

Active Electrosensing and Communication in MARL-trained Weakly Electric Fish Collectives

Satpreet H. Singh^{1,2*} Sonja Johnson-Yu^{1,2*} Zhouyang Lu^{1,5} Aaron Walsman² Federico Pedraja³
Denis Turcu³ Pratyusha Sharma⁴ Naomi Saphra² Nathaniel B. Sawtell³ Kanaka Rajan^{1,2}

¹Department of Neurobiology, Harvard Medical School, Boston, MA

²Kempner Institute for the Study of Natural and Artificial Intelligence at Harvard University, Boston, MA

³Columbia University, New York, NY, USA

⁴Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

⁵Brown University, Providence, RI, USA

Abstract

How complex collective behavior emerges from individual interactions is a fundamental scientific question, but experimental cost and difficulty of simultaneous multi-brain recordings limit direct study in animals. Here we introduce a novel computational framework modeling weakly electric fish-like agents with biophysically inspired electrosensing and actuation, trained to forage collectively via multi-agent reinforcement learning (MARL). Trained agents reproduce hallmarks of real fish, including curvilinear homing trajectories and heavy-tailed electric organ discharge (EOD) interval statistics, while exhibiting emergent active sensing, social foraging, dominance-like asymmetries, and aggression. We perform *in silico* interventions including sensor ablations, EOD silencing, and food distribution changes to identify causal drivers of social foraging. Analyses of recurrent neural dynamics further show robust encoding of task-relevant variables and social context. Our work has broad implications for the neuroethology of weakly electric fish and other social animals where extensive multi-individual neural recordings, and thus traditional data-driven modeling, remain challenging.

1 INTRODUCTION

Understanding how complex social behavior and neural activity emerge from interactions among individuals is a central problem in neuroscience, ethology, and artificial intelligence. Social behavior requires individuals to sense, communicate, and coordinate with others in dynamic environments, yet the mechanisms linking sensory signals, neural activity, and group behavior remain poorly understood. Progress is constrained from two directions. Experimentally, simultaneous neural recordings from multiple freely interacting animals are technically challenging, and large-scale behavioral assays are expensive, time-consuming, and difficult to manipulate^{1,2}. Computationally, existing agent-based and reinforcement-learning models of collective behavior have provided important insights into schooling, coordination, and emergent communication³⁻⁷, but typically rely on simplified sensory representations that are not grounded in the biophysics of any real sensory system. We pursued an approach that combines biophysical realism, closed-loop multi-agent decision-making, and full experimental controllability, which allowed us to study sensing, signaling, and collective behavior in a fully observable and controllable setting.

Weakly electric fish are a particularly well-suited model system for studying these questions, because they generate

brief electric organ discharges (EODs) that serve two coupled functions: active sensing of the environment and communication with conspecifics (members of the same species)⁸⁻¹⁵. Here, we use “communication” to refer operationally to EOD signals that are sensed by conspecifics via Knollenorgans and that influence their behavior. This coupling makes sensing and social signaling inseparable, uniquely positioning it for studying how sensory and communication signals jointly shape collective behavior. Weakly electric fish also display a rich repertoire of social foraging behaviors, including context-dependent EOD modulation, dominance hierarchies, and competition over food resources¹⁵⁻¹⁷. The biophysics of EOD generation and electroreception are well characterized^{14,18}, and experimental data on individual and dyadic behavior are available for model validation¹⁹⁻²¹. Simultaneously, multi-animal neural recordings from freely interacting fish are not yet feasible at scale, and group-level behavioral assays are difficult to control and interpret. As a result, major gaps remain in our understanding of how electrosensing and electrocommunication jointly support social foraging and collective behavior.

To address this, we introduce a simulation based framework in which fish-like agents are endowed with biophysically inspired electrosensing and motor capabilities and trained to forage collectively using multi-agent

* Contributed equally

† Correspondence: kanaka_rajan@hms.harvard.edu

reinforcement learning (MARL). In addition to providing an engineering benchmark, our framework serves as a scientific model of social sensing, closed-loop behavior, and inter-agent interaction. Prior work has modeled fish collective behavior^{3–6,22–24} and electrosensory processing^{14,18,25} separately, with collective models relying on simplified sensory representations^{6,22} and electrosensory models lacking closed-loop multi-agent decision-making^{18,25}. To our knowledge, no prior work has combined all three major mormyrid electroreceptor classes, active EOD generation, collective sensing, and closed-loop multi-agent decision-making in a single framework, nor produced electrocommunication patterns consistent with real fish from individual fitness incentives alone. By providing both validation against established biological data and the flexibility to perform interventions that are impractical or impossible *in vivo*, this model instantiates an emerging *virtual neuroscience* paradigm^{23,24,26–29} in the domain of electrosensory collective behavior. Full observability of each agent’s sensory inputs, neural activity, and actions across all agents simultaneously, combined with complete experimental control over sensing, signaling, and reward structure, enables a level of mechanistic analysis not yet accessible in real animals. More broadly, our framework illustrates how biophysically grounded virtual-agent models can address fundamental questions about embodied social intelligence and the consequences of disrupted sensing or communication for collective behavior.

Using this approach, we show that trained agents reproduce hallmark features of real weakly electric fish, including curvilinear homing trajectories and heavy-tailed EOD interval statistics, while exhibiting emergent social foraging, dominance-like asymmetries, and context-dependent EOD modulation. *In silico* sensor ablations and signaling manipulations reveal distinct causal contributions of short- and long-range electroreceptor classes to foraging efficiency, social spacing, and food-consumption inequality. Analyses of agents’ recurrent neural dynamics show that task-relevant variables and social context are encoded in recurrent activity, and that pairs of interacting agents show proximity-dependent correlated latent dynamics. Together, these results illustrate how sensing, signaling, and neural computation jointly support collective social behavior, and provide a platform for generating testable predictions for experiments in real fish and, more broadly, for understanding social intelligence in other multi-agent sensory systems.

2 BUILDING FISH-LIKE AGENTS WITH BIOPHYSICALLY INSPIRED ELECTROSENSING

To build our simulation, we developed a multi-agent reinforcement learning (MARL) environment in which fish-like agents move, emit electric organ discharges (EODs), sense electric fields, forage, and interact physically within a shared arena (Fig. 1A,C,G,H). Our framework

relies on a custom-built 2D physics simulator that manages the mechanical and electrical interactions between all objects in the simulated arena, including the generation and propagation of electric fields and how these fields are sensed by biomimetic electroreceptors (Fig. 1B,D–F; Methods).

Each agent is represented as a circular particle that chooses actions independently at every timestep. The multivariate action space spans continuous forward translation and angular rotation actions, and discrete EOD emission and bite commands. These collectively control locomotion, electrosensing, communication, and close-range aggression (Fig. 1A). Agent motion, collision handling, and the simulation timestep are bounded by empirical weakly electric fish data (Methods).

Each agent’s policy is a GRU-based³⁰ actor-critic network with feedforward action and value heads (Methods). To introduce heterogeneity, each agent is assigned a scalar *size* parameter that scales its maximum speed and is provided as a fixed context input in its observation vector per episode.

Agents are trained with a multi-agent variant of the Proximal Policy Optimization (PPO) algorithm, a policy gradient method known for its stability and efficiency^{31–33}. The task is specified by a reward function based on individual fitness, with rewards for successful foraging and asymmetric penalties during aggressive encounters: the larger (dominant) agent incurs a smaller penalty than the smaller (subordinate) agent.

No positive reward is assigned for communication, coordination, chasing, or aggression toward conspecifics; any coordination or communication that emerges arises from agents’ shared environment and individual foraging incentives. Here, we use “emergent” operationally to refer to behaviors that were not hand-coded or directly rewarded but arose through learning under the modeled sensory, motor, ecological, and reward constraints. We trained multiple independent seeds, performed a controlled behavioral assay on each and, similar to prior literature^{23,24}, selected one representative converged policy for further analysis. Seed selection used an ethologically motivated seed-selection criterion that balanced biological desiderata, including group foraging success, size-dependent resource asymmetry, and broad agent participation. Post-training analyses were then performed on this selected policy across several group-sizes, arena and food configurations, and sensory perturbations.

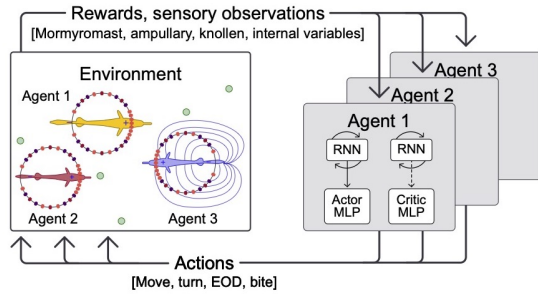
2.1 Electric field generation and sensation

In our simulator, electric fields arise from two sources: actively controlled EOD sources on agents, and intrinsic electric sources associated with agents and prey (Fig. 1D,E). These fields can be distorted as they interact with nearby objects. As a result, agents sense not only the presence of prey and conspecifics, but also how their own and others’ EODs are transformed by the shared environment.

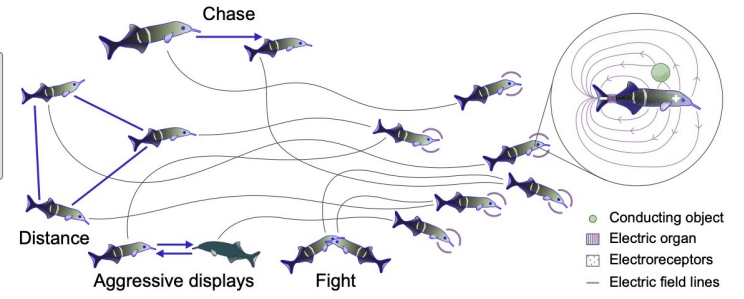
Agents are equipped with three types of biomimetic electroreceptors inspired by the well-characterized receptor

Training Weakly Electric Fish Agents using MARL

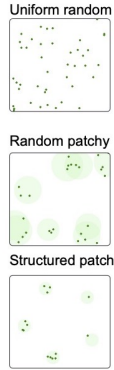
A. MARL Framework Overview



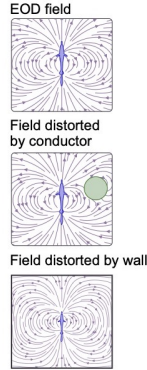
B. Social Behavior in Real Weakly Electric Fish



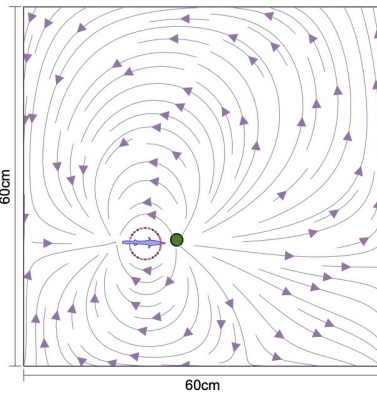
C. Food Distributions



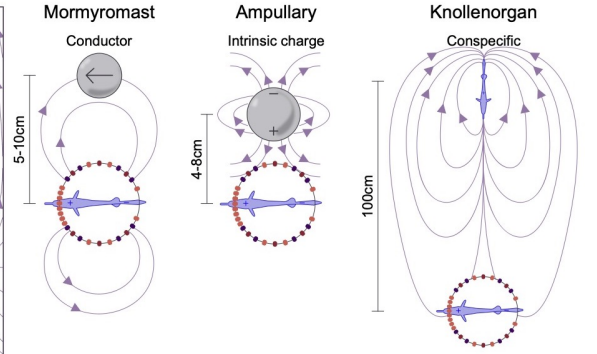
D. Electric Field Generation



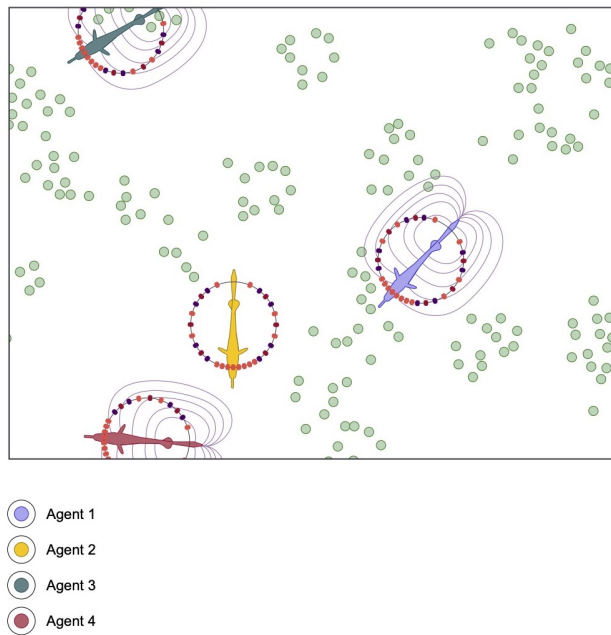
E. Electric Landscape



F. Electric Field Sensing



G. Multi-agent Simulation Schematic



H. Snapshot of Trained Foraging Agents and EOD Time Courses

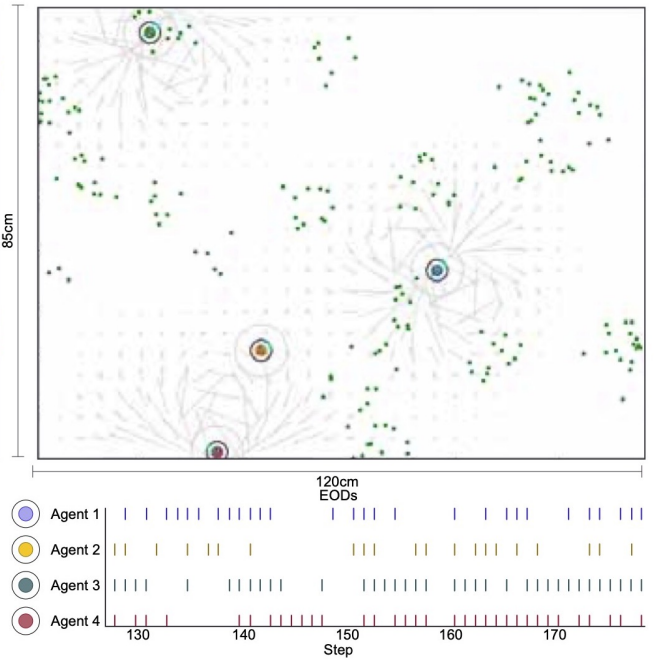


Fig. 1. MARL framework for training fish-like agents with biophysically inspired electrosensing: (A) Schematic of the MARL training loop, where agents interact within a simulated arena, emitting and sensing electric organ discharges (EODs) through weakly electric fish-inspired sensors. Rewards encourage successful foraging and penalize aggressive encounters. (B) Schematic of electrosensing and social behaviors in weakly electric fish, where individuals generate electric fields, detect distortions with electroreceptors, and use signals during chasing, spacing, and aggressive encounters. (C) Examples of arena food patch configurations used in training and evaluation. (D) Electric fields generated by sources and distorted by conducting objects and walls. (E) Electric landscape generated by a pair of monopoles modeling the EOD of a single agent. Field lines show effect of wall boundaries on the generated field. (F) Electrosensors on the circular body of each agent: Mormyromast (short-range, active), Ampullary (short-range, passive) and Knollenorgan (long-range, conspecific EOD only). (G) Schematic of all elements of multi-agent simulation shown together. (H) Snapshot of four fish-like artificial agents foraging in our simulated 2D arena (top), with electric organ discharges (EODs) from each over time (bottom).

classes of mormyrid weakly electric fish^{14,18,25,34,35}: Mormyromasts, Ampullary receptors, and Knollenorgans (Fig. 1F).

Mormyromasts support active electrolocation by detecting distortions in an agent’s own EOD field caused by nearby prey (≤ 5 cm) or agents (≤ 10 cm), generating a *self-image*; an internal cancellation signal suppresses the self-generated ("reafferent") component of this response^{14,36,37}. They also support collective sensing by detecting distortions associated with conspecific EODs, generating a *cons-image*¹⁵.

Ampullary receptors provide passive, short-range sensing of intrinsic electric fields from prey (≤ 4 cm) and conspecifics (≤ 8 cm)³⁵.

Knollenorgans detect conspecific EOD pulses and provide long-range (≤ 100 cm) directional social cues, with temporal coding conveying information about conspecific identity and waveform^{38,39}. In our model, these signals identify the direction of emitting agents and convey their size, but do not directly encode distance.

Full details of field equations, receptor placement, sensor calibration, action parameterization, and training are provided in Methods.

3 RESULTS

3.1 Electrosensing agents follow fish-like curvilinear, not straight-line trajectories to targets

Classic experiments¹⁹ show that weakly electric fish approach targets along smooth, curved trajectories rather than straight lines, a hallmark of their electric sensing modality. To validate our model, we tested whether agents reproduce this behavior. We trained agents in a minimal two-agent setup with reward terms for reaching the other agent, distance-decrease shaping, and a small time penalty. We evaluated agents in a 70×70 cm arena containing one fixed always-emitting "target" agent placed at a random location within the arena (Fig. 2A). A second "homing agent" was initialized at random positions in the arena across 100 trials. We compared these trajectories with a non-electric ray-casting baseline, in which radial line-of-sight rays supplied direct object-distance and object-type cues rather than electric-field measurements (Fig. 2B).

As shown in Fig. 2B, electrosensing agents followed curvilinear paths to targets, while ray-casting agents followed straight-line paths. The histogram (Fig. 2C) and time course (Fig. 2D) of the error angle, defined as the difference between the agent’s heading and the local direction of the electric field produced by the target agent, shows a bimodal distribution concentrated near parallel and anti-parallel alignment with field-lines, and progressive alignment during successful homing episodes. The resulting error-angle statistics qualitatively resemble those reported for real fish¹⁹.

Together, these results support the interpretation that our biophysically inspired sensing is sufficient to produce fish-like homing dynamics.

3.2 Trained agents develop flexible social foraging and context-dependent EOD modulation

Collective foraging in weakly electric fish involves both exploration and competition, with context-driven modulations of EOD signaling^{2,8,11}. To test whether our agents reproduce these patterns, we evaluated them in arenas with either patchy or uniform food distributions (Fig. 3). The multi-agent patchy panels used $70 \text{ cm} \times 70 \text{ cm}$ arenas with four agents initialized with random positions, orientations, and size parameters, repeated for 30 episodes; these capture group-level foraging and dominance dynamics. The inter-discharge interval (IDI), distance-modulation, and peri-event EOD panels used a two-agent wide assay in a $160 \text{ cm} \times 40 \text{ cm}$ arena, repeated for 100 evaluation episodes; this narrower protocol isolates pairwise EOD dynamics with higher statistical power.

In the four-agent patchy assay, agents alternated between exploration, feeding, and social interactions, with context-dependent changes in EOD rates visible across agents (Fig. 3A,B). In the two-agent wide assay, IDIs during foraging followed heavy-tailed distributions, as do those measured in real fish²¹ (Fig. 3C); both admit power-law-like fits⁴⁰ over the same range (Fig. 3D), though the exponents do not quantitatively match.

Agents flexibly modulated EOD emission based on spatial context. EOD rate varied systematically with distance to the nearest conspecific (Fig. 3E), distance to the closest food item (Fig. 3F), and distance to the nearest arena wall (Fig. 3G). Peri-event analyses revealed event-locked modulation of signaling: EOD rate showed a sharp transient increase around eating events (Fig. 3H), a build-up preceding and decline following the agent biting another fish (Fig. 3I), and an elevation around times in which the agent was bitten (Fig. 3J).

Training with inter-agent heterogeneity led to emergent dominance-like asymmetries: larger agents directed more biting toward smaller conspecifics, with win ratio increasing with body size (Fig. 3K,L). Larger agents also consumed more food per episode (Fig. 3M), and food-intake inequality scaled with body-size inequality across groups (Fig. 3N). Body size tended to correlate positively with EOD emission rate (Fig. 3O), and EOD rate advantage weakly tracked size advantage in close-range agent pairs (Fig. 3P).

Together, these results show that biophysically grounded sensory structure and simple individual incentives are sufficient for flexible, fish-like collective foraging across ecological conditions.

3.3 In silico perturbations reveal sensory and signaling contributions to collective foraging

Virtual interventions make it possible to causally isolate mechanisms that are often difficult or impossible to manipulate experimentally. We therefore performed manipulations targeting electroreceptor channels, active signaling, and food availability, in four-fish groups (ablations and environmental), a two-agent assay (EOD

Electrosensing Agents Follow Curvilinear Trajectories to Targets

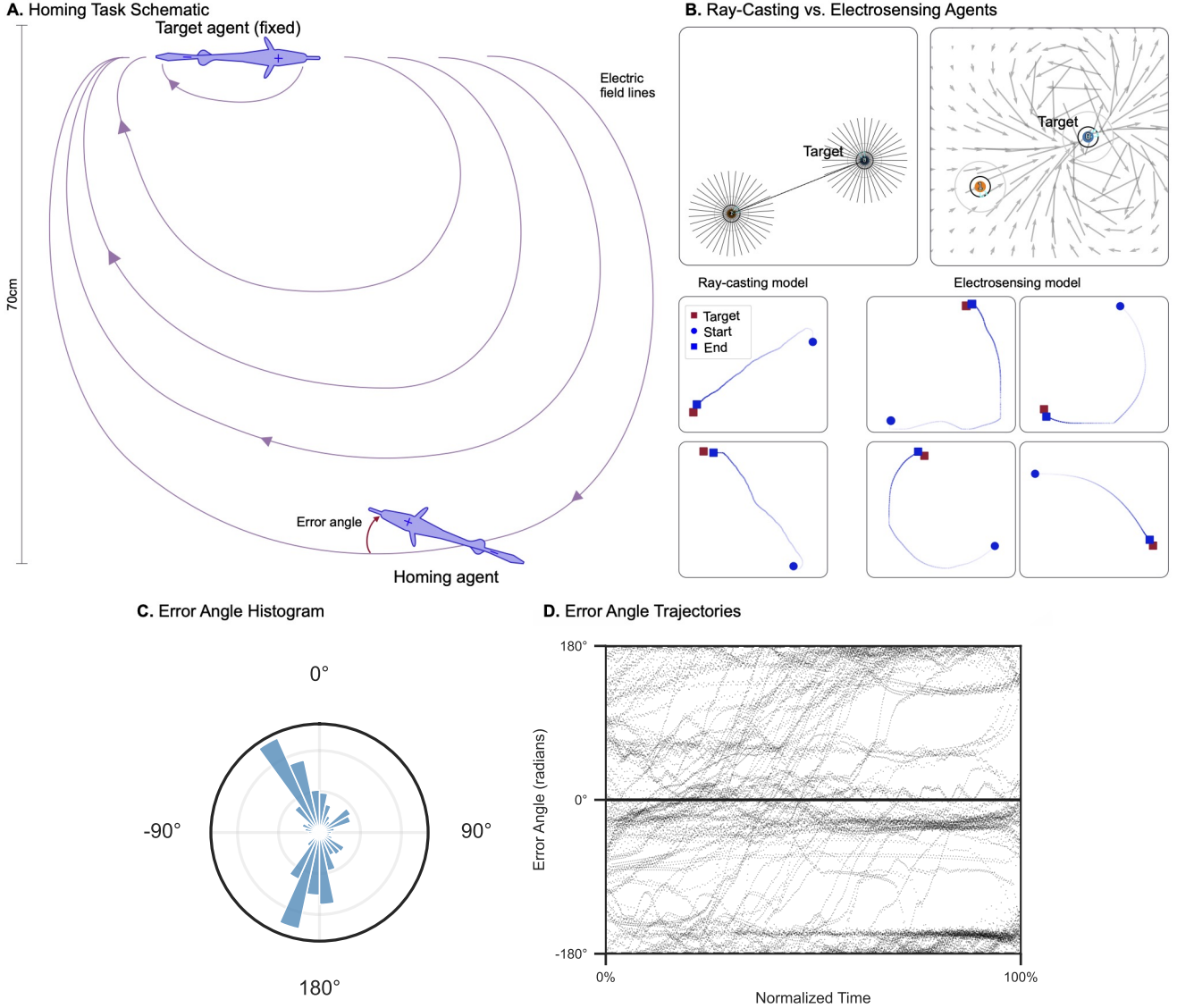


Fig. 2. Validation of electrosensing-based homing in artificial fish-like agents. (A) Homing task schematic. A mobile homing agent navigates toward a fixed, always-emitting conspecific that serves as the target. For the electric-field analysis, the error angle is defined as the angular difference between the homing agent's heading and the target's local electric-field direction. (B) Comparison of a non-electric ray-casting baseline, which supplies direct radial line-of-sight distance and object-type cues, and electrosensing-based homing. Unlike the direct, nearly straight paths produced by the ray-casting baseline, agents using the physics-based electric field model follow curved trajectories toward the target, producing qualitatively fish-like approach paths. (C) Distribution of error angles across trajectories. The polar histogram shows that homing agents preferentially orient relative to the target's local electric field direction, with peaks near both parallel and anti-parallel alignment, consistent with navigation along electric field lines. (D) Error-angle dynamics over normalized trial time. Error angles progressively align with anti-parallel approach over the course of successful homing episodes.

muting), and a solo-fish control.

Sensor ablations. We disabled each electroreceptor class in turn to identify causal contributions of short-range (Mormyromast, Ampullary) and long-range (Knollenorgan) channels (Fig. 4A–F). Short-range sensor ablations reduced food consumption in both the solo assay (Fig. 4A, single agent without conspecifics) and four-agent groups (Fig. 4B), indicating effects on individual foraging that persist independently of social interaction. Knollenorgan ablation left total food consumption intact but reshaped social

organization: EOD emission probability and biting rose while inter-agent spacing fell (Fig. 4C–F).

Collective sensing. Within the Mormyromast channel, we independently gated self- and conspecific-EOD inputs (Fig. 4G,H). Removing the conspecific-EOD image (self-image-only condition) drastically reduced food consumed (Fig. 4G), while removing the self-EOD image (cons-image-only condition) did not show a significant effect on foraging. EOD emission probability was lower in self-image-only agents and higher in cons-image-only

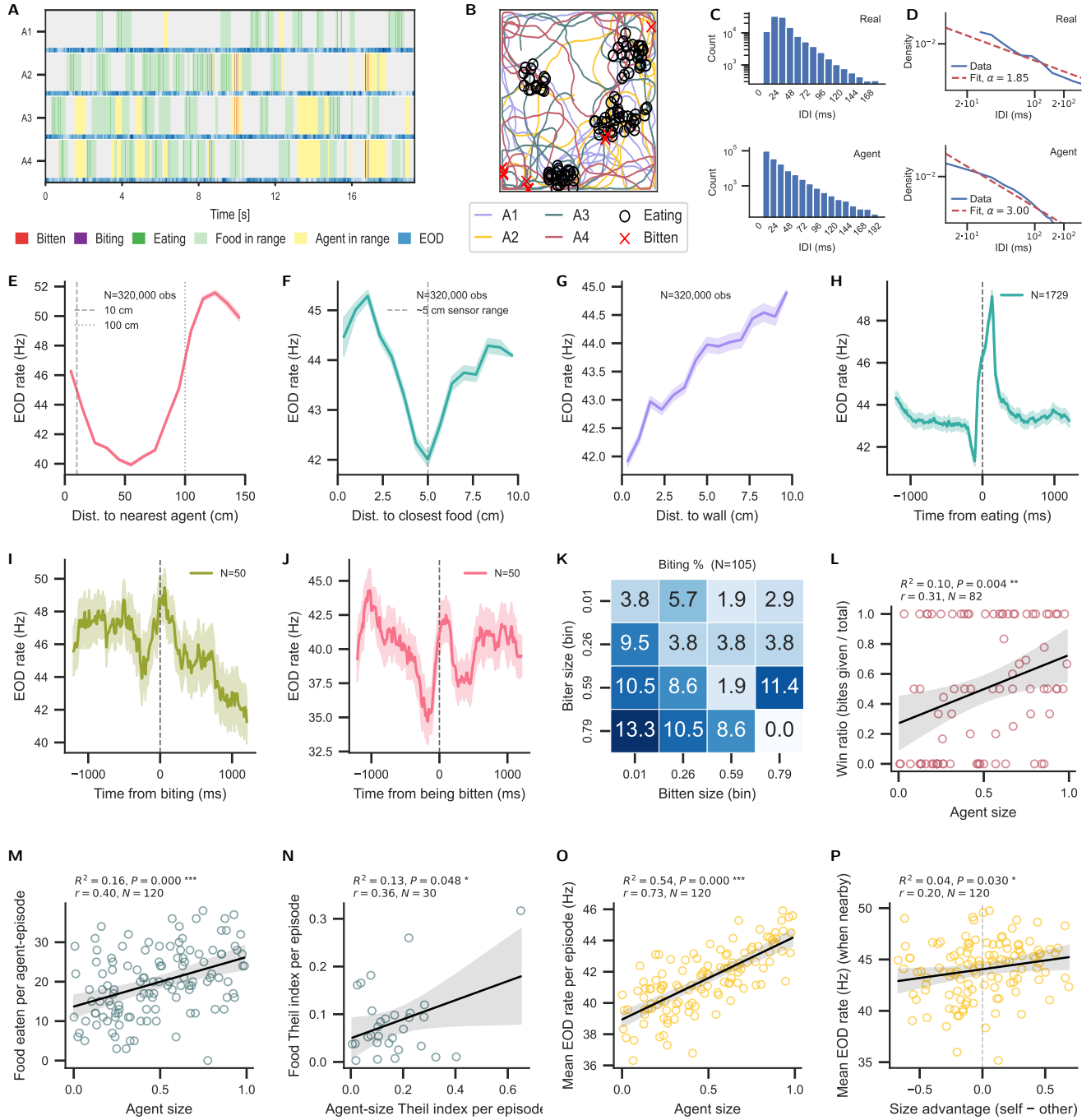


Fig. 3. Emergent social foraging in electrosensing agents. **(A)** Ethogram from a representative multi-agent episode showing exploration, feeding, and social interactions; thin blue bands indicate EOD rate for each agent. **(B)** Representative trajectories showing learned exploration and foraging in a shared arena with patchy food. **(C)** Inter-discharge interval (IDI) distributions for dyads of (top) real fish²¹ and (bottom) trained agents, showing heavy-tailed pulse timing statistics in both. **(D)** Power-law fits to IDI distributions for real fish and trained agents. **(E–G)** EOD rate is modulated by spatial context, varying with distance to the nearest agent, food, and arena wall. **(H–J)** Peri-event EOD dynamics around eating, biting another agent, and being bitten, showing event-locked modulation of signaling; all curves show mean $\pm$ SEM. **(K)** Biting heatmap shows that larger agents tend to bite smaller agents. **(L)** Win ratio (fraction of bites given / total) increases with body size. **(M)** Larger agents consume more food per episode. **(N)** Food consumption inequality (Theil index) scales with body-size inequality across groups. **(O)** Body size correlates positively with EOD emission rate. **(P)** EOD rate advantage tracks size advantage between paired agents when in close range (≤ 10 cm).

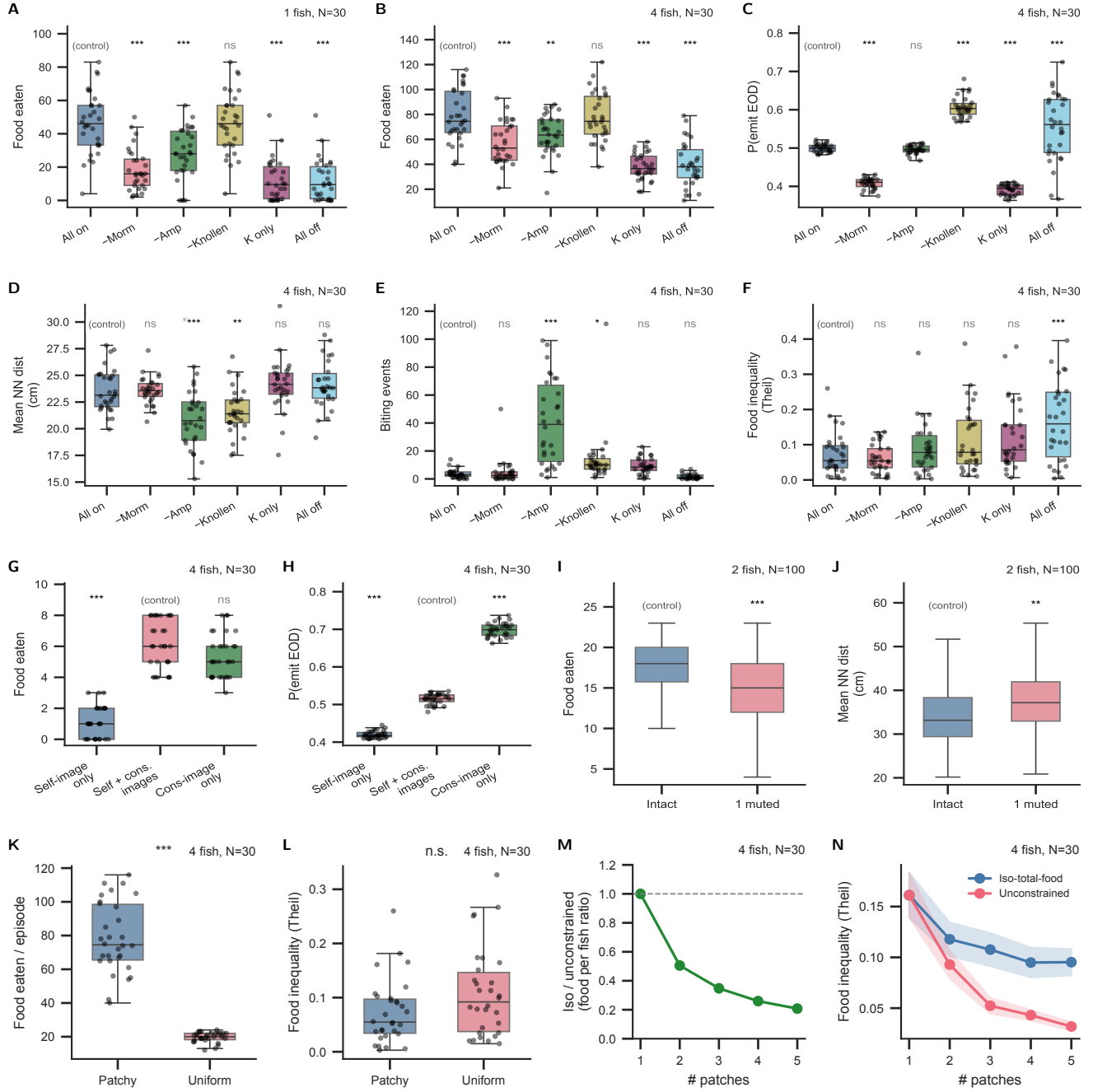


Fig. 4. In-silico interventions reveal potential causal factors underlying collective foraging. (A–F) Electrorceptor ablations compared against the full-sensor baseline using one-vs-control Dunnett tests. (A) Food eaten by a single isolated agent under each sensor ablation; isolates foraging from social effects. (B) Food eaten per agent in 4-fish groups under sensor ablation. (C) EOD emission probability under sensor ablation. (D) Mean nearest-neighbor spacing under sensor ablation. (E) Number of biting events under sensor ablation. (F) Food consumption inequality (Theil index) under sensor ablation. (G,H) Collective sensing manipulations (self-EOD versus cons-EOD inputs to Mormyromast channel): food eaten (G) and EOD emission probability (H). (I,J) EOD muting manipulations (intact group versus one-agent-muted group): food eaten (I) and mean nearest-neighbor distance (J). (K–N) Environmental manipulations. (K,L) Patchy versus uniform food distribution: total food eaten (K) and food consumption inequality (L). (M,N) Iso-food versus unconstrained patch-number sweep (1–5 patches): food-per-fish ratio of iso to unconstrained series (M) and food consumption (Theil) inequality (N) across patch counts. Bar and point summaries show mean $\pm$ SEM where error bars are present; significance annotations are panel-level exploratory summaries.

agents (Fig. 4H).

EOD muting. Silencing EOD emission in one of two agents reduced total food consumption and increased mean nearest-neighbor distance (Fig. 4I,J), indicating that active electrocommunication supports both effective foraging and social spacing.

Environmental manipulations. Patchy food distributions yielded substantially more food per episode than uniform layouts (Fig. 4K), with no significant difference in consumption inequality between the two conditions (Fig. 4L). We swept patch count from 1 to 5 under two conditions: an iso-food series (total food held constant, food-per-patch decreasing) and an unconstrained series (food-per-patch held constant, total food increasing). Per-fish intake fell progressively in the iso-food series relative to the unconstrained series (Fig. 4M); inequality fell with increasing patch count in both series, somewhat more steeply in the unconstrained condition (Fig. 4N).

Together, these manipulations identify distinct causal roles for short-range channels (foraging), Knollenorgan sensing (social spacing), collective sensing, and resource structure (inequality), yielding falsifiable predictions for ablation and competition experiments in real fish.

3.4 Dyadic assays reveal size and signal-dependent effects on foraging and interactions

Two-fish assays (150 cm $\times$ 150 cm) allow closer inspection of how EOD modulation and relative-size asymmetries shape competition at the smallest social scale. We evaluated pairs of agents in a one-patch dyadic task, with resident A initialized on a replenishing patch and intruder B initialized away from it, repeating 100 trials for each size condition (Fig. 5A).

Example trajectories illustrate how relative size shaped approach behavior: dominant intruders approached the patch directly and displaced the resident, same-size pairs often shared access, and subordinate intruders approached more indirectly with reduced patch access (Fig. 5B–D). Intruder B consumed significantly more food when A was present than when foraging alone (Fig. 5E), while resident A’s total food intake was not significantly affected by B’s presence (Fig. 5F). Consumption timecourses revealed a graded size-asymmetry effect: B accumulated food faster when dominant ($B > A$) and slower when subordinate ($A > B$), with A showing the complementary pattern (Fig. 5G,H).

Replacing the resident with a rule-based patch-confined random walker ("bot") isolated the contribution of EOD rate and movement state. B reached the patch more often when the bot was frozen than when moving (Fig. 5I). B’s patch-reach rate was generally higher when the bot emitted EODs at any nonzero rate than when silent, with success peaking at intermediate emission rates (Fig. 5J).

In separate two-agent square arenas (70 cm $\times$ 70 cm) with uniform food distributions, larger agents were more likely to act as chasers and smaller ones as chased (Fig. 5K,L). Similarly, biters tended to be larger than their victims

(Fig. 5M,N). Biting was concentrated in the first half of interaction windows, peaking at approximately 40% of the normalized interaction duration (Fig. 5O). Confronting interactions had the highest biting rate ($\approx 50\%$), followed by chasing, fleeing, and unaligned interactions (Fig. 5P) (See Methods for interaction definitions).

These assays identify relative body size and active EOD signaling as key determinants of competitive outcomes in dyadic interaction.

3.5 Recurrent dynamics encode task variables and social interaction state

Principal component analysis (PCA) of four-agent patchy foraging episodes yielded an effective rank of $D_{\text{eff}} = 81.3$ (304 PCs for 90% variance; Fig. 6A), well below the $D_H = 512$ nominal capacity.

Effective dimensionality scaled with group size when the Knollenorgan was intact, but remained flat near the solo baseline when the Knollenorgan was ablated (Fig. 6B), implicating long-range signaling as the primary driver of social-context-dependent dimensionality expansion. A subsampling analysis confirmed that the group-size scaling is visible even with partial unit recordings (Fig. 6C).

In the homing task, a linear decoder trained on RNN hidden states recovered the Knollenorgan-defined error angle (θ_{error}^K) with substantially higher performance than the straight-line angle to the target (θ_{agent} ; Fig. 6D), indicating that the network represents field-line geometry rather than Euclidean geometry.

In two-agent wide-assay episodes, collapsed across all timesteps, nearest-agent distance (d_{agent}) and Knollenorgan features (presence I_{agent}^K ; count n_{agent}^K) were the top-decoded variables, followed by log Mormyromast field magnitude ($\log |\vec{E}_M|$) and wall distance (d_{wall}) (Fig. 6E). Furthermore, decoding revealed strong context-dependence (Fig. 6F): distance to the conspecific (d_{agent}) was reliably decoded across both food-range conditions; food distance (d_{food}) and food bearing (α_{food}) tended to improve in decoding performance when food was within sensing range. Knollenorgan features (I_{agent}^K , n_{agent}^K) decoded at moderate levels in both conditions.

Partial least squares correlation (PLSC) revealed proximity-dependent correlated latent structure between paired agents (Fig. 6G,H). First-component correlation decreased across conspecific-distance bins (≤ 10 , 10–100, and > 100 cm; Fig. 6G). The number of statistically significant shared dimensions was similar for pairs within 100 cm (n.s. between the ≤ 10 and 10–100 cm bins) but collapsed to near zero beyond 100 cm (Fig. 6H).

RNN power-spectrum analyses showed that the presence of a nearby conspecific significantly reduced mean wideband RNN power (Fig. 6I), while subordinate (smaller) agents carried significantly higher mean wideband power than dominant (larger) agents (Fig. 6J).

Together, these analyses show that trained agents encode electrosensory task geometry and social context in

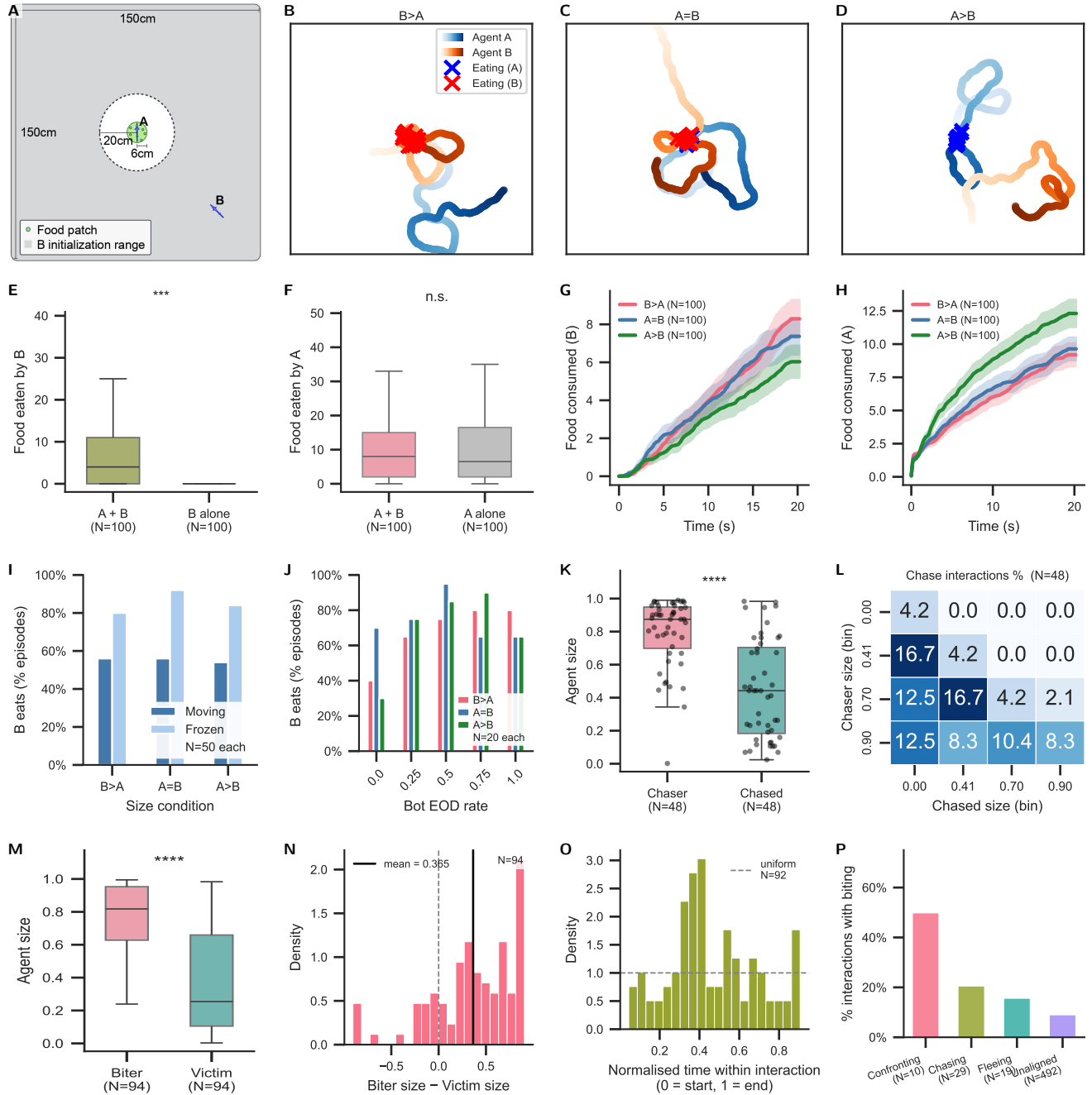


Fig. 5. Two-fish foraging assays reveal the role of signaling and dominance in social interactions. (A) Minimal two-agent single-patch foraging assay. Agent A is initialized within a replenishing food patch, while agent B is initialized at a random location within communication range. (B–D) Example trajectories showing B approaching a patch occupied by A; trajectories are colored from light (start) to dark (end). (B) When B is larger than A, B approaches the patch directly and displaces A. (C) When A and B are the same size, both agents access the patch. (D) When B is smaller than A, B's approach is more indirect and patch access is reduced. (E, F) Food consumed by B (E) and A (F) in social versus non-social (alone) conditions across size asymmetry classes. (G) Timecourse of food consumption by agent B across size asymmetry classes; curves show mean $\pm$ SEM. (H) Timecourse of food consumption by agent A across size asymmetry classes; curves show mean $\pm$ SEM. (I, J) Agent A is replaced by a rule-based random walker "bot" confined to the central patch. (I) B's patch-reach rate grouped by relative size, across all bot EOD rates, stratified by whether bot is moving or frozen in place. (J) B's patch-reach rate grouped by walker EOD emission rate, stratified by size condition and aggregated across bot-movement conditions. (K, L) In arenas with two agents and uniformly distributed food, larger agents are more likely to act as chasers: body size by chaser/chased role as a boxplot (K) and as a pairwise size heatmap (L). (M) Body size by biter/bitten role (boxplot); larger agents bite more frequently. (N) Size advantage of the biter over the bitten agent. (O) Timing of biting events within interaction windows. (P) Biting rate broken down by interaction class. Boxplot significance annotations use panel-level exploratory Mann-Whitney or Wilcoxon tests as appropriate.

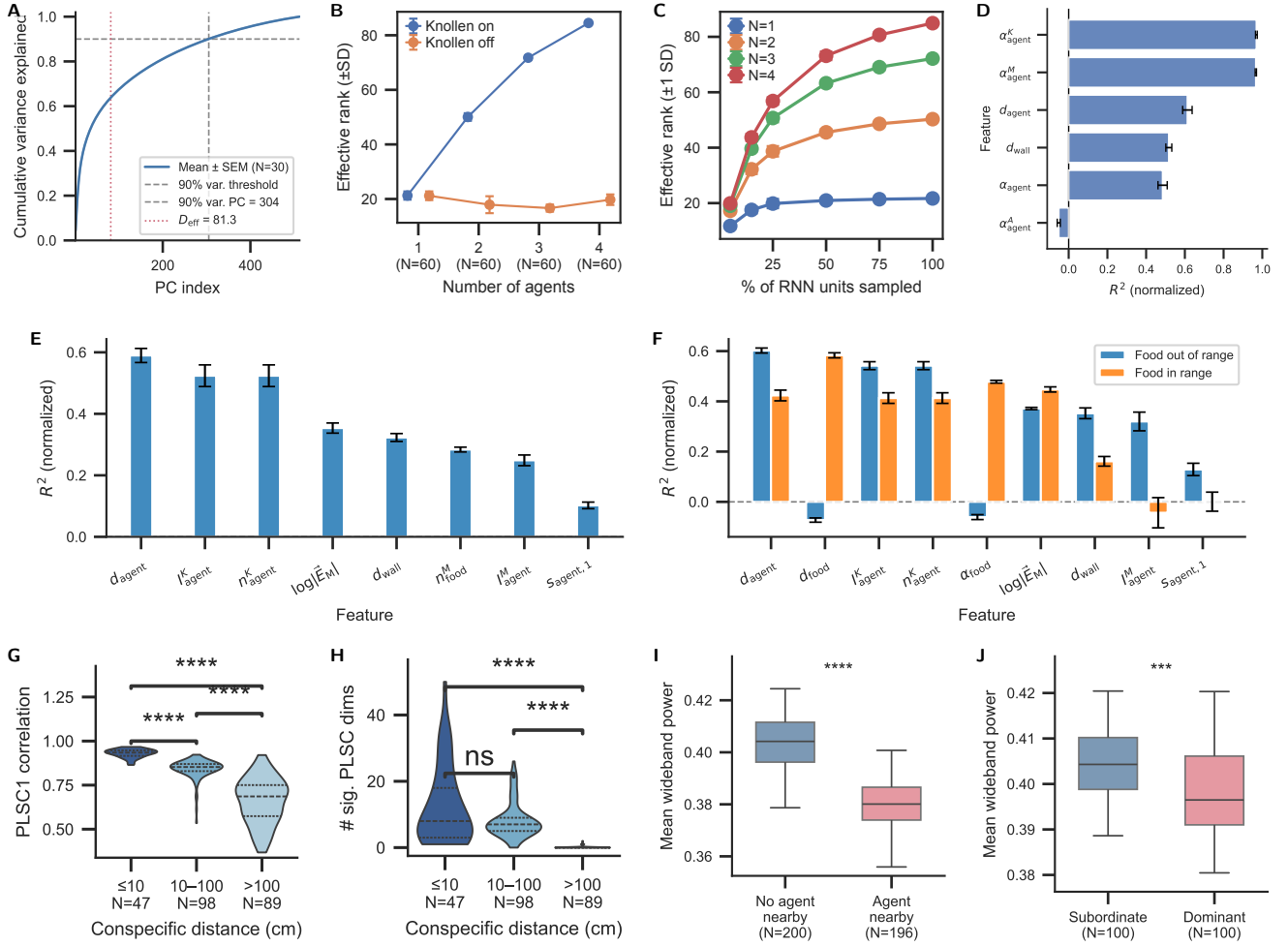


Fig. 6. Recurrent network activity encodes task, social, and behavioral context. (A) PCA of recurrent hidden-state activity shows that most variance is captured by a relatively low-dimensional subspace, with effective rank well below the full hidden-state size ($D_H = 512$). (B) Effective dimensionality (mean $\pm$ SD) increases with the number of interacting agents. When Knollenorgan is ablated, it stays steady around the effective dimensionality of the solo foraging setting. (C) Subsampling analysis shows that estimated effective rank increases with the fraction of recurrent units sampled, across group sizes, but even a small subsampling fraction can reveal differences across group sizes. (D) Homing-task linear decoding shows that the Knollenorgan-defined error angle (Θ_{error}^K) is decoded with higher test performance than the straight-line agent angle (Θ_{agent}), indicating stronger representation of electrosensory task geometry in recurrent activity. (E) Linear decoding performance for task-relevant variables across all timesteps (combined condition): nearest-agent distance (d_{agent}), Knollenorgan presence and count ($I_{\text{agent}}^K, n_{\text{agent}}^K$), log Mormyromast field magnitude ($\log|\vec{E}_M|$), and wall distance (d_{wall}). (F) Linear decoding performance split by food-proximity condition (food near vs. far): distance to conspecific (d_{agent}), food distance (d_{food}), food bearing (α_{food}), and Knollenorgan features ($I_{\text{agent}}^K, n_{\text{agent}}^K$); error bars show SEM across grouped cross-validation folds for both (E–F). (G,H) Partial least-squares correlation (PLSC) analysis: first-component correlation (G) and number of statistically significant shared dimensions (H) are both higher for agent pairs within interaction range than for out-of-range pairs; PLSC significance uses the circular-shift null described in Methods. (I) Episode-level mean wideband RNN power is lower for timesteps when a conspecific is nearby. (J) Episode-level mean wideband RNN power differs by relative size, with subordinate (smaller) agents showing higher wideband power than dominant (larger) agents.

structured, low-dimensional population dynamics, and that interacting agents exhibit proximity-dependent correlated latent dynamics.

4 DISCUSSION

Our results demonstrate that biophysically grounded agents trained with multi-agent reinforcement learning can qualitatively reproduce a broad range of known weakly electric fish behaviors and yield causal insights that would be difficult to obtain experimentally. By validating against established experiments, extending to novel conditions, and

conducting interventions that are otherwise impractical, the framework provides a useful complement to empirical research. It shows that biophysically grounded sensory structure, coupled with simple individual reward structures, is sufficient for rich, flexible group dynamics that mirror natural behavior. Our goal is not to provide a complete biophysical replica of weakly electric fish, nor to claim that the mechanisms discovered here are uniquely required in biological animals; rather, we use a deliberately simplified but biologically grounded closed-loop model to test whether realistic electrosensory structure, individual incentives, and

multi-agent interaction are *sufficient* to generate observed collective phenomena like electrosensing, foraging, and social dynamics.

4.1 Connections with known biology

Our agents reproduced a range of well-characterized behaviors of weakly electric fish, providing qualitative validation of the framework and suggesting mechanistic hypotheses for phenomena that remain difficult to study experimentally. Furthermore, virtual experiments in Figs. 4, 5 and 6 provide testable hypotheses that motivate future experimental follow-up.

Curvilinear homing trajectories. Without explicitly encoding homing geometry, agents followed smooth curved paths to targets, replicating the hallmark behavior reported in classic experiments on real fish^{9,19} (Fig. 2). Agents using a non-electric ray-casting baseline, which provided direct distance and object-type cues, followed straight-line paths instead (Fig. 2B), confirming that curved trajectories arise from the electrosensory modality rather than from goal-directed navigation in general. RNN decoding further showed that the network represents field-line geometry rather than Euclidean geometry (Fig. 6D), consistent with the electrosensory basis of the curvilinear trajectories.

Heavy-tailed and context-dependent EOD statistics. IDI histograms during foraging were heavy-tailed and context-dependent, qualitatively matching distributions measured in Mormyrid fish in laboratory and field settings^{20,34,39,41,42}. Our results suggest that this statistical structure need not be hard-wired; it can emerge from a reward-driven policy that flexibly modulates EOD emission based on foraging state, social context, and prior history. Event-locked analyses further revealed EOD modulation around eating, biting, and being-bitten events (Fig. 3H–J), rather than emitting at a fixed baseline rate. While we do not see a full cessation of EOD emission by subordinate individuals as observed in real fish^{12,43}, we do see a size-dependent reduction in EOD emission rate during interactions (Fig. 3P).

Emergent foraging dynamics. Agents trained under simple individual-fitness incentives spontaneously developed collective foraging strategies that varied flexibly with food distribution, group composition, and social context (Fig. 3). Agents demonstrated behavioral flexibility by alternating between exploration, targeted approach, and social interaction in ways consistent with observations of Mormyrid foraging^{22,44}, and with experimental evidence that social interactions improve foraging efficiency and compress income inequality in fish groups¹. This suggests that flexible collective foraging can emerge from the alignment of individual incentives with shared sensory ecology, without explicit group-level coordination signals.

Sensory contributions. Mirroring prior experimental work⁴⁵, ablating either short-range sensor (Mormyromasts or Ampullary) reduces foraging rate but does not stop foraging altogether (Fig. 4A–B).

Collective sensing. Within the Mormyromast channel, removing access to the conspecific-EOD image (self-image-only condition) drastically reduced food consumption, while removing the self-EOD image (cons-image-only condition) had little effect (Fig. 4G)^{14,15,46}. EOD emission probability was correspondingly lower in agents restricted to the self-generated image and elevated in those restricted to the conspecific image (Fig. 4H), implicating the cons-EOD image in both foraging efficiency and signaling regulation.

Long-range signaling. Classic *producer-scrounger* theory in behavioral ecology⁴⁷ tells us that individuals can exploit conspecific signals as public information at low cost, providing a functional basis for long-range sensory tuning. In our agents, Knollenorgan ablation left total food consumption largely intact while increasing EOD emission probability and close-range aggression, and reducing inter-agent spacing (Fig. 4C–F)¹⁵, suggesting this channel contributes primarily to social coordination rather than individual foraging efficiency. This generates a testable prediction: selectively blocking Knollenorgan-mediated signaling in real fish groups should increase agonistic interactions and reduce social spacing without substantially impairing individual foraging efficiency.

Active signaling. Silencing a single agent’s EOD emission reduced group-level food consumption and increased inter-agent spacing (Fig. 4I,J), indicating that active electrocommunication supports both foraging coordination and competitive spacing, distinct from the passive-reception role of the Knollenorgans².

Environmental structure. Patchy food distributions supported higher per-episode consumption than uniform distributions (Fig. 4K) without significantly affecting inequality (Fig. 4L), suggesting that spatial structure boosts collective foraging efficiency without systematically concentrating resources. Increasing patch count with fixed total food progressively reduced per-fish intake while decreasing inequality (Fig. 4M,N), consistent with the idea that resource subdivision reduces competitive skew by creating multiple simultaneous access points.

Aggression and size-dependent asymmetries. Training agents with size and speed heterogeneity produced emergent dominance-like asymmetries: larger agents consumed disproportionately more food and directed more aggression toward smaller conspecifics, consistent with ethological observations of size- and EOD-related dominance in weakly electric fish^{2,48}. The two-fish assays further showed that larger intruders were more likely to displace residents and monopolize food patches, mirroring behavioral observations of dyadic agonistic encounters in Mormyrids and size-dependent dominance dynamics in other weakly electric species^{48–50}. Notably, dominance-like behavior emerged from size-dependent movement constraints and asymmetric penalties for being bitten, without any explicit reward encouraging biting or chasing. In real weakly electric

fish, subordinate individuals sometimes temporarily silence their EODs during agonistic encounters as a submissive signal⁴³; whether a similar learned suppression emerges in agents trained under stronger social pressure or with richer communication structure is a direct test for future work. These assays generate testable predictions: larger intruders should systematically displace smaller residents, with displacement probability scaling with size difference; and intruders should reach occupied patches more readily when the resident is stationary or actively emitting, consistent with conspecific EODs serving as informative cues for patch discovery²¹ (Fig. 5A–J).

Neural dynamics and correlated recurrent-state structure. Effective dimensionality of population activity increased with the number of interacting agents, and this scaling was abolished when the Knollenorgan was ablated (Fig. 6B,C), implicating long-range electrocommunication as a driver of social-context-dependent state-space expansion. PLSC analysis showed that the first shared axis between two agents’ RNN states was significantly more correlated at close range than when agents were out of range, with shared dimensionality collapsing to near zero beyond 100 cm, showing a proximity-dependent pattern that qualitatively parallels the shared neural subspaces recently reported in interacting mice^{51,52}. Wideband RNN power was suppressed when a conspecific was nearby (Fig. 6I) and was elevated in subordinate relative to dominant agents (Fig. 6J), indicating proximity- and status-dependent modulation of recurrent dynamics.

4.2 Limitations and future work

We do not collect new experimental behavioral or neural data here. The model is instead intended to precede and guide future data collection by identifying regimes, perturbations, and readouts most informative to test experimentally. This is especially important in weakly electric fish collectives, where simultaneous behavioral tracking, EOD assignment, and multi-animal neural recordings remain technically challenging⁵³. Our framework provides a controlled setting in which signal sources, sensory access, neural states, and perturbations are fully known, allowing us to generate experimentally actionable predictions before committing to costly animal experiments.

Future work will use this framework in closer dialogue with experimental data. Beyond foraging, we will train agents on richer social tasks such as territorial defense, competition, and cooperation, and test whether the resulting EOD sequences and interaction motifs match newly collected behavioral and neural data⁵⁴. A further direction is to probe context-modulated electrocommunication through unsupervised machine-translation approaches⁵⁵, where virtual agents provide a fully observable ground truth for signal-state-action relationships while biological experiments test whether similar structure exists in real fish.

A broader limitation of RL/MARL agents is that out-of-distribution generalization can be brittle. Our

training regime partially mitigates this by exposing agents to substantial variability across initial positions, orientations, food configurations, social contexts, and stochastic rollouts, encouraging policies that generalize across a broad family of parameter regimes rather than memorizing a single arena configuration.

The model makes several simplifying assumptions. Agents are modeled as circular particles rather than elongated, three-dimensional fish bodies. The simulation does not model fluid dynamics or fluid-body interactions, and the electric field and sensory models assume electrostatic conditions in a time-discretized simulation with a 12 ms timestep, ignoring within-step electrodynamics. Since real weakly electric fish continuously integrate information across electrosensory, visual, and mechanosensory modalities, behaviors observed in our agents may differ quantitatively from those in real collectives, particularly at short range where hydrodynamic cues dominate. We do not model sex differences, which in real weakly electric fish drive rich EOD waveform variation and mate-preference dynamics^{12,56}. Future work could address these limitations through embodied 3D simulation with fluid dynamics, richer contact mechanics for aggression, sexual dimorphism, longer-timescale training, and larger arenas supporting more naturalistic exploration. Regardless, the framework demonstrates that virtual experiments can narrow the hypothesis space, accelerate experimental design, and uncover principles of electrosensory collective behavior that would be difficult or impossible to access through conventional experiments alone.

5 METHODS

5.1 Simulation details

Simulations consisted of N artificial weakly electric fish (WEF)-like agents, typically $N = 4$, interacting with each other and with simulated prey in a bounded 2D arena. At each simulation step, neural-network policies generated actions for each agent, and the simulator updated agent kinematics, collisions, prey interactions, EOD generation, electric-field interactions, and sensor observations.

The simulator used first-order command kinematics: policy outputs were mapped directly to bounded translational and rotational velocity commands. The simulator enforced non-overlap constraints and wall collisions, and ran at 83 frames per second to match the observed latency in the WEF’s social echo response (12 ms)⁵⁷, close to the minimum observed inter-pulse interval in this WEF species.

5.2 Agent design

Each agent was represented as a circular particle with diameter d , allocentric (ground-frame) orientation θ , and allocentric position $\mathbf{r} = (x, y)$.

Agents chose actions independently at every timestep from a multivariate action space comprising: (1) forward translation, (2) angular rotation, (3) electric organ discharge (EOD) emission, and (4) bite. These collectively

controlled locomotion, electrolocation, communication, and close-range aggression. The policy output was continuous; the environment interpreted movement outputs as velocity commands and thresholded the EOD and bite channels into binary events.

Observations included input from three types of biomimetic electrosensors, last action taken, and egocentric displacement vectors. We introduced inter-agent heterogeneity by assigning each agent a scalar “size” parameter, which scaled its maximum speed. An agent’s own size was included as a fixed value context input in the observation vector to allow size-dependent behavioral variation.

5.3 Electric field model

We followed the electrosensing model introduced by Chen et al.¹⁸ and built on a broader literature on electrosensory processing^{14,46,58}.

5.3.1 Field generation

Electric fields entered the simulation through two primary source classes. First, agents generated actively controlled monopole pairs that approximated pulsatile electric discharges from fish electric organs. Second, agents and prey carried always-on dipoles representing intrinsic electric sources.

In our agents, the electric organ was modeled as two internal monopoles with equal and opposite source strengths $\pm q$ (in coulombs) positioned at $\pm d/4$ along the body axis, avoiding singularities at sensor locations. The total electric field $\mathbf{E}_{\text{generated}}$ at position $\mathbf{r}$ combines monopole and dipole contributions:

$$\mathbf{E}_{\text{generated}} = \mathbf{E}_{\text{active}} + \mathbf{E}_{\text{intrinsic}}$$

$$\mathbf{E}_{\text{active}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{q_i(\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3} \quad (1)$$

$$\mathbf{E}_{\text{intrinsic}}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_i \frac{3(\mathbf{p}_i \cdot \hat{\mathbf{r}}_i)\hat{\mathbf{r}}_i - \mathbf{p}_i}{|\mathbf{r} - \mathbf{r}_i|^3} \quad (2)$$

where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m and $\hat{\mathbf{r}}_i$ is the unit vector from source i to the measurement point.

5.3.2 Field interactions with prey and agents

Prey and agent particles were modeled as conducting spheres on which dipole moments were induced by the external electric field. The induced dipole moment is given by:

$$\mathbf{p}_{\text{induced}} = 4\pi\epsilon_0 R^3 \chi \mathbf{E}_{\text{external}} \quad (3)$$

where $\chi = \left(\frac{\epsilon/\epsilon_0 - 1}{\epsilon/\epsilon_0 + 2} \right)$ ($\chi = -0.5$ per Chen et al 2005¹⁸), ϵ_0 is the permittivity of free space, ϵ is the permittivity of the dielectric material, R is the radius of the sphere, $\mathbf{E}_{\text{external}}$ is the total external electric field intersecting the object at its location. The negative sign signifies that the induced dipole is oriented opposite to the external field.

5.3.3 Field interactions with walls

The arena walls were modeled as non-conducting boundaries, and we used the method of images⁵⁹ to account for their effect on the electric field. This method introduced virtual image sources outside the domain to enforce the field boundary conditions. We reflected all active and induced sources and dipoles off each of the arena walls one time, resulting in images at the same distance from the boundary on the other side and with the same magnitude and sign (no sign-flip for non-conducting boundaries⁵⁹); this is a first-order image approximation, with higher-order reflections neglected.

5.4 Electroreception

The combined field available to sensors was:

$$\mathbf{E}_{\text{total}} = \mathbf{E}_{\text{active}} + \mathbf{E}_{\text{intrinsic}} + \mathbf{E}_{\text{induced}} + \mathbf{E}_{\text{reflected}}$$

Agents sensed this field through three biomimetic receptor classes: Mormyromasts, Knollenorgans, and Ampullary sensors. Sensor types were chosen to approximate the functional specialization, spatial arrangement, and relative receptor counts of real mormyrid electroreceptors.

5.4.1 Mormyromast sensors

These were the primary receptors for *active electrolocation*. In our model, Mormyromast channels encoded distortions caused by the agent’s own EOD, yielding a *self-image*, as well as distortions associated with conspecific EODs, yielding a *cons-image*. We simulated $N_m = 36$ Mormyromasts, 30% of which were concentrated near the front 60 degrees of the agent, since the “chin” on real WEF is known to be a foveal region in terms of receptor density and enhanced forward-facing sensing⁹. The remaining sensors were uniformly distributed along the circumference of the agent.

5.4.2 Ampullary sensors

These were modeled as *passive* sensors that in real WEF sense static or low-frequency electric fields (below 40 Hz³⁵) from nearby prey and conspecifics. We simulated $N_a = 24$ Ampullary sensors that were distributed uniformly along the circumference of the agent.

5.4.3 Knollenorgan sensors

These sensors provided passive detection of conspecific EOD communication signals. In real WEF, they are highly sensitive to the sharp, high-frequency EOD signals emitted by conspecifics. In our agents, these sensors were only sensitive to the electric fields generated by other agents’ EODs, and were insensitive to everything else. We simulated $N_k = 12$ uniformly spaced Knollenorgan sensors per conspecific. This allowed the agent to independently sense EODs from each conspecific within each simulation step (12 ms) without collisions in time. Knollenorgan sensors provide waveform intensity information whose magnitude is believed to carry information about conspecific size and sex⁶⁰. Conspecific size metadata was therefore

included in the corresponding observation channels, as described below.

5.4.4 Sensor processing

Field transduction. For each sensor, the electric field impinging on the sensor was converted to a raw measurement by projecting the field vector onto the sensor’s surface-normal vector. Each sensor’s normal vector was directed radially outwards from the center of the agent’s circular body. This projection captured the component of the field that the sensor was oriented to detect:

$$s = \mathbf{E}(\mathbf{r}_s) \cdot \mathbf{n}$$

where $\mathbf{E}(\mathbf{r}_s)$ is the electric field vector at the sensor’s position $\mathbf{r}_s$, and $\mathbf{n}$ is the sensor’s outward-facing normal vector.

Baseline subtraction. To prevent self-EOD and cons-EOD fields from saturating sensor channels, we subtracted precomputed sensor-specific baselines from raw field readings. Baselines were computed once at simulation start and differed by sensor type.

Ampullary sensors: We subtracted two baselines from each raw reading. First, the agent’s own intrinsic dipole field was always subtracted, suppressing static self-field components. Second, when the agent emitted an EOD, we subtracted the self-EOD corollary-discharge component.

Mormyromasts: For each Mormyromast, we always subtract the intrinsic and self-EOD baselines so the channel is sensitive to self-image distortions during active sensing. For cons-image sensing, whenever a conspecific emits an EOD we additionally compute a dynamic *cons-baseline*, i.e. the conspecific’s EOD field at the receiver’s sensors assuming no induced sources on nearby objects, and subtract it from the raw Mormyromast readings, leaving a residual that reflects only the object-induced field perturbations.

Knollenorgans: These channels were tuned to conspecific EOD transients and did not require explicit baseline subtraction.

Scaling and normalization. After baseline subtraction, each sensor reading s was mapped to a signed, unitless value in $[-1, 1]$ via sign-preserving, magnitude-only normalization over that sensor’s dynamic range $[|E_{\min}|, |E_{\max}|]$.

Let $\text{sgn}(s)$ denote the sign of s , and define the clipped magnitude

$$\tilde{m} = \text{clip}(|s|, E_{\min}, E_{\max}).$$

We then computed the normalized log-magnitude

$$\hat{m} = \frac{\log_{10}(\max(\tilde{m}, \epsilon)) - \log_{10}(E_{\min})}{\log_{10}(E_{\max}) - \log_{10}(E_{\min})},$$

where $\epsilon \leq E_{\min}$ is a small constant for numerical stability (default $\epsilon = 10^{-25}$). The final normalized reading is

$$\hat{s} = \text{sgn}(s)\hat{m} \in [-1, 1].$$

This preserves field polarity while compressing dynamic range for stable policy optimization.

5.4.5 Sensor calibration

We calibrated the intrinsic electric sources of agents and prey, and EOD monopole source strengths, to ensure that each class of electrosensor could detect objects within biologically plausible spatial ranges using commonly assumed sensitivity thresholds (Table 1).

As a starting point, we used analytical expressions for dipole fields and induced dipole interactions (Appendix .10), required for each sensor type to reach its respective sensitivity threshold at a target detection distance. We then tuned these values empirically using simulations of one or two fish and a single food object.

For Ampullary sensors (passive), we tuned intrinsic charges (using analytic expressions, see Appendix .10) on food and prey to detect low-frequency electric fields under 40 Hz from prey up to 4 cm away and from conspecifics up to 8 cm away.

For Mormyromasts, calibration was baseline- and dynamic-range based. **Self-images:** we first estimated the self-EOD operating baseline at each sensor, then set dynamic ranges to detect baseline perturbations from prey up to 5 cm and conspecifics up to 10 cm (lower bound $\sim \text{baseline} \times 10^{-6}$; upper bound $\sim \text{baseline} \times 0.25$).

Cons-images: the dynamic cons-baseline is recomputed at each timestep as the conspecific’s EOD field at each Mormyromast sensor, assuming no induced sources on nearby objects. Dynamic ranges for the residual cons-image are set to span $\text{baseline} \times 10^{-6}$ to $\text{baseline} \times 0.25$ relative to the self-EOD operating level.

Knollenorgans, which sense high-frequency EOD pulses from conspecifics, were tuned to detect signals up to 100 cm and encoded only the sign of the local electric field. We denote the resulting Mormyromast and Knollenorgan conspecific-sensing ranges by R_M (≈ 10 cm) and R_K (≈ 100 cm), respectively; these define the proximity conditions used in the decoding analyses (Results).

These calibration choices preserved the intended functional specialization of the three modeled receptor classes: Ampullary sensors provided broad passive sensitivity at short range, Mormyromasts provided spatially resolved active and collective electrolocation over shorter ranges, and Knollenorgans provided long-range directional cues restricted to conspecific EODs.

5.5 Agent observations and actions

The observation vector at each timestep comprised: (i) $N_m = 36$ Mormyromast readings, (ii) $N_a = 24$ Ampullary readings, (iii) $N_k \times (N_{\text{agents}} - 1) = 12 \times 3$ Knollenorgan readings plus a scalar encoding the emitting agent’s size for each conspecific, (iv) the agent’s last action vector, (v) egocentric displacement vectors, and (vi) the agent’s own size (scalar context input). This gives a fixed-length observation vector for the default $N = 4$ group; when evaluating with fewer agents the unused conspecific slots in the Knollenorgan block are zero-padded. Agents moved according to first-order command kinematics; movement

outputs directly set bounded translational and rotational velocity commands. The bounds on translational and rotational speed were derived from empirical tracking data on real weakly electric fish and were scaled by the agent’s size parameter. A bite event, available when the thresholded bite command was active and a conspecific was within the fixed biting radius, yielded a close-range aggression event with reward consequences. Food consumption was handled separately and did not result from biting another agent.

5.6 Training

Algorithm. Agents were trained jointly using a shared-parameter variant of Proximal Policy Optimization (PPO)^{31,33}, with Generalized Advantage Estimation (GAE) for variance reduction³². Training was conducted over 20 independent random seeds. After training, seeds were ranked using the ethologically motivated selection criterion described below, rather than by training reward alone. Hyperparameter details are given in Supplementary Section .4.1.

Reward structure. Each agent received an individual reward signal comprising: (1) a positive reward for consuming food; (2) a penalty for being bitten, scaled by the size difference between biter and victim in heterogeneous groups; (3) a small cost for successful biting; (4) a collision penalty; (5) a shaping term proportional to the decrease in distance to the nearest food item, normalized by arena perimeter; and (6) an effort-over-threshold penalty for large movement or turn commands. No explicit reward for social communication or coordination was provided; any emergent signaling arose solely from the shared environment and individual fitness incentives. The proximity-shaping term provided a dense learning signal for approach behavior under what would otherwise be a sparse-reward foraging task, and was included for sample-efficiency in line with standard RL practice on long-horizon navigation problems. Reward coefficient values are given in Supplementary Section .4.3.

Homing task. For the homing validation (Fig. 2), we separately trained agents in a minimal two-agent setup whose reward comprised a proximity reward for reaching the partner, a small distance-decrease shaping term, and a small time penalty, with no foraging or aggression terms. At evaluation, the target agent was fixed and always emitting while the homing agent approached from random initial positions. The target agent was also fixed during training, matching the fixed-target evaluation assay. For comparison in Fig. 2, the ray-casting baseline was an alternative non-electric perception model in which radial rays returned object identity and distance-to-intersection cues for food, walls, or other agents. Unlike the electrosensing model, this baseline did not compute electric fields or induced dipoles, and therefore provided direct geometric line-of-sight information rather than physics-based electroreceptor

measurements. For linear decoding analyses of homing performance (Fig. 6D), the Knollenorgan error angle Θ_{error}^K was defined as the signed angular difference between the agent’s heading and the direction of the electric field produced by the target agent at the homing agent’s sensor array.

Neural network architecture. Each agent had independent actor and critic networks, each comprising an MLP encoder followed by a single-layer GRU⁶¹ ($D_H = 512$), whose hidden state was carried across timesteps. The actor network’s hidden state drove a mixed continuous-discrete action distribution (translation, rotation, EOD emission probability, bite probability); the critic network’s hidden state produced a scalar state-value estimate. Full architecture details are given in Supplementary Section .4.4.

Seed selection. We trained 20 independent random seeds using the same architecture, environment, and reward function for 5M environment steps each, yielding variably performing agents at convergence of the high-dimensional stochastic nonconvex optimization⁶². After training, seeds were evaluated on a fixed assay of test rollouts that was not used for parameter updates. We selected a representative seed using a pre-specified ethologically motivated ranking procedure designed to identify policies that both performed well and expressed experimentally observed size-dependent resource asymmetries.

For each seed and rollout, we computed three summary metrics. First, we measured total food consumed by all agents, corresponding to social foraging performance and individual-fitness maximization. Second, we fitted a linear model relating food consumed by each agent to that agent’s size,

$$\text{food_eaten}_i = \beta_0 + \beta_1 \text{agent_size}_i + \varepsilon_i,$$

and summarized the seed by the median slope (β_1) across rollouts. This metric measures alignment with the experimental observation that larger agents obtain more food. Third, we computed the median (R^2) of this linear fit across rollouts, measuring the consistency of the size–food relationship.

To exclude degenerate policies in which one or more agents failed to forage, we filtered out seeds whose median intercept from the food-eaten versus agent-size fit was non-positive. Remaining seeds were ranked using the three evaluation metrics: total food consumed, median size–food slope, and median fit (R^2). As a consequence, the size–food relationship summarized in Fig. 3M is best seen as a quantitative confirmation, not an independent finding. The combined top-ranked seed was used for the detailed behavior, perturbation, and recurrent-dynamics analyses reported in the Results.

5.7 Evaluation and analysis

5.7.1 Evaluation protocol

Post-training evaluations were performed with fixed network weights. Evaluation settings were assay specific. The main 4-agent patchy and uniform foraging assays used 30 episodes in $70\text{ cm} \times 70\text{ cm}$ arenas. The homing assay used 100 episodes with pre-specified initial positions in a $70\text{ cm} \times 70\text{ cm}$ arena, with the homing agent initialized at least 20 cm from the target and both agents at least 5 cm from the walls. Two-agent square assays used $70\text{ cm} \times 70\text{ cm}$ arenas, the two-agent wide assay used a $160\text{ cm} \times 40\text{ cm}$ arena, and one-patch assays used a single centrally placed replenishing food patch. Initial positions, orientations, and size settings were randomized or fixed according to the protocol for each assay. Specific arena and food configurations for each analysis are described in the corresponding results subsection.

5.7.2 Inter-discharge interval extraction and fitting

The interdischarge intervals from real fish were derived from published dyadic behavior²¹. Published dyadic position labels were converted to interdischarge intervals using the 50 kHz sampling rate of the recordings.

IDs from artificial agents were derived from the two-agent wide assay used for Fig. 3C–J, with two agents in a $160\text{ cm} \times 40\text{ cm}$ arena and 100 evaluation episodes.

Powerlaw fits were calculated via the Python powerlaw package⁴⁰.

5.7.3 EOD rate and peri-event analyses

EOD rate was computed per timestep and binned by context variable (distance to nearest agent, food item, or wall) using 15 uniform bins spanning the relevant distance range; each bin reports mean $\pm$ SEM across observations. For peri-event analyses, EOD rate was aligned to eating and biting events; a symmetric window of ± 100 timesteps ($\pm 1200\text{ ms}$ at 83 Hz) around each event was extracted and averaged across events with a minimum of 3 qualifying events per agent.

5.8 Ablation and intervention experiments

Food abundance experiments. Agents were evaluated in $70\text{ cm} \times 70\text{ cm}$ arenas with patchy and uniform food distributions under the food-density settings used for each assay (Fig. 4). Food-scarcity sweeps included reduced patchy-food multipliers of $0.5\times$ and $0.25\times$ for the active sensor combinations used in the intervention analyses.

***N*-patch experiments.** In a second food-manipulation assay, agents were evaluated in $70\text{ cm} \times 70\text{ cm}$ arenas containing a fixed number of patches, $N \in \{1, 2, 3, 4, 5\}$. We used two sweeps: an iso-food sweep, where the per-patch food multiplier was scaled so that total food remained approximately fixed across N , and an unconstrained sweep, where each patch used the standard food multiplier so that total available food increased with N .

Sensor ablations. Sensor ablation experiments examined the effect of changing access to Mormyromast, Ampullary, and Knollenorgan channels. The intervention set includes single-class removals from the full model, mixed two-sensor controls, Knollenorgan-only and all-sensor-off controls. These evaluations were conducted in $70\text{ cm} \times 70\text{ cm}$ patchy arenas with an intermediate amount of food unless otherwise specified for a given assay. The reduced arena size ensured that agents remained within Knollenorgan range of one another. A parallel one-fish assay (single agent in isolation, no conspecifics present; Fig. 4A) was run alongside the four-agent ablations (Fig. 4B) so that sensor effects on individual foraging could be separated from those requiring social interaction.

EOD muting. EOD muting experiments assessed the role of active electrocommunication in group foraging and aggression. One of the four agents had its EOD emission silenced for the duration of the evaluation; all other policy components remained intact, so the muted agent could still sense and move but emitted no EODs, and its conspecifics received no EOD-related input from it. The intact group (all four agents emitting) served as the control condition. Experiments were conducted in the standard $70\text{ cm} \times 70\text{ cm}$ patchy arenas used for the sensor ablation assays.

Collective sensing ablations. Collective sensing ablations isolated the contributions of self-generated and conspecific-generated EOD images in the Mormyromast channel. Weakly electric fish have the capacity to sense both from self-generated EODs (self-EODs) as well as from conspecific EODs (cons-EODs)¹⁵. We manipulated the sensing capabilities of the mormyromasts to create three conditions: (1) control, with both self- and cons-EOD sensing enabled; (2) self-EOD only, where agents cannot benefit from other agents' EODs; and (3) cons-EOD only, where agents only received feedback through the mormyromast channel from EODs generated by other agents. These manipulations did not affect the ampullary channel. These experiments were conducted in a very small $40\text{ cm} \times 40\text{ cm}$ arena to ensure close proximity, as the cons-EOD benefits were expected only when agents were close to one another. Food was uniformly distributed at intermediate density. Each condition received 30 evaluation runs.

5.9 Two-agent experiments

Two-agent foraging. Two-agent experiments allowed closer examination of how EOD modulation and relative-size asymmetries shape pairwise competition. We evaluated two agents in a $70\text{ cm} \times 70\text{ cm}$ square arena with intermediate uniform food abundance for pairwise interaction and recurrent-state analyses. Agents were initialized with random positions, orientations, and size settings. We also evaluated agents in a *wide* $160\text{ cm} \times 40\text{ cm}$ arena to quantify long-range EOD modulation, IDI statistics, and attraction or

repulsion between two agents. Agents were initialized in a 160×40 cm arena, one on the left at (20, 20) and the other on the right side of the arena at (140, 20).

Bot (random-walker) control. To isolate social cues from food-associated gradients, in a subset of two-agent assays we replaced the resident agent with a rule-based bot confined to the food patch. The bot performed a random walk within the patch region and did not respond to its partner or to any sensory input. We ran conditions in which the bot was either stationary (*frozen*) or mobile, and in which it emitted EODs according to a simple stochastic schedule at a controlled rate (including a non-emitting, zero-rate condition). This provided a partner that supplied food-associated and EOD cues but no adaptive social behavior, and whose emission timing differed in temporal structure from that of trained agents.

One-patch dyadic assay. To assess resource-access competition between two agents differing in body size, we evaluated agents in a one-patch arena with a single centrally placed replenishing food patch (150 cm $\times$ 150 cm; see Supplementary Section .2). Agent A (resident) was initialized on the patch; agent B (intruder) was initialized at a random position within Knollenorgan detection range (≤ 100 cm from A). We tested three relative size conditions: $A < B$ (intruder larger), $A = B$ (size-matched), and $A > B$ (resident larger), plus bot-control variants in which the resident was replaced by the random-walking bot described above.

We defined pairwise interactions between agents A and B using their allocentric positions $\mathbf{r}_A(t), \mathbf{r}_B(t)$ and orientations $\theta_A(t), \theta_B(t)$ at each timestep t .

Interaction bouts. An interaction bout between two agents was defined as a maximal contiguous segment of time during which the agents remained within close spatial proximity. Bout start and end times were determined by the first and last timesteps of such proximity.

Distance thresholds. During each bout we tracked the inter-agent distance $d_{AB}(t)$. Two distance thresholds were used: a **contact threshold** ($d_{\text{contact}} = 2$ cm), indicating physical contact if $d_{AB}(t) \leq d_{\text{contact}}$ at any timestep, and a **reduced interaction threshold** ($d_{\text{reduce}} = 5$ cm), used when computing distance-restricted interaction statistics.

Geometric interaction labels. At each timestep we classified the spatial relationship between the two agents based on the alignment of each agent’s heading relative to the bearing toward the other agent.

Let $\phi_{A \rightarrow B}(t)$ denote the bearing from A to B, and $\phi_{B \rightarrow A}(t)$ the bearing from B to A. We computed the angular difference between each agent’s orientation and the bearing to its

partner:

$$\Delta_A(t) = \theta_A(t) - \phi_{A \rightarrow B}(t), \quad \Delta_B(t) = \theta_B(t) - \phi_{B \rightarrow A}(t).$$

Using angular thresholds $\theta_{\text{close}} = \pi/4$ and $\theta_{\text{opp}} = 3\pi/4$, interaction states were defined as: **Confronting** occurred when both agents faced toward each other ($|\Delta_A| < \theta_{\text{close}}$ and $|\Delta_B| < \theta_{\text{close}}$). **Chasing** occurred when agent A faced toward B while B faced away ($|\Delta_A| < \theta_{\text{close}}$ and $|\Delta_B| > \theta_{\text{opp}}$). **Fleeing** occurred when agent B faced toward A while A faced away ($|\Delta_B| < \theta_{\text{close}}$ and $|\Delta_A| > \theta_{\text{opp}}$). **Dispersing** occurred when both agents faced away from each other ($|\Delta_A| > \theta_{\text{opp}}$ and $|\Delta_B| > \theta_{\text{opp}}$). All other relative orientations were classified as **unaligned**.

Thus, in a *chasing* configuration agent A is the chaser and B the chased, whereas in a *fleeing* configuration the roles are reversed.

Bout-level summaries and biting. Each interaction bout was assigned a dominant interaction class defined as the most frequent timestep label during the bout. Aggressive encounters were defined as bouts in which either agent performed a bite action or was bitten by its partner at any point during the interaction.

5.10 Neural analyses

5.10.1 Effective dimensionality

We quantified the effective rank of RNN population activity using the participation ratio of PCA explained variance ratios:

$$D_{\text{eff}} = \frac{1}{\sum_i \tilde{\lambda}_i^2}$$

where $\tilde{\lambda}_i$ is the fraction of variance explained by the i -th principal component (eigenvalues normalized to sum to 1). D_{eff} was computed per episode by running PCA on the $T \times D_H$ hidden-state matrix (T timesteps, $D_H = 512$ units), then averaged across episodes and environments.

For group-size comparisons (Fig. 6B), the same trained policy was evaluated in separate rollouts with $N \in \{1, 2, 3, 4\}$ agents in 70 cm $\times$ 70 cm patchy arenas; D_{eff} was computed per episode and averaged across episodes within each condition. To verify that group-size differences in D_{eff} were not an artifact of the number of simultaneously recorded units, we performed unit subsampling (Fig. 6C): for each group size we repeatedly ($n = 30$ per fraction) drew random subsets of hidden units at fractions $\{0.05, 0.15, 0.25, 0.5, 0.75, 1.0\}$ and recomputed D_{eff} from the subsampled activity.

5.10.2 Feature extraction

We extracted a broad panel of hand-curated features spanning egocentric geometry (orientation, turn angle, displacement), electrosensory inputs (Knollenorgan, Mormyromast, and Ampullary field angles and magnitudes), distances to agents, food, and walls, temporal recency of key events (time since last EOD, feeding, or encounter), boolean

task features (nearby agents, bites, emissions, eating), and local counts of food or agents. Together, these circular, scalar, boolean, probability, and count features capture the sensory, spatial, and behavioral context available to the agent at each timestep. The full set of features is listed in Table 3.

5.10.3 Decoding analyses

We quantified how individual task and behavioral variables could be linearly decoded from the agents’ recurrent states. For each target, we fit ordinary least-squares linear models from recurrent hidden state to the target value and evaluated generalization on held-out episodes using grouped cross-validation. Scalar, boolean, probability, and count targets were treated as real-valued regression targets and scored with held-out R^2 . Circular targets were decoded by fitting parallel linear models to $\cos \phi$ and $\sin \phi$ and averaging the held-out R^2 across the two components. For each cross-validation fold, we subtracted the performance of a mean-prediction baseline fit on the training episodes, yielding a baseline-normalized test score. Reported error bars for decoding panels show the standard error across cross-validation folds.

5.10.4 Partial least squares correlation (PLSC)

We used partial least squares correlation (PLSC) to quantify shared population structure in the activity of pairs of RNNs. For every episode with at least one second of pairwise interaction (agents within the Mormyromast conspecific-sensing range R_M), we concatenated agent hidden states aligned by time and made balanced “in-range” and “out-of-range” datasets. After z-scoring hidden-state features, we fitted PLSC to find shared axes of variation and then computed the Pearson correlation between the paired canonical variates. Significance was assessed against a null distribution constructed by circularly shifting one agent’s hidden-state sequence in time ($n=30$ shifts), preserving temporal autocorrelation while disrupting alignment between agents. A component was counted as significant only when both its singular value and projected correlation exceeded the 97.5th percentile of the corresponding null distributions (one-sided test). The number of significant dimensions was counted contiguously from the first component until the first non-significant component. Additional details on PLSC are in Supplementary Section .5.2.

5.10.5 Power spectral density

We quantified wideband power in recurrent hidden-state activity for each episode, environment, and agent. Hidden-state time series were transformed with a short-time spectrogram using 20-timestep windows with 10-timestep overlap, and power was averaged across recurrent units, time windows, and frequencies to obtain one mean wideband-power value per agent episode. These episode-level power values were associated with timestep-level social labels to compare activity during

periods with and without a nearby conspecific, and with agent-level size status in two-agent assays. For size-status comparisons, the larger member of a pair was labeled as the larger agent and the smaller member as the smaller agent.

5.11 Statistical analyses and annotation

Statistical summaries were computed at the level appropriate to each assay: evaluation episodes, agent episodes, interaction bouts, agent pairs, events, or cross-validation folds. Food consumption inequality within groups was quantified using the Theil T index

$$T = \frac{1}{N} \sum_{i=1}^N \frac{x_i}{\mu} \ln \frac{x_i}{\mu},$$

where x_i is food consumed by agent i and μ is the group mean; $T = 0$ indicates perfect equality⁶³. In biting analyses, the win ratio for each agent is the fraction of biting events involving that agent in which it was the biter: win ratio = $n_{\text{given}} / (n_{\text{given}} + n_{\text{received}})$. For binned EOD-distance and peri-event EOD analyses, curves show mean $\pm$ SEM (standard error of the mean) across observations or events entering each bin or time offset. For all box-and-whisker plots, where ≤ 50 observations are being plotted, we have superimposed the jittered scatterplot on the plot. For intervention panels comparing each manipulation to a control condition, significance annotations use Dunnett one-vs-control tests. For two-group unpaired comparisons, we used two-sided Mann-Whitney tests. For paired role comparisons within the same interaction, we used two-sided Wilcoxon signed-rank tests. Linear decoding analyses report held-out, baseline-normalized R^2 with SEM across grouped cross-validation folds. PLSC significance was assessed with the circular-shift null described above. Unless otherwise specified, tests were treated as panel-level exploratory comparisons and no additional correction was applied across figure panels. Statistical annotations use the following convention:

Symbol	Significance range
ns	$5.00 \times 10^{-2} < p \leq 1.00$
*	$1.00 \times 10^{-2} < p \leq 5.00 \times 10^{-2}$
**	$1.00 \times 10^{-3} < p \leq 1.00 \times 10^{-2}$
***	$1.00 \times 10^{-4} < p \leq 1.00 \times 10^{-3}$
****	$p \leq 1.00 \times 10^{-4}$

DATA AND CODE AVAILABILITY

Training, simulation, and analysis code, as well as examples of trained policies, are available at <https://github.com/KempnerInstitute/wef>

ETHICS AND REPRODUCIBILITY

All experiments are in silico. Our framework reduces invasive experimentation on live animals by enabling in silico hypothesis generation and prioritization. All experimental data used in this paper were collected by

collaborators for previous neuroscientific studies of weakly electric fish.

AUTHOR CONTRIBUTIONS (CREDIT TAXONOMY)

S.H.S. and S.J.Y.: conceptualization, methodology, software, formal analysis, visualization, writing – original draft. Z.L., A.W., F.P., and D.T.: methodology, software. P.S. and N.S.: writing – review and editing. S.H.S., S.J.Y., N.B.S., and K.R.: conceptualization. N.B.S. and K.R.: supervision. K.R.: funding acquisition. All authors reviewed and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Eugene Vinitsky, Yohan John, Ann Huang, Gizem Ozdil, Nathan Wu, Roy Harpaz, Ella Batty, Kianté Brantley, Na (Lina) Li, Sam Gershman, Yuyang Zhang, Thomas Fel, Daphne Cornelisse, and members of the Rajan, Brantley and Gershman labs for helpful discussions. Funded by NIH (RF1DA056403), James S. McDonnell Foundation (220020466), Simons Foundation (Pilot Extension-00003332-02), McKnight Endowment Fund, CIFAR Azrieli Global Scholar Program, NSF (2046583), Harvard Medical School Dean’s Innovation Award, Harvard Medical School Neurobiology Lefler Small Grant Award, Alice and Joseph Brooks Fund Postdoctoral Fellowship (S.H.S.), Shanahan Family Foundation Fellowship at the Interface of Data and Neuroscience at the Allen Institute and University of Washington, supported in part by the Allen Institute (D.T.).

REFERENCES

- [1] Harpaz, R. & Schneidman, E. Social interactions drive efficient foraging and income equality in groups of fish. *eLife* **9**, e56196 (2020).
- [2] Arnegard, M. E. & Carlson, B. A. Electric organ discharge patterns during group hunting by a mormyrid fish. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1305–1314 (2005).
- [3] Couzin, I. D., Krause, J., James, R., Ruxton, G. D. & Franks, N. R. Collective memory and spatial sorting in animal groups. *Journal of theoretical biology* **218**, 1–11 (2002).
- [4] Katz, Y., Tunström, K., Ioannou, C. C., Huepe, C. & Couzin, I. D. Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences* **108**, 18720–18725 (2011).
- [5] Alageshan, J. K., Verma, A. K., Bec, J. & Pandit, R. Machine learning strategies for path-planning microswimmers in turbulent flows. *Physical Review E* **101**, 043110 (2020).
- [6] Löffler, R. C., Panizon, E. & Bechinger, C. Collective foraging of active particles trained by reinforcement learning. *Scientific Reports* **13**, 17055 (2023).
- [7] Wiczorek, T. J., Tchumatchenko, T., Wert-Carvajal, C. & Eggl, M. F. A framework for the emergence and analysis of language in social learning agents. *Nature Communications* **15**, 7590 (2024).
- [8] von der Emde, G. Electrolocation of capacitive objects in four species of pulse-type weakly electric fish: II. electric signalling behaviour. *Ethology* **92**, 177–192 (1992).
- [9] Von der Emde, G. Active electrolocation of objects in weakly electric fish. *Journal of experimental biology* **202**, 1205–1215 (1999).
- [10] Electoreception and communication in electric fish (2024).
- [11] Caputi, A. A. The electric organ discharge of pulse gymnotiforms: the transformation of a simple impulse into a complex spatio-temporal electromotor pattern. *Journal of Experimental Biology* **202**, 1229–1241 (1999).
- [12] Carlson, B. A. Electric signaling behavior and the mechanisms of electric organ discharge production in mormyrid fish. *Journal of Physiology-Paris* **96**, 405–419 (2002).
- [13] Jones, T. K., Allen, K. M. & Moss, C. F. Communication with self, friends and foes in active-sensing animals. *Journal of Experimental Biology* **224**, jeb242637 (2021).
- [14] Wallach, A. & Sawtell, N. B. An internal model for canceling self-generated sensory input in freely behaving electric fish. *Neuron* **111**, 2570–2582.e5 (2023).
- [15] Pedraja, F. & Sawtell, N. B. Collective sensing in electric fish. *Nature* **628**, 139–144 (2024).
- [16] Kramer, B. Communication Behavior and Sensory Mechanisms in Weakly Electric Fishes. In *Advances in the Study of Behavior*, vol. 23, 233–270 (Elsevier, 1994).

- [17] Hopkins, C. D. Neuroethology of Electric Communication .
- [18] Chen, L., House, J. L., Krahe, R. & Nelson, M. E. Modeling signal and background components of electrosensory scenes. *Journal of Comparative Physiology A* **191**, 331–345 (2005).
- [19] Hopkins, C. D., Shieh, K.-T., McBride Jr, D. W. & Winslow, M. A quantitative analysis of passive electrolocation behavior in electric fish; pp. 45-59. *Brain, Behavior & Evolution* **50** (1997).
- [20] Gebhardt, K., Alt, W. & Emde, G. v. d. Electric discharge patterns in group-living weakly electric fish, *Mormyrus rume* (Mormyridae, Teleostei) (2012).
- [21] Chrtkova, I. *et al.* Unsupervised electric signal separation for linking behavior and electrocommunication in *gnathonemus petersii*. *Scientific Reports* **15**, 33542 (2025).
- [22] Biswas, D. *et al.* Mode switching in organisms for solving explore-versus-exploit problems. *Nature Machine Intelligence* **5**, 1285–1296 (2023).
- [23] Merel, J. *et al.* Deep neuroethology of a virtual rodent. *arXiv preprint arXiv:1911.09451* (2019).
- [24] Singh, S. H., van Breugel, F., Rao, R. P. & Brunton, B. W. Emergent behaviour and neural dynamics in artificial agents tracking odour plumes. *Nature Machine Intelligence* **5**, 58–70 (2023).
- [25] Turcu, D., Zadina, A. N., Abbott, L. & Sawtell, N. B. An end-to-end model of active electrosensation. *Current Biology* **35**, 2295–2306 (2025).
- [26] Keller, R. *et al.* Autonomous behavior and whole-brain dynamics emerge in embodied zebrafish agents with model-based intrinsic motivation. *arXiv preprint arXiv:2506.00138* (2025).
- [27] Vaxenburg, R. *et al.* Whole-body physics simulation of fruit fly locomotion. *Nature* 1–3 (2025).
- [28] Lobato-Rios, V. *et al.* Neuromechfly, a neuromechanical model of adult drosophila melanogaster. *Nature Methods* **19**, 620–627 (2022).
- [29] Wang-Chen, S. *et al.* Neuromechfly v2: simulating embodied sensorimotor control in adult drosophila. *Nature Methods* **21**, 2353–2362 (2024).
- [30] Chung, J., Gulcehre, C., Cho, K. & Bengio, Y. Gated feedback recurrent neural networks. In *International conference on machine learning*, 2067–2075 (PMLR, 2015).
- [31] Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).
- [32] Ni, T., Eysenbach, B. & Salakhutdinov, R. Recurrent model-free rl is a strong baseline for many POMDPs. *arXiv preprint arXiv:2110.05038* (2021).
- [33] Yu, C. *et al.* The surprising effectiveness of PPO in cooperative multi-agent games. *Advances in Neural Information Processing Systems* **35**, 24611–24624 (2022).
- [34] Moller, P., Serrier, J. & Bowling, D. Electric Organ Discharge Displays during Social Encounter in the Weakly Electric Fish *Brienomyrus niger* L. (Mormyridae). *Ethology* **82**, 177–191 (1989).
- [35] Benda, J. The Physics of Electrosensory Worlds. In *The Senses: A Comprehensive Reference*, 228–254 (Elsevier, 2020).
- [36] Sawtell, N. B., Williams, A. & Bell, C. C. From sparks to spikes: information processing in the electrosensory systems of fish. *Current opinion in neurobiology* **15**, 437–443 (2005).
- [37] Wallach, A. & Sawtell, N. B. An Internal Model of Sensorimotor Context in Freely Swimming Electric Fish. preprint, Neuroscience (2022).
- [38] Baker, C. A., Kohashi, T., Lyons-Warren, A. M., Ma, X. & Carlson, B. A. Multiplexed temporal coding of electric communication signals in mormyrid fishes. *Journal of Experimental Biology* **216**, 2365–2379 (2013).
- [39] Carlson, B. A. & Hopkins, C. D. Stereotyped temporal patterns in electrical communication. *Animal Behaviour* **68**, 867–878 (2004).
- [40] Alstott, J., Bullmore, E. & Plenz, D. powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one* **9**, e85777 (2014).
- [41] Schuster, S. Count and spark? The echo response of the weakly electric fish *Gnathonemus petersii* to series of pulses. *Journal of Experimental Biology* **204**, 1401–1412 (2001).
- [42] Henninger, J., Krahe, R., Kirschbaum, F., Grewe, J. & Benda, J. Statistics of natural communication signals observed in the wild identify important yet neglected stimulus regimes in weakly electric fish. *Journal of Neuroscience* **38**, 5456–5465 (2018).
- [43] Moller, P. *Electric Fishes: History and Behavior*. Chapman & Hall fish and fisheries series (Springer Netherlands, 1995).
- [44] Friedman, M. A. & Hopkins, C. D. Tracking individual mormyrid electric fish in the field using electric organ discharge waveforms. *Animal Behaviour* **51**, 391–407 (1996).
- [45] Emde, G. V. D. & Bleckmann, H. Finding food: senses involved in foraging for insect larvae in the electric fish *gnathonemus petersii*. *Journal of Experimental Biology* **201**, 969–980 (1998).
- [46] Enikolopov, A. G., Abbott, L. & Sawtell, N. B. Internally Generated Predictions Enhance Neural and Behavioral Detection of Sensory Stimuli in an Electric Fish. *Neuron* **99**, 135–146.e3 (2018).
- [47] Giraldeau, L.-A. & Caraco, T. *Social foraging theory* (Princeton University Press, 2018).
- [48] Terleph, T. A. & Moller, P. Effects of social interaction on the electric organ discharge in a mormyrid fish, *gnathonemus petersii* (mormyridae, teleostei). *Journal of experimental biology* **206**, 2355–2362 (2003).

- [49] Worm, M. *et al.* Evidence for mutual allocation of social attention through interactive signaling in a mormyrid weakly electric fish. *Proceedings of the National Academy of Sciences* **115**, 6852–6857 (2018).
- [50] O’Shaughnessy, L., Izawa, T., Masai, I., Shaevitz, J. W. & Stephens, G. J. Dynamics of dominance in interacting zebrafish. *PRX Life* **2**, 043006 (2024).
- [51] Zhang, X. *et al.* Inter-brain neural dynamics in biological and artificial intelligence systems. *Nature* 1–11 (2025).
- [52] Redman, W. T., Dinc, F. D., Lin, X., Chan, M. G. & Alexander, A. S. Predictive pursuit emerges in high-dimensional recurrent neural networks. *bioRxiv* 2026–04 (2026).
- [53] Zheng, K. *et al.* Keypoint annotation for electrocommunication source separation with pikachu and raichu. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems workshop: AI for non-human animal communication* (2025).
- [54] Huang, A. *et al.* Inputds: Demixing then comparing recurrent and externally driven dynamics. *arXiv preprint arXiv:2510.25943* (2025).
- [55] Singh, S. H. *et al.* Proposal: Deciphering electrocommunication with marl and unsupervised machine translation. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems workshop: AI for non-human animal communication* (2025).
- [56] Zhou, M. & Smith, G. T. Structure and sexual dimorphism of the electrocommunication signals of the weakly electric fish, *adontosternarchus deventanzii*. *Journal of Experimental Biology* **209**, 4809–4818 (2006).
- [57] Heiligenberg, W. *Principles of electrolocation and jamming avoidance in electric fish: a neuroethological approach* (Springer Science & Business Media, 2012).
- [58] Muller, S. Z., Zadina, A. N., Abbott, L. & Sawtell, N. B. Continual Learning in a Multi-Layer Network of an Electric Fish. *Cell* **179**, 1382–1392.e10 (2019).
- [59] Griffiths, D. J. *Introduction to electrodynamics* (Cambridge University Press, 2023).
- [60] Von der Emde, G. & Schwarz, S. Imaging of objects through active electrolocation in *gnathonemus petersii*. *Journal of Physiology-Paris* **96**, 431–444 (2002).
- [61] Cho, K. *et al.* Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1724–1734 (2014).
- [62] Huang, A., Singh, S. H., Martinelli, F. & Rajan, K. Measuring and controlling solution degeneracy across task-trained recurrent neural networks. *arXiv preprint arXiv:2410.03972* (2024).
- [63] Conceição, P. & Ferreira, P. The young person’s guide to the theil index: Suggesting intuitive interpretations and exploring analytical applications (2000).
- [64] Stoddard, P. K., Tran, A. & Krahe, R. Predation and Crypsis in the Evolution of Electric Signaling in Weakly Electric Fishes. *Frontiers in Ecology and Evolution* **7**, 264 (2019).
- [65] Fukutomi, M. & Carlson, B. A. A History of Corollary Discharge: Contributions of Mormyrid Weakly Electric Fish. *Frontiers in Integrative Neuroscience* **14**, 42 (2020).
- [66] Engelmann, J., Wallach, A. & Maler, L. Linking active sensing and spatial learning in weakly electric fish. *Current Opinion in Neurobiology* **71**, 1–10 (2021).
- [67] Wallach, A., Melanson, A., Longtin, A. & Maler, L. Mixed selectivity coding of sensory and motor social signals in the thalamus of a weakly electric fish. *Current Biology* **32**, 51–63.e3 (2022).
- [68] Bod’Ová, K., Mitchell, G. J., Harpaz, R., Schneidman, E. & Tkačik, G. Probabilistic models of individual and collective animal behavior. *PloS one* **13**, e0193049 (2018).
- [69] Huang, C., Ling, F. & Kanso, E. Collective phase transitions in confined fish schools. *Proceedings of the National Academy of Sciences* **121**, e2406293121 (2024).
- [70] Harpaz, R. *et al.* Collective behavior emerges from genetically controlled simple behavioral motifs in zebrafish. *Sci. Adv.* **7**, eabi7460 (2021).
- [71] Li, L. *et al.* Reverse engineering the control law for schooling in zebrafish using virtual reality. preprint, In Review (2023).
- [72] Papaspyros, V. *et al.* Predicting the long-term collective behaviour of fish pairs with deep learning .
- [73] Haesemeyer, M., Schier, A. F. & Engert, F. Convergent temperature representations in artificial and biological neural networks. *Neuron* **103**, 1123–1134 (2019).
- [74] Chen, C., Hou, Y. & Ong, Y.-S. A conceptual modeling of flocking-regulated multi-agent reinforcement learning. In *2016 International Joint Conference on Neural Networks (IJCNN)*, 5256–5262 (IEEE, 2016).
- [75] Chipade, V. S. & Panagou, D. Multiagent planning and control for swarm herding in 2-d obstacle environments under bounded inputs. *IEEE Transactions on Robotics* **37**, 1956–1972 (2021).
- [76] Tsutsui, K., Tanaka, R., Takeda, K. & Fujii, K. Collaborative hunting in artificial agents with deep reinforcement learning. *Elife* **13**, e85694 (2024).
- [77] Goldshtein, A. *et al.* Reinforcement Learning Enables Resource Partitioning in Foraging Bats. *Current Biology* **30**, 4096–4102.e6 (2020).
- [78] McGraw, J. D., Lee, D. & Wood, J. N. Parallel development of social behavior in biological and artificial fish. *Nature Communications* **15**, 10613 (2024).
- [79] Lindsay, G. W. Grounding neuroscience in behavioral changes using artificial neural networks. *Current opinion in neurobiology* **84**, 102816 (2024).

- [80] Hasson, U., Nastase, S. A. & Goldstein, A. Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron* **105**, 416–434 (2020).
- [81] Nastase, S. A., Goldstein, A. & Hasson, U. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage* **222**, 117254 (2020).
- [82] Li, C., Kreiman, G. & Ramanathan, S. Discovering neural policies to drive behaviour by integrating deep reinforcement learning agents with biological neural networks. *Nature Machine Intelligence* **6**, 726–738 (2024).
- [83] Singh, S. H. Neuroprospecting with DeepRL agents. NeurIPS 2021 Workshop on AI for Science (2021).
- [84] Lazaridou, A. & Baroni, M. Emergent Multi-Agent Communication in the Deep Learning Era (2020). ArXiv:2006.02419 [cs].
- [85] Grupen, N. A., Lee, D. D. & Selman, B. Low-Bandwidth Communication Emerges Naturally in Multi-Agent Learning Systems (2020). ArXiv:2011.14890 [cs].
- [86] Rachum, R., Nakar, Y., Tomlinson, B., Alon, N. & Mirsky, R. Emergent Dominance Hierarchies in Reinforcement Learning Agents (2024). ArXiv:2401.12258 [cs].

SUPPLEMENTARY INFORMATION

.1 Related Work

Our work builds upon research spanning neuroethology, collective behavior, and artificial intelligence.

Weakly electric fish have been a key model system for understanding how animals use self-generated signals to sense their surroundings and communicate. Pulse-type weakly electric fish (*Mormyridae*) emit brief electric pulses known as electric organ discharges (EODs), whose timing and structure can be modulated depending on behavioral context^{9,11,12,16}. Their electrosensory system includes three major receptor classes: Mormyromasts for active electrolocation, Knollenorgans for passive detection of conspecific EODs, and Ampullary organs for low-frequency passive sensing. Together, these systems allow these nocturnal fish to sense their environment in the turbid, visually challenging conditions that characterize their natural habitat³⁵. Work on social and agonistic behavior has shown that EOD patterns carry information about identity, dominance, and aggression^{34,38,39,64}, while recent studies suggest that groups of weakly electric fish can benefit from one another’s signals through collective sensing¹⁵. At the physiological level, increasingly detailed biophysical and computational models are beginning to explain how electrosensory circuits process these signals^{14,18,25,36,65–67}. Our contribution is to place these biophysical foundations inside a closed-loop, multi-agent decision-making framework.

Classical agent-based models have shown that simple local rules can generate complex group-level phenomena in fish schools and bird flocks^{68–70}. Foundational rule-based approaches identified alignment, repulsion, and attraction zones as sufficient ingredients for schooling and milling³, while data-driven and deep-learning approaches have inferred interaction rules or predicted collective trajectories directly from data^{4,71,72}. More recently, MARL has been used to model collective locomotion and schooling^{5,6}. While these models have provided important insights into collective dynamics, they typically rely on hand-designed interaction rules and simplified sensory representations. As a result, they do not fully capture the closed-loop coupling between perception, motor control, communication, and ecological context that characterizes natural behavior.

Artificial neural network based RL agents have emerged as a tool for studying neural mechanisms underlying complex biological naturalistic behavior, both for individual behavior^{23,24,26–29,73} and collective behaviors such as pursuit, swarming, flocking, resource partitioning, and cooperative hunting^{51,52,74–78}. These efforts illustrate how biologically grounded RL frameworks can provide mechanistic hypotheses that complement experimental data, consistent with broader calls to embrace naturalistic, less tightly constrained models in neuroscience^{79–81}. Full observability and controllability of the artificial neural network enables mechanistic interpretation and in silico ablations^{24,82,83}.

A related literature studies emergent communication among artificial agents trained on cooperative or competitive tasks. In these settings, agents often develop signaling systems that improve coordination and task performance^{7,84}, including under low-bandwidth communication constraints⁸⁵ or in socially competitive environments⁸⁶. However, these studies typically consider explicit, often symbolic, communication channels rather than signals grounded in biophysically constrained sensing.

.2 Arena and Foraging Environment

The arena is a rectangular, bounded region with reflective walls. During training, width and height are independently sampled from [40, 300] cm, so the policy is exposed to a broad range of arena scales. Evaluation assays use fixed sizes: 70 cm × 70 cm for square foraging and homing, 160 cm × 40 cm for the wide two-agent assay, and 150 cm × 150 cm for dyadic patch-access assays.

Four food-distribution classes were used in reported experiments:

- **Uniform:** food pellets distributed uniformly across the full arena.
- **Patchy:** circular patches placed at a fixed spatial density; each patch radius is drawn from a truncated normal distribution.
- **N-Patch:** N circular patches of fixed radius arranged evenly around an ellipse centered in the arena, with a random starting angle.
- **One-Patch:** a single circular patch at the arena center.

In all foraging analyses, food is initialized at episode start via a Poisson draw with density ρ_0 (pellets cm⁻²) and neither replenishes nor decays during the episode. The one-patch dyadic assay is the exception: the patch replenishes continuously, sustaining competition across the episode.

.3 Dynamic Cons-Baseline Sensing Model

A key challenge in modeling collective sensing is that the conspecific (cons-) EOD field at a receiver's Mormyromast sensors is itself a dynamic, large-magnitude baseline that varies with inter-fish distance and orientation. The object-induced perturbation the receiver wants to detect (the *cons-image*) is orders of magnitude smaller than this direct cons-EOD field.

To isolate the cons-image, we implement a *dynamic cons-baseline* model inspired by the functional mechanism proposed by Pedraja and Sawtell¹⁵. Whenever a conspecific emits an EOD, we compute a *cons-baseline* at each receiver Mormyromast: the electric field that the conspecific's EOD would produce at that sensor location in the absence of any induced sources on nearby objects. Subtracting this cons-baseline from the total field reading leaves a residual that reflects only the perturbations induced on surrounding objects, i.e. the cons-image:

$$s_{\text{cons-image},k} = s_{\text{total},k} - s_{\text{cons-baseline},k}$$

where $s_{\text{total},k}$ is the raw reading at Mormyromast k and $s_{\text{cons-baseline},k}$ is the pre-subtracted direct cons-EOD contribution at that sensor, computed from the conspecific's EOD dipole field without object interactions.

The cons-baseline is recomputed at each timestep as the conspecific's position and emission state change, making it fully adaptive to inter-fish geometry.

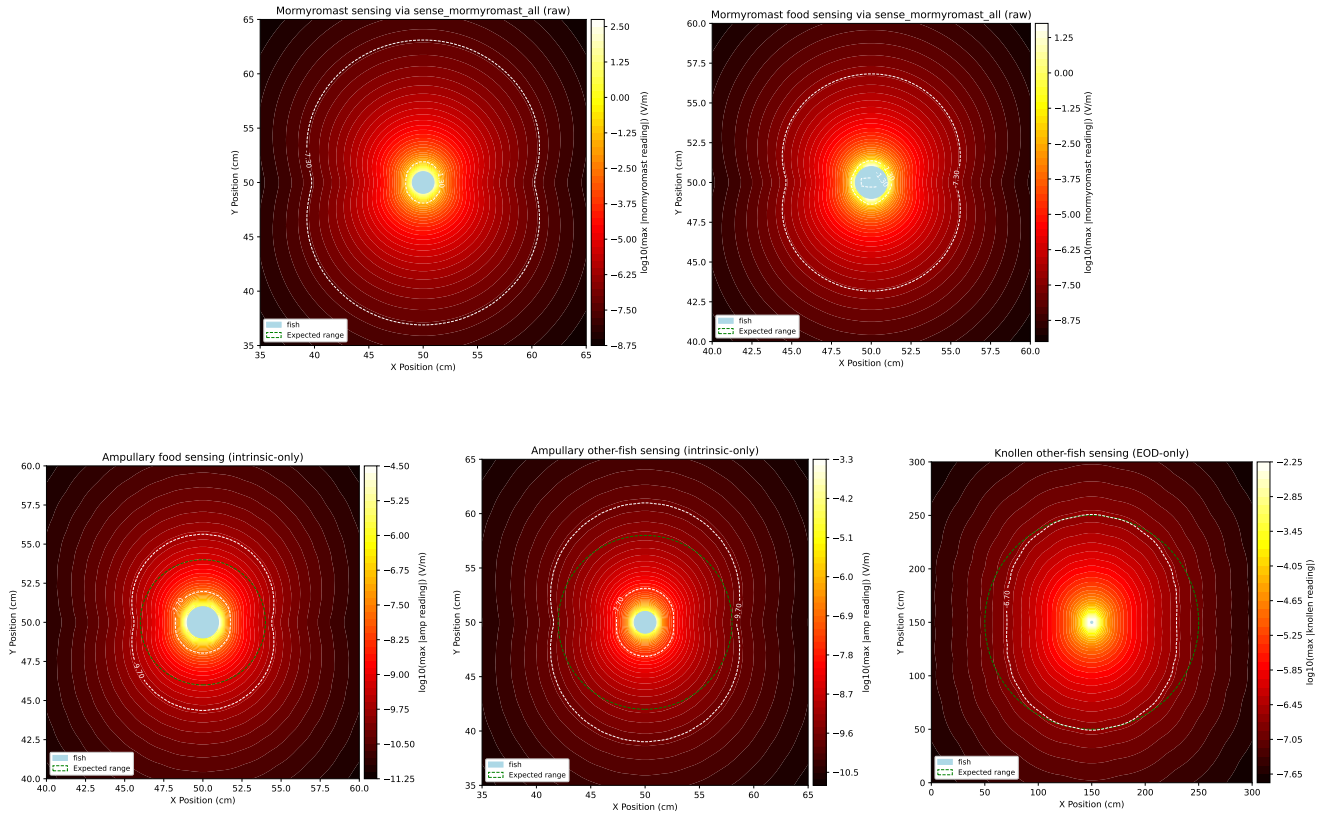


Fig. 7. Sensing ranges for different sensing modalities: Maximum intensity of **self-images** produced by food objects [Upper Left] and conspecifics [Upper Right] when actively sensed by **Mormyromasts**. Maximum intensity of intrinsic dipoles emanating from prey [Lower Left] and conspecifics [Lower Middle], sensed by **Ampullary** receptors. [Lower Right] Maximum intensity of passively sensed conspecific EODs using **Knollenorgans** (that binarize the received signal).

.4 Multi-Agent Reinforcement Learning

The foraging task is formulated as a decentralized partially observable Markov decision process (Dec-POMDP): at execution time each agent acts using only its local electric-sense observations with no explicit communication channel, so any emergent signaling arises solely through shared environmental consequences. During training, the critic receives the concatenated observations of all agents (centralized training, decentralized execution; CTDE), which is the standard value-estimation strategy of MAPPO³³. Each agent’s Knollenorgan observation includes a scalar encoding the emitting conspecific’s size, consistent with the biological finding that Knollenorgan waveform intensity carries information about conspecific size and sex⁶⁰.

.4.1 Training Algorithm

We trained 20 independent seeds for 5M environment steps each using Proximal Policy Optimization (PPO)³¹ extended to the recurrent, multi-agent setting^{32,33}. During training, arena width and height were each sampled independently from [40, 300] cm per episode, exposing the policy to a broad range of spatial scales and food-patch densities. Key algorithmic components:

- **Actor-critic architecture:** independent actor and critic networks, each with an MLP encoder, a recurrent trunk, and a linear output head (Section .4.4).
- **Generalized Advantage Estimation (GAE)** with $\lambda = 0.95$ for variance reduction.
- **Recurrent rollout buffer:** hidden states are carried across timesteps within an episode; truncated backpropagation through time (BPTT) is used with a fixed horizon to limit memory requirements.
- **Clipped surrogate objective** with clip parameter $\epsilon = 0.2$.
- **Entropy regularization** with coefficient c_H to encourage exploration.
- **Shared policy with agent-size-specific conditioning:** all N agents share a single set of weights; inter-agent heterogeneity is encoded via the scalar size context input in the observation.
- **Seed selection:** the best-performing model across 20 seeds is selected based on mean per-episode food consumption and the slope and goodness-of-fit (R^2) of food eaten regressed on agent size, evaluated on held-out episodes (Fig. 11).

Hyperparameters were taken directly from the original on-policy MARL codebase³³, which was tuned for continuous-action continuous-control tasks; no additional tuning was performed for this work.

Each random seed produces a slightly different solution, as expected for equilibrium-finding in non-convex multi-agent settings (Fig. 10, 11). Despite this variability, the top seeds share the same qualitative patterns: size-based dominance, structured EOD modulation, and low-dimensional RNN activity (Fig. 12). We selected seed S1 for the detailed analyses in the main text.

.4.2 Computational resources

Training and evaluation were run on a workstation with a single NVIDIA RTX 5090 GPU and on a GPU cluster using equivalent hardware. Each seed took approximately 6–9 hours to complete training and post-training evaluation.

.4.3 Reward Structure

The individual reward at each timestep is:

$$r_t = r_t^{\text{food}} + r_t^{\text{prox}} + r_t^{\text{bitten}} + r_t^{\text{bite}} + r_t^{\text{collision}} + r_t^{\text{effort}} \quad (4)$$

where:

- $r_t^{\text{food}} = +c_f$ whenever the agent successfully consumes a food item within its eating zone.
- $r_t^{\text{prox}} = c_p \cdot (d_{t-1}^{\text{food}} - d_t^{\text{food}}) / P_{\text{arena}}$: shaping reward proportional to the decrease in distance to the nearest food item, normalized by arena perimeter.
- r_t^{bitten} : penalty for being bitten. In heterogeneous groups, this penalty is scaled by $(1 + s_{\text{biter}} - s_{\text{victim}})$, so being bitten by a larger agent is penalized more strongly.
- r_t^{bite} : small cost applied when the agent successfully bites another agent.
- $r_t^{\text{collision}}$: penalty applied when the agent collides with another object or boundary.
- r_t^{effort} : penalty for movement or turn commands above the effort threshold.

The timestep and wall-proximity reward terms were zero in all reported experiments. No explicit reward for EOD emission, communication, or social coordination is provided.

.4.4 Neural Network Architecture

- **Actor network:** two-layer MLP encoder (ReLU activations, LayerNorm after each layer, hidden size 512) followed by a single-layer GRU⁶¹ ($D_H = 512$), with a linear head producing a continuous action distribution over movement commands, an EOD logit, and a bite logit. The environment thresholds the EOD and bite channels into binary events.
- **Critic network:** same MLP and GRU structure, with a linear head producing a scalar value estimate.

.5 Analyses

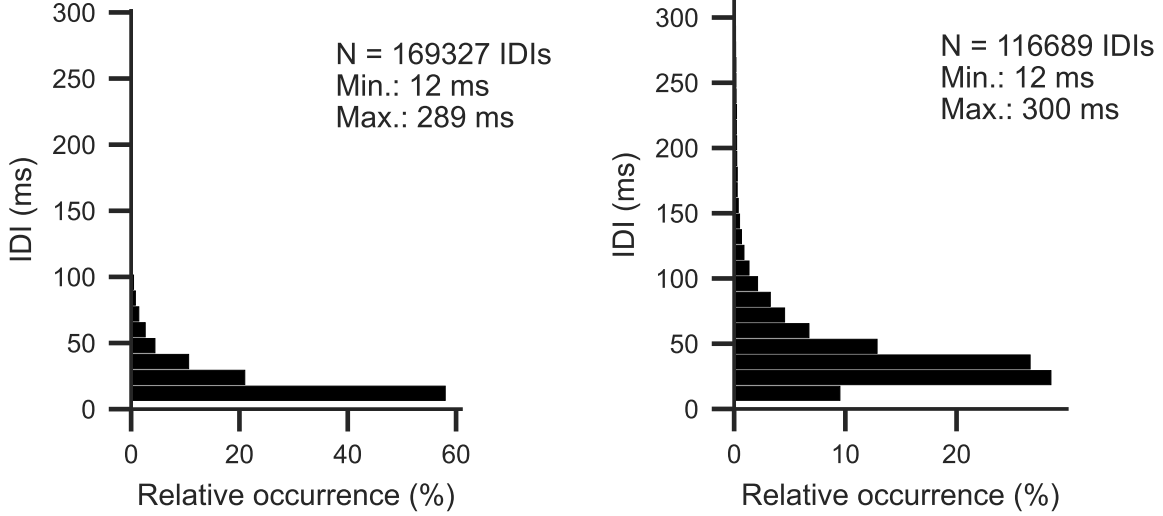


Fig. 8. Inter-discharge intervals from two freely foraging agents [Left], and from experimental data from two fish²¹ [Right]. These figures are inspired by Gebhardt et al. Fig. 4C²⁰ (and also resemble the data provided there.)

.5.1 Interaction definitions

See Fig. 9 for a schematic that summarizes the various types of dyadic agent interactions.

.5.2 PLSC details

Given two mean-centered data matrices $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$ (here, RNN hidden states of two agents across n timesteps), PLSC computes the cross-covariance matrix

$$C_{XY} = X^\top Y \in \mathbb{R}^{p \times q}.$$

Singular value decomposition (SVD) of this matrix,

$$C_{XY} = U \Sigma V^\top,$$

yields orthogonal weight vectors U and V for X and Y , respectively. Projecting X and Y onto these weight vectors produces paired latent variables (canonical variates), whose sample correlations quantify the strength of shared structure along each dimension (e.g. PLSC1, the first-component correlation).

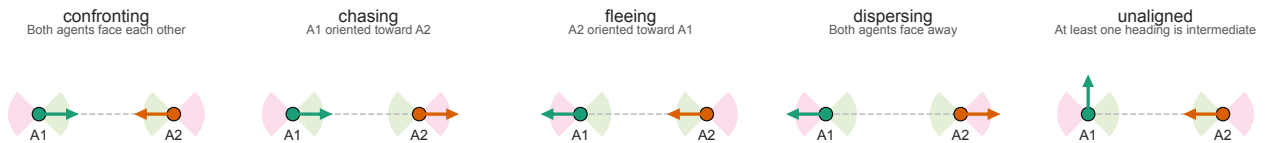


Fig. 9. Illustrative examples of interactions between two agents (A1 and A2). Green sector: $|\text{rel angle}| < \pi/4$ (toward other). Pink sector: $|\text{rel angle}| > 3\pi/4$ (away from other).

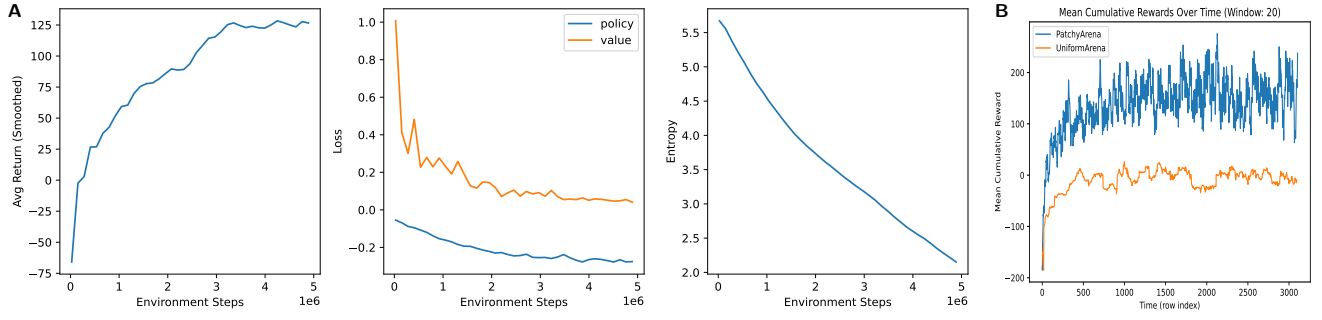


Fig. 10. Training diagnostics for the selected seed (S1). (A) Smoothed average episode return (left), policy and value loss (centre), and action entropy (right) over 5M environment steps, showing stable convergence. (B) Mean cumulative reward (moving window of 20 updates) separated by arena type sampled during training (patchy vs. uniform).

.6 Supplementary Figures

The following figures provide additional training diagnostics and inter-seed reproducibility evidence.

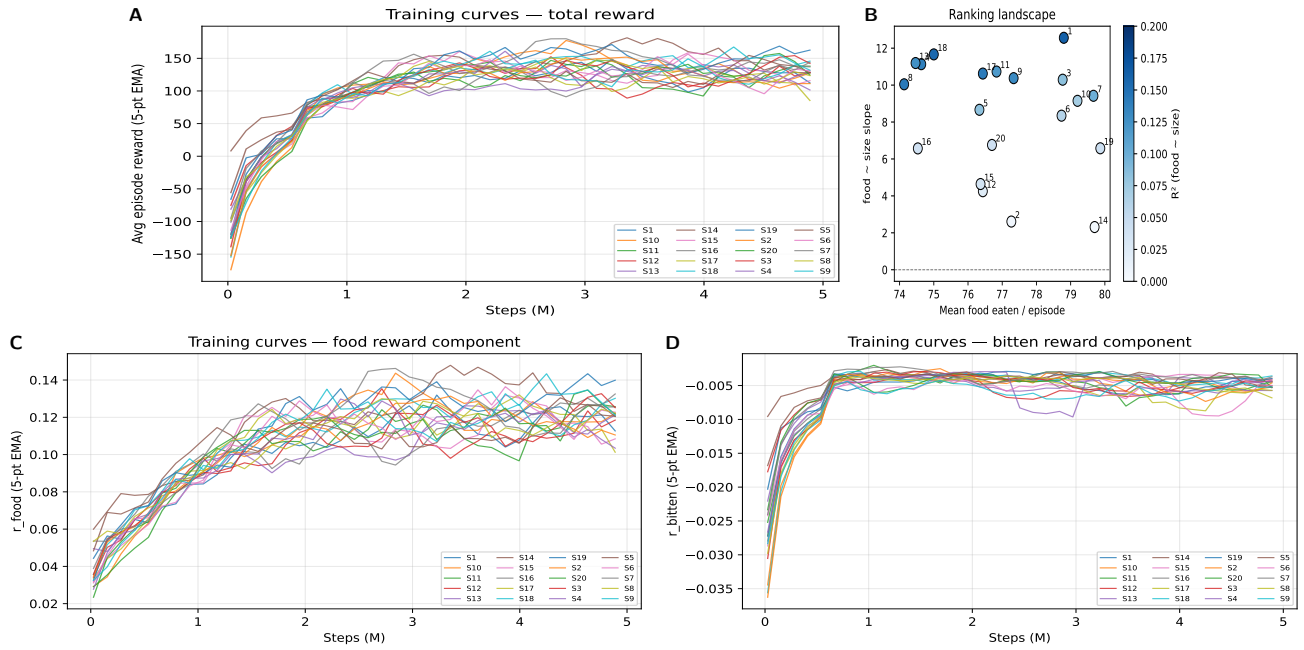


Fig. 11. Inter-seed reproducibility across 20 independent training runs (S1-S20, 5M steps each). (A) Total episode reward training curves (5-pt EMA) for all 20 seeds. All seeds converge reliably to a similar performance band (~ 100 – 160) by 5M steps with no catastrophic failures. (B) Seed ranking landscape: each point is one seed, positioned by mean food eaten per episode (x-axis, foraging efficiency) and the slope of a food~size linear regression (y-axis, strength of size-based dominance). Color encodes R^2 of food~size. Seeds in the upper-right quadrant (high efficiency, strong dominance structure) were preferred for downstream analysis; S1 was selected as the canonical seed. (C) Food reward component (r_{food} , 5-pt EMA) across all seeds, showing convergence to a fairly narrow band of foraging performance. (D) Bitten reward component (r_{bitten} , 5-pt EMA) across all seeds. The penalty is rapidly minimized from strongly negative values at initialization to near zero, indicating agents learn to avoid aggression costs over the course of training.

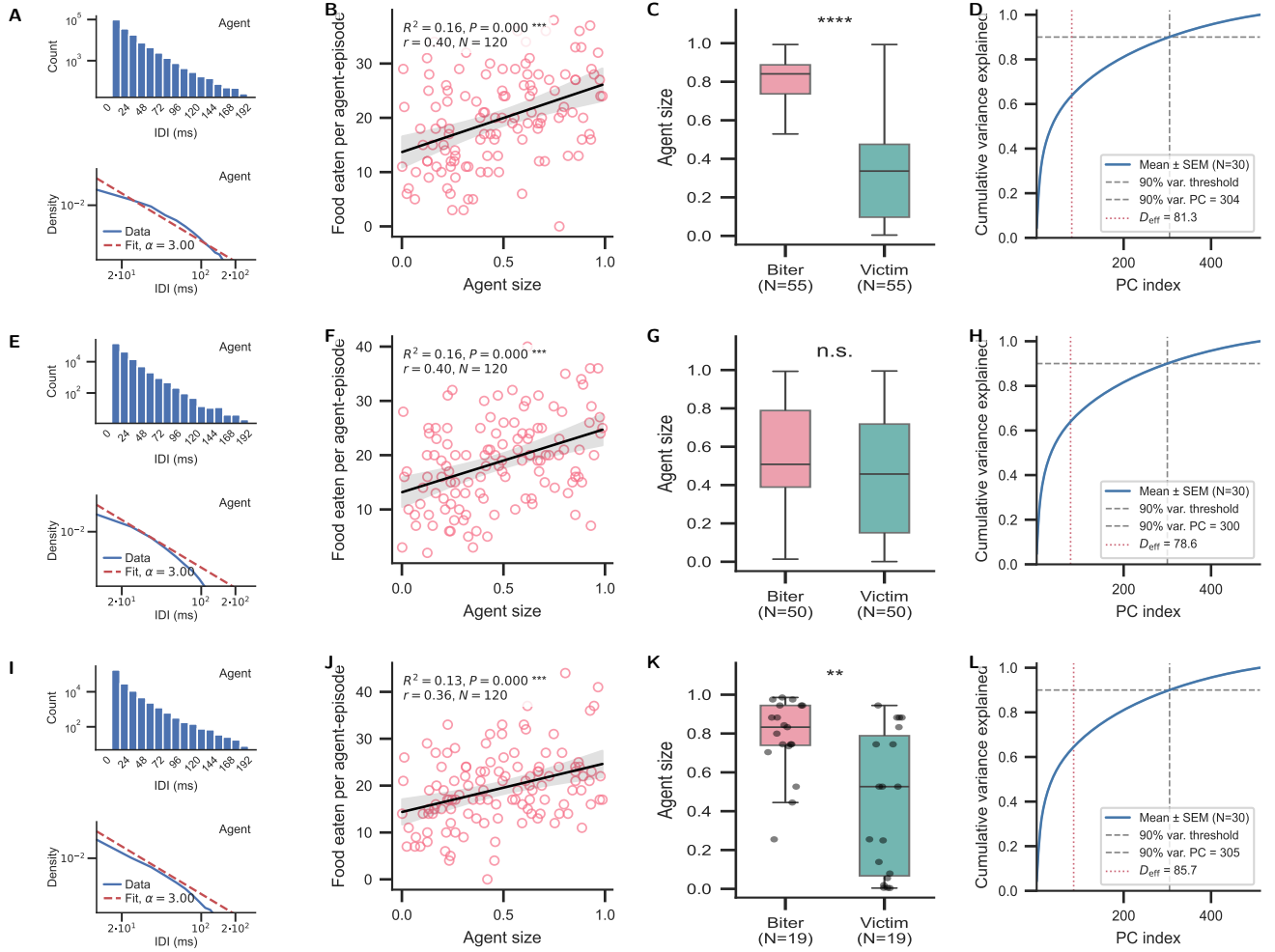


Fig. 12. Key result panels across the top-3 seeds (Rank 1: S1, Rank 2: S18, Rank 3: S9; rows A–D, E–H, I–L respectively). (A,E,I) Inter-discharge interval histogram and power-law fit (2-fish uniform-wide eval). (B,F,J) Food eaten vs. agent size (4-fish patchy eval; Fig. 3M). (C,G,K) Biter vs. bitten size by role boxplot (2-fish uniform-wide eval; Fig. 5M). (D,H,L) RNN cumulative variance explained per episode, showing effective dimensionality (4-fish patchy eval; Fig. 6A). Note that Rank 1 was used in the main manuscript.

.7 Simulation and Calibration Parameters

Parameter	Symbol	Unit	Value
Simulation Configuration			
Prey radius	R_{prey}	m	0.0025
Agent radius	R_{agent}	m	0.01
Conductor contrast factor	χ	–	-0.5
Default training arena range	–	m	0.40–3.00
One-patch arena size	–	m	1.50×1.50
Homing training arena range	–	m	0.75–1.50
Simulation rate	–	Hz	83
Electric wall-image scale	–	–	0.95
Sensor Thresholds			
Ampullary min. threshold	$E_{\text{A,min}}$	V/m	2.0×10^{-10}
Ampullary max. threshold	$E_{\text{A,max}}$	V/m	2.0×10^{-8}
Mormyromast min. threshold	$E_{\text{M,min}}$	V/m	5.0×10^{-8}
Mormyromast max. threshold	$E_{\text{M,max}}$	V/m	5.0×10^{-2}
Knollenorgan min. threshold	$E_{\text{K,min}}$	V/m	2.0×10^{-7}
Knollenorgan log-saturation scale	$E_{\text{K,log max}}$	V/m	2.0×10^{-4}
Active (Self-Image) Detection Ranges			
Mormyromast prey detection range	$x_{\text{M,prey}}$	m	0.05
Mormyromast agent detection range	$x_{\text{M,agent}}$	m	0.10
Passive Intrinsic Detection Ranges			
Ampullary prey detection range	$x_{\text{A,prey}}$	m	0.04
Ampullary agent detection range	$x_{\text{A,agent}}$	m	0.08
Passive Cons-EOD Detection Ranges			
Knollenorgan agent detection range	$x_{\text{K,agent,consEOD}}$	m	1.00
Calculated/Estimated Quantities			
Agent EOD monopole source	q_{EOD}	C	$\pm 1.11 \times 10^{-15}$
Agent EOD source separation	–	cm	1.0
Prey intrinsic dipole moment	$\mathbf{p}_{\text{prey}}$	C·m	$(0, 1.11 \times 10^{-24})$
Agent intrinsic dipole moment	$\mathbf{p}_{\text{agent}}$	C·m	$(1.11 \times 10^{-23}, 0)$
Cons-image scaling factor	k_{cons}	-	100

Table 1. Physics simulation and calibrated parameters used for the dynamic-sensing model. Distances are reported in meters unless otherwise specified. Electric-field thresholds are reported in V/m after unit conversion.

.8 Summary of evaluation assays

Figure panels	Assay	Notes
Fig. 2A–D	Homing to a fixed emitting target in a 70 cm × 70 cm arena	100 precomputed trials; compared with a non-electric ray baseline
Fig. 3A,B,K–P	Four-agent patchy foraging in a 70 cm × 70 cm arena	30 episodes; heterogeneous agent sizes
Fig. 3C–J	Two-agent wide foraging in a 160 cm × 40 cm arena	100 episodes; IDI and distance-dependent EOD analyses
Fig. 4A–F	Four-agent sensor-ablation foraging in patchy arenas	Single-sensor removals and reduced-sensor controls
Fig. 4G,H	Collective-sensing manipulations in a 40 cm × 40 cm arena	Self-EOD and conspecific-EOD components separated in the Mormyromast channel
Fig. 4I,J	EOD muting manipulations	EOD emission silenced in selected agents; food consumed and nearest-neighbor distance compared against baseline
Fig. 4K,L	Patchy versus uniform food distributions	Same trained policy evaluated under altered food layout
Fig. 4M,N	Iso-food patch-number sweep	Patch count 1–5; iso series holds total food constant; free series holds food-per-patch constant
Fig. 5A–H	Two-agent one-patch competition assay	100 trials per trained-agent size condition
Fig. 5I,J	Patch-confined bot control	300-condition grid crossing size, EOD rate, movement state, and initial position
Fig. 5K–P	Two-agent square interaction assay	70 cm × 70 cm arena; pairwise approach, chasing, and biting analyses
Fig. 6A	Four-agent recurrent-state PCA	Hidden states from patchy foraging episodes
Fig. 6B,C	Group-size recurrent dimensionality assay	Same policy evaluated with different numbers of agents
Fig. 6D	Homing recurrent-state decoding	Knollenorgan-defined and straight-line target angles decoded from hidden states
Fig. 6E,F	Two-agent recurrent-state decoding	Split by sensory range and pooled across timesteps
Fig. 6G,H	Two-agent shared-state PLSC	In-range versus out-of-range pairwise comparisons
Fig. 6I,J	Two-agent recurrent-state power analysis	Episode-level wideband power associated with social proximity and relative size

Table 2. Condensed summary of the evaluation assays used for each main figure.

.9 List of recurrent-state decoding targets

Target	Symbol	Type	Description
X Position	x	Scalar	Allocentric x position
Y Position	y	Scalar	Allocentric y position
Distance to Nearest Agent	d_{agent}	Scalar	Distance to closest conspecific
Distance to Closest Food	d_{food}	Scalar	Distance to nearest food item
Angle to Closest Agent	α_{agent}	Circular	Bearing to closest conspecific
Angle to Closest Food	α_{food}	Circular	Bearing to nearest food item
Actual Turn	ω	Scalar	Realized angular change per timestep
Agent Size	s_{agent}	Scalar	Scalar size context of the focal agent
Energy	e	Scalar	Internal energy state, when present in the rollout table
Agents in Knollen Range	n_{agent}^K	Count	Number of conspecifics within Knollenorgan range
Size of Nearest Agent	$s_{\text{agent},1}$	Scalar	Size metadata for the nearest conspecific
Distance to Second Nearest Agent	$d_{\text{agent},2}$	Scalar	Distance to the second-nearest conspecific
Size of 2nd Nearest Agent	$s_{\text{agent},2}$	Scalar	Size metadata for the second-nearest conspecific
Food Count Centered (5cm)	n_{food}^M	Count	Number of food items within the centered 5 cm local window
Food Count Front (5cm)	$n_{\text{food},\text{front}}$	Count	Number of food items in the 5 cm front sector
Food Count Behind (5cm)	$n_{\text{food},\text{back}}$	Count	Number of food items in the 5 cm rear sector
Food Count Left (5cm)	$n_{\text{food},\text{left}}$	Count	Number of food items in the 5 cm left sector
Food Count Right (5cm)	$n_{\text{food},\text{right}}$	Count	Number of food items in the 5 cm right sector
Agent Nearby	I_{agent}^M	Boolean	Whether a conspecific is within the nearby-agent threshold
Was Bitten	I_{bitten}	Boolean	Whether the agent was bitten on that timestep
Agent in Knollen Range	I_{agent}^K	Boolean	Whether any conspecific is within Knollenorgan range
Mormyromast Field Magnitude	$\log \vec{E}_M $	Scalar	Log-transformed Mormyromast field magnitude, when center-field columns are available
Ampullary Field Magnitude	$\log \vec{E}_A $	Scalar	Log-transformed Ampullary field magnitude, when center-field columns are available
Knollen Error Angle	α_{agent}^K	Circular	Orientation error relative to the Knollenorgan field direction
Distance to Wall	d_{wall}	Scalar	Distance from the focal agent to the nearest arena wall
Mormyromast Error Angle	α_{agent}^M	Circular	Homing-task error angle for the Mormyromast field
Ampullary Error Angle	α_{agent}^A	Circular	Homing-task error angle for the Ampullary field
Knollen Error Angle to Nearest Agent	θ_{error}^K	Circular	Homing-task error angle for the nearest-agent Knollen field

Table 3. Recurrent-state decoding targets used in the analyses. Targets absent from a given assay were omitted from that assay-specific decoding fit.

.10 Sensor Calibration Calculations

.10.1 Dipole Moment for a Given Maximum Field at Distance x

The electric field ($\mathbf{E}$) of a dipole at a point $\mathbf{r}$ (where $x = |\mathbf{r}|$) is given by the general vector formula:

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left[\frac{3(\mathbf{p} \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{p}}{x^3} \right]$$

where $\mathbf{p}$ is the dipole moment vector and $\hat{\mathbf{r}}$ is the unit vector in the direction of $\mathbf{r}$. Alternatively, in spherical coordinates the magnitude of the electric field at a point (x, θ) is:

$$E(x, \theta) = \frac{p}{4\pi\epsilon_0 x^3} \sqrt{3\cos^2 \theta + 1}$$

where θ is the angle between the dipole moment vector $\mathbf{p}$ and the position vector $\mathbf{r}$.

The magnitude of the electric field (E) is maximized when $\cos \theta = \pm 1$, which corresponds to an angle of $\theta = 0^\circ$ or $\theta = 180^\circ$, i.e. directly along the axis of the dipole. The field magnitude along the axis is then:

$$E_{\max} = \frac{p}{2\pi\epsilon_0 x^3}$$

Therefore, the dipole moment p required to generate a maximum electric field $E_{\max}$ at a distance x is:

$$p = 2\pi\epsilon_0 x^3 E_{\max}$$

where p is the dipole moment (C·m), x is the distance from the dipole where the field is maximized (m), $E_{\max}$ is the required electric field magnitude (V/m), and ϵ_0 is the permittivity of free space (8.854×10^{-12} F/m).

.10.2 Dipole Moment Required to Induce a Specific Field at Its Own Location

Active sensing involves three steps: a source dipole ($\mathbf{p}$) creates a field, this field induces a new dipole ($\mathbf{p}'$) on a conductor (other agent or prey), and this induced dipole in turn creates a field that the source's sensors detect.

To calibrate the strength of the EOD source, we find the source dipole moment (p) such that the maximum induced field at the source's location, E_{induced} , is equal to a desired value, E_{max} .

The electric field at the conductor due to the source dipole is given by the on-axis formula:

$$E_{\text{source}} = \frac{p}{2\pi\epsilon_0 x^3}$$

This source field induces a dipole moment ($\mathbf{p}'$) on the conductor, which is assumed to be a sphere of radius R :

$$p' = 3\epsilon_0 V \cdot \chi \cdot E_{\text{source}}$$

where $V = \frac{4}{3}\pi R^3$ is the volume of the conductor.

Substituting the expression for E_{source} and simplifying yields:

$$p' = 3\epsilon_0 \left(\frac{4}{3}\pi R^3 \right) \cdot \chi \cdot \frac{p}{2\pi\epsilon_0 x^3} = \frac{2R^3 \cdot \chi \cdot p}{x^3}$$

The induced dipole ($\mathbf{p}'$) then creates its own on-axis electric field at the location of the original source dipole. The magnitude of this induced field is:

$$E_{\text{induced}} = \frac{p'}{2\pi\epsilon_0 x^3}$$

Substituting the simplified expression for p' into this equation, we get:

$$E_{\text{induced}} = \frac{1}{2\pi\epsilon_0 x^3} \left(\frac{2R^3 \cdot \chi \cdot p}{x^3} \right) = \frac{R^3 \cdot \chi \cdot p}{\pi\epsilon_0 x^6}$$

To calibrate the source EOD dipole strength, we set this induced field equal to the sensor's minimum detection threshold, E_{max} and solve for the required source dipole moment, p :

$$p = \frac{E_{\text{max}} \pi \epsilon_0 x^6}{R^3 \cdot \chi}$$

where, p is the required dipole moment of the source (C·m), E_{max} is the target electric field magnitude at the source dipole (V/m), x is the separation between the source dipole and the conductor (m), R is the conductor's radius (m), χ is the conductor contrast factor, ϵ_0 is the permittivity of free space (8.854×10^{-12} F/m).

.10.3 Self-EOD vs. Induced-dipole field strength at a Mormyromast sensor (on-axis, collinear)

Consider a source dipole at $(0,0)$ with moment $\mathbf{p} = p\hat{\mathbf{y}}$, a spherical object of radius R centered at $(0, x_{\text{obj}})$, and a Mormyromast-type sensor on the source body at $(0, r)$ with outward normal $\hat{\mathbf{y}}$. All positions are collinear along $\hat{\mathbf{y}}$ and fields are evaluated on-axis.

Self field at the sensor: On-axis self-EOD field magnitude at $(0, r)$:

$$E_{\text{self}} = \left| \frac{1}{2\pi\epsilon_0} \frac{p}{r^3} \right|.$$

Induced field at the sensor: Field at the object's center:

$$E_{\text{source} \rightarrow \text{obj}} = \left| \frac{1}{2\pi\epsilon_0} \frac{p}{x_{\text{obj}}^3} \right|.$$

Assuming object is a sphere with contrast factor χ the induced dipole moment is

$$p' = 4\pi\epsilon_0 R^3 \chi E_{\text{source} \rightarrow \text{obj}} = \frac{2R^3 \chi}{x_{\text{obj}}^3} p.$$

The field magnitude at $(0, r)$ due to the induced dipole at $(0, x_{\text{obj}})$ is

$$E_{\text{ind}} = \left| \frac{1}{2\pi\epsilon_0} \frac{p'}{|x_{\text{obj}} - r|^3} \right| = \frac{1}{2\pi\epsilon_0} \frac{1}{|x_{\text{obj}} - r|^3} \left(\frac{2R^3 \chi}{x_{\text{obj}}^3} p \right).$$

Ratio (self vs. induced). Taking the ratio of magnitudes,

$$\frac{E_{\text{self}}}{E_{\text{ind}}} = \frac{\frac{p}{2\pi\epsilon_0 r^3}}{\frac{1}{2\pi\epsilon_0} \frac{2R^3 \chi}{x_{\text{obj}}^3} \frac{p}{|x_{\text{obj}} - r|^3}} = \boxed{\frac{x_{\text{obj}}^3 |x_{\text{obj}} - r|^3}{2\chi R^3 r^3}}.$$

In the common regime $r \ll x_{\text{obj}}$, $|x_{\text{obj}} - r| \approx x_{\text{obj}}$ and

$$\frac{E_{\text{self}}}{E_{\text{ind}}} \approx \frac{x_{\text{obj}}^6}{2\chi R^3 r^3}.$$

For a prey object that is $x_{\text{obj}} = 5$ cm away, with $R = 0.25$ cm radius,

$|\chi| = 0.5$, and agent radius $r = 1$ cm, $\frac{E_{\text{self}}}{E_{\text{ind}}} \approx 10^6$

For a post-baseline-subtraction sensing dynamic range of 10 (“one order of magnitude”) on E_{ind} , i.e.: $E_{\text{ind}}^B / E_{\text{ind}}^A = 10$, implies a distance range of $x_{\text{obj}}^A / x_{\text{obj}}^B = 10^{1/6} = 1.47 = 1/(0.68)$

.10.4 Self-EOD vs. Cons-EOD field strength at a Mormyromast sensor (on-axis, collinear)

Given two agents, one at $(0,0)$ and the conspecific agent at $(0, x_{\text{cons}})$, we will calculate the ratio between the field due to self-EOD $E_{\text{self-EOD}}$ and due to cons-EOD $E_{\text{cons-EOD}}$ at a Mormyromast-type sensor located at $(0, r)$ with a radially-outward pointing normal.

For simplicity, we assume that a point dipole $\mathbf{p} = p\hat{\mathbf{y}}$ at the first agent's center generates the *self-EOD*. When a conspecific emits (receiver is not emitting), the conspecific dipole (assumed collinear and aligned) at $y = x_{\text{cons}} > r$ on the $+\hat{\mathbf{y}}$ axis.

The on-axis dipole field is given by

$$E_y^{(\text{dip})}(\Delta y) = \frac{1}{2\pi\epsilon_0} p \frac{\text{sgn}(\Delta y)}{|\Delta y|^3},$$

$E_{\text{self-EOD}}$, the magnitude of the self-EOD field at the sensor is given by:

$$E_{\text{self-EOD}} = \left| \frac{1}{2\pi\epsilon_0} \frac{p_{\text{self}}}{r^3} \right|.$$

$E_{\text{cons-EOD}}$, the magnitude of the cons-EOD field, due to a dipole at $(0, x_{\text{cons}})$, at the same sensor is given by:

$$E_{\text{cons-EOD}} = \left| \frac{1}{2\pi\epsilon_0} \frac{p_{\text{cons}}}{|x_{\text{cons}} - r|^3} \right|.$$

Assuming same dipole strength, their ratio is given by

$$\frac{E_{\text{self-EOD}}}{E_{\text{cons-EOD}}} = \frac{|x_{\text{cons}} - r|^3}{r^3}$$

For agents separated by $x_{\text{cons}} \approx 10$ cm and a agent body radius $r \approx 1$ cm, $E_{\text{self-EOD}}/E_{\text{cons-EOD}} = 9^3 = 729$.