Social Information Quality Determines Learned Collective Foraging Strategies

Valerii Chirkov[1,2,*] and Pawel Romanczuk[1,2,3]

[1]Institute for Theoretical Biology, Humboldt Universität zu Berlin, Berlin, Germany
[2]Science of Intelligence, Cluster of Excellence, Berlin, Germany
[3]Bernstein Center for Computational Neuroscience, Berlin, Germany
[*]Corresponding author: valerii.chirkov@hu-berlin.de

Abstract

Collectively foraging animals navigate a fundamental trade-off between private exploration and using social information. This social information can be multidimensional and diverse, ranging from simple positional cues to complex payoff signals. How the types of available social cues and environmental volatility shape collective foraging behavior, however, remains unclear. We address this using a spatially explicit agent-based model in which agents learn to track a mobile resource via multi-agent reinforcement learning. Agents choose between random exploration, costly private tracking, and social attraction. We systematically varied resource volatility and the type of available social cues to analyze their effect on emergent foraging behavior.

Our results show that the quality of social information dictates the group's collective behavior. Low-quality social cues (e.g., position, action) result in a fragile "Cohesive Tracking" strategy that is effective in stable environments but fails as volatility increases. Conversely, high-quality social information enables behavioral diversity. Agents selectively copy others and perform individual tracking or exploration depending on the environmental context.

Together, our findings demonstrate that the interplay between information quality and ecological context is a fundamental mechanism governing the emergence of distinct forms of collective intelligence from individual decision rules.

Keywords: Collective Behavior; Social Foraging; Social Learning; Multi-Agent Reinforcement Learning; Agent-based Simulations

1 Introduction

Animals in social foraging contexts must constantly balance the costs of discovering resources through private exploration with the benefits of using socially acquired information from their peers [1, 2]. This trade-off between exploration and exploitation is a central challenge in behavioral ecology, as relying on others can provide a valuable shortcut but also risks the propagation of outdated or erroneous information [3, 4].

To explain how animals solve this problem, classic models often rely on a fixed set of heuristics like "copy the successful" [5, 6]. While useful, these rules fail to capture how animals flexibly integrate the rich, multidimensional social information available in nature. The information animals acquire can be broadly classified into discrete social cues (e.g., an alarm call indicating presence of a predator) and graded public information (e.g., observing foraging success) [7, 8]. For instance, a foraging bird might observe a neighbor's **presence** (positional cue), see it actively **pecking** (action cue), or witness it successfully **extracting** grubs (payoff-based public information) [9]. The adaptive value of each cue is highly context-dependent. In a stable environment, simple cues may suffice. However, in a volatile environment where resources are ephemeral—such as birds tracking thermal updrafts [10] or predators hunting mobile prey [11]—an animal may need high-fidelity public information before committing to a costly behavior based on it, such as abandoning a known foraging patch to join a peer [12]. A significant gap remains in understanding how animals learn to weigh these distinct cue types based on their reliability and the surrounding ecological context.

Deep multi-agent reinforcement learning (MARL) offers a powerful framework to address this gap and challenge the dominant view that social learning relies on fixed heuristics. By treating agents as "smart particles" whose decisions and movements are optimized together [13], MARL allows for the emergence of flexible, adaptive policies from reward maximization, supporting recent efforts to move toward more holistic views on collective intelligence [14]. Prior work has shown that MARL agents can learn biologically plausible behaviors, such as using social information adaptively [15], converging on collective foraging patterns [16, 17], and spontaneously acquiring social learning heuristics [18]. Our work builds on this by providing a principled simulation-based experimental paradigm to investigate how the *quality* of social information interacts with environmental dynamics to shape emergent collective foraging behavior.

To address this, we model a social foraging scenario where agents track a mobile resource performing correlated random walk (see Fig. 1). Their behavioral repertoire is inspired by fundamental strategies observed across foraging species. Agents can engage in a random walk (**exploration**), a widely observed search pattern used when animals lack reliable directional cues [19]. Alternatively, agents can use their private information to **track** the resource. This action is modeled as inherently costly, resulting in a slower

effective tracking speed than other actions. This cost operationalizes the energetic and cognitive demands of actively sensing and pursuing a resource, such as a predator stalking prey [20], and reflects behavioral constraints like pausing to sense or orient, a constraint central to gradient-following strategies like infotaxis [21, 22]. Finally, agents can move toward a peer (**social attraction**), a core mechanism of social learning where individuals join others to exploit a resource discovered by the group [9, 23]. This action set allows us to probe the fundamental trade-off between relying on costly individual sensing and using socially acquired information. To investigate the "best possible" coordinated strategies that emerge from these trade-offs, our primary MARL approach uses centralized training (where agents share neural network weights) for the main experiments, as it is known to effectively find such group-level solutions. We also explored a fully decentralized version (where each agent learns independently, using its own distinct network) to verify the robustness of our findings. We systematically vary both the quality of social cues available (from neighbor distance to noisy payoff signals) and the environmental volatility (resource speed), allowing us to isolate how their interplay drives emergent collective foraging behavior.

Our simulations reveal that the quality of available social information dictates the group's behavioral repertoire, with the adaptive value of social cues being shaped by environmental volatility. When limited to low-quality social cues (neighbor proximity or action), agents are locked into a single strategy: **Cohesive Tracking**. This strategy is effective in stable, slow-changing environments, where agents achieve near-maximal performance (Fig. 2) by predominantly relying on costly private tracking while using social cues to maintain a cohesive group cluster (Figs. 3 and 4). However, this strategy is fragile, degrading significantly as environmental volatility increases and simple cues become unreliable. In these volatile contexts, high-quality payoff information becomes critical to sustain performance, granting agents the flexibility to adopt distinct strategies. When private tracking is viable (e.g., in slow environments or when tracking cost is low), agents adopt a **Track-or-Copy** strategy, using payoff cues to selectively follow more successful peers – a form of payoff selectivity known to be critical for collective success [6, 24, 25]. However, when private tracking becomes ineffective (due to high resource volatility or high tracking cost), the group seamlessly transitions to an **Explore-or-Copy** strategy. This latter strategy corresponds to a form of distributed collective sensing [10, 26, 27], where agents abandon costly tracking, explore randomly, and use payoff cues to aggregate at temporary information hubs created by successful explorers, dynamically switching between exploration and following successful peers (Figs. 3 and 4). Together, these results generate testable predictions for behavioral research and demonstrate that the interplay between information quality and ecological context fundamentally governs how individual decision rules give rise to distinct, adaptive forms of collective intelligence.

2 Materials and Methods

2.1 The Collective Foraging Model

2.1.1 Environment and Resource Dynamics

To investigate the interplay between environmental volatility and social information use, we frame collective foraging as a *resource-tracking* problem where agents must track a mobile resource (Fig. 1A). The task takes place in a continuous 20×20 unit square arena A with reflective boundaries. A group of N identical agents tracks a single mobile resource for $T = 1000$ discrete time steps.

Let $\mathbf{x}_R(t) \in A$ be the resource's position at time t. To simulate resources with dynamics relevant to ecological contexts, the resource is mobile and performs a random walk. Turning angles $\Delta\theta$ relative to the previous direction are sampled uniformly from $[0, 2\pi)$. Step lengths ℓ are drawn from a truncated Pareto distribution with a shape parameter $\alpha = 1$ and $1 \leq \ell \leq 100$. The resource travels this distance at a constant speed $v_{resource}$ and value controls the task's overall difficulty (see Section 2.3). This truncated power-law distribution generates occasional long relocations, mixing local diffusion with long-distance movements. While not a formal Lévy walk (as truncation ensures finite moments), this pattern characterizes real-world animal and resource movements across wide range of taxa [28–30]. Tracking this erratic target encapsulates the ecological challenge of following spatiotemporally dynamic resources that underlies migration, foraging, and hunting [20, 31].

All agents and the resource were initialized randomly for each training and evaluation episode within a 10×10 area in the center of the environment to prevent agents and the resource from starting near the arena boundaries.

2.1.2 Agent Dynamics and Reward

Let $\mathbf{x}_i(t) \in A$ be the position of agent i at time t. Agents can move at a maximum speed of $v_{\max} = 0.05$ units per step. The absolute value of this speed is less critical than its value relative to the resource speed (Section 2.3.1) and the effective tracking speed (Section 2.2.1). At each time step, agent i receives an individual reward $r_i(t)$ based on its proximity to the resource:

$$r_i(t) = \frac{1}{1 + d_i(t)^2}, \tag{1}$$

where $d_i(t) = \|\mathbf{x}_i(t) - \mathbf{x}_R(t)\|_2$ is the Euclidean distance to the resource. This function provides a strong, graded incentive for agents to minimize their distance to the resource.

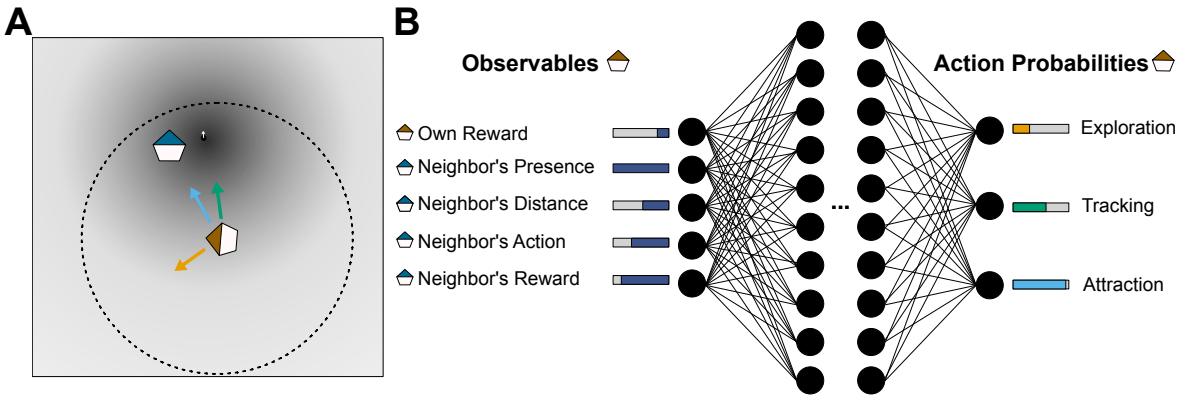

Figure 1: Multi-agent social foraging task. A: Schematic of the social foraging scenario. Agents aim to localize and track a continuously moving resource (white arrow) in a 2D environment. At each time step, the focal agent (brown) selects one of three discrete actions: random exploration (orange arrow), slow tracking toward the resource (green arrow), or social attraction toward another agent (blue arrow). The focal agent can observe one randomly selected peer (blue agent) within its vision range (dashed circle), and receives rewards based on proximity to the resource (indicated by a background gradient). **B**: Schematic of the actor network for a single agent. The network maps local, partial observations to action probabilities. Observations include the agent's own reward, a binary indicator of whether a neighbor is visible, the distance to the observed neighbor, the neighbor's current action, and the neighbor's reward. Available social features depend on the current social information condition (see Methods for details).

2.2 Agent Behavioral Model

2.2.1 Agent Actions

To probe the fundamental trade-off between private exploration and social exploitation, we designed agents choose from a repertoire of three discrete actions, $a_i(t) \in$ {Exploration, Tracking, Social Attraction} (see Fig. 1B). The chosen action dictates the agent's movement direction, i.e. the direction its current velocity vector $\mathbf{v}_i(t)$. The position is updated in discrete time by $\mathbf{x}_i(t + \Delta t) = \mathbf{x}_i(t) + \mathbf{v}_i(t)\Delta t$ with $\Delta t = 1$, subject to reflection at arena boundaries. The actions represent distinct foraging strategies observed in natural systems:

1. **Exploration** models random search behavior when reliable directional cues are lacking. The agent performs a correlated random walk at maximum speed ($v_{\max}$). The movement direction ϕ_t is updated by $\phi_t = \phi_{t-1} + \Delta\phi$, where the turning angle $\Delta\phi$ is sampled from a von Mises distribution (a circular analogue of the Gaussian distribution commonly used to model animal orientation; [see e.g. 32, 33]), $\Delta\phi \sim \text{vonMises}(\mu = 0, \kappa = 2)$. The resulting movement vector is $\mathbf{v}_i(t) = v_{\max} \cdot [\cos\phi_t, \sin\phi_t]$.
2. **Tracking** represents costly exploitation based on private information. The agent attempts to move directly toward the resource. Let $\mathbf{u}_{\text{track}}(t) = (\mathbf{x}_R(t) - \mathbf{x}_i(t))/d_i(t)$ be the unit vector toward the resource. The movement is stochastic to operationalize

the costs associated with acquiring and acting on private information (e.g., pausing to sense, orient, handle prey, and related to strategies like infotaxis [21]):

$$\mathbf{v}_i(t) = \begin{cases} v_{\max} \cdot \mathbf{u}_{\text{track}}(t) & \text{with probability } p_{\text{track}} \\ \mathbf{0} & \text{with probability } 1 - p_{\text{track}} \end{cases} \tag{2}$$

The probability $1-p_{\text{track}}$ of remaining stationary can be interpreted as the temporal cost of selecting the tracking action; a higher cost (larger $1-p_{\text{track}}$) directly reduces the agent's effective movement speed when tracking. This yields an effective average tracking speed $v_{\text{tracking}} = p_{\text{track}} \times v_{\max}$. For our main experiments, $p_{\text{track}} = 0.1$, resulting in $v_{\text{tracking}} = 0.1 \times v_{\max}$. The cost parameter $1 - p_{\text{track}}$ was varied in robustness analyses (see Section 2.5).

3. **Social attraction** represents leveraging social information by moving towards peers, potentially avoiding the cost of slow tracking, a core mechanism of social foraging [2]. The agent moves at maximum speed ($v_{\max}$) toward its observed peer j. Let $\mathbf{u}_{\text{social}}(t) = (\mathbf{x}_j(t) - \mathbf{x}_i(t))/\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|_2$ be the unit vector toward the peer. The movement vector is $\mathbf{v}_i(t) = v_{\max} \cdot \mathbf{u}_{\text{social}}(t)$.

The "Tracking" and "Exploration" actions were always available. "Social Attraction" was only meaningful if a peer was visible. To maintain a consistent action space while preventing the policy network from processing undefined inputs when social cues were absent, we introduced a penalty mechanism. If an agent selected "Social Attraction" when no peer was visible (due to visual range limits or the experimental condition lacking social cues), it received a large penalty $r_{\text{penalty}} = -1$, and its movement direction remained unchanged ($\mathbf{v}_i(t) = \mathbf{v}_i(t-1)$). This ensures the policy learns to avoid inapplicable actions. Such invalid selections are rare, typically occurring only early in training or in edge cases with limited visibility, and this penalty mechanism had negligible impact on the final learned strategies presented.

2.2.2 Agent Observations

Agents operate under partial observability, receiving only local information. To model the different types of social cue (distance, action, payoff), each agent i perceives its environment through a five-dimensional observation vector $\mathbf{o}_i(t)$, showed schematically in Fig. 1B. This is constructed based on a visual range $r_{\text{vis}} = 15$ units (unless varied in robustness checks, see Section 2.5).

Let $V_i(t) = \{j \neq i \mid \|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|_2 \leq r_{\text{vis}}\}$ be the set of peers visible to agent i at time t. To model perceptual and attentional constraints often observed in animal groups [e.g. 34], and to simplify the analysis by avoiding complex social information aggregation rules, if $V_i(t)$ is not empty, agent i attends to *exactly one* peer j, sampled uniformly at random from $V_i(t)$ at each time step. If $V_i(t)$ is empty, no social information is gathered.

Because the focal neighbor is resampled independently at each step, the attention process is ergodic: given enough time, each pair of agents will be observed with equal frequency, provided that the visibility graph remains connected. This corresponds to an effective averaging of interactions across neighbors over time. Thus, our qualitative results are representative of a setting where each agent observes *all* neighbors simultaneously in the limit of a slowly changing environment, analogous to results on synchronization in blinking network models [35, 36].

The observation vector $\mathbf{o}_i(t) = (o_1, o_2, o_3, o_4, o_5)$ components represent different information types:

1. $o_1 = r_i(t)$: Private information about own success (reward).
2. $o_2 = \mathbb{I}(V_i(t) \neq \emptyset)$: Basic social cue indicating neighbor presence.
3. $o_3 = d_{ij}(t)$: Positional social cue (distance to observed peer j). Set to 0 if $o_2 = 0$.
4. $o_4 = a_j(t-1)$: Behavioral social cue (observed peer j's previous action). Set to 0 if $o_2 = 0$.
5. $o_5 = \tilde{r}_j(t)$: Payoff-based public information (observed peer j's reward), corrupted by additive Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$ to model imperfect observation. $\tilde{r}_j(t) = r_j(t) + \epsilon$. Set to 0 if $o_2 = 0$.

The availability of social components (o_3, o_4, o_5) and the noise level σ on the payoff observation (o_5) were key manipulations in our experimental design.

2.3 Experimental Design

To directly address the central question of how environmental dynamics interact with the quality of social information, we systematically varied two factors: Resource Speed (proxy for environmental volatility) and Social Information Quality (controlling the type and fidelity of available social cues).

2.3.1 Factor 1: Resource Speed

To simulate environments ranging from stable to volatile, we varied resource speed (v_{resource}) relative to the agents' maximum speed ($v_{\max}$). The three levels tested were:

- **Slow:** Resource speed 0.1 ($v_{\text{resource}} = 0.1 \times v_{\max}$)
- **Medium:** Resource speed 0.3 ($v_{\text{resource}} = 0.3 \times v_{\max}$)
- **Fast:** Resource speed 0.5 ($v_{\text{resource}} = 0.5 \times v_{\max}$)

2.3.2 Factor 2: Social Information Quality

To explore the value of different social cues (distance, action, payoff) and the fidelity of the payoff signal, we designed seven experimental conditions that progressively increased the information available in the agent's observation vector ($\mathbf{o}_i(t)$). Starting from purely

private information, features corresponding to different social cues were added incrementally:

1. **Private:** Only o_1 (own reward) available.
2. **+Distance:** o_1, o_2, o_3 available (reward, presence, distance).
3. **+Action:** o_1, o_2, o_3, o_4 available (reward, presence, distance, peer action).
4. **+Payoff (High Noise):** All components $o_1, \ldots, o_5$ available, with high noise on peer reward o_5 ($\sigma = 0.1$).
5. **+Payoff (Medium Noise):** All components available, with medium noise on o_5 ($\sigma = 0.05$).
6. **+Payoff (Low Noise):** All components available, with low noise on o_5 ($\sigma = 0.01$).
7. **+Payoff (No Noise):** All components available, with **no noise** on o_5.

2.4 Multi-Agent Reinforcement Learning (MARL)

2.4.1 Training Algorithm

We adopt a MARL approach to train agents to maximize their individual rewards, allowing collective strategies to emerge. Specifically, we used Multi-Agent Proximal Policy Optimization (MAPPO) [37, 38], a widely used algorithm based on the centralized training–decentralized execution (CTDE) paradigm. CTDE is well-suited for partially observable multi-agent problems; the centralized critic leverages global information during training to stabilize learning and improve coordination [39], while the decentralized actors learn policies based only on local observations, ensuring scalability during execution. All agents shared the parameters of a single actor policy network and a single centralized critic network.

2.4.2 Policy and Critic Architecture

Both the actor (policy) and critic (value) networks were multi-layer perceptrons (MLPs) with two hidden layers of 256 units each, using tanh activation functions. The actor network mapped local observations $\mathbf{o}_i(t)$ to probabilities for the three discrete actions via a final sigmoid layer and a categorical distribution (see schematic illustration in Fig. 1B). The critic network mapped the joint state (approximated using all agents' local observations during centralized training) to a scalar state-value estimate. We utilized standard PPO components: a clipped surrogate objective for the policy loss, a mean squared error loss for the value function, and an entropy bonus to encourage exploration. Advantages were calculated using Generalized Advantage Estimation (GAE) [40] with advantage normalization and value clipping applied. Key training hyperparameters are detailed in Supplementary Information Table S2.

2.4.3 Training and Implementation Details

The simulation environment was implemented using the Vectorized Multi-Agent Simulator `VMAS` v1.4.3 [41]. All MARL training was conducted using `PyTorch` v2.5.1 [42] and the `TorchRL` library v0.6.0 [39], using its `TensorDict` v0.6.2 data structures. A separate policy was trained from scratch for each unique combination of Resource Speed, Social Information Quality, and the training seed. Each training run encompassed 480 iterations, gathering 500,000 environment frames per iteration (equivalent to 500 episodes of 1000 steps). This resulted in a total of 240 million environment interactions per run. Training was performed on a single NVIDIA A100 or V100S GPU, with an average runtime of 17.9 hours (std=1.8). All runs used distinct random seeds for initialization and data generation. Training metrics were logged using `wandb` v0.19.2 [43]. Figures were generated using `Matplotlib` v3.9.0 [44] and `Seaborn` v0.13.2 [45]. See the details of the third-party software in Table S1.

2.5 Robustness and Sensitivity Analyses

To assess the robustness of the emergent strategies and the sensitivity of our findings to specific modeling choices (results presented in Section 3.4), we conducted several additional sets of experiments. For the Tracking Cost and Decentralized Training analyses, the training procedure, number of seeds (seven), evaluation methodology, and strategy consistency analysis followed the same protocols described for the main experiments.

- **Tracking Cost / Effective Speed:** We systematically varied the tracking cost $1 - p_{\text{track}}$ (Eq. (2)) to explore how the intrinsic cost of private information affects strategy emergence. We tested levels $1 - p_{\text{track}} \in \{0.7, 0.8, 0.95\}$, corresponding to effective tracking speeds $v_{\text{tracking}} \in \{0.3 \times v_{\max}, 0.2 \times v_{\max}, 0.05 \times v_{\max}\}$.
- **Decentralized Training:** To ensure our results were not solely dependent on the centralized training setup, we replicated the main experiments using a fully decentralized training algorithm (Independent PPO, IPPO [46]). In IPPO, each agent learns independently using its own actor and critic networks, with no parameter sharing or access to other agents' information during training.
- **Visual Range:** To test sensitivity to perceptual limitations, we repeated the main experiments with reduced visual ranges, setting $r_{\text{vis}} \in \{10, 5, 1\}$ units, compared to the default $r_{\text{vis}} = 15$. Performance was unchanged for $r_{\text{vis}} = 10$ and $r_{\text{vis}} = 5$, and deteriorated only at the minimal range of $r_{\text{vis}} = 1$ (results detailed in Supplementary Information Section S3).

2.6 Data Collection and Evaluation Procedure

2.6.1 Performance Evaluation

Following best practices for RL reproducibility [47], each experimental condition was trained using seven independent random seeds. Learned policies were evaluated every 20 training iterations during the final phase of training (iterations 400 to 480) to assess performance near convergence. Each evaluation consisted of 1000 independent rollout episodes per seed. This resulted in 5 evaluation points $\times$ 1000 episodes/point $=$ 5000 evaluation episodes per seed. Final performance metrics (e.g., average reward) reported in the Results are averaged across these 5000 episodes and then across the 7 seeds. Error bars represent 95% confidence intervals computed across the 7 independent seeds only.

2.6.2 Strategy Consistency Analysis

To confirm that the training process consistently converged to similar behavioral strategies across different random seeds within the same condition, we performed a post-hoc analysis of behavioral strategy similarity. For each condition, we first calculated the average action composition vector (proportion of time spent in Exploration, Tracking, Social Attraction, analogous to the behavioral analysis described in the Results Section 3.2) for each of the 7 seeds, averaging across all 5000 evaluation episodes for that seed. We then computed the pairwise cosine similarity between these 7 strategy vectors.

Agglomerative hierarchical clustering was performed on the cosine distance matrix ($1-$ similarity) derived from these vectors. We examined the first bifurcation (split at the root) of the clustering tree. If the average cosine similarity *between* the two resulting clusters fell below a threshold of 0.9, we flagged the smaller cluster as representing a divergent or potentially unconverged strategy. In such cases, the runs belonging to the smaller cluster were excluded from the main analysis to ensure reported results reflect the dominant, consistent behavioral mode. This conservative pruning procedure, implemented using `SciPy` [48] and `Scikit-learn` [49], affected only 5 out of 147 total training runs (approx. 3.5%), mostly in non-social conditions. Overall consistency within the retained dominant clusters was very high (average cosine similarity > 0.95), confirming the robustness of the emergent strategies reported (details in Supplementary Information Table S3).

3 Results

We trained reinforcement learning agents to collectively track a mobile resource, systematically varying the quality of social information and the volatility of the environment. Our results reveal a fundamental principle: the quality of social information determines a group's behavioral flexibility. With only low-quality cues, agents are locked into a sin-

gle, cohesive tracking strategy. With access to high-quality information, however, the group gains the ability to flexibly switch between distinct strategies—either tracking and copying successful peers, or exploring and copying them—depending on the ecological context. We show that this strategic choice is governed by a simple trade-off between an agent's effective tracking speed (v_{tracking}) and the resource's own speed (v_{resource}). As the agents' maximum speed ($v_{\max}$) was held constant, all speeds are reported relative to this value, which does not affect comparisons between conditions.

To present these findings, this section is structured as follows. We first establish that the adaptive value of any given social cue is context-dependent, with high-quality payoff signals becoming essential as environmental volatility increases. We then explore how these performance differences arise from shifts in individual agent behavior, leading to the emergence of collective tracking strategies. Finally, we confirm that these emergent strategies are robust outcomes of the agent-environment interaction by testing their stability against changes in both intrinsic costs and the learning algorithm itself.

3.1 The value of social information depends on environmental volatility

To understand how information quality and environmental volatility shape performance, we first established their interactive effect on foraging success. As predicted, access to higher quality social information consistently improved agent performance, but the magnitude of this benefit was dictated by environmental volatility (Fig. 2). We denote resource speed by v_{resource}, the normalized reward by R, and changes in the reward by ΔR. Unless stated otherwise, we report medians with 95% bootstrap confidence intervals in brackets; full statistics can be found in Supplementary Information Table S4.

In stable environments with a slow-moving resource ($v_{\text{resource}} = 0.1 \times v_{\max}$), agents gained the most from low-quality social cues. Reward jumped substantially when distance was added ($\Delta R = 0.409[0.367, 0.424]$), and again with a neighbor's action state ($\Delta R = 0.202[0.139, 0.214]$). Higher-quality payoff signals offered only modest further improvements, suggesting that in predictable environments, simple social cues are sufficient for effective coordination.

This pattern reversed in volatile environments. For fast-moving resources ($v_{\text{resource}} = 0.5 \times v_{\max}$), the limitations of low-quality cues became apparent. While neighbor distance still provided an initial benefit ($\Delta R = 0.291[0.285, 0.294]$), information about a neighbor's action was of little value ($\Delta R = 0.049[0.041, 0.060]$). Instead, the most substantial performance gains came from access to high-quality payoff signals, with foraging success increasing by $\Delta R = 0.252[0.250, 0.256]$ when payoff noise was reduced from high ($\sigma = 0.1$) to moderate ($\sigma = 0.05$). This highlights that as environmental unpredictability increases, the ability to accurately assess the success of others becomes critical. Interestingly, in

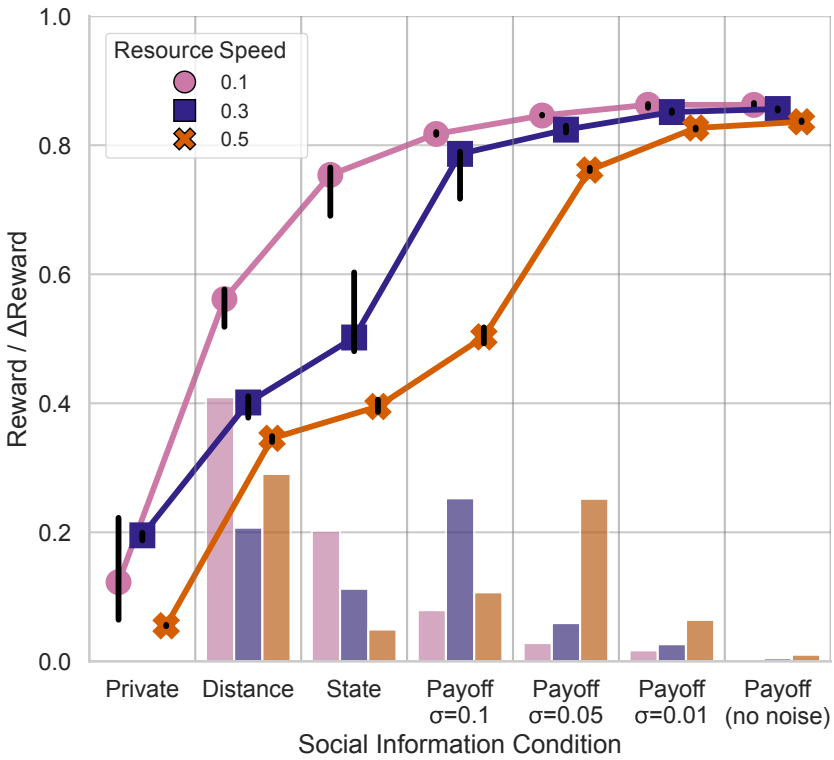

Figure 2: Effects of social information quality on agents' performance across resource speeds. Median normalized rewards (±95% CI) are shown for agents operating under seven social information conditions, across three resource speeds. Point markers indicate median reward, while bars represent the relative improvement in reward compared to the immediately preceding social information condition. Social information conditions **incrementally** add features: starting from private information only, followed by access to (1) distance to a neighbor, (2) neighbor's current action (state), and (3) neighbor's reward signal with decreasing noise ($\sigma = 0.1$, $\sigma = 0.05$, $\sigma = 0.01$, and no noise). Marker shapes and bar colors indicate resource speeds (relative to the constant $v_{\max}$).

all environments, performance saturated before payoff signals became noiseless (minimal gains for $\sigma < 0.05$), demonstrating that the learned strategies are robust to the moderate levels of informational noise characteristic of natural systems [1, 27].

3.2 Adaptive shifts in individual foraging strategy

The performance gains documented above were driven by fundamental shifts in the agents' individual behavioral strategies. We examined how agents allocated time among the three available actions—tracking, exploration, and social attraction—and found that their collective behavior emerged from context-dependent changes in these individual decisions (Fig. 3). Unless stated otherwise, we report means with 95% bootstrap confidence intervals in brackets; full statistics can be found in Supplementary Information Table S5.

In stable, slow-moving environments ($v_{\text{resource}} = 0.1$), tracking was the dominant behavior across all conditions. However, as environmental volatility increased, the cost-benefit ratio of private tracking shifted. This prompted agents to rely more on other

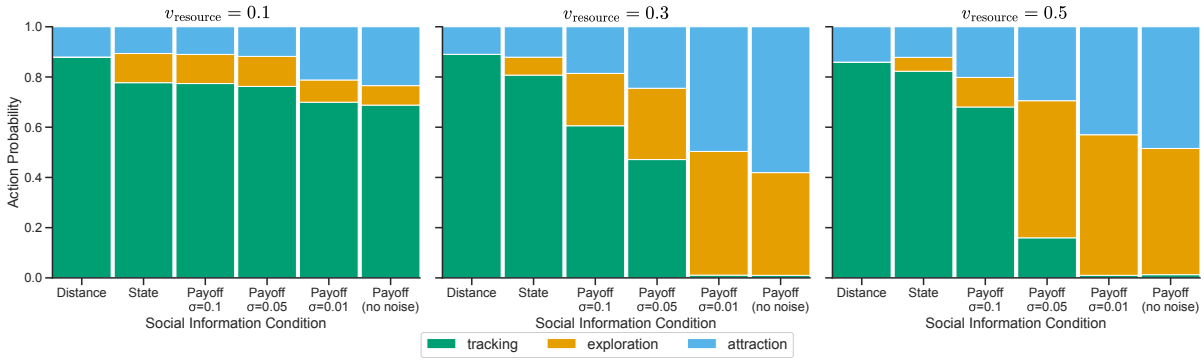

Figure 3: Effect of social information conditions and resource speeds of behavioral strategy. Bars show the average proportion of time agents spent in each behavioral mode—tracking (green), exploration (orange), and social attraction (blue)—under different combinations of social information and resource speed. Social conditions incrementally add features: starting from private information only, followed by access to (1) distance to a neighbor, (2) neighbor's current action (state), and (3) neighbor's reward signal with decreasing noise ($\sigma = 0.1$, $\sigma = 0.05$, $\sigma = 0.01$, and no noise). Resource speeds are relative to the constant $v_{\max}$.

behaviors in their repertoire. The specific type of social information available determined how this shift occurred. With access to only neighbor distance, agents completely abandoned exploration (mean exploration $= 0[0, 0]$; Fig. 3, leftmost bars). We interpret this striking behavioral shift as an emergent solution to the problem of social noise; by collectively ceasing to explore, the group effectively enhances the reliability of the social signal, ensuring that an agent's position is more strongly correlated with the resource location and thus increasing the value of simple proximity cues [50].

When high-quality payoff signals were available, a more dramatic transformation occurred. Agents sharply reduced their reliance on costly private tracking, instead adopting a mixed strategy of exploration and social attraction. In the most volatile environment ($v_{\text{resource}} = 0.5$) with noiseless payoff signals, agents spent almost no time tracking ($1.3\%[0.010, 0.016]$), dividing their efforts almost equally between exploration ($50.3\%[0.470, 0.536]$) and social attraction ($48.5\%[0.449, 0.518]$). This strategic shift enabled a fundamentally different, and more effective, collective response to environmental circumstances.

3.3 Emergence of distinct collective behaviors

The shifts in individual behavior give rise to distinct, group-level strategies. As summarized by the conceptual flowchart in Fig. 5A, the quality of social information is the primary determinant of the group's behavioral flexibility. Low-quality cues constrain the group to a single strategy, whereas high-quality cues enable a flexible response based on the viability of private tracking. Below, we detail the three major emergent strategies, focusing on the most volatile environment ($v_{\text{resource}} = 0.5 \times v_{\max}$).

Cohesive Tracking When agents have access only to low-quality social information (e.g., neighbor distance), they are locked into a single, inflexible strategy: Cohesive Tracking. In this mode, agents predominantly rely on private tracking while using social attraction non-selectively to maintain group cohesion. The underlying rule is simple: an agent follows a peer unless it is already very close to the resource or the peer is within immediate proximity (Fig. 4, top panels; Supplementary Videos 1). This results in a compact, mobile cluster that moves as a coherent unit, reminiscent of collective movements where leaders "pull" followers along [51]. While effective in predictable environments, the strategy's strong reliance on private tracking means its performance is vulnerable when conditions change, such as when the resource becomes more volatile or the costs of tracking increase.

Track-or-Copy When high-quality payoff information is available and private tracking is viable, a more straightforward Track-or-Copy strategy emerges. Here, agents still default to private tracking but now use social information selectively. By observing a neighbor's payoff at each time step, they identify opportunities to copy a peer that is having higher success (Fig. 5B, bottom-left panel; Supplementary Videos 2). This represents a more efficient use of social information than simple cohesion.

Explore-or-Copy and Distributed Sensing When high-quality payoff information is available but private tracking is rendered ineffective by high environmental volatility or tracking costs, the group flexibly switches to an Explore-or-Copy strategy. While this strategy uses the same social rule as Track-or-Copy—copying a more successful peer—it is fundamentally different in its default asocial behavior. Here, agents abandon costly tracking altogether and instead engage in random exploration. They use social information opportunistically, following a rule akin to the "copy the successful" heuristic [6]: explore randomly, but if a more successful neighbor with a higher payoff is detected, copy them (Fig. 4, bottom panels; Fig. 5B, bottom-right panel Supplementary Videos 3). This individual rule gives rise to a powerful emergent behavior that functions as a form of distributed collective sensing [10, 26, 27]. Successful explorers become temporary information hubs, attracting other agents and allowing the group as a whole to effectively track the resource in a decentralized manner.

3.4 Robustness of Emergent Strategies

To further test the robustness of our findings, we conducted two additional sets of experiments where we manipulated key aspects of our model: the intrinsic cost of tracking and the training algorithm itself.

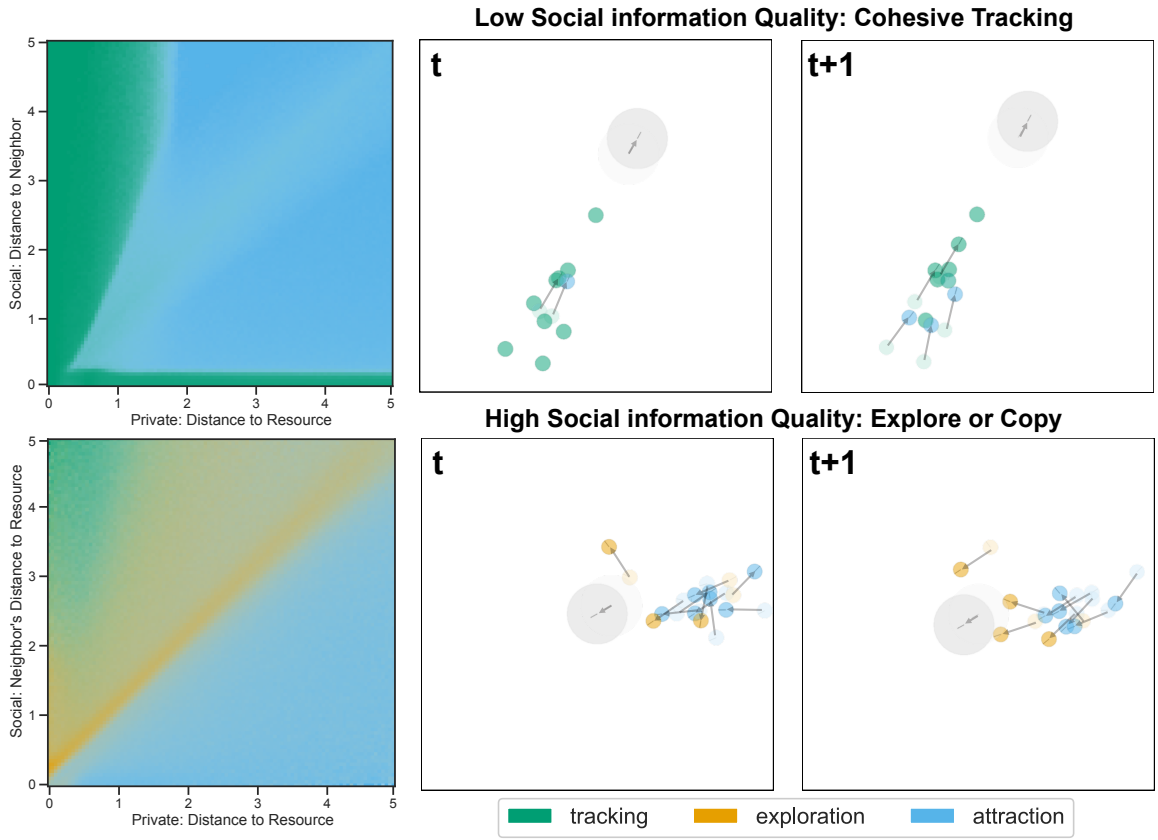

Figure 4: Emergent behavioral rules and group-level patterns. This figure shows the emergent strategies in the most volatile environment ($v_{\text{resource}} = 0.5 \times v_{\max}$). The **top row** corresponds to the low-quality "distance-only" condition, showing the Cohesive Tracking strategy. The **bottom row** corresponds to the high-quality "noiseless payoff" condition, showing the Explore-or-Copy strategy. **Left column**: Averaged action probability maps as a function of private distance to resource (x-axis) and social feature (y-axis). Color intensity reflects the probability of each action: tracking (green), exploration (orange), and social attraction (blue). **Middle and right columns**: Successive frames (t and $t{+}1$). Agent colors mark current behavioral states; vectors from their centers show velocity. The large gray disk is the moving resource. The $t{-}1$ and t frames (semitransparent) are superimposed on the t and $t{+}1$ frames, respectively, to highlight motion (arrows). Frames are cropped to the active region of the environment. See Supplemental Materials for full videos.

Effective Tracking Speed In our main results, we used a fixed *effective tracking speed* $v_{\text{tracking}} = 0.1 \times v_{\max}$ (resulting from $p_{\text{track}} = 0.1$). To test the robustness of our findings, we systematically varied the effective tracking speed, testing levels $v_{\text{tracking}} \in \{0.3 \times v_{\max}, 0.2 \times v_{\max}, 0.05 \times v_{\max}\}$ (corresponding to $p_{\text{track}} \in \{0.3, 0.2, 0.05\}$). Not surprisingly, a lower effective tracking speed (i.e., higher tracking cost) disproportionately harms performance in low-quality information conditions ("distance" and "action"), as agents in these scenarios rely heavily on the Cohesive Tracking strategy (Fig. S2). As the effective tracking speed is reduced (i.e., tracking cost is increased), tracking is progressively "pushed out" of the agents' learned policies and replaced by exploration and attraction (Fig. 5B). With high-quality information, this leads to a complete strategic shift: for a high tracking speed ($v_{\text{tracking}} = 0.3 \times v_{\max}$), Track-and-Copy is preferred, but for a low tracking speed ($v_{\text{tracking}} = 0.05 \times v_{\max}$), agents switch entirely to Explore-and-Copy (Fig. S3). This confirms that the trade-off between tracking viability and social information use is a primary driver of strategy selection, mirroring the behavioral transition observed when environmental volatility increases (see Fig. 3).

Decentralized Training The main results presented used a centralized training procedure (MAPPO) known to foster coordination. To test if our findings depend specifically on this, we reproduced our results using a fully decentralized algorithm (IPPO) where each agent learned fully independently (see Methods section for more information). The same fundamental strategies emerged (Fig. S5B-C), confirming the robustness of the core relationships between information quality, environmental volatility, and strategy selection. However, decentralized training led to a noticeable drop in performance, particularly in low-quality information conditions (Fig. S4) and an increased reliance on social cues across all settings (Fig. S5A). Interestingly, the learned policies in the 'distance-only' condition under decentralized training showed a stronger prioritization of tracking over social attraction compared to the centralized case (compare Fig. 4 top-left and Fig. S5B). This suggests that the Cohesive Tracking strategy, which relies on maintaining a tight cluster, may require a degree of cooperation facilitated by centralized training, while decentralized learning appears to favor individual tracking, potentially at the expense of group cohesion. These findings highlight that Cohesive Tracking is sensitive not only to ecological conditions, but also to learning dynamics.

Together, these manipulations reveal a key principle about the adaptive value of high-quality social information. When only low-quality cues are available, the group is effectively locked into the Cohesive Tracking strategy. This makes the collective system fragile; its success is highly sensitive to changes in the viability of tracking, as determined by its intrinsic costs or the training conditions. In contrast, access to high-quality social information makes the entire system more flexible and robust. With access to reliable payoff cues, agents are no longer constrained to a single strategy. They can flexibly switch

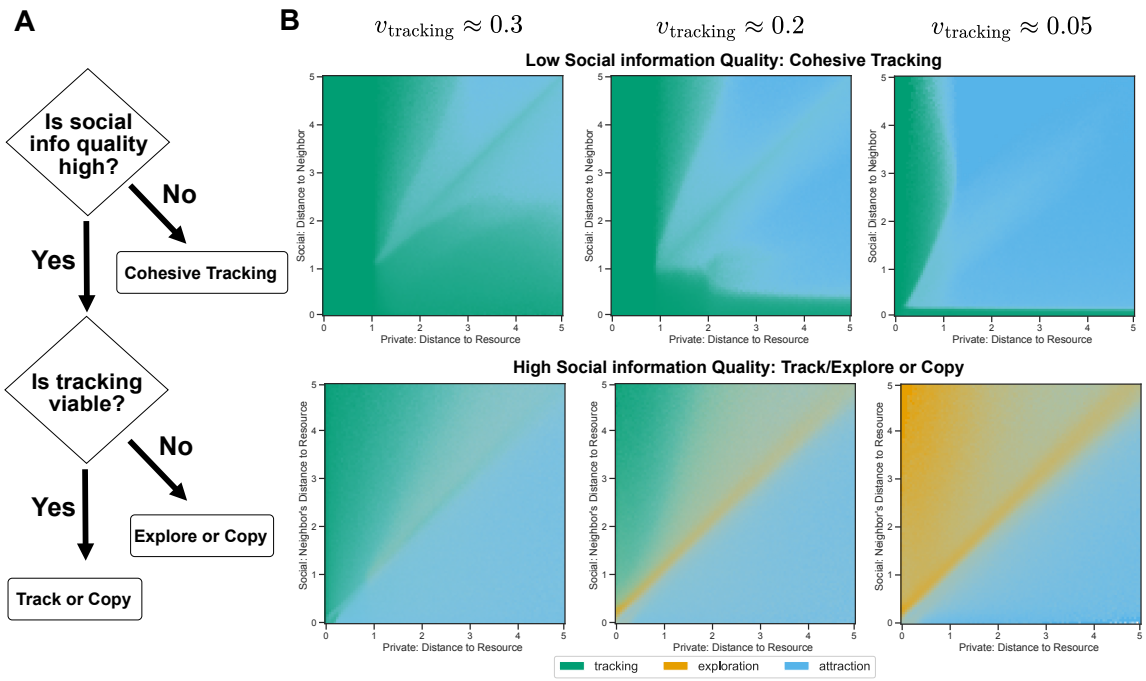

Figure 5: Ecological Conditions Determine the Emergence of Collective Foraging Strategies. A: Schematic illustration of how the interplay of social information quality and the viability of private tracking determines which collective strategy emerges. **B**: The plots show the averaged action probabilities for agents in the most volatile environment ($v_{\text{resource}} = 0.5 \times v_{\max}$) under varying effective tracking speeds ($v_{\text{tracking}} = 0.3 \times v_{\max}, 0.2 \times v_{\max}$, and $0.05 \times v_{\max}$ from left to right). Colors indicate the probability of tracking (green), exploration (orange), or social attraction (blue). The **top row** (low-quality cues) shows that agents remain locked in the Cohesive Tracking strategy, increasing their use of social attraction as tracking becomes less viable. The **bottom row** (high-quality cues) demonstrates behavioral flexibility, showing a clear transition from a Track-or-Copy strategy (left) to an Explore-or-Copy strategy (right) as tracking viability decreases.

between Track-and-Copy and Explore-and-Copy depending on the ecological context. For example, if tracking becomes too costly, the group seamlessly transitions to distributed sensing without a substantial loss in performance. Therefore, high-quality social information does not just enable a single, better strategy; it allows the group to maintain a repertoire of effective strategies and deploy the appropriate one as conditions change, ensuring the collective foraging outcome remains successful.

4 Discussion

Our study demonstrates how environmental dynamics and the availability of social information interactively shape the emergence of collective foraging strategies. By systematically varying the type and fidelity of social features in a simulated social foraging task, we showed that agents adaptively integrate social cues based on their predictive value within specific ecological contexts. We identified two distinct behavioral regimes determined

by information quality. The first is a fragile, cohesive strategy reliant on low-quality cues. The second is flexible and adaptive, emerging only when high-quality information is available.

With access to only low-quality social cues (neighbor proximity or action), agents converged on a Cohesive Tracking strategy. This strategy, effective in stable, slow-moving resource environments, relies on all agents predominantly using costly private tracking while using social cues to maintain a compact group formation. However, this strategy was not robust. Its success was highly sensitive to ecological context; performance degraded significantly as environmental volatility (resource speed) increased or as the intrinsic cost of private tracking increased (i.e., lower effective v_{tracking}), as showed by our robustness analyses (Fig. 5B).

In sharp contrast, access to high-quality payoff information allowed the group to be more flexible in their behavior. This enabled them to switch strategies based on the viability of private tracking. When private tracking was effective (slow resource or low tracking cost), agents adopted a Track-or-Copy strategy, using social information to selectively copy more successful peers. When private tracking became ineffective (fast resource or high tracking cost), the group evolve an Explore-or-Copy strategy. This emergent behavior functions as a form of distributed sensing, where agents abandon costly tracking, explore randomly, and use payoff cues to aggregate at temporary information hubs created by successful explorers (Fig. 4).

The emergence of the "Explore-or-Copy" strategy highlights a flexible solution to the classic Producer-Scrounger (P-S) dilemma [2, 52–54]. In our model, agents are not locked into fixed "Producer" (Exploration/Tracking) or "Scrounger" (Social Attraction) roles. Instead, the availability of high-fidelity payoff information allows them to dynamically adopt the most profitable role at each moment. An agent can act as a Producer by exploring, and the instant it observes a more successful peer, it switches to a Scrounger role to copy them. This dynamic switching is possible only because high-quality social information allows agents to accurately differentiate successful peers from unsuccessful ones. This selectivity makes strategies based on high-quality social information not only robust to ecological factors (like volatility) but also robust to social uncertainty. The group becomes effectively "immune" to the dilemma's primary pitfall: the risk of scroungers copying other unsuccessful scroungers, leading to negative information cascades. Because successful agents are reliably identified, the incentive to explore (produce) is maintained. Conversely, when payoff information is noisy, this filtering capacity breaks down, making agents vulnerable to such copying errors and degrading performance. This contrasts with the Cohesive Tracking strategy, which appeared to require significant group-level coordination to maintain a tight formation, forcing agents to prioritize group cohesion over immediate, individual tracking. Our decentralized training experiments supported this, showing that performance in this low-quality social information regime dropped

significantly without a mechanism to establish this group-wide behavioral consensus.

From a broader theoretical perspective, the learned strategies functionally encode the value of information. The agents' emergent policies reflect an adaptive valuation of information based on its utility in reducing uncertainty, a cornerstone of behavioral ecology [1, 55]. In our model, simple, low-bandwidth cues – analogous to an animal observing a neighbor's mere presence – were sufficient to reduce uncertainty in stable environments, leading to the Cohesive Tracking strategy. However, as environmental volatility increased this uncertainty, these cues became unreliable. High-bandwidth, high-fidelity signals – analogous to an animal observing a peer's actual foraging success – became necessary to reduce uncertainty and sustain performance, enabling the robust Explore-or-Copy strategy. Our results thus provide a mechanism for how such valuation can be learned, allowing agents to develop adaptive, context-dependent strategies that move beyond the simplified social learning strategies – such as "copy the successful" – that are often proposed [5, 6].

Here we investigated optimal social information usage in a simplified environment with a single moving resource. Although there are multiple ways to extend this scenario, we believe several directions are particularly promising. First, exploring multi-resource scenarios could provide valuable insights into collective shepherding [56, 57] and collective hunting [31] behaviors. Second, systematically investigating heterogeneous policies could allow for beneficial specialization among agents [58]. While our decentralized training experiments allowed for such heterogeneity, this primarily led to a decrease in coordination for the Cohesive Tracking strategy. A promising future direction would be to explore whether heterogeneity could instead lead to more efficient task distribution, such as explicit Producer and Scrounger roles, under different environmental or informational conditions.

Third, our agents operated with significant informational constraints: they were memory-less (reacting only to the current time step) and perceived only one peer at a time. Introducing memory, for instance, would allow agents to integrate information over time to learn resource trajectories or build reputations [59, 60]. Similarly, extending the model to allow for multi-sensor integration—such as combining visual and audio cues [61–63] or aggregating information from multiple peers simultaneously [64]—would be a critical step toward understanding more complex attentional mechanisms and collective decision-making, better capturing the rich social dynamics observed in natural systems.

Data Accessibility. All the code necessary to reproduce the training, evaluation, and analysis is available on GitHub: https://github.com/vagechirkov/Social-Information-Quality-D

etermines-Learned-Collective-Foraging-Strategies.git

Authors' Contributions. V.C. and P.R. conceptualized the study. V.C. performed the simulations, curated the data, conducted the analysis, and prepared the figures under the supervision of P.R. V.C. wrote the original manuscript draft, and all authors contributed to reviewing and editing the final version.

Competing Interests. The authors declare that they have no competing interests.

Funding. This research has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2002/1 "Science of Intelligence". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements. We are grateful to Dr. Mohsen Raoufi and Dr. Valentin Lecheval for fruitful discussions regarding the simulation design and for their insightful comments on the manuscript.

References

[1] Sasha R. X. Dall, Luc-Alain Giraldeau, Ola Olsson, John M. McNamara, and David W. Stephens. Information and its use by animals in evolutionary ecology. *Trends in Ecology & Evolution*, 20(4): 187–193, April 2005. ISSN 0169-5347. doi: 10.1016/j.tree.2005.01.010.

[2] Luc-Alain Giraldeau and Thomas Caraco. *Social Foraging Theory*, volume 73 of *Monographs in Behavior and Ecology*. Princeton University Press, Princeton, New Jersey, 2000. ISBN 978-0-691-04877-2.

[3] Marcus W. Feldman, Kenichi Aoki, and Jochen Kumm. Individual Versus Social Learning: Evolutionary Analysis in a Fluctuating Environment. *Anthropological Science*, 104(3):209–231, 1996. doi: 10.1537/ase.104.209.

[4] Wataru Toyokawa, Andrew Whalen, and Kevin N. Laland. Social learning strategies regulate the wisdom and madness of interactive crowds. *Nature Human Behaviour*, 3(2):183–193, 2019. ISSN 2397-3374. doi: 10.1038/s41562-018-0518-x.

[5] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and K. N. Laland. Why Copy Others? Insights from the Social Learning Strategies Tournament. *Science*, 328(5975):208–213, 2010. doi: 10.1126/science.1184719.

[6] William Hoppitt and Kevin N. Laland. *Social Learning: An Introduction to Mechanisms, Methods, and Models*. Princeton University Press, 2013. ISBN 978-0-691-15070-3.

[7] Thomas J. Valone and Jennifer J. Templeton. Public information for the assessment of quality: A widespread social phenomenon. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 357(1427):1549–1557, November 2002. ISSN 0962-8436. doi: 10.1098/rstb.2002.1064.

[8] Etienne Danchin, Etienne Danchin, Luc-Alain Giraldeau, Thomas J. Valone, and Richard H. Wagner. Public information: From nosy neighbors to cultural evolution. *Science*, 305(5683):487–491, July 2004. doi: 10.1126/science.1098254.

[9] Bennett G. Galef and Luc-Alain Giraldeau. Social influences on foraging in vertebrates: Causal mechanisms and adaptive functions. *Animal Behaviour*, 61(1):3–15, January 2001. ISSN 0003-3472. doi: 10.1006/anbe.2000.1557.

[10] Máté Nagy, Iain D. Couzin, Wolfgang Fiedler, Martin Wikelski, and Andrea Flack. Synchronization, coordination and collective sensing during thermalling flight of freely migrating white storks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1746):20170011, March 2018. doi: 10.1098/rstb.2017.0011.

[11] Christopher T. Monk, Matthieu Barbier, Pawel Romanczuk, James R. Watson, Josep Alós, Shinnosuke Nakayama, Daniel I. Rubenstein, Simon A. Levin, and Robert Arlinghaus. How ecology shapes exploitation: A framework to predict the behavioural response of human and animal foragers along exploration–exploitation trade-offs. *Ecology Letters*, 21(6):779–793, June 2018. ISSN 1461-023X, 1461-0248. doi: 10.1111/ele.12949.

[12] Rachel L. Kendal, Neeltje J. Boogert, Luke Rendell, Kevin N. Laland, Mike Webster, and Patricia L. Jones. Social Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive Sciences*, 22(7):651–665, 2018. ISSN 1879-307X. doi: 10.1016/j.tics.2018.04.003.

[13] Valentin Lecheval and Richard P. Mann. Smart self-propelled particles: A framework to investigate the cognitive bases of movement. *Journal of The Royal Society Interface*, 20(204):20230127, July 2023. doi: 10.1098/rsif.2023.0127.

[14] Mirta Galesic, Daniel Barkoczi, Andrew M. Berdahl, Dora Biro, Giuseppe Carbone, Ilaria Giannoccaro, Robert L. Goldstone, Cleotilde Gonzalez, Anne Kandler, Albert B. Kao, Rachel Kendal, Michelle Kline, Eun Lee, Giovanni Francesco Massari, Alex Mesoudi, Henrik Olsson, Niccolo Pescetelli, Sabina J. Sloman, Paul E. Smaldino, and Daniel L. Stein. Beyond collective intelligence: Collective adaptation. *Journal of The Royal Society Interface*, 20(200):20220736, March 2023. doi: 10.1098/rsif.2022.0736.

[15] Kamal Ndousse, Douglas Eck, Sergey Levine, and Natasha Jaques. Emergent Social Learning via Multi-agent Reinforcement Learning. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 7991–8004. PMLR, 2021. URL https://proceedings.mlr.press/v139/ndousse21a.html.

[16] Carsten Hahn, Fabian Ritz, Paula Wikidal, Thomy Phan, Thomas Gabor, and Claudia Linnhoff-Popien. Foraging Swarms using Multi-Agent Reinforcement Learning. In *Artificial Life Conference Proceedings 32*, pages 333–340. MIT Press, 2020. doi: 10.1162/isal_a_00267.

[17] Robert C. Löffler, Emanuele Panizon, and Clemens Bechinger. Collective foraging of active particles trained by reinforcement learning. *Scientific Reports*, 13(1):17055, 2023. doi: 10.1038/s41598-023-44268-3.

[18] Seungwoong Ha and Hawoong Jeong. Social learning spontaneously emerges by searching optimal heuristics with deep reinforcement learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 12319–12338. PMLR, 2023. URL https://proceedings.mlr.press/v202/ha23a.html.

[19] Edward A Codling, Michael J Plank, and Simon Benhamou. Random walk models in biology. *Journal of The Royal Society Interface*, 5(25):813–834, April 2008. doi: 10.1098/rsif.2008.0014.

[20] Briana Abrahms, Ellen O. Aikens, Jonathan B. Armstrong, William W. Deacy, Matthew J. Kauffman, and Jerod A. Merkle. Emerging Perspectives on Resource Tracking and Animal Movement Ecology. *Trends in Ecology & Evolution*, 36(4):308–320, 2021. doi: 10.1016/j.tree.2020.10.018.

[21] Massimo Vergassola, Emmanuel Villermaux, and Boris I. Shraiman. 'Infotaxis' as a strategy for searching without gradients. *Nature*, 445(7126):406–409, January 2007. ISSN 1476-4687. doi: 10.1038/nature05464.

[22] Julian Rode, Maja Novak, and Benjamin M. Friedrich. Information Theory of Chemotactic Agents Using Both Spatial and Temporal Gradient Sensing. *PRX Life*, 2(2):023012, June 2024. doi: 10.1103/PRXLife.2.023012.

[23] Mike M Webster and Kevin N Laland. Social information use and social learning in non-grouping fishes. *Behavioral Ecology*, 28(6):1547–1552, November 2017. ISSN 1045-2249. doi: 10.1093/beheco/arx121.

[24] Ketika Garg, Christopher T Kello, and Paul E Smaldino. Individual exploration and selective social learning: Balancing exploration–exploitation trade-offs in collective foraging. *Journal of The Royal Society Interface*, 19:20210915, 2022. doi: 10.1098/rsif.2021.0915.

[25] Valerii Chirkov, Ralf Kurvers, Pawel Romanczuk, and Dominik Deffner. Payoff selectivity unlocks collective intelligence in naturalistic human groups. *[Preprint]*, September 2025. URL https://osf.io/zrhwt.

[26] Iain Couzin. Collective minds. *Nature*, 445(7129):715–715, February 2007. ISSN 1476-4687. doi: 10.1038/445715a.

[27] Andrew Berdahl, Colin J. Torney, Christos C. Ioannou, Jolyon J. Faria, and Iain D. Couzin. Emergent Sensing of Complex Environments by Mobile Animal Groups. *Science*, 339(6119):574–576, February 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1225883.

[28] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415, 1996. ISSN 1476-4687. doi: 10.1038/381413a0.

[29] V. Zaburdaev, S. Denisov, and J. Klafter. Lévy walks. *Reviews of Modern Physics*, 87(2):483–530, 2015. doi: 10.1103/RevModPhys.87.483.

[30] Andy M. Reynolds. Current status and future directions of Lévy walk research. *Biology Open*, 7 (1):bio030106, 2018. ISSN 2046-6390. doi: 10.1242/bio.030106.

[31] Matthew J. Hansen, Paolo Domenici, Palina Bartashevich, Alicia Burns, and Jens Krause. Mechanisms of group-hunting in vertebrates. *Biological Reviews*, 98(5):1687–1711, 2023. doi: 10.1111/brv.12973.

[32] Brett T. McClintock, Roland Langrock, Olivier Gimenez, Emmanuelle Cam, David L. Borchers, Richard Glennie, and Toby A. Patterson. Uncovering ecological state dynamics with hidden Markov models. *Ecology Letters*, 23(12):1878–1903, 2020. ISSN 1461-0248. doi: 10.1111/ele.13610.

[33] Ron R. Togunov, Andrew E. Derocher, Nicholas J. Lunn, and Marie Auger-Méthé. Characterising menotactic behaviours in movement data using hidden Markov models. *Methods in Ecology and Evolution*, 12(10):1984–1998, 2021. doi: 10.1111/2041-210X.13681.

[34] Weijia Wang, Ramón Escobedo, Stéphane Sanchez, Clément Sire, Zhangang Han, and Guy Theraulaz. The impact of individual perceptual and cognitive factors on collective states in a data-driven fish school model. *PLOS Computational Biology*, 18(3):e1009437, 2022. doi: 10.1371/journal.pcbi.1009437.

[35] Daniel J. Stilwell, Erik M. Bollt, and D. Gray Roberson. Sufficient Conditions for Fast Switching Synchronization in Time-Varying Network Topologies. *SIAM Journal on Applied Dynamical Systems*, 5(1):140–156, 2006. doi: 10.1137/050625229.

[36] Fatemeh Parastesh, Karthikeyan Rajagopal, Sajad Jafari, Matjaž Perc, and Eckehard Schöll. Blinking coupling enhances network synchronization. *Physical Review E*, 105(5):054304, 2022. doi: 10.1103/PhysRevE.105.054304.

[37] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. In *Advances in Neural Information Processing Systems*, volume 35, pages 24611–24624. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/9c1535a02f0ce079433344e14d910597-Paper-Datasets_and_Benchmarks.pdf.

[38] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *CoRR*, abs/1707.06347, 2017.

[39] Albert Bou, Matteo Bettini, Sebastian Dittert, Vikash Kumar, Shagun Sodhani, Xiaomeng Yang, Gianni De Fabritiis, and Vincent Moens. TorchRL: A data-driven decision-making library for PyTorch. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=QxItoEAVMb.

[40] John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-Dimensional Continuous Control Using Generalized Advantage Estimation. In *The Fourth International Conference on Learning Representations*, San Juan, Puerto Rico, 2016. URL http://arxiv.org/abs/1506.02438.

[41] Matteo Bettini, Ryan Kortvelesy, Jan Blumenkamp, and Amanda Prorok. VMAS: A Vectorized Multi-agent Simulator for Collective Robot Learning. In Julien Bourgeois, Jamie Paik, Benoît Piranda, Justin Werfel, Sabine Hauert, Alyssa Pierson, Heiko Hamann, Tin Lun Lam, Fumitoshi Matsuno, Negar Mehr, and Abdallah Makhoul, editors, *Distributed Autonomous Robotic Systems*, pages 42–56, Cham, 2024. Springer Nature Switzerland. doi: 10.1007/978-3-031-51497-5_4.

[42] Jason Ansel, Edward Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, Peter Bell, David Berard, Evgeni Burovski, Geeta Chauhan, Anjali Chourdia, Will Constable, Alban Desmaison, Zachary DeVito, Elias Ellison, Will Feng, Jiong Gong, Michael Gschwind, Brian Hirsh, Sherlock Huang, Kshiteej Kalambarkar, Laurent Kirsch, Michael Lazos, Mario Lezcano, Yanbo Liang, Jason Liang, Yinghai Lu, CK Luk, Bert Maher, Yunjie Pan, Christian Puhrsch, Matthias Reso, Mark Saroufim, Marcos Yukio Siraichi, Helen Suk, Michael Suo, Phil Tillet, Eikan Wang, Xiaodong Wang, William Wen, Shunting Zhang, Xu Zhao, Keren Zhou, Richard Zou, Ajit Mathews, Gregory Chanan, Peng Wu, and Soumith Chintala. PyTorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM, April 2024. doi: 10.1145/3620665.3640366.

[43] Lukas Biewald. Experiment Tracking with Weights and Biases, 2020. URL https://www.wandb.com/.

[44] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3): 90–95, 2007. doi: 10.1109/MCSE.2007.55.

[45] Michael L. Waskom. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6 (60):3021, April 2021. ISSN 2475-9066. doi: 10.21105/joss.03021.

[46] Christian Schroeder de Witt, Tarun Gupta, Denys Makoviichuk, Viktor Makoviychuk, Philip H. S. Torr, Mingfei Sun, and Shimon Whiteson. Is Independent Learning All You Need in the StarCraft Multi-Agent Challenge?, November 2020.

[47] Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep Reinforcement Learning That Matters. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2018. doi: 10.1609/aaai.v32i1.11694.

[48] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der

Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, March 2020. ISSN 1548-7105. doi: 10.1038/s41592-019-0686-2.

[49] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85): 2825–2830, 2011. ISSN 1533-7928. URL http://jmlr.org/papers/v12/pedregosa11a.html.

[50] R. A. Johnstone, S. R. X. Dall, Luc-Alain Giraldeau, Thomas J. Valone, and Jennifer J. Templeton. Potential disadvantages of using socially acquired information. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1427):1559–1566, November 2002. doi: 10.1098/rstb.2002.1065.

[51] Ariana Strandburg-Peshkin, Damien R. Farine, Iain D. Couzin, and Margaret C. Crofoot. Shared decision-making drives collective movement in wild baboons. *Science*, 348(6241):1358–1361, June 2015. doi: 10.1126/science.aaa5099.

[52] C.J. Barnard and R.M. Sibly. Producers and scroungers: A general model and its application to captive flocks of house sparrows. *Animal Behaviour*, 29(2):543–550, May 1981. ISSN 00033472. doi: 10.1016/S0003-3472(81)80117-0.

[53] Guy Beauchamp. A spatial model of producing and scrounging. *Animal Behaviour*, 76(6):1935–1942, December 2008. ISSN 0003-3472. doi: 10.1016/j.anbehav.2008.08.017.

[54] Ralf H. J. M. Kurvers, Steven Hamblin, and Luc-Alain Giraldeau. The Effect of Exploration on the Use of Producer-Scrounger Tactics. *PLOS ONE*, 7(11):e49400, November 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0049400.

[55] John R. Krebs and David W. Stephens. *Foraging Theory*. Princeton University Press, Princeton, 2019. ISBN 978-0-691-20679-0. URL https://muse.jhu.edu/pub/267/monograph/book/71820.

[56] Italo Napolitano, Andrea Lama, Francesco De Lellis, and Mario di Bernardo. Emergent Cooperative Strategies for Multi-Agent Shepherding via Reinforcement Learning. *[Preprint]*, 2024. doi: 10.48550/arXiv.2411.05454.

[57] Andrea Lama and Mario di Bernardo. Shepherding and herdability in complex multiagent systems. *Physical Review Research*, 6(3):L032012, July 2024. doi: 10.1103/PhysRevResearch.6.L032012.

[58] Matteo Bettini, Ryan Kortvelesy, and Amanda Prorok. Controlling Behavioral Diversity in Multi-Agent Reinforcement Learning. In *Forty-First International Conference on Machine Learning*, June 2024. URL https://openreview.net/forum?id=qQjUgItPq4.

[59] John M. McNamara and Alasdair I. Houston. Memory and the efficient use of information. *Journal of Theoretical Biology*, 125(4):385–395, April 1987. ISSN 00225193. doi: 10.1016/S0022-5193(87)80209-6.

[60] Andrew M. Berdahl, Albert B. Kao, Andrea Flack, Peter A. H. Westley, Edward A. Codling, Iain D. Couzin, Anthony I. Dell, and Dora Biro. Collective animal navigation and migratory culture: From theoretical models to empirical evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1746):20170009, May 2018. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2017.0009.

[61] Sarah Partan and Peter Marler. Communication Goes Multimodal. *Science*, 283(5406):1272–1273, February 1999. doi: 10.1126/science.283.5406.1272.

[62] Kelly L. Ronald, Esteban Fernández-Juricic, and Jeffrey R. Lucas. Mate choice in the eye and ear of the beholder? Female multimodal sensory configuration influences her preferences. *Proceedings of the Royal Society B: Biological Sciences*, 285(1878):20180713, May 2018. doi: 10.1098/rspb.2018.0713.

[63] Wouter Halfwerk, Judith Varkevisser, Ralph Simon, Ezequiel Mendoza, Constance Scharff, and Katharina Riebel. Toward Testing for Multimodal Perception of Mating Signals. *Frontiers in Ecology and Evolution*, 7, April 2019. ISSN 2296-701X. doi: 10.3389/fevo.2019.00124.

[64] David Mezey, Dominik Deffner, Ralf H. J. M. Kurvers, and Pawel Romanczuk. Visual social information use in collective foraging. *PLOS Computational Biology*, 20(5):1–24, 2024. doi: 10.1371/journal.pcbi.1012087.