there exists an agent $i \in \mathcal{N}$ such that

$$c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \min_{\mathbf{y}_i \in \mathcal{X}_i} c_i(\mathbf{y}_i, \mathbf{x}_{-i}) > \delta h. \tag{25}$$

Hold the actions of all agents other than i fixed at $\mathbf{x}_{-i}$. By Assumption 1, whenever the current action of agent i is not a best response, Algorithm 1 admits a finite local-search path to a strict improvement. Since the random exploration probability is strictly positive, each such path can be realized with positive probability.

Moreover, Lines 6–9 of Algorithm 1 do not terminate the local search after a strict improvement is found. Instead, the incumbent is updated, the solution list is reset around the new incumbent, and the search continues. Thus, on a favorable realization, strict unilateral improvements can be successively generated while $\mathbf{x}_{-i}$ remains fixed. Since $\mathcal{X}_i$ is finite and the cost of agent i strictly decreases at every incumbent update, after finitely many such improvements agent i reaches a best response. Denote the resulting joint profile by

$$\mathbf{x}' = (\mathbf{x}_i', \mathbf{x}_{-i}), \qquad \mathbf{x}_i' \in \arg\min_{\mathbf{y}_i \in \mathcal{X}_i} c_i(\mathbf{y}_i, \mathbf{x}_{-i}).$$

It follows from (25) that

$$c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - c_i(\mathbf{x}_i', \mathbf{x}_{-i}) > \delta h.$$

Since Γ is δ-near the exact potential game $\widehat{\Gamma}$,

$$\begin{aligned} \widehat{\Phi}(\mathbf{x}) - \widehat{\Phi}(\mathbf{x}') &\ge c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - c_i(\mathbf{x}_i', \mathbf{x}_{-i}) - \delta \\ &> \delta h - \delta. \end{aligned} \tag{26}$$

Hence, for $h > 1$,

$$\widehat{\Phi}(\mathbf{x}') < \widehat{\Phi}(\mathbf{x}).$$

Now suppose instead that $\mathbf{x} \in \mathcal{X}_{\delta h} \setminus \mathcal{X}_0$. By Assumption 4, there exists an agent i such that

$$c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \min_{\mathbf{y}_i \in \mathcal{X}_i} c_i(\mathbf{y}_i, \mathbf{x}_{-i}) > \delta. \tag{27}$$

Applying exactly the same finite local-search construction while holding $\mathbf{x}_{-i}$ fixed, there is a positive-probability realization in which agent i successively improves its incumbent until reaching a best response $\mathbf{x}'_i$. Therefore,

$$c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - c_i(\mathbf{x}'_i, \mathbf{x}_{-i}) > \delta.$$

The near-potential property then yields

$$\begin{aligned} \widehat{\Phi}(\mathbf{x}) - \widehat{\Phi}(\mathbf{x}') &\geq c_i(\mathbf{x}_i, \mathbf{x}_{-i}) - c_i(\mathbf{x}'_i, \mathbf{x}_{-i}) - \delta \\ &> 0. \end{aligned} \tag{28}$$

Thus, also in this case, there exists a finite positive-probability DLS realization leading to a profile with strictly smaller potential.

Consequently, from every $\mathbf{x} \notin \mathcal{X}_0$, there exists a finite positive-probability DLS realization leading to some $\mathbf{x}'$ satisfying $\widehat{\Phi}(\mathbf{x}') < \widehat{\Phi}(\mathbf{x})$. and therefore repeating this construction whenever the resulting profile is not in $\mathcal{X}_0$ produces, along a favorable realization,

$$\widehat{\Phi}(\mathbf{x}^0) > \widehat{\Phi}(\mathbf{x}^1) > \widehat{\Phi}(\mathbf{x}^2) > \cdots.$$

Since $\mathcal{X}$ is finite, the strictly decreasing sequence cannot continue indefinitely. Let $\mathbf{x}^K$ denote its terminal profile. Suppose, by contradiction, that $\mathbf{x}^K \notin \mathcal{X}_0$. If $\mathbf{x}^K \notin \mathcal{X}_{\delta h}$, the first case above provides a finite favorable realization ending at some $\mathbf{x}^{K+1}$ such that

$$\widehat{\Phi}(\mathbf{x}^{K+1}) < \widehat{\Phi}(\mathbf{x}^K).$$

If instead $\mathbf{x}^K \in \mathcal{X}_{\delta h} \setminus \mathcal{X}_0$, Assumption 4 and the second case above yield the same conclusion. In either case, $\mathbf{x}^K$ admits a further strict potential decrease, contradicting its terminality. Therefore, we conclude $\mathbf{x}^K \in \mathcal{X}_0$. Hence every favorable finite descent path terminates at a pure Nash equilibrium. By the uniform finite-horizon probability bounds, there exist $H < \infty$ and $\rho > 0$ such that, whenever $\mathbf{x}_t \notin \mathcal{X}_0$,

$$\mathbb{P}(\mathbf{x}_s \in \mathcal{X}_0 \text{ for some } s \in \{t+1, \ldots, t+H\} \mid \mathcal{F}_t) \geq \rho.$$

Thus, with

$$\tau_0 := \inf\{t \geq 0 : \mathbf{x}_t \in \mathcal{X}_0\},$$

we have

$$\mathbb{P}(\tau_0 > mH) \leq (1-\rho)^m \to 0,$$

so $\tau_0 < \infty$ almost surely.

It remains to show stabilization. Let $\mathbf{x}^\star \in \mathcal{X}_0$ be a visited pure Nash equilibrium. Since $\mathbf{x}^\star$ is strict and the cost extensions are continuous, there exists $\xi^\star > 0$ such that sufficiently accurate local beliefs satisfy

$$c_i(\mathbf{x}_i^\star, \widehat{\mathbf{f}}^i_{-i}) < c_i(\mathbf{y}_i, \widehat{\mathbf{f}}^i_{-i}), \qquad \text{for all } i, \text{ for all } \mathbf{y}_i \neq \mathbf{x}_i^\star.$$

Hence Algorithm 1 cannot move away from $\mathbf{x}^\star$ once the beliefs enter this neighborhood. By inertia, there is a positive probability that all agents keep $\mathbf{x}^\star$ fixed for a sufficiently long finite horizon. By Lemma 1, conditional on such a holding period, the local estimates become arbitrarily close to $\mathbf{x}^\star$. Hence, with positive probability,

$$\|\widehat{\mathbf{f}}^i_{-i}(t) - \mathbf{x}^\star_{-i}\| < \xi^\star, \qquad \text{for all } i \in \mathcal{N},$$

while the joint action remains at $\mathbf{x}^\star$. By strictness of $\mathbf{x}^\star$, Algorithm 1 then admits no strict improvement, so $\mathbf{x}^\star$ is preserved. Applying Lemma 1 thereafter gives

$$\mathbf{f}_j(t) \to \mathbf{x}_j^\star, \qquad \widehat{\mathbf{f}}^i_j(t) \to \mathbf{x}_j^\star, \qquad \text{for all } i, j \in \mathcal{N}.$$

If stabilization does not occur at a visit to $\mathcal{X}_0$, the preceding hitting argument implies another visit almost surely. Moreover, by Assumption 3 and the conditional Borel–Cantelli argument used in Lemma 1, successful communication blocks occur infinitely often almost surely, even though their probability lower bounds need not be uniform. Together with the positive probability of holding the joint action by inertia, these repeated opportunities eventually place the local beliefs inside the strictness neighborhood of a visited pure Nash equilibrium. Once this occurs, Algorithm 1 admits no strict improvement, so the equilibrium is preserved thereafter. Consequently, for some almost surely finite random time τ and some $\mathbf{x}^\star \in \mathcal{X}_0$, it holds $\mathbf{x}_t = \mathbf{x}^\star$, for all $t \geq \tau$,

□

B PROBLEM DEFINITIONS

This appendix gives the full definitions and parameters of the problems P1–P8 of Section 6.1.

Throughout, $x_{ik} \in \{0,1\}$ indicates that agent i selects item k (a target, project, channel or route) and $n_k = \sum_i x_{ik}$ is the load of item k. In all problems except P8 the cost c_i is linear in $\mathbf{x}_i$ for fixed $\mathbf{x}_{-i}$, so the agent's problem is a binary program with a linear objective; the centralized social-cost minimization $\min_{\mathbf{x}} \sum_i c_i(\mathbf{x})$ is at least quadratic in every case.

P1: sensor coverage. Each of N sensors selects up to B of K targets and pays $c_i(\mathbf{x}) = \sum_k x_{ik}\big(\beta \sum_{j\neq i} x_{jk} - v_{ik}\big)$ with player-specific values $v_{ik} = v_k + \varepsilon_{ik}$, $v_k \sim U(1,5)$, $\varepsilon_{ik} \sim 0.15\mathcal{N}(0,1)$, and $\mathcal{X}_i = \{\sum_k x_{ik} \leq B\}$. The social cost $-\sum_{ik} v_{ik}x_{ik} + \beta \sum_k n_k(n_k-1)$ is quadratic in $\mathbf{x}$; since it depends on the loads only, it has an exact MILP reformulation through the epigraph $p_k \geq m\, n_k - m(m+1)/2$, $m = 1,\dots,N-1$, of $\binom{n_k}{2}$. With a common β the game is an exact potential game with potential $-\sum_{ik} v_{ik}x_{ik} + \beta \sum_k \binom{n_k}{2}$. We report results for instances with $(N,K,\beta,B) \in \{(4,8,1,2),(6,10,1.5,3),(8,12,2,3),(32,24,1.5,3),(100,32,2,3)\}$.

P2: sensor coverage with energy budget and region caps. P1 with three additional rows in $\mathbf{A}_i\mathbf{x}_i \leq \mathbf{b}_i$: an energy budget $\sum_k e_{ik}x_{ik} \leq E$ with integer costs $e_{ik} \in \{1,\dots,5\}$ and $E = 3B-1$ (a knapsack row), and a cap of $\lceil B/2 \rceil$ targets in each of two regions (halves of the targets). The agent's problem is a small knapsack; the centralized problem is the same MILP with the extra rows.

P3: exact-cardinality assignment. P1 with $\sum_k x_{ik} = B$, represented by the two constraints $\mathbf{1}^\top\mathbf{x}_i \leq B$ and $-\mathbf{1}^\top\mathbf{x}_i \leq -B$. From a feasible decision no single bit flip is feasible: every improvement is a swap, i.e. a path of length two through a decision with violation one (Assumption 1).

P4: competing project selection. N firms choose projects under a knapsack budget, $\mathcal{X}_i = \{\sum_k w_{ik}x_{ik} \leq W\}$ with $w_{ik} \in \{1,\dots,5\}$ and $W = 8$, and pay $c_i(\mathbf{x}) = \sum_k x_{ik}\big(\sum_{j\neq i} \beta_k(1 + \eta a_{ij})x_{jk} - r_{ik}\big)$ with $r_{ik} \sim U(2,8)$, $\beta_k \sim U(0.3,1.2)$ and an antisymmetric competition pattern $a_{ij} = -a_{ji} \in \{-1,+1\}$. For $\eta = 0$ this is an exact potential game; for $\eta = 0.5$ it is a genuinely near-potential game. Because a is antisymmetric the perturbation cancels in the social cost, which remains the load-based quadratic of P1 (exact MILP), while the equilibria depend on η. Instances reported: $N = 8, 32, 100$, with $K = 24$.

P5: wireless channel allocation. N transmitters at random positions select channels under a power budget $\sum_k p_{ik}x_{ik} \leq P$ ($p_{ik} \in \{1,\dots,5\}$, $P = 8$), with cost $c_i(\mathbf{x}) = \sum_k x_{ik}\big(\sum_{j\neq i} G_{ij}x_{jk} - v_{ik}\big)$, where $G_{ij} = h_j/d_{ij}^2$ (transmit power $h_j \sim U(0.5,2)$, distance d_{ij}; mean coefficient normalized to one) and $v_{ik} \sim U(1,5)$. Since $G_{ij} \neq G_{ji}$ the game is near-potential and the social cost $-\sum v_{ik}x_{ik} + \sum_k \sum_{i\neq j} G_{ij}x_{ik}x_{jk}$ is a non-convex binary quadratic program with $N(N-1)K$ product terms. Instances reported: $N = 8, 32, 100$ with $K = 24$.

P6: cubic interference. P5's setting with a symmetrized pairwise term Q_{ij}, a single capacity row ($B = 3$, $K = 16$), and a three-way intermodulation term: $c_i(\mathbf{x}) = \sum_k x_{ik}\big(-v_{ik} + \sum_{j\neq i} Q_{ij}x_{jk} + \sum_{j<l,\, j,l\neq i} C_{ijl}\, x_{jk}x_{lk}\big)$ with symmetric $C_{ijl} = \kappa(Q_{ij}Q_{il}Q_{jl})^{1/3}$, $\kappa = 0.5$. With the other agents' decisions fixed, the cost remains linear in agent i's own decision, while the social cost is a *cubic* pseudo-Boolean function; its exact linearization ($y_{jlk} \geq x_{jk} + x_{lk} - 1$, $w_{ijlk} \geq x_{ik} + x_{jk} + x_{lk} - 2$) needs $\binom{N}{3}K$ auxiliary variables (79,000 at $N = 32$, 670,000 at $N = 64$). This is an exact potential game. Instances reported: $N = 8, 16, 32, 64, 100$.

P7: autonomous vehicle–target assignment. The game of Arslan et al. (2007): vehicle i eliminates target k with probability $p_{ik} \sim U(0,1)$, a team engaging k eliminates it with probability $1 - \prod_{i\in S_k}(1-p_{ik})$, the global utility is $U_g(\mathbf{x}) = \sum_k V_k\big(1 - \prod_i (1-p_{ik})^{x_{ik}}\big)$ with $V_k \sim U(0,1)$, and the social cost is $-U_g$. Vehicle i may only engage the targets in its range $\mathcal{A}_i$ ($|\mathcal{A}_i| = 8, 15, 20$), at most $B = 2$ of them, within a fuel budget $\sum_k d_{ik}x_{ik} \leq F_i$ ($d_{ik} \in \{1,\dots,10\}$, $F_i = \lceil 1.5B\bar{d}_i \rceil$); the range restriction is the single row $\sum_{k\notin\mathcal{A}_i} x_{ik} \leq 0$. Two of the utility designs of Arslan et al. (2007) are used: the "wonderful-life utility" WLU, $c_i(\mathbf{x}) = -\sum_k x_{ik}p_{ik}V_k \prod_{l\neq i}(1-p_{lk}x_{lk})$, which is the

marginal contribution to U_g and yields an exact potential game with potential $-U_g$, and the "equally-shared utility" ESU, $c_i(\mathbf{x}) = -\sum_k x_{ik} U_k(\mathbf{x})/n_k(\mathbf{x})$, which is not aligned with U_g and gives a near-potential game. Both are linear in $\mathbf{x}_i$ with coefficients that are products over the other vehicles; with estimates the products are taken over $\hat{f}^i_{lk}$. The centralized problem is an integer program with exponential terms, solved by Gurobi. Instances reported: $(N, K) = (20, 20), (50, 50), (100, 100)$.

P8: own-target interactions. P1 with $K = 500$ targets, $B = 25$, $\beta = 1$, and a dense symmetric interaction matrix S^i between a sensor's own targets (joint monitoring of two targets is cheaper or costlier), with zero diagonal and $S^i_{kl} = S^i_{lk} \sim U(-0.6, 0.6)$ for $k < l$: $c_i(\mathbf{x}) = \sum_k x_{ik}\big(\beta \sum_{j \neq i} x_{jk} - v_{ik}\big) + \mathbf{x}_i^\top S^i \mathbf{x}_i$, where the last term equals $2\sum_{k<l} S^i_{kl} x_{ik} x_{il}$. The added term depends on $\mathbf{x}_i$ only, so the game remains an exact potential game, but the agent's problem becomes a cardinality-constrained QUBO (a non-convex binary quadratic program with $\binom{500}{2}$ terms): an exact solver returns nothing within 0.05 s, is 65–71% off after 0.2 s and needs about 1 s per call to be within 2% of its 60 s incumbent, whereas Algorithm 1 only evaluates neighbor costs ($O(K)$ per flip). No centralized optimum is computable at this size. Instances reported: $N = 8$ and $N = 32$.

C Nash gaps for competing project selection

Table 3 supplements the social costs and exact-equilibrium counts for P4 in Table 1. Each entry is the mean over 50 seeds of the final Nash gap: the largest cost reduction available to any single agent through a feasible unilateral deviation, with all other agents' true final decisions held fixed. These gaps are evaluated using exact best responses. A zero gap characterizes a Nash equilibrium; a positive gap measures the remaining incentive to deviate. The mean is taken over all runs, including those ending at an equilibrium, and is not an upper bound on every run's gap. Failure to reach an exact equilibrium within the experimental horizon does not establish that the game has no pure Nash equilibrium.

Table 3: Mean final Nash gaps over 50 seeds for competing project selection with asymmetric competition, $\eta = 0.5$ (P4). Values are rounded to one decimal place.

instance	DLS	FLS	EBR
$N = 8$	0.0	7.9	0.0
$N = 32$	1.5	4.3	1.5
$N = 100$	8.6	7.9	8.3

D Connectivity ablation

Assumption 3 only requires that, over blocks of B steps, information from every agent reaches all others with probabilities $p_{j,m}$ satisfying $\sum_m p_{j,m} = \infty$: a strongly connected base graph whose links are active with any constant probability $p > 0$ satisfies it, a disconnected base graph does not, and Theorem 1 gives no convergence rate. We measure how sparse communication affects convergence on P1 and P5 at $N = 32$ (DLS, 50 seeds, $T = 60N$) with three sweeps: (A1) the activation probability $p \in \{0.02, \ldots, 1\}$ on the complete graph; (A2) the base graph at $p = 0.6$, using Erdős–Rényi graphs $G(N, q)$ around the connectivity threshold $q^* \approx \ln N/N \approx 0.11$ and Watts–Strogatz graphs (Watts & Strogatz, 1998) with four neighbors and rewiring probability β; and (A3) vanishing probabilities $p(t) = \min\{1, 0.6(1 + t/100)^{-\gamma}\}$ over $2T$ iterations, for which $\sum_t p(t)$ diverges if and only if $\gamma \leq 1$. Every $T/20$ iterations we record the Nash gap of the true decisions; we report the number of runs ending at an exact equilibrium and the median settling time (the first checkpoint from which the profile remains an exact equilibrium).

Table 4 summarizes the results; Figure 1 plots them against each swept parameter. For every $p \geq 0.05$, 49–50 of 50 seeds reach an equilibrium with the same final cost as at $p = 0.6$; sparser links only lengthen the median hitting time (768 steps at $p = 1$ and 1,248 at $p = 0.05$ on P1, against 192 with perfect information). The base graph must be connected: the ER base graphs with $q \leq 0.15$ used here are disconnected (at $q = 0.15$ one agent is isolated), and no seed reaches an equilibrium

for $q \leq 0.15$, whereas $q \geq 0.2$ recovers 49–50 of 50. At fixed degree, rewiring a ring into a small world raises the equilibria reached on P1 from 20 to 46 of 50. With vanishing probabilities, $\gamma \leq 1$ behaves like constant p, while $\gamma = 2$ freezes the estimates before the decisions settle (1 of 50 seeds at an equilibrium on P1, 29 of 50 on P5), consistent with the condition $\sum_m p_{j,m} = \infty$ of the tracking lemma.

Table 4: Connectivity ablation at $N = 32$. Each cell shows a/b: a is the number of the 50 seeds ending at an exact Nash equilibrium and b the median hitting time in steps ("–" when fewer than half of the seeds reach one); the swept parameters are defined above.

sweep	problem	exact NE of 50 / median hitting time, per parameter value					
A1: activation probability p		$p = 1$	0.6	0.2	0.1	0.05	0.02
	P1	50/768	50/864	50/960	50/960	49/1248	20/–
	P5	49/672	50/576	50/576	50/672	50/816	45/1344
A2: ER base graph, edge probability q ($p = 0.6$)		$q = 1$	0.3	0.2	0.15	0.1	0.05
	P1	50/864	50/960	49/1152	0/–	0/–	0/–
	P5	50/576	50/672	49/864	0/–	0/–	0/–
A2: WS base graph, rewiring β ($p = 0.6$)		$\beta = 1$	0.5	0.1	0.01	0	
	P1	46/1344	40/1632	28/1872	22/–	20/–	
	P5	49/960	48/960	42/1152	44/1248	41/1296	
A3: vanishing $p(t) \propto (1 + t/100)^{-\gamma}$, horizon $2T$		$\gamma = 0$	0.5	1	1.5	2	
	P1	50/960	50/960	50/960	39/2112	1/–	
	P5	50/576	48/672	50/768	48/864	29/2400	

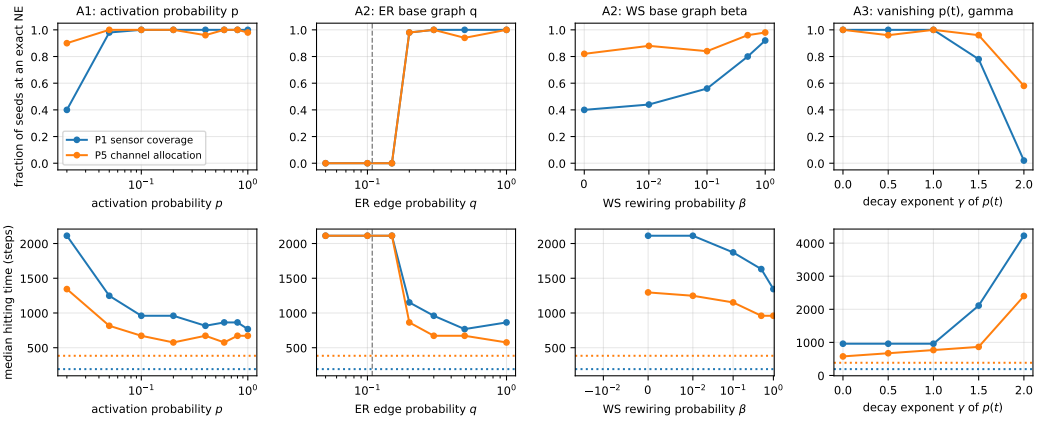


Figure 1: Connectivity ablation (P1 and P5 at $N = 32$, 50 seeds): fraction of seeds at an exact Nash equilibrium (top) and median hitting time with the perfect-information reference dotted (bottom; runs that do not reach an equilibrium are plotted at $1.1T$) against the activation probability p, the Erdős–Rényi edge probability q (dashed: $\ln N/N$), the Watts–Strogatz rewiring probability β and the decay exponent γ of $p(t)$.

E Additional experiments

P9: routing with player-specific beliefs. N players each choose one of $R = 8$ routes (one-hot decision, $\sum_r x_{ir} = 1$); route r with load σ_r costs $c_r(\sigma_r) = \phi_r \sigma_r^2 + \beta_r \sigma_r + \kappa_r$ with $\phi_r = 1/(1+2(r-1))$, $\beta_r = \kappa_r = 1$. Players hold private beliefs $\hat{\phi}_{ir} = \phi_r(1+\eta\xi_{ir})$, $\hat{\kappa}_{ir} = \kappa_r(1+\eta\xi'_{ir})$, $\xi, \xi' \sim U(-1,1)$, so $\hat{c}_i(\mathbf{x}) = \sum_r x_{ir}\left[\hat{\phi}_{ir}(1+n_{-i,r})^2 + \beta_r(1+n_{-i,r}) + \hat{\kappa}_{ir}\right]$: linear in $\mathbf{x}_i$ with coefficients quadratic in the others' load (with estimates the second moment of the load is used). For $\eta = 0$ this is a congestion game with Rosenthal potential; for $\eta = 0.3$ it is a congestion game with player-specific cost functions, which has pure equilibria but in general no potential (Milchtaich, 1996); it is a near-potential game arising from parameter uncertainty. The social cost is cubic in the binaries and is minimized exactly by a MILP with load indicators. Instances reported: $N = 20, 50, 100$ ($\eta = 0.3$) and $N = 50$ ($\eta = 0$).

Table 5 reports the results. On P9 with player-specific beliefs every rule reaches an exact equilibrium in all seeds although no potential exists, and the price of anarchy shrinks with N (4.0%, 1.4%, 0.5%); at $N = 100$ the decentralized equilibrium is better than the centralized greedy.

Table 5: Routing game with quadratic route costs and player-specific beliefs (P9).

instance	greedy	optimum	DLS	FLS	EBR	exact NE
η=0 N=50	667.535	667.535	667.535 ± 0.000	667.535 ± 0.000	667.535 ± 0.000	50/50/50
η=0.3 N=20	86.930	86.551	89.980 ± 0.787	89.670 ± 0.648	89.670 ± 0.648	50/50/50
η=0.3 N=50	606.533	599.117	607.775 ± 2.681	607.190 ± 2.351	607.190 ± 2.351	50/50/50
η=0.3 N=100	3197.507	3175.333	3190.410 ± 4.995	3190.244 ± 4.974	3190.244 ± 4.974	50/50/50

F SOCIAL-COST TRAJECTORIES OF THE DECENTRALIZED RULES

Tables 1, 2 and 5 report only the end points of the dynamics. This appendix shows the whole path: Figures 2–10 plot, for every instance of P1–P9 in the order of the tables, the social cost of the current decisions against the iteration counter of Algorithm 2.

Social cost. The social cost of a decision profile $\mathbf{x}$ is the total cost of all agents,

$$W(\mathbf{x}) = \sum_{i \in \mathcal{N}} c_i(\mathbf{x}), \tag{29}$$

evaluated with the agents' true cost functions (for P7, $W = -U_g$, the negative global utility). It is the social cost a central planner minimizes (the objective of the greedy and optimum references), whereas every agent in the dynamics minimizes its own cost c_i only; a Nash equilibrium is a rest point of the individual costs, not a minimizer of W. The trajectories therefore show how the welfare of the population evolves while the agents move toward an equilibrium; their end points are the social-cost entries of the tables, and whether the end point is an equilibrium can be read only from the tables.

For every rule (DLS, FLS and EBR; on P8 also DLS-300 and EBR-s) the social cost $W(\mathbf{x}(t))$ of the *true* decisions is recorded after every iteration $t = 0, \ldots, T$ of Algorithm 2, where $\mathbf{x}(0)$ is the random feasible initial profile (common to all rules for a given seed) and $T = 500$ for $N \leq 8$, $T = 60N$ otherwise. The line is the mean over the 50 seeds and the band one standard deviation around it. Horizontal lines mark the centralized references of the tables: greedy (orange, dashed) and optimum (red, dashed; the 600 s incumbent where the table marks it with an asterisk, absent on P8 and on P6 for $N \geq 64$). The random initial profiles cost far more than any equilibrium (4,618 against -62 at the end on P1 with $N = 100$; 20,508 against -149 on P6), so the vertical axis is clipped to the neighborhood of the final levels and the first iterations leave the panel at the top. When every curve has settled (its mean within 1% of its total decrease) long before T, the horizontal axis is cut at four times the settling time; beyond it all curves are constant.

P1: sensor coverage (Figure 2). DLS, FLS and EBR are almost indistinguishable on all five instances and settle within about a hundred iterations on the three instances with $N \leq 8$, a few hundred at $N = 32$ and about a thousand at $N = 100$. The common level lies far above the optimum; EBR, which uses exact best responses, ends at the same level, so the distance to the optimum is a property of the game's equilibria. At $N = 100$ the mean is not monotone: after a first drop to about -10 at $t \approx 200$ it rises to $+120$ at $t \approx 300$ before settling at -62, a transient in which many agents react simultaneously to estimates that lag behind the decisions.

P2: energy budget and region caps (Figure 3). At $N = 4$ FLS flattens at -24.7 while DLS and EBR reach -26.3; the gap between the curves is the 28 of 50 seeds at which FLS is stuck at a non-equilibrium. At $N = 6$ and $N = 8$ FLS ends marginally *lower* in cost than DLS and EBR (-24.8 against -24.5, -22.5 against -22.1), yet 4 of its 50 final profiles at $N = 6$ are not equilibria: a lower welfare along the path says nothing about stability, which is why Table 1 reports the equilibrium count. DLS and EBR coincide.

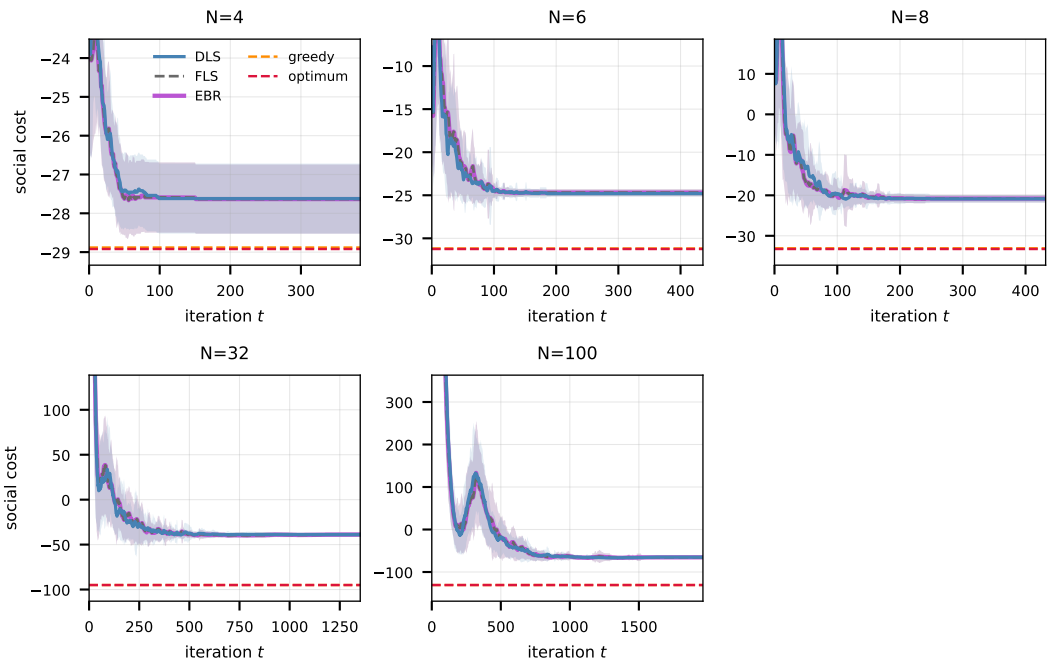


Figure 2: Sensor coverage (P1): social cost along the iterations of Algorithm 2 (mean $\pm$ std over 50 seeds) for DLS, FLS and EBR on the five instances of Table 1, with the centralized references as horizontal lines.

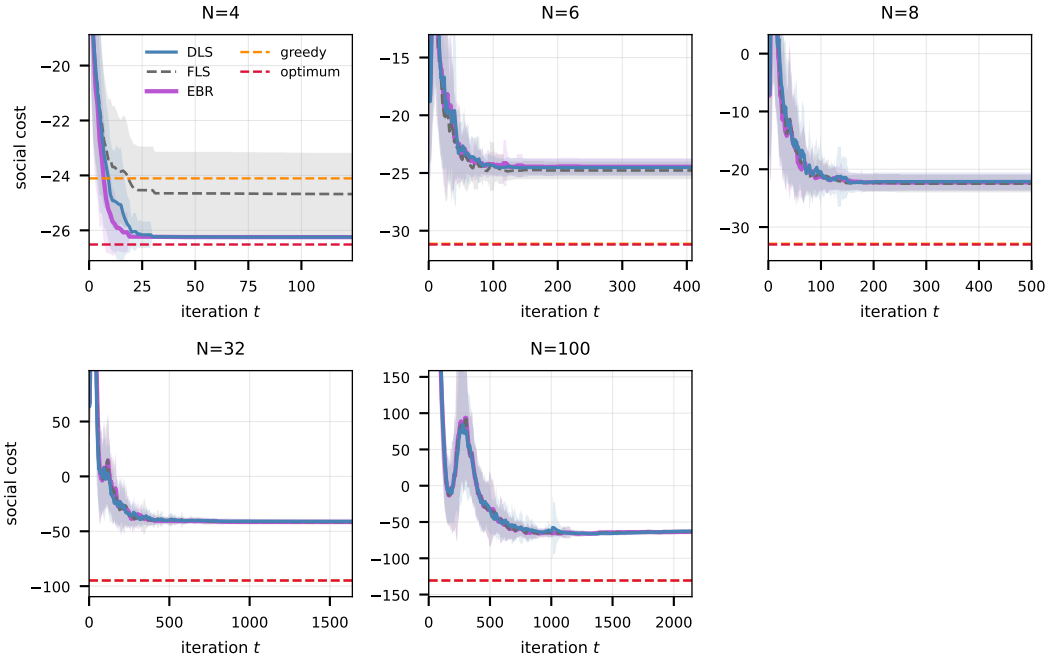


Figure 3: Sensor coverage with energy budget and region caps (P2): social-cost trajectories on the instances of Table 1.

P3: exact cardinality (Figure 4). With $\sum_k x_{ik} = B$ every improvement is a swap through an infeasible decision, and the descent is slower than on P1: the curves settle after 100–160 iterations on the instances with $N \leq 8$ and about 650 at $N = 32$, and at $N = 100$ they are still decreasing at

$T = 6{,}000$ (from 3,996 at $t = 2{,}000$ to 3,987 at the end, toward the greedy level 3,940), which is the incomplete equilibrium count of Table 1. The three rules follow the same path.

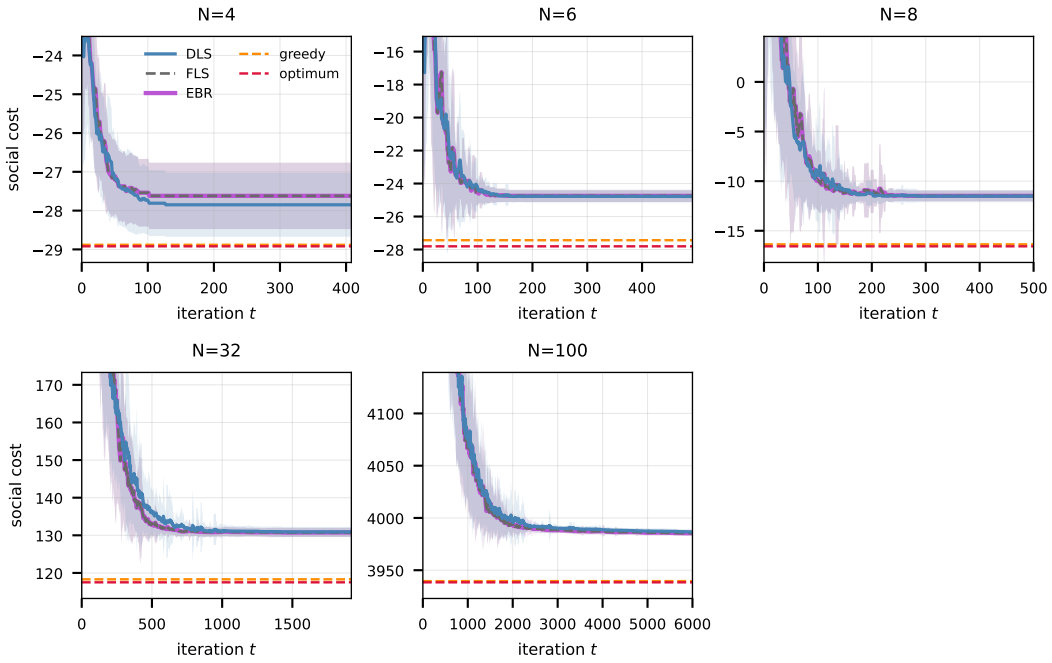


Figure 4: Exact-cardinality assignment (P3): social-cost trajectories on the instances of Table 1.

P4: competing project selection (Figure 5). At $N = 8$ FLS flattens at -215 after 60 iterations while DLS and EBR continue to -224, close to the optimum -233; none of the FLS end points is an equilibrium. At $N = 32$ and $N = 100$ ($\eta = 0.5$, the near-potential regime in which no rule reaches an exact equilibrium) the means flatten after 200–350 iterations while the profiles keep moving, and at $N = 32$ FLS ends visibly lower (-480) than DLS (-461) and EBR (-465): the feasible-only search descends further in welfare and still ends at a profile with a mean Nash gap of 4.3 (Table 1).

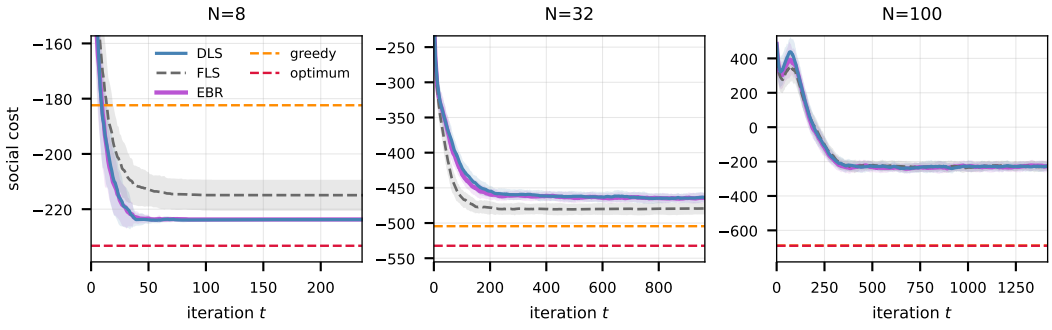


Figure 5: Competing project selection, $\eta = 0.5$ (P4): social-cost trajectories on the instances of Table 1.

P5: wireless channel allocation (Figure 6). At $N = 8$ FLS stops at -130 against -138 for DLS and EBR (optimum -140). At $N = 32$ and $N = 100$ FLS reaches the same level as DLS and EBR, although 0 and 13 of its 50 end points are equilibria (Table 1); the curves of DLS and EBR coincide and settle after about 400 and 650 iterations.

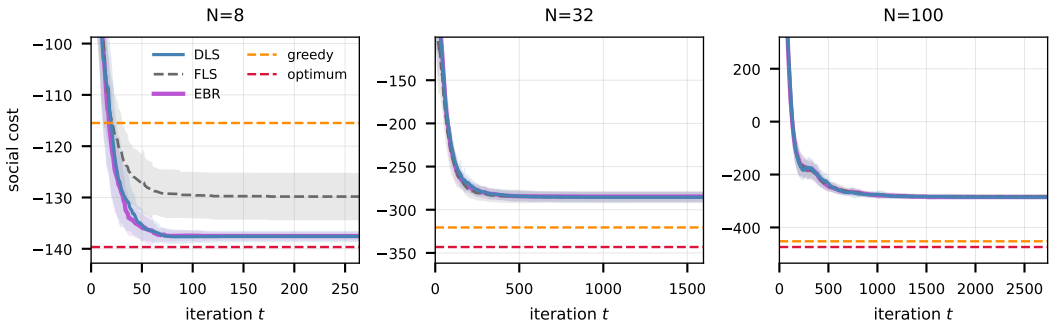


Figure 6: Wireless channel allocation (P5): social-cost trajectories on the instances of Table 1.

P6: cubic interference (Figure 7). Random profiles pile the transmitters onto the same channels, so the initial cost is huge (734 at $N = 32$, 20,508 at $N = 100$, against final levels of -167 and -149) and the first 50–100 iterations remove almost all of it; the remaining approach to the equilibrium takes another 100–200 iterations at every size. The three rules coincide. For $N \geq 64$ only the greedy reference exists (Table 1).

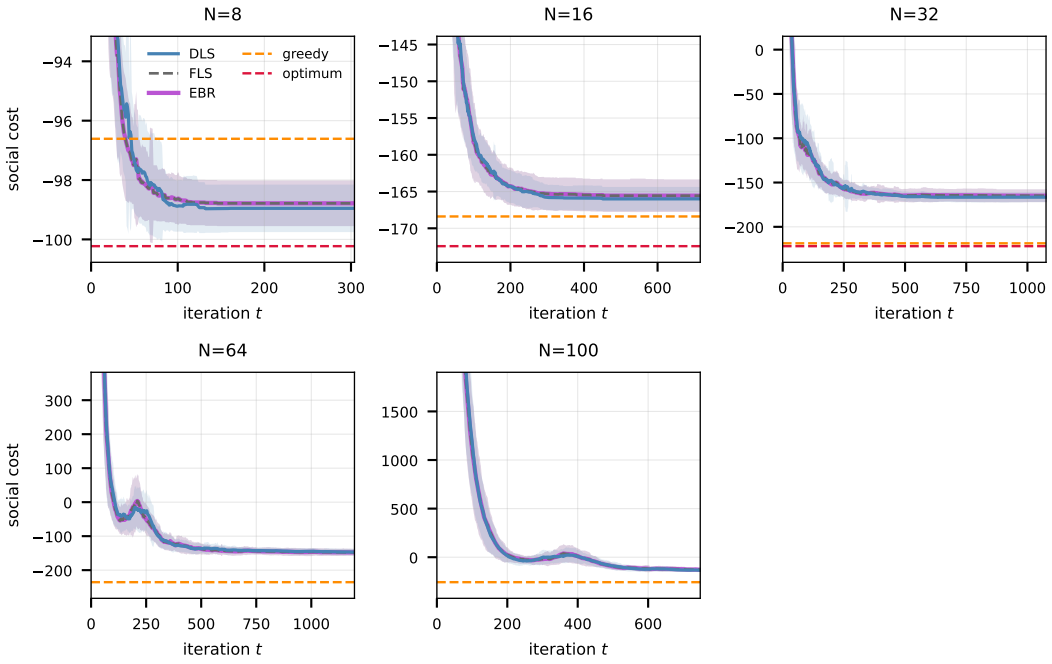


Figure 7: Cubic interference (P6): social-cost trajectories on the instances of Table 1.

P7: vehicle–target assignment (Figure 8). With the aligned utility WLU the three rules coincide and end within 0.1–0.3 of the centralized optimum of $-U_g$ at all three sizes. With the misaligned utility ESU the same rules, following the same path, settle at -8.3, -24.6 and -46.1 against optima of -9.0, -26.5 and -49.5, about 7% higher cost, and take longer to do so (about 800 instead of 750 iterations at $N = 100$, 270–350 instead of 140–155 at $N = 20$): the utility design, not the learning rule, sets the welfare of the equilibrium (Table 1).

P8: own-target interactions (Figure 9). The only problem on which the rules separate in steady state. EBR-0.2 never leaves the initial level: the solver returns no improved solution within 0.2 s and the sensors keep their decisions. EBR-0.5 and EBR-1 reach the lowest cost ($-1{,}803$ and $-1{,}810$ at $N = 8$, $-6{,}355$ and $-6{,}375$ at $N = 32$); DLS settles 3–5% above them ($-1{,}748$ and $-6{,}085$)

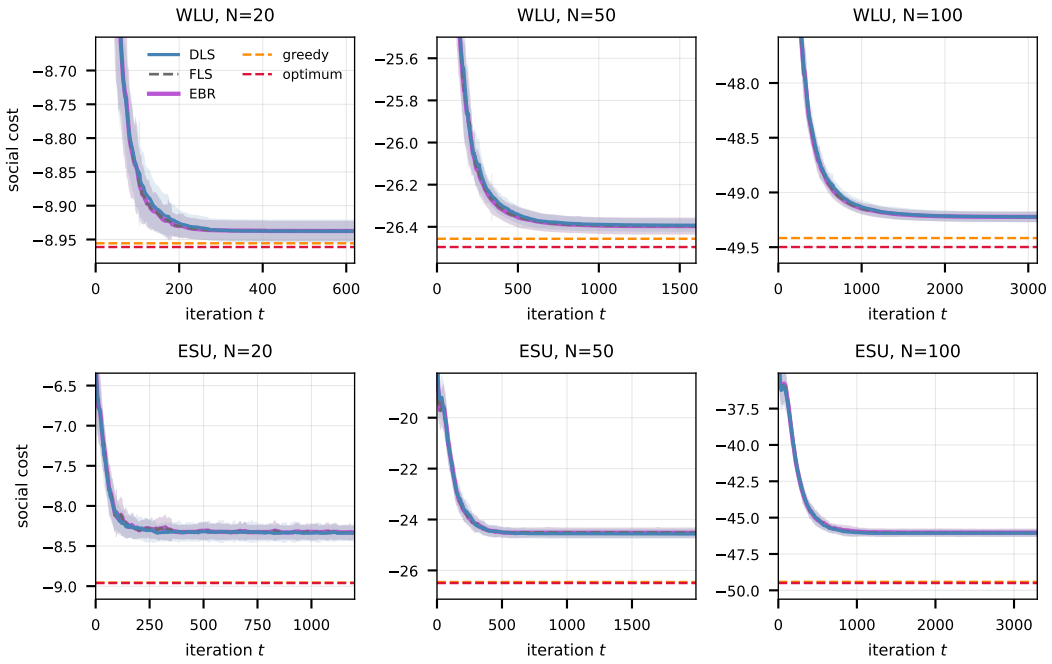


Figure 8: Vehicle–target assignment (P7) with the wonderful-life utility WLU (top) and the equally-shared utility ESU (bottom): social cost $-U_g$ along the iterations on the instances of Table 1.

and FLS slightly above DLS. DLS-300 reaches the same level as DLS about three times faster (50 instead of 140 iterations at $N = 8$, 220 instead of 630 at $N = 32$). Every rule except EBR-0.2 ends below the centralized greedy line ($-1{,}707$ and $-5{,}888$).

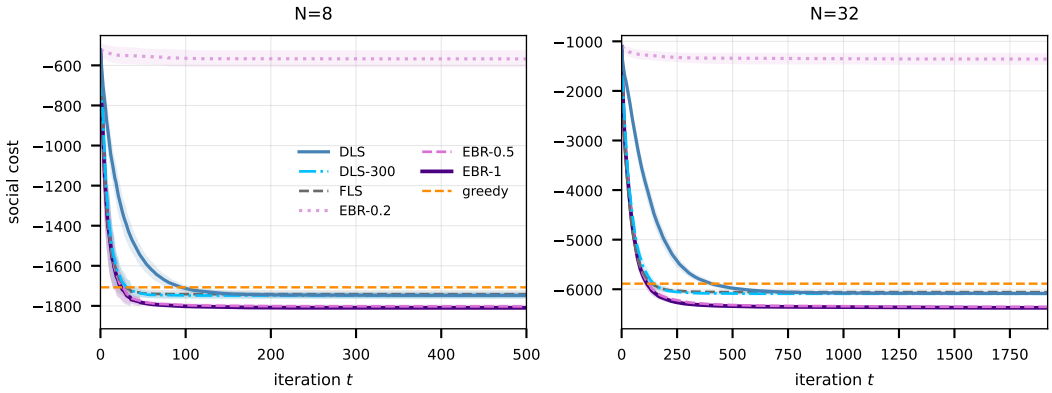


Figure 9: Own-target interactions, $K = 500$ (P8): social-cost trajectories of DLS, DLS-300, FLS and the time-limited exact best responses EBR-s on the instances of Table 2; no optimum is computable.

P9: routing with player-specific beliefs (Figure 10). For $\eta = 0$ the curves reach the optimum 667.5 (in this case also attained by greedy algorithm). For $\eta = 0.3$ the social costs settle 0.5–4% above the optimum (90.0 against 86.6 at $N = 20$, 607.8 against 599.1, 3,190 against 3,175), at the greedy level. DLS is slightly slower than FLS and EBR (about 970 against 840 iterations at $N = 100$) because the randomized search evaluates a few non-improving flips before each move, but ends at the same level (Table 5).

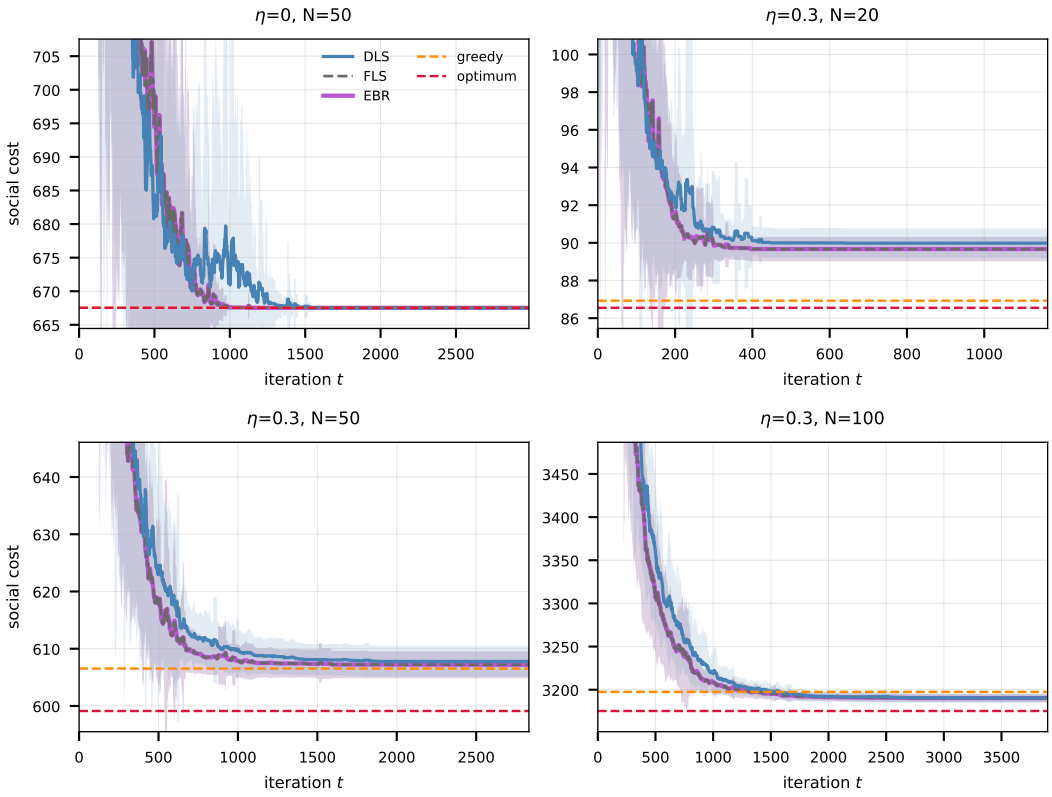


Figure 10: Routing with player-specific beliefs (P9): social-cost trajectories on the instances of Table 5.