

Policy heterogeneity improves collective olfactory search in three-dimensional turbulence

Lorenzo Piro^{1,*}, Robin A. Heinonen^{1,2}, Maurizio Carbone^{3,4}, Luca Biferale¹, and Massimo Cencini^{3,4}

¹*Department of Physics & INFN, Tor Vergata University of Rome, Via della Ricerca Scientifica 1, 00133 Rome, Italy*

²*Machine Learning Genoa Center (MaLGA) & Department of Civil, Chemical and Environmental Engineering, University of Genoa, Genoa, Italy*

³*Istituto dei Sistemi Complessi, CNR, Via dei Taurini 19, 00185 Rome, Italy*

⁴*INFN “Tor Vergata”, Via della Ricerca Scientifica 1, 00133 Rome, Italy*



(Received 14 April 2025; accepted 5 March 2026; published 2 April 2026)

We examine how heterogeneous swarms, mixing exploratory and exploitative agents with distinct decision rules, consistently outperform homogeneous ones where each agent balances exploration and exploitation individually, performing experiments in two contrasting deployment scenarios. Using odor fields from state-of-the-art direct numerical simulations of the 3D Navier-Stokes equations, we find that policy diversity typically allows the group to reach the source of the odor more efficiently by mitigating the detrimental effects of spatiotemporal turbulent correlations. These findings provide insights into collective search behavior and offer promising strategies for the design of robust, bioinspired search algorithms in engineered systems.

DOI: [10.1103/6zls-m67c](https://doi.org/10.1103/6zls-m67c)

I. INTRODUCTION

Looking for a target is an ubiquitous task facing living things, from insects to humans. In the absence of any cue, it is best to devise an efficient random search [1–3]. In contrast, when cues are available, searchers need to optimally exploit the information and find the target quickly. In particular, many biological and robotic agents face the challenge of locating an odor source in turbulent environments [4–7], where intermittent and patchy signals make simple strategies, such as gradient-climbing chemotaxis [8], ineffective [9–11] (see also Fig. 1). Living organisms have evolved remarkable search strategies whose key to success is the balance between information gathering (exploration) and capitalizing on high-confidence cues (exploitation) [12,13]. This has inspired heuristic policies for individual agents, based on Bayesian inference [14], that embody the exploration-exploitation trade-off, combining infotaxis [15], which seeks to maximize information gathering thus shifting the balance toward exploration, with policies biasing the search towards the likely source location [16,17]. In particular, space-aware infotaxis (SAI) [17] refines this trade off by blending infotaxis with a greedy policy that aims to minimize the expected distance from the source [18]. The robust performances achieved by SAI across diverse environments have established it as a state-of-the-art baseline for single-agent olfactory search [17,19–21].

In nature, it has been widely observed that individuals significantly benefit from sharing information with their conspecifics [22,23], which inspired multiagent extensions of heuristic search strategies [24–27], so far mainly explored in simplified settings. Understanding the exploration-exploitation trade off for multiagent systems in real-world turbulent environments is still an open problem, crucial to swarm robotics [28–30] and applications ranging from environmental monitoring to rescue operations [31,32].

Studies on animal behavior, foraging, and navigation have shown that groups of agents, ranging from small insects to large mammals, generally benefit from a division of labor, with some agents focusing on exploration, while others exploit the information collected by the group [33–36]. In this article, we introduce policy heterogeneity as a way to balance exploration and exploitation in a group of cooperating agents searching for an odor source in a turbulent flow. Specifically, we show that heterogeneous swarms, combining infotactic and greedy agents, outperform homogeneous groups where all agents use SAI. Our findings reveal that, under realistic environmental and initial conditions, an optimal fraction of exploitative agents systematically speeds up source localization, also reducing the probability of being lost, as compared to homogeneous swarms of SAI agents.

II. COLLECTIVE OLFACTORY SEARCH

We consider a group of N agents moving in a grid world Ω of size $L_x \times L_y \times L_z$ and lattice spacing Δ . Their goal is to find a source, located in $\mathbf{r}_s$, that sparsely emits odor particles whose trajectories are evolved with state-of-the-art DNS (see [37,38] for details) of the incompressible Navier-Stokes equations under turbulent conditions ($\text{Re}_\lambda \simeq 150$) with a uniform mean wind $\mathbf{U}$ along the $-\hat{x}$ direction (Fig. 1). The particle trajectories are coarse-grained within the 3D arena, thus resulting in the concentration field $c(\mathbf{r}, t)$ displayed in Fig. 1.

*Contact author: lorenzo.piro@roma2.infn.it

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

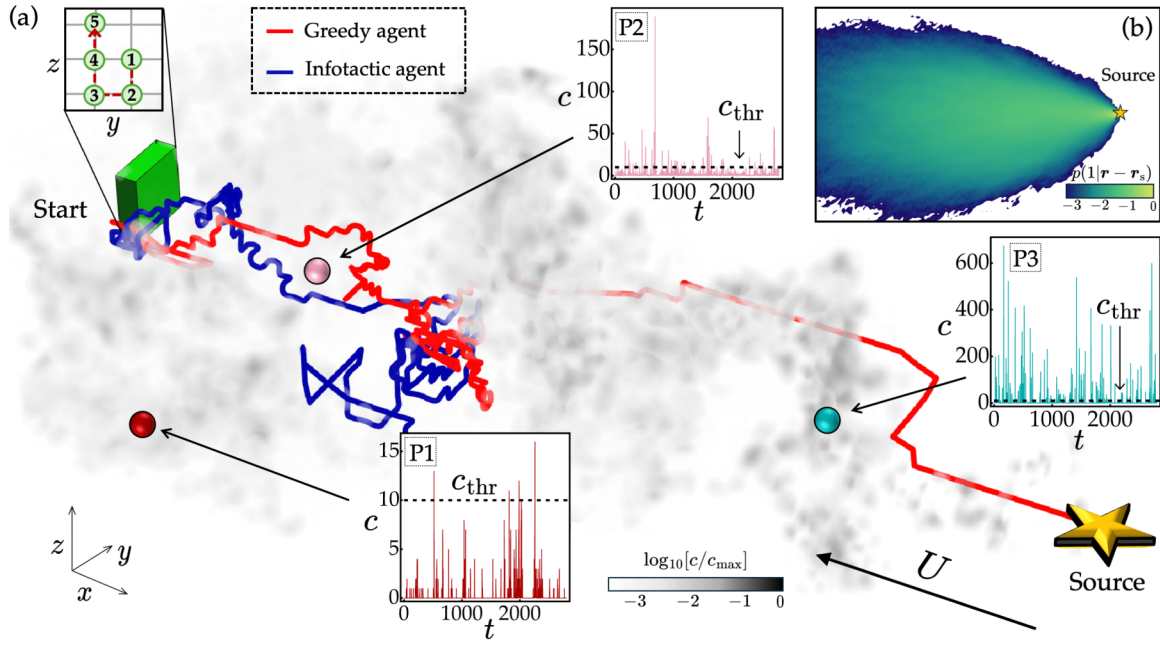


FIG. 1. (a) Snapshot of the concentration of odor particles (logarithmic grayscale) emitted by a localized source (yellow star) in a turbulent flow. The blue/red curves correspond to the typical trajectories of an infotactic/greedy agent starting from a localized region (green box) and cooperating to find the source. Agents are initially placed close to each other, as depicted in the top left inset. Plots (P1–P3) display the intermittent time evolution of the odor signal at three different locations, and the dashed black line indicates the concentration threshold $c_{thr} = 10 \approx 2.4\bar{c}$ above which the agents can detect odor particles. Unless otherwise specified, hereafter, the mean wind strength U with respect to the intensity of the turbulent fluctuations u_{rms} is equal to $U/u_{rms} = 2.05$. (b) Two-dimensional cut of the empirical likelihood of making a detection, $p(1|\mathbf{r} - \mathbf{r}_s)$, in the plane parallel to the mean wind U and containing the source.

An underlying assumption of our study is that the agents are small enough to leave the flow unperturbed but still capable of not being carried away by it. This is reasonable for flying insects [4–6] and for insectlike robots [39]. Agents have a prescribed sensitivity to the odor and make binary observations, i.e., a detection occurs only when the highly fluctuating and intermittent local concentration (see panels P1–P3 in Fig. 1) exceeds a given threshold c_{thr} (see the SM for details on the numerics and Fig. 1 for a graphical summary of the search setup when agents are initially spatially clustered).

All agents share a probability spatial map (a “belief”) about the source position defined over Ω , $b(\mathbf{r}) \equiv \text{Prob}(\mathbf{r}_s = \mathbf{r})$. Assuming no prior knowledge, the belief is initialized to a uniform distribution and set to zero at the agents’ positions. Within the Bayesian framework, agents must rely on a model of the environment to interpret their measurements and localize the source, reflecting the idea that agents possess prior knowledge of the odor statistics. We therefore assume that each agent i knows its position, $\mathbf{r}_i$, and the likelihood $p(h_i^{(t)}|\mathbf{r}_i - \mathbf{r})$ of making an observation, $h_i^{(t)} \in \{0, 1\}$, at time t obtained from the time average of the coarse-grained concentration. This model of the environment [see Fig. 1(b) for a 2D projection] only captures the single-time odor statistics, as it neglects the spatiotemporal correlations exhibited by the odor plume, and is therefore not exact. However, it represents the best tractable model of the turbulent odor signal generated from the DNS and will hereafter be referred to as the empirical likelihood.

At each time step, agents measure the odor concentration in synchrony at their positions and update the shared belief by the Bayes rule [14]

$$b^{(t)}(\mathbf{r}) = \frac{1}{\mathcal{N}} \prod_{i=1}^N p(h_i^{(t)}|\mathbf{r}_i - \mathbf{r}) b^{(t-1)}(\mathbf{r}), \quad (1)$$

where $\mathcal{N}$ is a normalization factor. Note that a shared belief can be maintained by agents exchanging only their positions and binary detections—data easily transmitted via wireless links in swarm robotics [40], which is far more efficient than sharing full belief maps. The cost of communication is therefore greatly reduced with respect to communicating the entire belief. Upon updating it via (1), at each time step, each agent chooses an action to move in one of the allowed directions $\{\pm\hat{x}, \pm\hat{y}, \pm\hat{z}\}$ by a lattice spacing Δ .

Given the high dimensionality of multi-agent olfactory search, identifying an optimal collective protocol by solving the Bellman equation for the corresponding partially observable Markov decision process is infeasible [19,20]. Thus, in the following, we focus on comparing heuristics that embody the exploration-exploitation trade off of the group of agents in two different ways. In particular, we consider a heterogeneous swarm divided into two groups; N_I agents following a risk-averse infotactic strategy [15], which prioritizes exploration by selecting actions that maximize the expected information gain about the source location; and N_G agents following a strongly exploitative greedy policy [18], meant

to reduce the expected Manhattan distance to the source. We shall compare the performance of our heterogeneous (HET) group of agents $N = N_I + N_G$ made up of different fractions of greedy/infotactic players against a baseline where all agents manage the exploration-exploitation trade off at the individual level, considering a homogeneous swarm where all agents adopt the same SAI policy [17,27], which combines infotaxis and greedy policy.

In detail, depending on its state $s_i \equiv (\mathbf{r}_i, b(\mathbf{r}))$, defined by its position, the shared belief and its policy, π_i , the i th agent selects the action that minimizes a policy-dependent cost function $\mathcal{C}_{\pi_i}$:

$$a_i^* = \operatorname{argmin}_a \mathcal{C}_{\pi_i}(s_i, a), \quad \text{with} \quad (2)$$

$$\mathcal{C}_{\pi_i}(s_i, a) = \begin{cases} \frac{1}{2}(2^{H(s_i|a)} - 1), & \text{if } \pi_i = \text{infotaxis} \\ D(s_i|a), & \text{if } \pi_i = \text{greedy} \\ \mathcal{C}_{\text{Infotaxis}} + \mathcal{C}_{\text{Greedy}}, & \text{if } \pi_i = \text{SAI}, \end{cases} \quad (3)$$

where $H(s_i) \equiv -\sum_{\mathbf{r}} b(\mathbf{r}) \log_2 b(\mathbf{r})$ is the Shannon entropy of the shared belief, and $D(s_i) \equiv \sum_{\mathbf{r}} b(\mathbf{r}) \|\mathbf{r}_i - \mathbf{r}\|_1$ the expected Manhattan distance from the source (see also [15,17,20,21,27]). All agents move synchronously, measure the odor concentration, and update the shared belief again, repeating until one of them satisfies $\mathbf{r}_i = \mathbf{r}_s$, indicating that the source has been found. If the search exceeds a predefined time limit ($T_{\max}$), the episode ends with the agents considered lost.

Although agents do not coordinate their actions—to avoid an exponential increase in computational complexity with the number of agents [24,41]—cooperation is achieved via the shared belief [27,41]. Furthermore, an action is forbidden if it would move the agent outside the domain or result in an overlap with another agent. If no permissible actions are available, the agent remains in place. See the SM for more details on the numerical implementation of the collective search algorithm and on the heuristic policies introduced above.

III. THE BENEFIT OF POLICY HETEROGENEITY

We first focus on the case where the agents are initially clustered in a position extracted from the empirical likelihood (Fig. 1) and start their search when at least one agent detects the odor signal. Figure 2(a) shows the cumulative distribution functions (CDF) of the first agent arrival time conditioned on successful episodes, i.e., those in which agents do not get lost ($T < T_{\max} \simeq 15T_\Omega$, with T_Ω the time to cross the arena Ω), for HET and SAI swarms made up of $N=2$ and $N=10$ agents, respectively. For each N , we find that an optimal fraction of greedy agents exists that maximizes the overall search efficiency (see the inset for $N=10$), making them achieve shorter first arrival times to the source compared to swarms comprising only SAI agents. Indeed, as shown in Fig. 2(b), HET swarms consistently show $\gtrsim 25\%$ improvement in terms of mean first arrival time with respect to SAI agents and, remarkably, the performance achieved with ten SAI agents can be achieved with a HET group of just five agents ($N_I=3$ and $N_G=2$). Moreover, the latter always manages to reach the source in a finite time, whereas SAI swarms have a finite probability of getting lost during the search [inset of Fig. 2(b)].

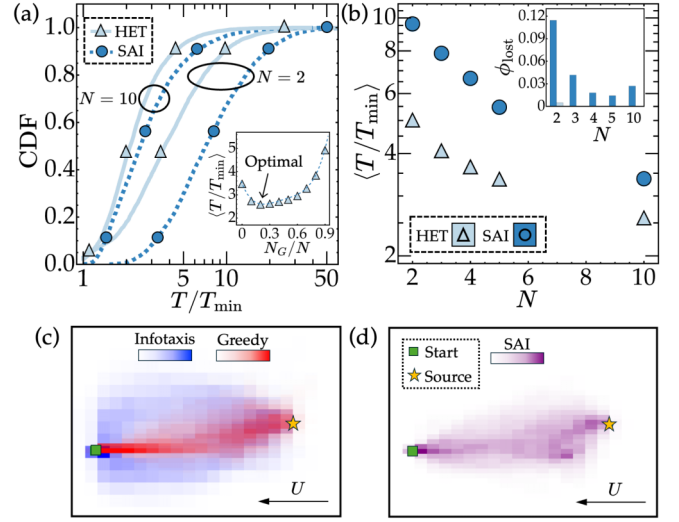


FIG. 2. (a) Cumulative distribution function (CDF) of first agent arrival times T normalized with the minimum search time $T_{\min}$ (Manhattan distance from the source) when using $N = 2$ or 10 agents. The results of the HET swarms correspond to those obtained with the optimal fraction of greedy agents as shown in the inset for the case $N = 10$. (b) Mean first arrival times as a function of N for SAI swarm or the optimal HET configuration. The statistics are conditioned on successful episodes, while the inset shows the fraction of episodes ϕ_{lost} in which the search fails ($T > T_{\max}$). Note that for HET swarms $\phi_{\text{lost}} < 0.01$. Error bars corresponding to the statistical error on the mean are within the symbols. (c), (d) Heatmaps of the spatial trajectories for ten HET agents starting from the green square and finding the source (yellow star) when $N_I = N_G = 5$ (c), and using SAI (d). The arrow indicates the mean wind direction. For the sake of clarity in the presentation, such maps have been obtained in the 2D setup (see the SM [43] for details).

Consistently, HET agents have a much lower probability to be trapped in long search episodes as shown by the reduced tails in the PDF of arrival times (see Fig. S1 in the SM [43]).

IV. ROLE OF TURBULENT CORRELATIONS

A key factor behind the improvement of HET over SAI is how the two groups manage spatiotemporal correlations of olfactory signals. In realistic turbulent flows, odor plumes tend to form filaments and puffs, Fig. 1, potentially leading to highly correlated detections between nearby agents. These events are not included in the model likelihood, Fig. 1(b), as it only accounts for single-point statistics. By measuring the probability of making simultaneous detections, we found that SAI agents systematically experience a higher frequency of concurrent detections compared to HET, with such events mostly happening when the agents are just a few lattice points away from each other. The 2D heat maps shown in Figs. 2(c) and 2(d) qualitatively demonstrate how HET swarms tend to spread out more, better covering the space than SAI agents (compare panels (c) and (d) and see 3D movies in the SM [43]). SAI groups are more prone to clustering, resulting in more redundant detections due to correlations and consequently inefficient exploration. Clustering likely arises because, as their common belief collapses,

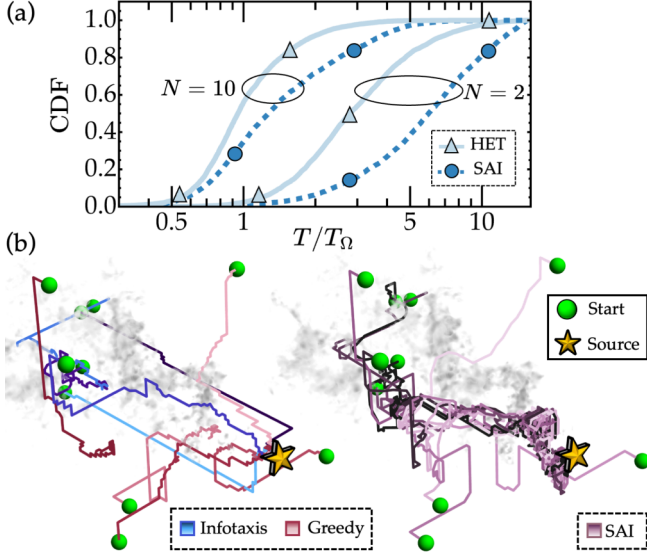


FIG. 3. Results with initially sparse swarms. (a) CDF of first arrival times T (in units of the time T_Ω to cross the arena) conditioned on success when using $N = 2$ or 10 agents with either SAI or the optimal fraction $N_G/N = 0.5$ of greedy agents for the HET group under the same initial conditions. (b) Typical episode in which $N = 10$ HET agents (left) beat their SAI counterparts (right), showing agent paths from random start positions (green spheres) until the source (yellow star) is found.

SAI agents' greedy bias naturally draws them together and then, due to the policy homogeneity and their proximity, they will choose similar moves, thus reinforcing group formation. In contrast, a mix of greedy and infotactic agents promotes diversity, prevents unnecessary clustering, and improves the overall efficiency of the search. In Fig. S2 of the SM [43], we further support these findings through the statistical analysis of blanks, i.e., the time intervals between two consecutive odor detections, emphasizing the impact of odor intermittency. Indeed, when detections are drawn directly from the empirical likelihood, removing correlations in the signal, the performance gap between SAI and HET significantly narrows and tends to vanish as the group size increases. These findings demonstrate that policy heterogeneity serves as an effective mechanism to compensate for imperfections in environmental modeling, particularly when agents interact with realistic, turbulent odor signals. Moreover, they underscore the importance of odor intermittency in driving the emergent search dynamics.

V. SPARSE INITIAL CONDITIONS

A crucial ingredient that limits the performance of SAI agents is the clustering induced by the policy. To investigate whether this detrimental effect depends on the starting conditions, we tested when they are randomly uniformly positioned across the domain. In this case, reflecting practical deployment scenarios, we let the agents start searching immediately, without requiring an initial detection. Figure 3(a) displays the CDF of arrival times for the first agent finding the source. Remarkably, even under these dispersed starting conditions, SAI groups exhibit longer arrival times and higher lost fractions,

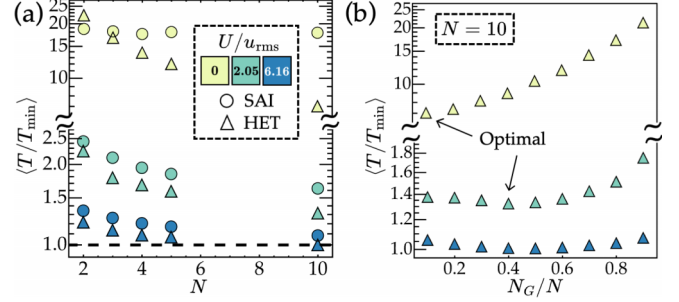


FIG. 4. (a) Normalized mean arrival times $\langle T/T_{\min} \rangle$ conditioned on successful episodes as a function of the number of agents N . Colors distinguish the strength of the mean wind U with respect to the u_{rms} of the turbulent flow as in the legend. The dashed black line denotes the theoretical optimum $\langle T/T_{\min} \rangle = 1$. (b) Mean normalized arrival time as a function of the greedy agents N_G/N for HET with $N = 10$. Error bars corresponding to the statistical error on the mean are within the symbols.

whereas HET swarms maintain a clear advantage in search speed and still feature negligible losses (see also Table I). The trajectories in Fig. 3(b) reveal that SAI agents (right), although initially dispersed, tend to converge into clusters, again making them more susceptible to the effects of turbulent correlations. In contrast, the HET group (left) maintains a wider spatial distribution and ensures a better spatial coverage.

VI. COMPARISON ACROSS DIFFERENT ENVIRONMENTS

To scrutinize the robustness of our findings against environmental conditions, we investigate how different wind conditions influence the effectiveness of HET and SAI swarms for the case of initially clustered agents. We do this by restricting the search to 2D, thus allowing the agents to move within a thin slab containing the source and aligned with the mean wind direction, when present. The results obtained under three different wind intensities are summarized in Fig. 4 (Fig. S3 in the SM [43] shows the empirical likelihood in the three scenarios).

Increasing the wind strength simplifies the search by providing agents with stronger directional cues about the source location and restricting the detection cone. Although this tends to diminish the advantage of policy heterogeneity, HET swarms still systematically outperform SAI ones, even managing (for $N = 10$) to find the source in the shortest possible time $T_{\min}$ [dashed black line in Fig. 4(a)]. Conversely, in the absence of a mean wind, the correlations are stronger and less information is provided by the empirical likelihood, leading to an increased performance gap between HET and SAI swarms. In fact, deploying more SAI agents reduces neither the mean arrival time [Fig. 4(a)] nor the probability of being lost (see Table I). In contrast, the mean arrival time of HET swarms improves threefold when increasing from two to ten agents, with a vanishing lost fraction, similar to the 3D case. These results depend only weakly on the sensitivity of the agents to the odor signal and remain qualitatively unchanged when varying the detection threshold c_{thr} (Fig. S4 in the SM [43]).

TABLE I. Fraction of episodes in which search fails for SAI and HET swarms under all tested configurations. Results are shown for clustered and sparse initial conditions in the 3D setup ($U/u_{\text{rms}} = 2.05$), and for clustered initial conditions in 2D varying the mean wind strength ($U/u_{\text{rms}} = \{0, 2.05, 6.16\}$) with $N = 2$ and 10 agents. The smaller value of ϕ_{lost} for each setup is highlighted in bold.

Initial conditions U/u_{rms} N	Clustered (3D)		Sparse (3D)		Clustered (2D)					
	2.05		2.05		0		2.05		6.16	
	2	10	2	10	2	10	2	10	2	10
$\phi_{\text{lost}}(\text{SAI})$	0.114	0.027	0.158	0.010	0.137	0.127	0.011	0.007	0.010	0.007
$\phi_{\text{lost}}(\text{HET})$	0.005	0	0.010	0	0.144	0.001	0.001	0	0.001	0

VII. EFFECTS OF ALEATORIC ERRORS

Motivated by the technical challenges faced in real-world scenarios, we also investigate the impact of introducing an error in either the odor measurement or proprioception (see Fig. S6 in the SM [43]). We find that moderate noise levels degrade absolute performance, increasing the percentage of search failures; however, the relative advantage of HET over SAI swarms remains substantial. We also note that, in real swarms, agents often communicate only with nearby peers [44–46], requiring individual belief updates and increasing computational complexity. At the same time, they may share less information than the whole belief map and rely on simpler computations to infer others' behavior [16,24]. These constraints diversify search strategies and can significantly influence performance, warranting dedicated future studies.

VIII. DISCUSSION

We showed that policy heterogeneity significantly improves the efficiency of collective olfactory search in realistic turbulent environments, both in terms of search time and robustness, by changing the (i) initial agent distribution, (ii) ratio between turbulent fluctuations and the mean wind, (iii) detection sensitivity, and (iv) measurement errors. As a key insight, we find that heterogeneity mitigates the challenges posed by turbulent intermittency and correlations, enabling swarms to balance exploration and exploitation more effectively than homogeneous groups that resolve the trade off at the individual level by weighting the infotactic and greedy cost in their action selection [42]. This leads HET agents to exhibit superior performance in particularly challenging search tasks, as evidenced by a lower probability of extremely large arrival times (see Fig. S1 in the SM [43]). Significantly, heterogeneity benefits both individual agents and the group as a whole by reducing the mean arrival times of both the first and last agent (see Fig. S5 in the SM [43]), thereby improving overall swarm efficiency and coordination and ensuring that all agents reach the source more rapidly.

Finally, we remark that, in the clustered setup, episodes start with a detection, so Figs. 2 and 4 already condition on identical first-detection times, whereas in the sparse-start case, HET swarms outperform SAI even without this conditioning (cf. Fig. 3), making a detailed study of time-to-first-detection statistics an interesting avenue for future work.

Our results align with biological observations suggesting that policy heterogeneity might be a natural optimization mechanism for animals [47–50]. Likewise, these insights could inform the design of artificial swarm systems for real-world applications such as environmental monitoring or disaster response [51,52].

Future research should explore adaptive learning mechanisms, allowing agents to develop mixed strategies [53] to dynamically switch between exploratory and exploitative policies in response to real-time environmental feedback. Preliminary tests (not shown) of dynamic policy switching in our scenarios do not show any significant advantage. Furthermore, while model-free methods have recently been applied to the single-agent problem [54,55], it remains an interesting open question to allow the swarm to dynamically adapt its behavior via multiagent reinforcement learning [56,57], for which our approach would serve as a baseline. Lastly, while in the present study agents neither perturbed the flow nor were advected by it, relaxing these simplifying assumptions (as in [58]) may have important effects; their investigation is deferred to future studies.

ACKNOWLEDGMENTS

We acknowledge the useful collaboration of Chiara Calascibetta. We also thank Antonio Celani, Mauro Sbragaglia, and Massimo Vergassola for helpful discussions and their insightful comments on the first draft of this article. We acknowledge financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on 2.2.2022 by the Italian Ministry of University and Research (MUR), funded by the European Union—NextGenerationEU—Project Title Equations informed and data-driven approaches for collective optimal search in complex flows (CO-SEARCH), Contract 202249Z89M—CUP B53-D23003920006 and E53-D23001610006. This work was also supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreements No. 882340 and No. 101002724) and by the National Institute of Health under Grant No. R01DC018789.

DATA AVAILABILITY

The data that support the findings of this article are openly available [59].

- [1] O. Bénichou, C. Loverdo, M. Moreau, and R. Voituriez, Intermittent search strategies, *Rev. Mod. Phys.* **83**, 81 (2011).
- [2] V. Tejedor, R. Voituriez, and O. Bénichou, Optimizing persistent random searches, *Phys. Rev. Lett.* **108**, 088103 (2012).
- [3] A. Chechkin and I. M. Sokolov, Random search with resetting: A unified renewal approach, *Phys. Rev. Lett.* **121**, 050601 (2018).
- [4] J. Murlis, J. S. Elkinton, and R. T. Cardé, Odor plumes and how insects use them, *Annu. Rev. Entomol.* **37**, 505 (1992).
- [5] T. D. Wyatt, *Pheromones and Animal Behaviour: Communication by Smell and Taste* (Cambridge University Press, Cambridge, 2003).
- [6] K. L. Baker, M. Dickinson, T. M. Findley, D. H. Gire, M. Louis, M. P. Suver, J. V. Verhagen, K. I. Nagel, and M. C. Smear, Algorithms for olfactory search across species, *J. Neurosci.* **38**, 9383 (2018).
- [7] A. Francis, S. Li, C. Griffiths, and J. Sienz, Gas source localization and mapping with mobile robots: A review, *J. Field Robot.* **39**, 1341 (2022).
- [8] H. C. Berg, *Random Walks in Biology* (Princeton University Press, Princeton, NJ, 1993).
- [9] J. P. Crimaldi and J. R. Koseff, High-resolution measurements of the spatial and temporal scalar structure of a turbulent plume, *Exp. Fluids* **31**, 90 (2001).
- [10] E. Balkovsky and B. I. Shraiman, Olfactory search at high Reynolds number, *Proc. Natl. Acad. Sci. USA* **99**, 12589 (2002).
- [11] A. Celani, E. Villermaux, and M. Vergassola, Odor landscapes in turbulent environments, *Phys. Rev. X* **4**, 041015 (2014).
- [12] C. A. Hernandez-Reyes, S. Fukushima, S. Shigaki, D. Kurabayashi, T. Sakurai, R. Kanzaki, and H. Sezutsu, Identification of exploration and exploitation balance in the Silkmoth olfactory search behavior by information-theoretic modeling, *Front. Comput. Neurosci.* **15**, 629380 (2021).
- [13] G. Reddy, V. N. Murthy, and M. Vergassola, Olfactory sensing and navigation in turbulent environments, *Annu. Rev. Condens. Matter Phys.* **13**, 191 (2022).
- [14] G. E. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (John Wiley & Sons, Hoboken, NJ, 2011).
- [15] M. Vergassola, E. Villermaux, and B. I. Shraiman, ‘Infotaxis’ as a strategy for searching without gradients, *Nature (London)* **445**, 406 (2007).
- [16] J.-B. Masson, Olfactory searches with limited space perception, *Proc. Natl. Acad. Sci. USA* **110**, 11261 (2013).
- [17] A. Loisy and C. Eloy, Searching for a source without gradients: how good is infotaxis and how to beat it, *Proc. R. Soc. London A* **478**, 20220118 (2022).
- [18] J. L. Fernández, R. Sanz, R. G. Simmons, and A. R. Diéguez, Heuristic anytime approaches to stochastic decision processes, *J. Heuristics* **12**, 181 (2006).
- [19] A. Loisy and R. A. Heinonen, Deep reinforcement learning for the olfactory search POMDP: A quantitative benchmark, *Eur. Phys. J. E* **46**, 17 (2023).
- [20] R. A. Heinonen, L. Biferale, A. Celani, and M. Vergassola, Optimal policies for Bayesian olfactory search in turbulent flows, *Phys. Rev. E* **107**, 055105 (2023).
- [21] R. A. Heinonen, L. Biferale, A. Celani, and M. Vergassola, Exploring Bayesian olfactory search in realistic turbulent flows, *Phys. Rev. Fluids* **10**, 064614 (2025).
- [22] A. M. Berdahl, A. B. Kao, A. Flack, P. A. H. Westley, E. A. Codling, I. D. Couzin, A. I. Dell, and D. Biro, Collective animal navigation and migratory culture: from theoretical models to empirical evidence, *Philos. Trans. R. Soc. B* **373**, 20170009 (2018).
- [23] M. Nagy, A. Horicsányi, E. Kubinyi, I. D. Couzin, G. Vásárhelyi, A. Flack, and T. Vicsek, Synergistic benefits of group search in rats, *Curr. Biol.* **30**, 4733 (2020).
- [24] E. D. Karpas, A. Shklarsh, and E. Schneidman, Information socialtaxis and efficient collective behavior emerging in groups of information-seeking agents, *Proc. Natl. Acad. Sci. USA* **114**, 5589 (2017).
- [25] C. Song, Y. He, B. Ristic, L. Li, and X. Lei, Multi-agent collaborative infotaxis search based on cognition difference, *J. Phys. A* **52**, 485202 (2019).
- [26] M. Durve, L. Piro, M. Cencini, L. Biferale, and A. Celani, Collective olfactory search in a turbulent environment, *Phys. Rev. E* **102**, 012402 (2020).
- [27] E. Panizon and A. Celani, Seeking and sharing information in collective olfactory search, *Phys. Biol.* **20**, 065001 (2023).
- [28] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo, Swarm robotics: A review from the swarm engineering perspective, *Swarm Intelligence* **7**, 1 (2013).
- [29] H. L. Kwa, J. Leong Kit, and R. Bouffanais, Balancing collective exploration and exploitation in multi-agent and multi-robot systems: A review, *Front. Robot. AI* **8**, 771520 (2022).
- [30] M. M. Shahzad, Z. Saeed, A. Akhtar, H. Munawar, M. H. Yousaf, N. K. Baloach, and F. Hussain, A review of swarm robotics in a NutShell, *Drones* **7**, 269 (2023).
- [31] U. K. Verfuss, A. S. Aniceto, D. V. Harris, D. Gillespie, S. Fielding, G. Jiménez, P. Johnston, R. R. Sinclair, A. Sivertsen, S. A. Solbø, R. Størvoold, M. Biuw, and R. Wyatt, A review of unmanned vehicles for the detection and monitoring of marine fauna, *Mar. Pollut. Bull.* **140**, 17 (2019).
- [32] T. Dang, F. Mascarich, S. Khattak, H. D. Nguyen, H. Nguyen, S. Hirsh, R. Reinhart, C. Papachristos, and K. Alexis, Autonomous search for underground mine rescue using aerial robots, in *Proceedings of the 2020 IEEE Aerospace Conference* (IEEE, Piscataway, NJ, 2020), pp. 1–8.
- [33] T. D. Seeley, Adaptive significance of the age polyethism schedule in honeybee colonies, *Behav. Ecol. Sociobiol.* **11**, 287 (1982).
- [34] B. Hölldobler and E. O. Wilson, *The Ants* (Harvard University Press, Cambridge, MA, 1990).
- [35] D. Kengyel, H. Hamann, P. Zahadat, G. Radspieler, F. Wotawa, and T. Schmickl, in *PRIMA 2015: Principles and Practice of Multi-Agent Systems*, edited by Q. Chen, P. Torroni, S. Villata, J. Hsu, and A. Omicini (Springer International Publishing, Cham, 2015), pp. 201–217.
- [36] E. Goodale and R. D. Magrath, Species diversity and interspecific information flow, *Biol. Rev.* **99**, 999 (2024).
- [37] L. Piro, R. A. Heinonen, M. Cencini, and L. Biferale, Many wrong models approach to localise an odour source in turbulence with static sensors, *J. Turbul.* **26**, 153 (2025).
- [38] L. Biferale, F. Bonaccorso, N. Cocciaglia, R. A. Heinonen, and L. Piro, TURB-Smoke. A database of Lagrangian pollutants emitted from point sources in turbulent flows with a mean wind, *Sci. Data* **13**, 428 (2026).

- [39] G. de Croon, J. Dupeyroux, S. Fuller, and J. Marshall, Insect-inspired AI for autonomous robots, *Sci. Robot.* **7**, eabl6334 (2022).
- [40] A. Girma, N. Bahadori, M. Sarkar, T. G. Tadewos, M. R. Behnia, M. N. Mahmoud, A. Karimodini, and A. Homaifar, IoT-enabled autonomous system collaboration for disaster-area management, *IEEE/CAA J. Autom. Sinica* **7**, 1249 (2020).
- [41] J.-B. Masson, M. B. Bechet, and M. Vergassola, Chasing information to search in random environments, *J. Phys. A* **42**, 434009 (2009).
- [42] Although we varied these internal weights, no homogeneous SAI variant could match the performance of the heterogeneous mixture.
- [43] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/6zls-m67c> for supplementary movies and figures.
- [44] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic, Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study, *Proc. Natl. Acad. Sci. USA* **105**, 1232 (2008),.
- [45] J. Gautrais, F. Ginelli, R. Fournier, S. Blanco, M. Soria, H. Chaté, and G. Theraulaz, Deciphering interactions in moving animal groups, *PLoS Comput. Biol.* **8**, e1002678 (2012).
- [46] R. Olfati-Saber, J. A. Fax, and R. M. Murray, Consensus and cooperation in networked multi-agent systems, *Proc. IEEE* **95**, 215 (2007).
- [47] J. F. Traniello and R. B. Rosengaus, Ecology, evolution and division of labour in social insects, *Anim. Behav.* **53**, 209 (1997).
- [48] A. Duarte, I. Pen, L. Keller, and F. J. Weissing, Evolution of self-organized division of labor in a response threshold model, *Behav. Ecol. Sociobiol.* **66**, 947 (2012).
- [49] J. W. Jolles, A. J. King, and S. S. Killen, The role of individual heterogeneity in collective animal behaviour, *Trends Ecol. Evol.* **35**, 278 (2020).
- [50] M. Fröhlich, C. Boeckx, and C. Tennie, The role of exploration and exploitation in primate communication, *Proc. R. Soc. Lond. B* **292**, 20241665 (2025).
- [51] M. Dorigo, D. Floreano, L. M. Gambardella, F. Mondada, S. Nolfi, T. Baaboura, M. Birattari, M. Bonani, M. Brambilla, A. Brutschy, D. Burnier, A. Campo, A. L. Christensen, A. Decugniere, G. Di Caro, F. Ducatelle, E. Ferrante, A. Forster, J. M. Gonzales, J. Guzzi, *et al.*, Swarmanoid: A novel concept for the study of heterogeneous robotic swarms, *IEEE Robot. Autom. Mag.* **20**, 60 (2013).
- [52] B. Khaldi, F. Harrou, and Y. Sun, Collaborative swarm robotics for sustainable environment monitoring and exploration: Emerging trends and research progress, *Energy Nexus* **17**, 100365 (2025).
- [53] D. J. Barraclough, M. L. Conroy, and D. Lee, Prefrontal cortex and decision making in a mixed-strategy game, *Nat. Neurosci.* **7**, 404 (2004).
- [54] S. H. Singh, F. van Breugel, R. P. N. Rao, and B. W. Brunton, Emergent behaviour and neural dynamics in artificial agents tracking odour plumes, *Nat. Mach. Intell.* **5**, 58 (2023).
- [55] M. Rando, M. James, A. Verri, L. Rosasco, and A. Seminara, Q-learning with temporal memory to navigate turbulence, *eLife* **13**, 102906 (2025).
- [56] L. Canese, G. C. Cardarilli, L. Di Nunzio, R. Fazzolari, D. Giardino, M. Re, and S. Spanò, Multi-agent reinforcement learning: A review of challenges and applications, *Appl. Sci.* **11**, 4948 (2021).
- [57] S. V. Albrecht, F. Christianos, and L. Schäfer, *Multi-agent Reinforcement Learning: Foundations and Modern Approaches* (MIT Press, Cambridge, MA, 2024).
- [58] S. Colabrese, K. Gustavsson, A. Celani, and L. Biferale, Flow navigation by smart microswimmers via reinforcement learning, *Phys. Rev. Lett.* **118**, 158004 (2017).
- [59] L. Piro, M. Carbone, R. Heinonen, L. Biferale, and M. Cencini, Data for “Policy heterogeneity improves collective olfactory search in three-dimensional turbulence” [Data set], Zenodo (2026), <https://doi.org/10.5281/zenodo.19109495>.