



Daring few, patient many: Division of labor in decentralized foraging collectives

Hyunjoong Kim^{a,1} , Zachary P. Kilpatrick^{b,c,1} , and Krešimir Josić^{d,e,f,1}

Affiliations are included on p. 8.

Edited by Simon Asher Levin, Princeton University, Princeton, NJ; received February 12, 2026; accepted May 23, 2026

How can social animals divide labor to forage effectively without a leader? Effective foraging requires balancing individual exploration costs against collective information gains, but without central coordination. This balance must emerge from the distributed decisions of group members. We address this challenge using a collective foraging model in which individuals share information and rewards but each must choose whether to bear the cost of exploring or to remain idle. We show that decentralized collectives can match the performance of centrally controlled groups through a division of labor with a small exploratory minority bearing the cost of foraging in lean times, continuously gathering information to enable a synchronized majority to exploit favorable conditions. This division of labor is inherently adaptive because fixed individual thresholds produce flexible collective behavior without central adjustment of roles. Information redundancy causes the optimal number of explorers to grow logarithmically with group size, so larger groups need proportionally fewer explorers. We find that the ideal level of group heterogeneity is maximized at intermediate ecological pressures, whereas optimal groups are homogeneous under extreme conditions. Collective responses to environmental changes are asymmetric and detecting improving conditions is slower than detecting deteriorating ones, as the exploratory minority needed to signal recovery is costly to maintain. We thus show how a division of labor based on simple individual threshold rules leads to optimal collective performance without central coordination, and predict when ecological pressures favor heterogeneous vs. uniform group composition.

collective foraging | division of labor | informed minority | decentralized coordination | exploration–exploitation

How animal groups balance individual exploration costs and collective information gains remains a central question in our understanding of ecological decision making (1–4). Individual foragers bear energetic and survival costs when venturing into uncertain environments (5, 6), yet the information about the environmental state resulting from these forays can outweigh any immediate rewards (2, 7, 8). This tension raises a fundamental question: How should groups be organized to balance these costs and benefits?

One solution to this problem is centralized leadership, in which a small number of individuals guide collective action. In many wild wolf packs, for example, the breeding pair tends to initiate movements and coordinates traveling, resting, and hunting (9). Human history offers parallel examples. Admiral Zheng He commanded fleets across the Indian Ocean in the early fifteenth century on imperial orders, spurring Chinese emigration and trade throughout Southeast Asia (10). A century later, royal sponsorship supported Magellan's circumnavigation, opening the first western sea route to the Spice Islands and accelerated global trade and contact between distant regions (11). In each case, a small number of designated explorers, acting under a central authority, bore enormous personal risk, while the societies that sent them reaped lasting rewards. However, most social animals, including honeybees, ants, and fish schools, lack fixed leaders and coordinate through local interactions (2, 3, 7, 12–14). Despite extensive empirical study of decentralized animal groups (3, 12, 15, 16), it remains unclear how closely decentralized coordination can match centralized control, how performance depends on diversity in individual strategies, and under what conditions heterogeneous strategies offer an advantage in leaderless groups.

To address these questions, we develop and analyze a stochastic foraging model (17–24) to understand how collectives can achieve efficient exploration. We show that well-calibrated collectives exhibit two characteristic behavioral regimes: during times

Significance

From honeybee swarms to fish schools, leaderless animal groups must divide labor to forage effectively without central control. Using a model of collective foraging in changing environments, we describe the most advantageous division of labor: a small number of risk-tolerant individuals persistently explore the environment, while a coordinated majority forages only when conditions are favorable. This division of labor emerges from individuals following simple threshold-based rules, rather than the directives of a leader. We show that the optimal number of explorers grows sublinearly with group size, so larger groups need proportionally fewer explorers. Our findings provide a framework for understanding how decentralized collectives balance exploration costs against information benefits, and predict when ecological pressures favor heterogeneous vs. uniform group compositions.

Author contributions: H.K., Z.P.K., and K.J. designed research; H.K. performed research; H.K. contributed new reagents/analytic tools; H.K., Z.P.K., and K.J. analyzed data; and H.K., Z.P.K., and K.J. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

¹To whom correspondence may be addressed. Email: hyunjoong.kim@uc.edu, zpkipat@colorado.edu, or josic@math.uh.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2605309123/-DCSupplemental>.

Published June 23, 2026.

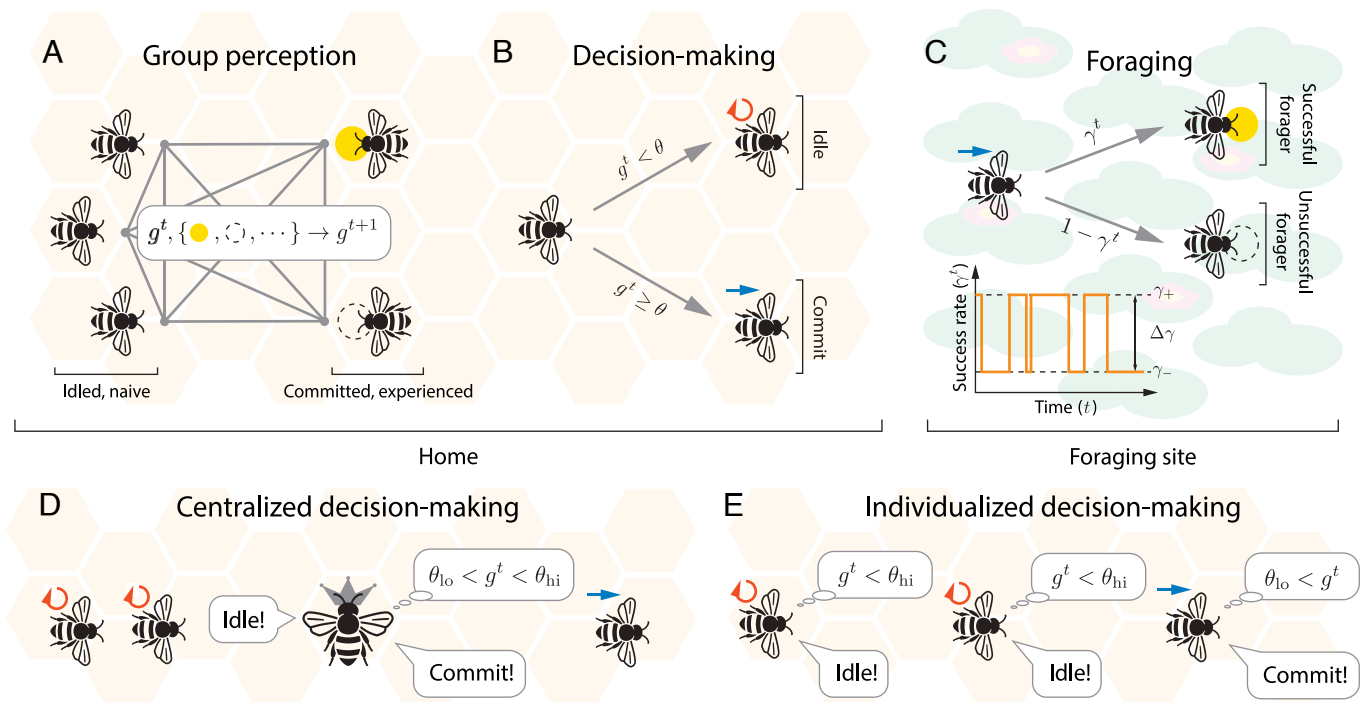


Fig. 1. Overview of model and setup. Individuals in a collective observe returning foragers to update their belief about the environmental state and select actions. (A) The belief, g^t , is updated based on the observed outcomes of foraging attempts. (B) Individuals forage when their belief that the environment is in a high-reward state exceeds a threshold, θ ; these thresholds can differ between individuals. (C) Foraging is successful with probability γ^t , which switches between high and low values with the environmental state. (D) Under centralized control, a planner decides on the optimal number of foragers to maximize returns. (E) With decentralized control, coordination emerges from individual foraging decisions.

of plenty, risk-averse individuals venture out together, amplifying collective gains. When resources are scarce, most individuals refrain from foraging, but a small subset of risk-tolerant explorers continues foraging to rapidly detect environmental improvements. Composition is critical as populations dominated by cautious individuals fail to discover new resources, whereas those dominated by risk-tolerant foragers exhaust shared reserves. The number of risk-tolerant individuals grows logarithmically with population size, a nonintuitive result demonstrated by explicit optimization, making a balanced division of labor between exploration and exploitation essential for efficient foraging (14, 25). Remarkably, collectives with a well-calibrated mixture of cautious and bold individuals can perform as well as groups governed by a central decision-maker. The most heterogeneous compositions emerge under intermediate ecological pressures, when exploration costs and environmental variability are balanced. We thus show how a division of labor based on simple individual threshold rules leads to optimal collective performance, and predict when ecological pressures favor heterogeneous vs. uniform group composition.

A Model of Collective Foraging under Uncertainty

Collective decisions about foraging in large, decentralized collectives face a fundamental challenge: How do groups balance individual risk and collective benefit in changing environments without central coordination? To address this question, we consider a minimal model designed to examine whether decentralized decision-making mechanisms can yield optimal collective long-term returns, even when individual expected returns are negative. Here we provide an overview of the

model; mathematical details can be found in *SI Appendix, section 1*.

Foraging Agents in a Dynamic Environment. We consider a collective of N agents who forage for resources. Time is discrete, and on each time step an agent chooses either to leave the “hive” and forage, or to stay. Each foraging attempt is either successful, resulting in a positive reward, or unsuccessful, resulting in no reward. The environment alternates between a high- and low-rewarding state, so that in the high (low) state a foraging attempt is successful with probability γ_+ (γ_-). To model energetic expenditure and environmental risk, we assume that each foraging attempt incurs a fixed cost, λ , and that the expected return from an attempt is positive in the high-reward state and negative in the low-reward state. Transitions between the states are memoryless, so the environmental state evolves as a two-state discrete-time Markov process with symmetric switching probability, ϵ (Fig. 1C). While real environments fluctuate across a continuum of states, assuming two states provides a minimal framework that captures the essential trade-off between exploration and exploitation under uncertainty (19–21, 24).

Collective Perception. We assume that the community coordinates its behavior through shared information about the outcomes of foraging attempts. Returning foragers are observed by the entire community, so all individuals know how many foraging attempts have been successful (Fig. 1A), an idealization of group-level information flow in social collectives (2, 3, 26). Individuals are Bayesian observers (27–29) and compute the probability that the environmental state is in a high-reward state based on the observed number of successful foragers and knowledge about the environment’s volatility. We refer to this

subjective probability as their “belief” about the environmental state. Because all individuals have access to the same information, they share a common belief that guides their foraging decisions. The belief is updated according to

$$g^{t+1} = (1 - \epsilon)U(x^t|g^t, a^t) + \epsilon(1 - U(x^t|g^t, a^t)), \quad [1]$$

where x^t and a^t are the number of successful and committed foragers at time t , and $U(x^t|g^t, a^t)$ is the Bayesian posterior that the environment is high-rewarding given the observed outcomes (SI Appendix, Eq. S1).

Decision Making with a Centralized Planner. Maximizing returns requires optimizing forager number: deploying too many foragers when conditions are poor wastes effort; too few when conditions are good misses opportunities. To establish an upper bound on collective performance, we derive a theoretical benchmark based on a centralized planner that optimally determines foraging effort required to maximize long-term collective returns (Fig. 1D). Under this centralized policy, all individuals follow the directives of a planner, which infers the environmental state by observing foraging successes and then prescribes the best number of foragers. This optimal policy can be computed using dynamic programming (18, 30, 31) and provides an upper bound on achievable performance. The value function $V(g)$ satisfies

$$V(g) + \rho = \max_{a=0,1,\dots,N} \left\{ r(g|a) + \sum_{g'} V(g')p(g'|g, a) \right\}, \quad [2]$$

where $r(g|a) = a(\bar{\gamma}(g) - \lambda)$ is the expected immediate reward for deploying a foragers given belief g , ρ is the long-run average reward, and $p(g'|g, a)$ is the belief transition probability (SI Appendix, Eq. S2). We compare this benchmark with outcomes from agents who make individual decisions based on shared information but without central coordination.

Decisions in the Absence of a Centralized Planner. With no central decision-maker, each individual decides whether to forage based on their estimate of the environmental state. We assume that each agent chooses to forage if their belief that the environment is in the high-reward state exceeds a private threshold (Fig. 1B and E). The distribution of these thresholds across individuals determines the behavior of the decentralized group (32, 33). Such threshold-based decision rules are consistent with empirical work on quorum/acceptability thresholds in social insects and with theory showing how heterogeneity in thresholds shapes collective decisions (34–39). As we show, an appropriate distribution of such thresholds allows the collective to match the performance of a central optimizer, despite no individual having knowledge of others’ thresholds or any group-level objective.

Despite its abstraction, this model captures the ecological tradeoffs between exploration cost, environmental uncertainty, and collective return. It also allows us to ask whether decentralized decision making itself entails a performance cost, separate from any limitation imposed by imperfect information sharing. By holding information fixed across the centralized and decentralized settings, we isolate the cost of autonomous decision making and ask how closely individual threshold rules can approximate jointly optimized collective behavior.

Results

Using our model of foraging and belief updating, we ask how a collective translates shared environmental information

into effective foraging decisions. We first establish a normative benchmark based on a centralized decision-maker that maximizes expected payoff. We then show that a leaderless group can achieve optimal performance using heterogeneous individual decision thresholds and analyze how environmental conditions shape the optimal threshold distribution in a decentralized collective.

Robust Collective Coordination Through Structured Heterogeneity. A central decision-maker can select an optimal number of foragers to maximize expected payoffs on each turn. In a decentralized collective, individuals share a common belief about the environmental state but make individual choices based on a private threshold. If all individuals choose a threshold that maximizes their own expected return, then all thresholds are identical. Since they share the same belief, everyone makes the same decision: forage when the expected reward is positive, and remain idle when the expected reward is negative. As a result, a collective can quickly detect when an environment switches from a high- to a low-rewarding state. However, when the collective believes the reward state is likely to be low, no individuals forage, and the collective will not detect a switch back to a high-rewarding state. The resulting failure to explore thus leads to starvation even when conditions improve. Conversely, if all individuals adopt low thresholds and forage often, the collective can quickly infer the environmental state but incurs large energetic losses in lean times. Effective collective performance therefore requires a division of labor between a minority that bears the cost of foraging in lean times and a majority that waits for favorable conditions.

With an appropriate distribution of individual decision thresholds, a decentralized collective can achieve the same expected returns as an optimally controlled group. We show this by constructing a threshold distribution that reproduces the number of foragers deployed under the optimal centralized policy. We do not suggest that collectives compute or implement a centralized solution; rather, this construction demonstrates that simple local threshold rules can recover centralized allocations, allowing us to analyze how ecological parameters shape the structure and robustness of the resulting strategy. Even when all individuals observe the same foraging outcomes, each agent acts on a private threshold with no joint optimization of roles, so the centralized benchmark measures the gap attributable to decentralized decision making alone. The optimal strategy of a central planner can be represented by a function, $\pi(g)$, that maps the collective belief about the state of the environment, g , to the number of individuals sent out to forage, $n = \pi(g)$ (Fig. 2A). As the belief that the environment is in a high-reward state increases, the planner commits more foragers (17, 40). The optimal policy, $\pi(g)$, is nondecreasing in the belief g , since higher expected rewards never justify deploying fewer foragers. We can therefore define a sequence of belief thresholds,

$$\theta_n = \inf\{\theta \in [0, 1] : \pi(\theta) \geq n\}, \quad n = 1, 2, \dots, N, \quad [3]$$

where θ_n is the smallest belief at which deploying at least n foragers is optimal. Multiple thresholds may coincide, and in large groups the majority of thresholds equal the critical belief θ_c (Fig. 3) at which the expected foraging return is zero.

The strategy of a decentralized collective is fully determined by a set of private thresholds. Hence, a collective of size N whose individuals have thresholds $\{\theta_n\}_{n=1}^N$ will send out exactly n foragers when the belief satisfies $\theta_n \leq g < \theta_{n+1}$, with the convention $\theta_{N+1} = 1$ (Fig. 2B). This choice of thresholds thus results in a number of foragers equal to that dictated by the optimal policy, $\pi(g)$ for any belief g . Importantly,

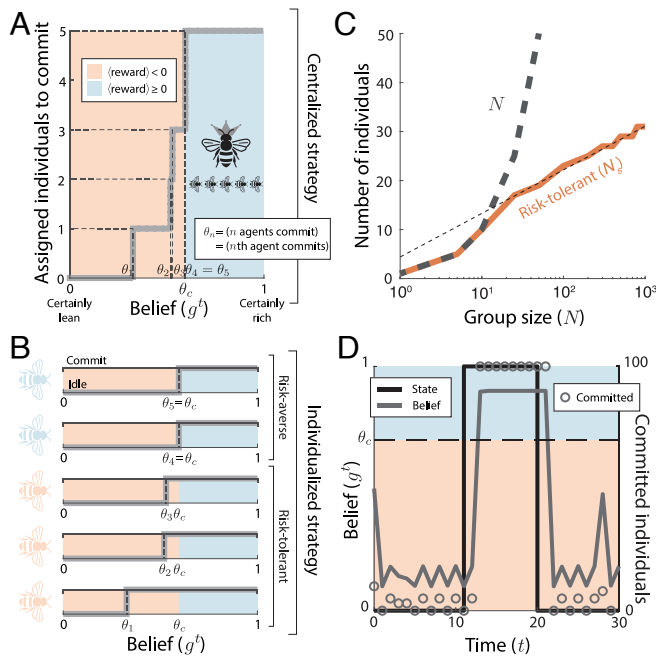


Fig. 2. Division of labor enables centralized and decentralized coordination. (A) In decentralized collectives, individuals have different foraging thresholds. Risk-averse individuals share a common threshold, θ_c , and forage only when the expected reward is positive. Risk-tolerant explorers have lower, heterogeneous thresholds and forage even when the expected reward is negative. (B) Under centralized control, a planner implementing the optimal policy, $\pi(g)$, deploys an increasing number of foragers as the belief increases. An appropriate distribution of individual thresholds allows a decentralized collective to match this allocation (compare panels A and B). (C) The optimal number of risk-tolerant explorers, N_s , grows sublinearly with population size, N , so an increasingly small fraction of explorers suffices for efficient foraging. (D) The collective belief, g^t , tracks environmental fluctuations because a subset continues exploring during lean periods, so the collective can rapidly re-engage when conditions improve.

individual decisions are based on a shared belief paired with private thresholds, rather than on orders from a central planner. The resulting distribution of foraging strategies could arise from evolved or learned differences in risk sensitivity (41–43), and is consistent with classic response-threshold models of division of labor (32, 44). We show that such a distribution can emerge from a homogeneous population via selection on group-level foraging returns, with no central authority assigning thresholds (SI Appendix, Fig. S1).

Unlike centralized collectives that concentrate decision making in a single coordinator, decentralized collectives distribute responsibility across individuals with heterogeneous strategies, making them more robust to decision noise and environmental perturbations, and potentially promoting evolutionary stability (2, 45). As long as phenotypic differences across the population generate an appropriate distribution of risk tolerances, the collective can achieve optimal foraging performance (25, 41). Structured heterogeneity thus allows decentralized groups to match the allocations of a central planner while maintaining robust performance in dynamic environments. We next analyze the structure of the optimal population and ask how ecological pressures shape this distribution of strategies.

Minority Exploration and Majority Synchronization. An optimal collective balances risk-tolerant individuals, who forage under uncertainty, with risk-averse individuals, who forage only when the expected reward is positive. Empirical studies show a consistent

pattern in many social species: a small fraction of individuals, such as “scouts” in honeybee populations, persistently engages in exploration, while most colony members adopt safer strategies and exploit known resources (14, 41, 46, 47). We show that the optimal decentralized strategy yields a similar pattern. To do so, we use the critical threshold, θ_c , and classify individuals with decision thresholds below θ_c as risk-tolerant, and those with thresholds equal to or exceeding θ_c as risk-averse.

In small optimal collectives, all individuals are risk-tolerant, because each foraging outcome is highly informative and has a large impact on the collective belief about the environmental state (Fig. 2C). As group size increases, performance improves, but the optimal division of labor changes. The expected cost of foraging grows linearly with the number of active foragers, whereas additional foraging outcomes provide redundant information and thus become less valuable. A mathematical analysis shows that the optimal number of risk-tolerant individuals grows logarithmically with group size (dashed line in Fig. 2C), reflecting a precise balance between the diminishing informational returns of additional explorers and the linearly accumulating costs of exploration. This sublinear scaling is robust across a wide range of environmental switching rates, foraging costs, and reward structures (SI Appendix, Fig. S2). Consequently, larger groups comprise a growing majority of risk-averse individuals, while the fraction of risk-tolerant explorers declines. This prediction is consistent with empirical and theoretical studies showing that effective animal collectives rely on a small exploratory minority whose size grows more slowly than group size (2, 48). Details of the asymptotic analysis underlying this result are provided in SI Appendix, section 2.

While the decision thresholds of risk-tolerant individuals are distributed, all risk-averse individuals share a common threshold, θ_c . Indeed, the entire collective commits to foraging once the expected reward becomes positive (SI Appendix, section 2 and Fig. S3). This synchronization of foraging bouts is a consequence of the linear increase of total reward with the number of committed individuals, implying that full commitment is optimal when returns are expected to be favorable. Thus, an optimal collective consists of a heterogeneous minority that maintains environmental awareness through continued sampling, and a synchronized majority that maximizes collective returns through coordinated exploitation once conditions are deemed favorable.

Optimal collective foraging strategies exhibit a characteristic cascade of activity in response to environmental changes (Fig. 2D). When conditions deteriorate, widespread foraging produces few successes, quickly shifting the collective belief toward low expected rewards. Most individuals stop foraging, leaving a small subset of risk-tolerant explorers that continue sampling. When conditions deteriorate severely, even limited exploration ceases to minimize costs. Nevertheless, in the absence of new information, collective belief gradually relaxes toward uncertainty, prompting renewed exploration. If some of these sparse foraging attempts are successful, the belief shifts toward high expected rewards, and more foragers are recruited. Once the collective belief crosses the threshold θ_c , coordinated collective foraging is triggered again, and the process repeats. This cascade is asymmetric: deteriorating conditions are detected rapidly by the whole group, whereas improving conditions must first be detected by the costly exploratory minority before triggering group-wide recruitment. Heterogeneity in decision thresholds enables the population to function as a distributed sensor, maintaining sensitivity to environmental changes while balancing the cost of exploration.

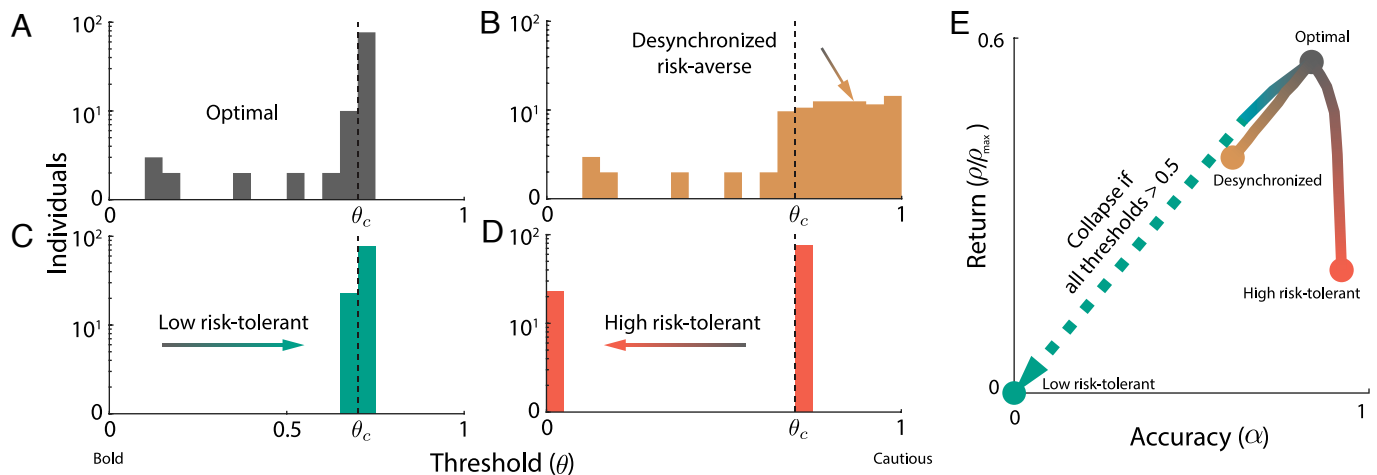


Fig. 3. Effects of perturbing the distribution of decision thresholds. (A) Optimal threshold distribution for a collective of $N = 100$ individuals. (B–D) Perturbed strategy distributions: (B) increasing the spread of thresholds in the risk-averse majority; (C) risk-tolerant individuals sharing a low threshold. Colors indicate the different perturbation types. (E) Normalized return vs. accuracy for the optimal strategy and its perturbations shows that returns are maximized at intermediate accuracy (gray dot; panel A), reflecting weakened coordinated exploitation under desynchronization (brown-gray curve; panel B), improved accuracy at increased cost under moderate heterogeneity (pink-gray curve; panel C), and degraded returns under excessive risk tolerance (blue-gray curve; panel D). Returns are normalized by the maximum possible return $\rho_{\max}$, defined as the return obtained when the entire collective forages only in the high-reward state.

Risk-tolerant individuals thus act as functional leaders by providing information about the environment, rather than by imposing their choices on others. Their exploratory actions inform the more cautious majority, resulting in a form of passive leadership that emerges from threshold heterogeneity (49–51), rather than social dominance or intentional coordination. The resulting collective dynamics operate on two timescales: slow environmental detection by the exploratory minority and rapid group-wide responses once conditions are confirmed favorable (Fig. 2D and SI Appendix, Fig. S4).

Uniform Strategies Degrade Collective Performance. A division of labor based on diverse risk tolerances is therefore crucial for balancing the costs and benefits of exploration (Fig. 3A). We next compare the performance of such optimal collectives to those with suboptimal strategy compositions. We quantify collective performance using two metrics (Fig. 3E). The first is the long-term average return, ρ , normalized by the maximum possible return $\rho_{\max} = N(\gamma_+ - \lambda)/2$ achieved when all individuals forage only when conditions are favorable. The second is accuracy, α , defined as the time-averaged proportion of individuals committed to foraging when the environment is in the high-reward state.

When all thresholds are high, the population is overly cautious, since even relatively risk-tolerant individuals forage too infrequently to detect environmental improvements. As a result, the collective responds slowly when good times return, reducing long-term average reward (Fig. 3C and E). In the extreme case where all individuals require excessive certainty to forage ($\theta_n > 1/2$ for all n), unsuccessful foraging drives the collective belief below $1/2$, halting further exploration. In the absence of new observations, the belief relaxes back toward uncertainty ($g = 1/2$). Since all decision thresholds exceed this value, the collective remains inactive indefinitely.

Conversely, a collective with too many risk-takers fails in the opposite way: eager explorers can maintain high accuracy, but only by sustaining near-continuous sampling. The marginal information gain of these explorers is outweighed by the energetic cost of foraging, depressing long-term returns

(Fig. 3D and E). In contrast, unstructured deviations from a shared risk-averse threshold primarily impair exploitation rather than information acquisition, desynchronizing the cautious majority and preventing coordinated surges of foraging when conditions are favorable (Fig. 3B and E).

Optimal collectives thus exhibit functional specialization, with a synchronized risk-averse majority sharing threshold θ_c responsible for coordinated exploitation, and a heterogeneous risk-tolerant minority that sustains exploration through a broad distribution of thresholds in large groups (Fig. 3 and SI Appendix, Fig. S5). Together, these results reveal a collective exploration–exploitation trade-off: heterogeneity among risk-tolerant individuals sustains exploration under uncertainty, while synchronization of the risk-averse majority maximizes returns once conditions are favorable. This exploration–exploitation trade-off requires only that the right proportions of explorers and exploiters are maintained for given ecological conditions; the model does not prescribe how this balance is achieved, leaving open the possibility that individuals can rotate between roles over time.

Ecological Extremes Suppress Division of Labor. We next ask how changes in environmental conditions impact the optimal division of labor between risk-tolerant and risk-averse individuals. In particular, we show that the cost of foraging, λ (energetic expenditure and risk), the difference between expected reward during good and bad times, $\Delta\gamma = \gamma_+ - \gamma_-$; and environmental variability, ϵ (the probability of state transitions) strongly shape the structure of a well-calibrated collective. We quantify these effects using the SD of the optimal threshold distribution as a measure of heterogeneity of strategies in the collective.

Group heterogeneity varies nonmonotonically with foraging cost, peaking at intermediate values (Fig. 4A and D and SI Appendix, Fig. S6). When costs are low ($\theta_c(\lambda) < \epsilon$), resources are easily accessed and all individuals forage, resulting in a homogeneous population of foragers. Conversely, when costs are high ($1 - \epsilon < \theta_c(\lambda)$), exploration becomes prohibitively expensive, and all individuals refrain from foraging. Between these extremes ($\epsilon \leq \theta_c(\lambda) \leq 1 - \epsilon$), a maximal variability

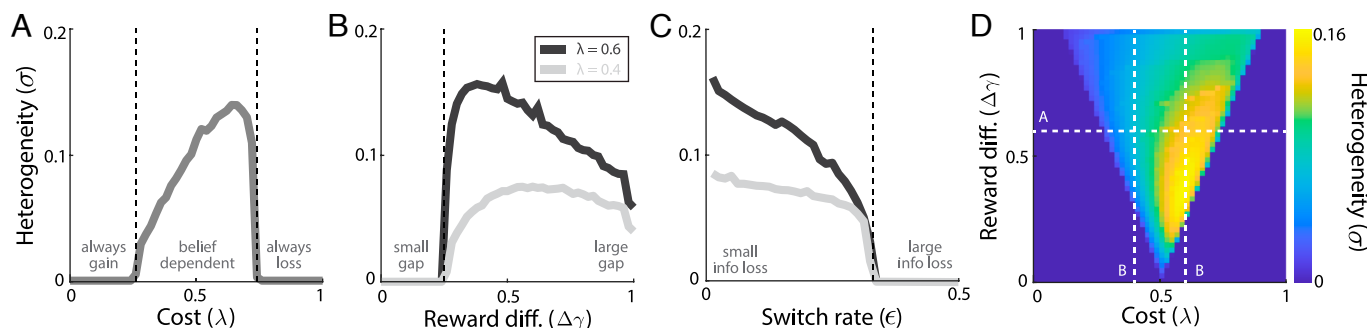


Fig. 4. Ecological modulation of collective heterogeneity. (A) Heterogeneity of optimal decision thresholds varies nonmonotonically with commitment cost, λ , peaking at intermediate values where foraging decisions depend sensitively on belief; extreme costs lead to uniform inactivity (high cost) or uniform foraging (low cost). (B) Heterogeneity also peaks at intermediate reward rate contrasts, $\Delta\gamma$, (shown for two commitment costs): small changes in reward rate are both hard to detect and not worth detecting. Large changes are easily detected by a few foragers. Heterogeneity thus saturates in both limits. (C) Rapid environmental switching suppresses heterogeneity by limiting achievable inference accuracy. (D) Joint variation of cost and reward contrast reveals a wedge of intermediate ecological conditions in which division of labor and structured heterogeneity are optimal.

of risk-preference is attained. In this intermediate regime, moderate increases in foraging cost selectively suppress marginal exploration by cautious individuals, while favoring continued sampling by a smaller risk-tolerant subset, thereby amplifying the diversity of thresholds used (wedge-shaped region of high heterogeneity in Fig. 4D).

Group heterogeneity also varies nonmonotonically with the difference between the expected reward rate in the two environmental states, $\Delta\gamma = \gamma_+ - \gamma_-$ (Fig. 4B and D and SI Appendix, Fig. S6). When this difference is small ($\theta_c(\Delta\gamma) < \epsilon$ or $\theta_c(\Delta\gamma) > 1 - \epsilon$), the environment is nearly deterministic, and the population follows a homogeneous strategy. Again, all individuals forage or remain idle depending on whether the average reward exceeds the commitment cost. As the difference increases, distinguishing favorable from unfavorable conditions becomes more valuable, as early detection yields large gains in expected return. This leads to the emergence of risk-tolerant individuals and greater strategic heterogeneity. At extreme contrasts in reward, transitions to a favorable state are easily inferred without extensive exploration, because even sparse successes strongly shift belief, reducing the benefits of maintaining many risk-tolerant individuals in the population. Heterogeneity is amplified when commitment costs are high, as risk-averse individuals require greater certainty to forage, increasing the informational value of continued sampling by a small exploratory minority.

Rapid environmental changes suppress heterogeneity, favoring homogeneous risk aversion over exploratory strategies (Fig. 4C). When environments fluctuate rapidly, maintaining risk-tolerant individuals becomes counterproductive as information gained through exploration quickly becomes obsolete. In this regime, the timescale of environmental changes is shorter than the time needed to coordinate a response. As a result, fast-changing environments suppress heterogeneity, favoring uniform populations: all individuals forage when costs are low relative to the time-averaged reward, or all refrain when costs are high. In contrast, when environmental changes are slow and predictable relative to the timescale of exploration and response, information accumulates and remains reliable over long periods, so the group can profit from exploratory foraging in future decisions. Under these conditions, risk-tolerant individuals can repeatedly detect state changes, stabilizing collective belief and enabling timely collective commitment and making a structured division of labor both possible and advantageous.

Discussion

We have shown that leaderless collectives can achieve optimal foraging performance through a division of labor based on heterogeneous risk tolerances. A small fraction of risk-tolerant individuals, whose number increases logarithmically with group size, maintains environmental awareness by exploring even when expected rewards are negative, while a synchronized risk-averse majority maximizes collective gains by committing only when conditions are favorable. This asymmetric structure reflects fundamental principles of collective information gathering: as individuals sample independently, the information value of additional observations diminishes due to redundancy, whereas exploration costs accumulate linearly with the number of explorers, so only a small fraction is needed for accurate environmental tracking in large groups.

Optimal collective foraging exhibits characteristic decision cascades in response to environmental changes. Minority-driven exploration triggers rapid group-wide recruitment when conditions improve, enabling swift responses to opportunities while limiting costly false alarms. Exploration loses its value under extreme ecological pressures or in rapidly changing environments where information becomes obsolete before it can guide actions. This causes collectives to favor uniform risk aversion over strategic diversity. Hence, the ideal breadth of heterogeneity in decision thresholds is most pronounced under intermediate foraging costs and environmental contrasts (Fig. 4).

Risk-tolerant individuals function as informative leaders in decentralized groups, guiding collective behavior by maintaining environmental awareness during periods of uncertainty (2, 49, 50). Whether such leadership is persistent or transient depends on its origin. Stable phenotypic variation in energetic state or temperament can produce consistent leaders (49, 52, 53), whereas probabilistic decision making by otherwise homogeneous individuals allows leadership to rotate while preserving near-optimal collective performance (51, 54, 55) (see SI Appendix, section 3 for proof). Notably, even a population initialized with identical thresholds at the greedy level can spontaneously develop the requisite heterogeneity: agents with poor reward histories gradually resample their thresholds, diversifying the distribution toward the optimal structure through purely local responses to collective experience (SI Appendix, Fig. S1).

While comparable divisions of labor can be imposed by centralized or dominance-based control in some groups (9, 55), we focused on how informational leadership can emerge in

leaderless collectives via heterogeneity in exploration preference under uncertainty (4, 7). A minority that samples when payoffs are ambiguous can distribute group sensing to regulate majority engagement, consistent with broader views of collective computation and information flow in animal groups (26). This perspective suggests that if exploratory roles reflect stable phenotypic differences, individuals should show cross-context consistency in risk-taking and influence (52, 53, 56); if instead exploration is due to stochastic choices among otherwise similar individuals, leadership should rotate without loss of group-level performance (54, 57). Correspondingly, altering exploration costs or environmental persistence is expected to shift the degree and stability of leadership and the strength of division of labor (8, 58). These predictions are experimentally accessible in social insects and other tractable systems where ecological timescales and costs can be manipulated and collective responses quantified (47, 59, 60).

A sublinearly increasing number of risk-tolerant explorers appears in diverse collective decision-making systems: informed minorities can effectively guide naïve majorities in moving animal groups (2), a small fraction of hasty explorers can optimize group decision efficiency when observed by a large, deliberate majority (61), and a few scouts can reliably inform nest-site choices in social insects (48). Although leadership by risk-tolerant minorities has been studied experimentally in groups of fixed size (62–64), how the number of leaders should scale with group size and how this scaling emerges from individual decision rules remains largely unexplored. Related work shows a similar sublinear scaling of active individuals with colony size, arising from a balance of social activation and inhibition rather than diminishing informational returns (65, 66), suggesting the pattern may reflect a general principle with multiple mechanistic origins. Recent work in swarm robotics reaches qualitatively similar conclusions, showing that heterogeneity in individual thresholds improves collective foraging performance and that the degree of optimal heterogeneity depends on group size (67–69). Our work complements this literature by deriving the theoretical optimum directly via dynamic programming and proving constructively that it can be matched by a decentralized collective, providing a principled benchmark against which approximate strategies can be evaluated.

The role of heterogeneity in group composition has been extensively studied using threshold-based models of division of labor, which typically assume symmetric, normally distributed response thresholds across individuals (32, 33, 44, 70, 71). However, the structure of threshold distributions itself has been largely overlooked. Skewness alone can generate complex relationships between group composition and collective performance. Division of labor in social insects can emerge from multiple interacting factors, beyond threshold variation (71), yet the contribution of the distribution structure remains unexplored. We hypothesize that natural groups may evolve skewed threshold compositions to optimize collective task performance.

One limitation of our model is the assumption of perfect information sharing as all individuals instantaneously observe collective foraging outcomes. In many natural systems not all potential foragers observe every returning individual, and information is transmitted locally or indirectly (3, 26, 61). Such local transmission can generate modular organization, in which distinct subpopulations have unequal access to exploratory returns. Even when information is available, individuals may differ in perceptual or decision noise (e.g., noisy belief updates or variable thresholds) (28), resulting in heterogeneous beliefs and

stochastic actions. Under such conditions, individuals may rely not only on explicit signals but also on implicit social information inferred from the behavior of others (57, 72, 73). How partial information sharing and social inference reshape the optimal division of labor remains an open question (44). We show that the core organizational structure identified here, a minority of explorers along with a synchronized majority, remains robust under imperfect communication, although achieving comparable performance requires a somewhat larger exploratory cohort to compensate for information loss (*SI Appendix*, Fig. S7).

Our results also raise questions about how finite energetic reserves constrain collective decision making. While our model implicitly assumes unbounded accumulation of returns, biological systems operate under both lower and upper energetic limits. Near starvation, collectives may be forced into riskier strategies to avoid extinction. How such transitions are distributed across heterogeneous groups remains unclear. If risk heterogeneity reflects hunger-driven phenotypic variation, individuals with depleted reserves may become disproportionately risk-tolerant (34), amplifying the influence of an informed minority during critical periods (49). Conversely, when storage approaches saturation, continued foraging becomes unnecessary, potentially suppressing exploration and diminishing the role of exploratory individuals altogether. Understanding how energetic constraints reshape the emergence, persistence, and influence of informed minorities is therefore essential for linking individual state, division of labor, and collective responsiveness of natural systems.

In human society, the parallel to the risk-tolerant explorer is thus not the state-sponsored adventurer like Zheng He or Magellan, but the self-motivated innovator: the garage inventor, the citizen scientist, the open-source builder, the high-risk entrepreneur, and the avant-garde artist. These individuals are driven by curiosity and passion and have lower thresholds for action than others. They often bear substantial personal costs, yet when they succeed, they spark movements that drive collective behavior at scale.

Materials and Methods

Bellman Equation and Numerical Solution. To determine the optimal collective foraging strategy under a centralized planner, we formulate the problem as a Markov decision process in which the system state is given by the collective belief about the environmental state. The goal is to find the strategy that maximizes the long-term average collective return. Using dynamic programming, we derive the Bellman equation characterizing the optimal policy, $\pi(g)$, that maps belief states g to the number of agents committed to foraging. The derivation of the Bellman equation is provided in *SI Appendix*, section 1.

We compute the optimal policy numerically by discretizing the belief space on the interval $[0, 1]$ using a uniform grid of 201 points and solving the discretized Bellman equation via fixed-point iteration. Iteration is continued until the value function converges to tolerance 10^{-6} . The resulting policy serves as a benchmark for collective performance and as a reference for constructing decentralized strategies.

Asymptotic Analysis and Large Deviation Theory. To understand how the optimal collective strategy, $\pi(g)$, changes with group size N , we perform asymptotic analysis of the Bellman equation in the limit of large N . This analysis characterizes the leading-order behavior of the optimal policy and shows that only a sublinear number of agents explore (forage) under unfavorable conditions, i.e., when $g < \theta_c$. In our model setup, this scaling is logarithmic in N .

Rare events associated with future rewards and belief transitions are characterized using large deviation theory. These estimates allow us to quantify

the contribution of the current number of committed agents to long-term returns and to derive bounds on the optimal allocation of exploratory effort for large groups. Technical assumptions and full derivations are presented in [SI Appendix, section 2](#).

Perturbations of the Optimal Policy. To perturb the optimal policy in [Fig. 3](#), we construct strategies by interpolating between the optimal policy $\pi(g)$ and three extreme reference policies $\pi_{\text{ext}}(g)$ corresponding to uniformly bold, uniformly cautious, and desynchronized cautious strategies. More precisely, the perturbed policy is given by $\pi_{\lambda}(g) = \lfloor (1 - \lambda)\pi(g) + \lambda\pi_{\text{ext}}(g) \rfloor$ for $\lambda \in [0, 1]$, where $\lfloor z \rfloor$ denotes the largest integer less than or equal to z . These interpolations systematically alter the degree of heterogeneity and synchronization in foraging behavior. Performance under perturbed policies is evaluated using the same metrics as for the optimal strategy (return and accuracy).

Model Parameters. All figures in the main text are generated using a baseline parameter set with group size $N = 100$, reward probabilities $\gamma_{\pm} = 0.5 \pm 0.3$, switch probability $\epsilon = 0.1$, and commitment cost $\lambda = 0.62$ (and thus $\theta_c = (\lambda - \gamma_-)/(\gamma_+ - \gamma_-) = 0.7$) chosen to represent typical model behavior unless otherwise specified. In [Fig. 3](#), we vary the interpolation parameter while holding other parameters fixed at baseline values. In [Fig. 4](#), we choose $\gamma_{\pm} = (1 \pm \Delta\gamma)/2$.

Monte Carlo Simulations. To compute performance metrics for given model parameters, we employed Monte Carlo simulations of the stochastic foraging process. For each parameter set and (perturbed) policy, we simulate the collective dynamics over long time horizons to estimate the performance metrics. Each data point in [Fig. 3](#) is computed from 10^3 independent realizations, each simulated for 2×10^4 time steps.

Data, Materials, and Software Availability. All scripts for simulations, analysis, and figure generation are available in a public GitHub repository ([74](#)).

ACKNOWLEDGMENTS. We thank Benjamin Hayden, Joshua Gold, Ahmed El Hady, and Tahra Eissa for providing helpful feedback on manuscript drafts. H.K. gratefully acknowledges support from the Charles Phelps Taft Research Center. Z.P.K. gratefully acknowledges support from the NSF (DMS-2207700, DMS-2527338). K.J. gratefully acknowledges support from the NSF (DMS-2207647 and EFRI-OI-2515431).

Author affiliations: ^aDepartment of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221; ^bDepartment of Applied Mathematics, University of Colorado, Boulder, CO 80309; ^cInstitute of Cognitive Science, University of Colorado, Boulder, CO 80309; ^dDepartment of Mathematics, University of Houston, Houston, TX 77204; ^eDepartment of Biology and Biochemistry, University of Houston, Houston, TX 77204; and ^fNSF-Simons National Institute for Theory and Mathematics in Biology, Chicago, IL 60611

1. L. Conradt, T. J. Roper, Group decision-making in animals. *Nature* **421**, 155–158 (2003).
2. I. D. Couzin, J. Krause, N. R. Franks, S. A. Levin, Effective leadership and decision-making in animal groups on the move. *Nature* **433**, 513–516 (2005).
3. D. Sumpter, The principles of collective animal behaviour. *Philos. Trans. R. Soc. B: Biol. Sci.* **361**, 5–22 (2005).
4. I. D. Couzin, Collective cognition in animal groups. *Trends Cogn. Sci.* **13**, 36–43 (2009).
5. D. W. Stephens, J. R. Krebs, *Foraging Theory* (Princeton University Press, 1986), vol. 6.
6. T. J. Wolf, P. Schmid-Hempel, Extra loads and foraging life span in honeybee workers. *J. Anim. Ecol.* **58**, 943 (1989).
7. A. M. Hein *et al.*, The evolution of distributed sensing and collective computation in animal populations. *eLife* **4**, e10955 (2015).
8. S. Dall, L. Giraldeau, O. Olsson, J. McNamara, D. Stephens, Information and its use by animals in evolutionary ecology. *Trends Ecol. Evol.* **20**, 187–193 (2005).
9. L. D. Mech, Alpha status, dominance, and division of labor in wolf packs. *Can. J. Zool.* **77**, 1196–1203 (1999).
10. L. Levathes, *When China Ruled the Seas: The Treasure Fleet of the Dragon Throne* (Open Road Media, 2014), pp. 1405–1433.
11. L. Bergreen, *Over the Edge of the World: Magellan's Terrifying Circumnavigation of the Globe* (William Morrow and Company, 2003).
12. S. Pratt, E. Mallon, D. Sumpter, N. Franks, Quorum sensing, recruitment, and collective decision-making during colony emigration by the ant *Leptothorax albigipennis*. *Behav. Ecol. Sociobiol.* **52**, 117–127 (2002).
13. T. D. Seeley, P. K. Visscher, Quorum sensing during nest-site selection by honeybee swarms. *Behav. Ecol. Sociobiol.* **56**, 594–601 (2004).
14. T. D. Seeley, *Honeybee Democracy* (Princeton University Press, 2011).
15. I. Giardina, Collective behavior in animal groups: Theoretical models and empirical studies. *Hfsp J.* **2**, 205–219 (2008).
16. D. M. Gordon, The ecology of collective behavior. *PLoS Biol.* **12**, e1001805 (2014).
17. M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (Wiley, 1994).
18. S. Mahadevan, Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Mach. Learn.* **22**, 159–195 (1996).
19. J. I. Gold, M. N. Shadlen, The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
20. Z. P. Kilpatrick, J. D. Davidson, A. El Hady, Uncertainty drives deviations in normative foraging decision strategies. *J. R. Soc. Interface* **18**, 20210337 (2021).
21. L. Blum Moysé, A. El Hady, Social patch foraging theory in an egalitarian group. *PRX Life* **3**, 033007 (2025).
22. A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications* (Springer, Berlin/Heidelberg, Germany, 2010).
23. P. C. Bressloff, *Probability Theory and Martingales* (Springer International Publishing, 2021), pp. 681–703.
24. N. W. Barendregt, J. I. Gold, K. Josić, Z. P. Kilpatrick, Information-seeking decision strategies mitigate risk in dynamic, uncertain environments. *J. R. Soc. Interface* **23**, 20251308 (2026).
25. D. Charbonneau, A. Dornhaus, Workers 'specialized' on inactivity: Behavioral consistency of inactive workers and their role in task allocation. *Behav. Ecol. Sociobiol.* **69**, 1459–1472 (2015).
26. D. Sumpter, C. Buhl, D. Biro, I. Couzin, Information transfer in moving animal groups. *Theory Biosci.* **127**, 177–186 (2008).
27. A. Wald, J. Wolfowitz, Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* **19**, 326–339 (1948).
28. R. Bogacz, E. Brown, J. Moehlis, P. Holmes, J. D. Cohen, The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
29. A. Veliz-Cuba, Z. P. Kilpatrick, K. Josić, Stochastic models of evidence accumulation in changing environments. *SIAM Rev.* **58**, 264–289 (2016).
30. R. Bellman, The theory of dynamic programming. *Bull. Am. Math. Soc.* **60**, 503–515 (1954).
31. R. Sutton, A. Barto, Reinforcement learning: An introduction. *IEEE Trans. Neural Networks* **9**, 1054–1054 (1998).
32. E. Bonabeau, G. Theraulaz, J. L. Deneubourg, Quantitative study of the fixed threshold model for the regulation of division of labour in insect societies. *Proc. R. Soc. London, Ser. B: Biol. Sci.* **263**, 1565–1569 (1996).
33. E. Bonabeau, Fixed response thresholds and the regulation of division of labor in insect societies. *Bull. Math. Biol.* **60**, 753–807 (1998).
34. G. Theraulaz, E. Bonabeau, J. N. Deneubourg, Response threshold reinforcements and division of labour in insect societies. *Proc. R. Soc. London, Ser. B: Biol. Sci.* **265**, 327–332 (1998).
35. N. R. Franks, A. Dornhaus, J. P. Fitzsimmons, M. Stevens, Speed versus accuracy in collective decision making. *Proc. R. Soc. London, Ser. B: Biol. Sci.* **270**, 2457–2463 (2003).
36. E. J. H. Robinson, N. R. Franks, S. Ellis, S. Okuda, J. A. R. Marshall, A simple threshold rule is sufficient to explain sophisticated collective decision-making. *PLoS ONE* **6**, e19981 (2011).
37. L. Conradt, Models in animal collective decision-making: Information uncertainty and conflicting preferences. *Interface Focus* **2**, 226–240 (2012).
38. R. Jeanson, A. Weidenmüller, Interindividual variability in social insects—Proximate causes and ultimate consequences. *Biol. Rev.* **89**, 671–687 (2014).
39. N. Masuda, T. A. O'Shea-Wheller, C. Doran, N. R. Franks, Computational model of collective nest selection by ants with heterogeneous acceptance thresholds. *R. Soc. Open Sci.* **2**, 140533 (2015).
40. D. M. Topkis, *Supermodularity and Complementarity* (Princeton University Press, 1998).
41. T. D. Seeley, Division of labor between scouts and recruits in honeybee foraging. *Behav. Ecol. Sociobiol.* **12**, 253–259 (1983).
42. C. Dreier, Division of labor between scouts and recruits: Genetic influence and mechanisms. *Behav. Ecol. Sociobiol.* **43**, 191–196 (1998).
43. Z. S. Liang *et al.*, Molecular determinants of scouting behavior in honey bees. *Science* **335**, 1225–1228 (2012).
44. S. N. Beshers, J. H. Fewell, Models of division of labor in social insects. *Annu. Rev. Entomol.* **46**, 413–440 (2001).
45. J. M. Jandt *et al.*, Behavioural syndromes and social insects: Personality at multiple levels. *Biol. Rev.* **89**, 48–67 (2013).
46. M. Beekman, D. J. T. Sumpter, F. L. W. Ratnieks, Phase transition between disordered and ordered foraging in pharaoh's ants. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9703–9706 (2001).
47. A. Dornhaus, N. R. Franks, Colony size affects collective decision-making in the ant *Temnothorax albigipennis*. *Insectes Soc.* **53**, 420–427 (2006).
48. A. L. Cronin, M. C. Stumpe, Ants work harder during consensus decision-making in small groups. *J. R. Soc. Interface* **11**, 20140641 (2014).
49. A. J. King, D. D. Johnson, M. Van Vugt, The origins and evolution of leadership. *Curr. Biol.* **19**, R911–R916 (2009).
50. A. Strandburg-Peshkin, D. Papageorgiou, M. C. Crofoot, D. R. Farine, Inferring influence and leadership in moving animal groups. *Philos. Trans. R. Soc. B: Biol. Sci.* **373**, 20170006 (2018).
51. J. W. Jolles, A. J. King, S. S. Killen, The role of individual heterogeneity in collective animal behaviour. *Trends Ecol. Evol.* **35**, 278–291 (2020).
52. P. A. Biro, J. A. Stamps, Are animal personality traits linked to life-history productivity? *Trends Ecol. Evol.* **23**, 361–368 (2008).
53. J. W. Jolles, N. J. Boogert, V. H. Sridhar, I. D. Couzin, A. Manica, Consistent individual differences drive collective behavior and group functioning of schooling fish. *Curr. Biol.* **27**, 2862–2868 (2017).

54. I. D. Couzin *et al.*, Uninformed individuals promote democratic consensus in animal groups. *Science* **334**, 1578–1580 (2011).
55. L. Conradt, T. J. Roper, Consensus decision making in animals. *Trends Ecol. Evol.* **20**, 449–456 (2005).
56. A. J. King, G. Fehlmann, D. Biro, A. J. Ward, I. Fürtbauer, Re-wilding collective behaviour: An ecological perspective. *Trends Ecol. Evol.* **33**, 347–357 (2018).
57. R. P. Mann, Collective decision making by rational individuals. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E10387–E10396 (2018).
58. S. L. Lima, L. M. Dill, Behavioral decisions made under the risk of predation: A review and prospectus. *Can. J. Zool.* **68**, 619–640 (1990).
59. T. D. Seeley, Honey bee foragers as sensory units of their colonies. *Behav. Ecol. Sociobiol.* **34**, 51–62 (1994).
60. J. D. Davidson, R. P. Arauco-Aliaga, S. Crow, D. M. Gordon, M. S. Goldman, Effect of interactions between harvester ants on forager decisions. *Front. Ecol. Evol.* **4**, 115 (2016).
61. B. Karamched *et al.*, Heterogeneity improves speed and accuracy in social networks. *Phys. Rev. Lett.* **125**, 218302 (2020).
62. S. G. Reebs, Can a minority of informed leaders determine the foraging movements of a fish shoal? *Anim. Behav.* **59**, 403–409 (2000).
63. M. Beekman, R. L. Fathke, T. D. Seeley, How does an informed minority of scouts guide a honeybee swarm as it flies to its new home? *Anim. Behav.* **71**, 161–171 (2006).
64. J. R. Dyer, A. Johansson, D. Helbing, I. D. Couzin, J. Krause, Leadership, consensus decision making and collective behaviour in humans. *Philos. Trans. R. Soc. B: Biol. Sci.* **364**, 781–789 (2008).
65. M. Porfiri, N. Abaid, S. Garnier, Socially driven negative feedback regulates activity and energy use in ant colonies. *PLoS Comput. Biol.* **20**, e1012623 (2024).
66. M. Napoli, R. Sipahi, M. Porfiri, The role of colony size on activity rhythms of ants. *J. Theor. Biol.* **617**, 112286 (2026).
67. W. Liu, A. F. T. Winfield, Modelling and optimization of adaptive foraging in swarm robotic systems. *Int. J. Rob. Res.* **29**, 1743–1760 (2010).
68. A. Martinoli, K. Easton, W. Agassounon, Modeling swarm robotic systems: A case study in collaborative distributed manipulation. *Int. J. Rob. Res.* **23**, 415–436 (2004).
69. M. S. Talamali *et al.*, Sophisticated collective foraging with minimalist agents: A swarm robotics test. *Swarm Intell.* **14**, 25–56 (2020).
70. T. A. O'Shea-Wheller, N. Masuda, A. B. Sendova-Franks, N. R. Franks, Variability in individual assessment behaviour and its implications for collective decision-making. *Proc. R. Soc. B: Biol. Sci.* **284**, 20162237 (2017).
71. Y. Ulrich *et al.*, Response thresholds alone cannot explain empirical patterns of division of labor in social insects. *PLoS Biol.* **19**, e3001269 (2021).
72. R. P. Mann, Collective decision-making by rational agents with differing preferences. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10388–10396 (2020).
73. H. Suganuma, K. Katahira, H. Ohtsuki, T. Kameda, How social learning enhances—Or undermines—Efficiency and flexibility in collective decision-making under uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2516827122 (2025).
74. H. Kim, Z. P. Kilpatrick, K. Josić, Code for "Daring few, patient many: Division of labor in decentralized foraging collectives." GitHub. <https://github.com/hkimmathbio/division-of-labor>. Deposited 14 April 2026.



Supporting Information for

Daring few, patient many: division of labor in decentralized foraging collectives

Hyunjoong Kim, Zachary P. Kilpatrick, Krešimir Josić

Correspondence to: hyunjoong.kim@uc.edu (H.K.), zpkilpat@colorado.edu (Z.P.K.), josic@math.uh.edu (K.J.)

This PDF file includes:

Supporting text
Figs. S1 to S7
SI References

Supporting Information Text

1. Detailed model description

Collective foraging in a changing environment. We consider a group of N agents with a single central home, which is situated at a distance from a single foraging site. The placement of the home site is motivated by dwellings of socially foraging animal groups who bring food back to a central location (e.g., honeybees, ants, and many other social insects), serving as a base from which information and resources are collected. At each time $t = 1, 2, \dots$, each agent chooses to either idle or commit. Individuals who choose to idle remain at the home site, avoiding the expense and danger of exploration, but also foregoing potential rewards. In contrast, agents who choose to commit to the foraging site will leave the home site during a time step and may receive a reward but incur some cost for exploring, λ . Each agent can choose whether to commit or idle at each step. Rewards are resources that benefit the collective, such as food that is immediately consumed or nectar that is brought back to the home site. The cost of exploring includes both expended energy and risks of foraging such as predation.

When visiting the site, agents receive a reward with a probability determined by the hidden state of the environment at each time t . When the environment is in the high- (low-)rewarding state $s^t = s_+$ ($s^t = s_-$), a single forager receives a reward with probability γ_+ (γ_-),

$$\Pr(x^t = 1 | s^t = s_{\pm}, a^t = 1) = \gamma_{\pm}, \quad \gamma_- < \gamma_+,$$

where $x^t = 1$ if a reward is received in the special case $a^t = 1$, and $x^t = 0$ otherwise. Agents that do not visit the site do not obtain rewards.

The environmental state changes probabilistically as a two-state Markov chain with symmetric switching (hazard) rates

$$\Pr(s^{t+1} \neq s^t | s^t) = \epsilon,$$

where $\epsilon \in [0, 1/2]$. Our results can be extended to the case in which transitions out of either state occur at different rates, $\Pr(s^{t+1} = s_{\pm} | s^t = s_{\mp}) = \epsilon_{\pm}$, leading to unequal average dwell times (rather than the symmetric case considered here), but this substantially complicates the resulting expressions.

Collective perception of environment. We assume agents combine their own observations with social information shared across the group to compute the posterior probability that the environment is in the high rewarding state (1). We also assume that all agents share their observations with all others once home, at which point everyone in the group has the same information and the same *belief* about the environmental state. For example, in a sufficiently small colony of honeybees, all members can observe each other's waggle dances, which communicate information about foraging sites (2–4). We derive a recursive expression for the probability that the environment is one of the two states given the observations of the collective, corresponding to the belief shared by all agents.

An agent infers the state, s^t , using the outcome of all foraging attempts made by other agents, and knowledge about the volatility of the environment, ϵ . Let a^t be the number of agents choosing to forage at time t . Among these agents, x^t agents successfully receive rewards by finding food at the foraging site. The probability that x^t out of a^t agents receive rewards if the state is s^t is given by

$$\Pr(x^t | s^t = s_{\pm}, a^t) = \mathcal{B}(x^t | a^t, \gamma_{\pm}) \equiv \binom{a^t}{x^t} \gamma_{\pm}^{x^t} (1 - \gamma_{\pm})^{a^t - x^t}.$$

We will discuss later how agents ideally choose to take specific actions based on their belief in Sect. 1, focusing on how the number a^t of foragers at time step t is chosen based on prior reward $x^{1:t-1} = \{x^1, \dots, x^{t-1}\}$ and foraging $a^{1:t-1} = \{a^1, \dots, a^{t-1}\}$ outcomes. The best number a^t of foragers to send out is in turn determined by the subjective probability (belief) that the current state is high-rewarding, $s^t = s_+$, based on these previous observations. We denote this belief by

$$g^t = \Pr(s^t = s_+ | x^{1:t-1}, a^{1:t-1}).$$

Once a^t agents commit, $x^t \in \{0, 1, \dots, a^t\}$ is sampled from a binomial distribution with parameters determined by the present state s^t . Subsequently all agents *update* their belief based on x^t and a^t . We can thus compute this update iteratively by accounting for this conditional dependence of x^t and the remaining prior contributions to the belief:

$$\begin{aligned} \Pr(s^t | a^{1:t}, x^{1:t}) &\propto \Pr(x^t | s^t, a^t, a^{1:t-1}, x^{1:t-1}) \Pr(s^t, a^t | x^t, a^{1:t-1}, x^{1:t-1}) \\ &\propto \Pr(x^t | s^t, a^t) \Pr(s^t | a^{1:t-1}, x^{1:t-1}). \end{aligned}$$

Since there are only two potential values for the state, $s_{\pm}$, conservation of probability ensures

$$\begin{aligned} \Pr(s^t = s_+ | a^{1:t}, x^{1:t}) &= \frac{\mathcal{B}(x^t | a^t, \gamma_+) g^t}{\mathcal{B}(x^t | a^t, \gamma_+) g^t + \mathcal{B}(x^t | a^t, \gamma_-) (1 - g^t)} \\ &= \frac{g^t}{g^t + (1 - g^t) \mathcal{C}(x^t, a^t)} \equiv \mathcal{U}(x^t | g^t, a^t), \end{aligned}$$

where $\mathcal{C}(x, a) = (\gamma_-/\gamma_+)^x((1-\gamma_-)/(1-\gamma_+))^{a-x}$ is the ratio of the conditional binomial probabilities in either state. After this update, each agent *predicts* the next state, s^{t+1} , from the obtained posterior probability (belief) of the current state. The Chapman-Kolmogorov equation for the state yields

$$\begin{aligned} g^{t+1} &= \Pr(s^{t+1}|a^{1:t}, x^{1:t}) = \sum_{s^t} \Pr(s^{t+1}|s^t, a^{1:t}, x^{1:t}) \Pr(s^t|a^{1:t}, x^{1:t}) \\ &= \sum_{s^t} \Pr(s^{t+1}|s^t) \Pr(s^t|a^{1:t}, x^{1:t}), \end{aligned}$$

according to the assumption that s^{t+1} only depends on s^t . It follows that

$$g^{t+1} = (1 - \epsilon)\mathcal{U}(x^t|g^t, a^t) + \epsilon(1 - \mathcal{U}(x^t|g^t, a^t)) \equiv \mathcal{G}(x^t|g^t, a^t). \quad [1]$$

We introduced a Bayesian model of collective perception in a changing environment. In this framework, both the agents' returns and their beliefs are shaped by their actions, a^t . This raises a central question: given a belief, g^t , what is the optimal number of agents that should forage at each time step? In Sect. 1, we address this question by using dynamic programming to find the optimal policy of collective agents to maximize their return, and discuss how individuals without a coordinator optimize their decisions as a collective.

Normative foraging strategy. The allocation of agents impacts the immediate reward, and determines the information gathered (Section 1), and hence future rewards received by the group. The *group action* at time t is the number of foraging agents, denoted as a^t . Since the group size is constant, the number of idlers is $N - a^t$. Therefore, the group action (deploy a^t foragers) is chosen from the action space:

$$a^t \in \mathcal{A} = \{0, 1, \dots, N\}.$$

Let x^t represent the number of agents who receive a reward at time t and let g denote the probability that $s^t = s_+$ (belief). The expected immediate reward (per unit time), $\bar{r}^t$, for a given action is then expressed as:

$$\begin{aligned} \bar{r}^t(g^t|a) &= \mathbb{E}[r^t|a^t = a] \\ &= \mathbb{E}[x^t - \lambda a^t|a^t = a] \\ &= a(\gamma_+g^t + \gamma_-(1-g^t) - \lambda) \equiv a\bar{r}_1(g^t), \end{aligned}$$

where λ is the commitment cost a single agent incurs during the trip, and $\bar{r}_1(g)$ is the expected reward for a single agent given belief g . We assume that idling incurs no cost for simplicity, but could compute idling cost effects by shifting $\bar{r}^t$ and scaling λ appropriately. From the viewpoint of a single agent, a *greedy* strategy is one that always selects the action that gives the best immediate reward. Let θ_c be the *critical threshold* where $\bar{r}_1(\theta_c) = 0$, which is given by

$$\theta_c = \frac{\lambda - \gamma_-}{\gamma_+ - \gamma_-}.$$

The greedy agent chooses to forage if $g > \theta_c$, and otherwise chooses to idle.

In addition to potentially receiving a reward, the presence or absence of the reward also provides noisy evidence about the state of the environment. The anticipated value of the information obtained from returning agents should also be taken into account when computing the value of an action. We do so by obtaining the probability of a future belief, g^{t+1} , resulting from sending out a^t agents, given that the current belief is g^t . Given g^t and a^t , there are at most $a^t + 1$ possible values for g^{t+1} since $x^t \in \{0, 1, \dots, a^t\}$. At time t , if $a^t = a$ agents have gone out to forage and $x^t = x$ of these agents receive rewards, Eq. (1) gives the belief transition from $g^t = g$ to $g^{t+1} = g'$ as follows:

$$p(g'|g, a) = \mathcal{B}(x|a, \gamma_+)g + \mathcal{B}(x|a, \gamma_-)(1-g), \quad g' = \mathcal{G}(x|g, a). \quad [2]$$

Hence, the belief transition probability is determined completely by the probability of receiving x rewards when sending out a agents, given the current belief g .

We next determine the optimal foraging strategy, $\pi : [0, 1] \rightarrow \mathcal{A}$ determining how many agents should be deployed given a belief, $g^t = g \in [0, 1]$? To measure foraging efficiency, we introduce the *average reward* (5, 6). Denote by $r^{t\pi}$ the collective reward at time t , given that the planner uses policy π . Then the average reward in the long time limit is given by

$$\rho^\pi = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{t=1}^T r^{t\pi} \right],$$

which is independent of the initial action if the underlying Markov decision process is ergodic (6). This average reward satisfies the Bellman equation,

$$V^\pi(g) + \rho^\pi = r(g|\pi(g)) + \mathcal{P}V^\pi(g|\pi(g)), \quad \mathcal{P}V(g|a) = \sum_{g'} V(g')p(g'|g, a), \quad [3]$$

which incorporates the adjusted value function

$$V^\pi(g) = \mathbb{E} \left[\sum_{s=t}^{\infty} r^{s\pi} - \rho^\pi \middle| g^t = g \right].$$

On the right-hand side of Eq. (3), the first term represents the expected immediate reward and the second term accounts for the expected future rewards if $a = \pi(g)$ agents are sent out to forage under policy π . The probability $p(g'|g, a)$ is defined by Eq. (2), and the sum is over the $N + 1$ possible outcomes of the belief, g' , at the next timestep. Using dynamic programming, we can determine the optimal number of agents that should commit to the environment given the current belief g , denoted by the policy $\pi(g)$, which satisfies Bellman's optimality equation:

$$V(g) + \rho = \max_{a=0,1,\dots,N} \{ r(g|a) + \mathcal{P}V(g|a) \}. \quad [4]$$

2. Asymptotic analysis of Bellman equation

Two-step look-ahead approximation. We now consider the two-step look-ahead model, whose value function is given by

$$V_{\text{two-step}}(g|a) = ar(g) + \mathcal{P}V_{\text{one-step}}(g|a), \quad V_{\text{one-step}}(g) = Nr(g)H(g - \theta_c), \quad [5]$$

where $H(g)$ denotes the Heaviside function, equal to 1 if $g \geq 0$, and 0 otherwise. Here the first term represents the immediate reward, and the second term represents the expected return by looking two-step ahead. The one-step-ahead value function is determined by the greedy strategy. The optimal composition in the two-step-ahead model satisfies

$$\pi_{\text{two-step}}(g) = \arg \max_{a=0,1,\dots,N} V_{\text{two-step}}(g|a).$$

We next determine the first two leading-order terms of the second term in the value function, $\mathcal{P}V_{\text{one-step}}(g|a)$ by expanding as

$$\begin{aligned} \frac{1}{N} \mathcal{P}V_{\text{one-step}}(g|a) &= \sum_{\mathcal{G}(x|g,a) \geq \theta_c} (\mathcal{G}(x|g,a) - \theta_c) p(x|g,a) \\ &= (1 - \epsilon - \theta_c)g \sum_{\mathcal{G}(x|g,a) \geq \theta_c} \mathcal{B}(x|a, \gamma_+) - (\theta_c - \epsilon)(1 - g) \sum_{\mathcal{G}(x|g,a) \geq \theta_c} \mathcal{B}(x|a, \gamma_-) \\ &:= (1 - \epsilon - \theta_c)g(1 - \mathcal{R}_+(a|g)) - (\theta_c - \epsilon)(1 - g)\mathcal{R}_-(a|g), \end{aligned}$$

based on the identity

$$\mathcal{G}(x|g,a)p(x|g,a) = (1 - \epsilon)g\mathcal{B}(x|a, \gamma_+) + \epsilon(1 - g)\mathcal{B}(x|a, \gamma_-).$$

We analyze the remainders $\mathcal{R}_\pm$ using large deviation theory (7, 8). For simplicity, we again assume $\gamma_+ + \gamma_- = 1$ and denote $\gamma_+ = \gamma$. Let $X_{a,\gamma}$ denote a binomial random variable with size a and probability γ . We focus on the non-trivial regime $\epsilon < \theta_c < 1 - \epsilon$, so that

$$\begin{aligned} \mathcal{R}_+(a|g) &= \Pr(\mathcal{G}(X_{a,\gamma_+}|g,a) < \theta_c) \\ &= \Pr\left(X_{a,\gamma_+} < \frac{a}{2} + z(g)\right), \quad z(g) = \frac{\log \frac{\theta_c - \epsilon}{1 - \epsilon - \theta_c} - \log \frac{g}{1 - g}}{\log \frac{\gamma}{1 - \gamma}}, \\ &= \Pr\left(X_{a,\gamma_-} > \frac{a}{2} - z(g)\right). \end{aligned}$$

Similarly,

$$\mathcal{R}_-(a|g) = \Pr\left(X_{a,\gamma_-} \geq \frac{a}{2} + z(g)\right).$$

By Cramer's theorem, the binomial random variable satisfies the large deviation principle:

$$\lim_{a \rightarrow \infty} \frac{1}{a} \ln \Pr(X_{a,\gamma} > \xi a) = -\mathcal{I}(\xi|\gamma), \quad \mathcal{I}(\xi|\gamma) = \xi \ln \frac{\xi}{\gamma} + (1 - \xi) \ln \frac{1 - \xi}{1 - \gamma}, \quad [6]$$

for $\xi > \gamma$, where $\mathcal{I} > 0$ is the rate function for the binomial random variable. For sufficiently large a and any $0 < \zeta < \frac{1}{2} - (1 - \gamma) = \gamma - 1/2$, the following inequality holds

$$\left(\frac{1}{2} - \zeta\right)a < \frac{a}{2} - |z(g)| \leq \frac{a}{2} + |z(g)| < \left(\frac{1}{2} + \zeta\right)a$$

which implies that

$$\Pr\left(X_{a,1-\gamma} > \left(\frac{1}{2} + \zeta\right)a\right) \leq \mathcal{R}_\pm(a|g) \leq \Pr\left(X_{a,1-\gamma} > \left(\frac{1}{2} - \zeta\right)a\right).$$

Applying the binomial rate function gives

$$-\mathcal{I}\left(\frac{1}{2} - \zeta \middle| 1 - \gamma\right) \leq \lim_{a \rightarrow \infty} \frac{1}{a} \ln \mathcal{R}_{\pm}(a|g) \leq -\mathcal{I}\left(\frac{1}{2} + \zeta \middle| 1 - \gamma\right).$$

Since $\zeta > 0$ is arbitrary, it follows that

$$\lim_{a \rightarrow \infty} \frac{1}{a} \ln \mathcal{R}_{\pm}(a|g) = -\mathcal{I}_0 := -\mathcal{I}\left(\frac{1}{2} \middle| 1 - \gamma\right).$$

Thus, the operator $\mathcal{P}$ admits the asymptotic form:

$$\frac{1}{N} \mathcal{P} V_{\text{one-step}}(g|a) \sim \mathcal{B}_0(g) - c\mathcal{B}_1(g)e^{-a\mathcal{I}_0}, \quad [7]$$

for sufficiently large a , where c is a positive constant and

$$\mathcal{B}_0(g) = g(1 - \epsilon - \theta_c) - (1 - g)(\theta_c - \epsilon), \quad \mathcal{B}_1(g) = g(1 - \epsilon - \theta_c) + (1 - g)(\theta_c - \epsilon).$$

Note that $\mathcal{B}_1 > 0$ for any $g \in (\epsilon, 1 - \epsilon)$.

We finally optimize the value function in Eq. (5) by substituting the expansion from Eq. (7):

$$V_{\text{two-step}}(g|a) = ar(g) + N(\mathcal{B}_0(g) - c\mathcal{B}_1(g)e^{-a\mathcal{I}_0}).$$

If $g \geq \theta_c$, the value function $V_{\text{two-step}}(g|a)$ increases with a , and is therefore maximized at $a = N$. Otherwise, if $g < \theta_c$, there exists a value $a < N$ satisfying the critical condition

$$r(g) + cN\mathcal{I}_0\mathcal{B}_1(g)e^{-a\mathcal{I}_0} = 0.$$

Solving for a gives

$$a = \frac{\log N}{\mathcal{I}_0} + \log \mathcal{I}_0 + \log \frac{c\mathcal{A}_1(g)}{\theta_c - g} + \dots,$$

Therefore, the optimal composition in the two-step-ahead model is

$$\pi(g) \sim \begin{cases} \mathcal{O}(\log N) & , g < \theta_c \\ N & , g \geq \theta_c \end{cases}.$$

for sufficiently large N . Put simply, nearly all agents should refrain from foraging unless the belief $g \geq \theta_c$ and a small, logarithmically scaled minority should forage at lower beliefs.

First-order optimal policy. The optimal composition of individual strategies satisfies Bellman's equation given in Eq. (3):

$$V(g) + \rho = \max_{a=0,1,\dots,N} \{ar(g) + \mathcal{P}V(g|a)\},$$

where $\mathcal{P}$ denotes the operator for the expected future value,

$$\mathcal{P}V(g|a) = \sum_{x=0}^a (V \circ \mathcal{G})(x|g, a)p(x|g, a).$$

Here $\mathcal{G}(x|g, a)$ denotes the updated belief given current belief g , the number of committed foragers a , and the number of successful foragers x . The term $p(x|g, a)$ represents the corresponding probability. For simplicity, we assume $\gamma_+ + \gamma_- = 1$ and denote $\gamma_+ = \gamma$. Taking $a \rightarrow \infty$, we have

$$\begin{aligned} \lim_{a \rightarrow \infty} \mathcal{G}(x|g, a) &= \lim_{a \rightarrow \infty} \frac{(1 - \epsilon)g + \epsilon(1 - g)C_{x,a}}{g + (1 - g)C_{x,a}}, \quad C_{x,a} = \left(\frac{\gamma}{1 - \gamma}\right)^{a-2x}, \\ &\rightarrow \begin{cases} \epsilon & , x < a/2 \\ (1 - \epsilon)g + \epsilon(1 - g) & , x = a/2 \\ 1 - \epsilon & , x > a/2 \end{cases}. \end{aligned} \quad [8]$$

We can also compute the probability limit

$$\lim_{a \rightarrow \infty} \sum_{x < a/2} p(x|g, a) = \lim_{a \rightarrow \infty} g \sum_{x < a/2} \mathcal{B}(x|a, \gamma_+) + (1 - g) \sum_{x < a/2} \mathcal{B}(x|a, \gamma_-) = 1 - g.$$

Similarly,

$$\lim_{a \rightarrow \infty} \sum_{x > a/2} p(x|g, a) = g.$$

Using these limits, the operator converges to

$$\begin{aligned} \lim_{a \rightarrow \infty} \mathcal{P}V(g|a) &= \lim_{a \rightarrow \infty} V(\epsilon) \sum_{x < a/2} p(x|g, a) + V(1 - \epsilon) \sum_{x > a/2} p(x|g, a) \\ &= gV(1 - \epsilon) + (1 - g)V(\epsilon) := \mathcal{P}_0V(g). \end{aligned}$$

Introducing the expansion

$$a = \pi(g) = N\pi_0(g) + \dots, \quad \pi_0(g) \in [0, 1],$$

we find that the leading-order behavior of the operator is given by

$$\mathcal{P}V(g|a) = \mathcal{P}_0V(g) + \dots$$

We further introduce the expansions

$$V(g|a) = NV_0(g|a) + \dots, \quad \rho = N\rho_0 + \dots$$

Substituting those expansions into Bellman's equation yields

$$NV_0(g|a) + N\rho_0 + \dots = \max \{N\pi_0(g)r(g) + \mathcal{P}_0V_0(g) + \dots\}.$$

Collecting $\mathcal{O}(N)$ terms gives the leading-order equation

$$\begin{aligned} V_0(g|a) + \rho_0 &= \max \{ \pi_0(g)r(g) + \mathcal{P}_0V_0(g) \} \\ &= \max \{ \pi_0(g)r(g) \} + \mathcal{P}_0V_0(g), \end{aligned}$$

since $\mathcal{P}_0$ is independent of a . Since $r(g) \geq 0$ when $g \geq \theta_c$, the right-hand side is maximized at $\pi_0(g) = 1$. Similarly, it is maximized at $\pi_0(g) = 0$ if $g < \theta_c$. Thus, the leading-order optimal proportion of committed foragers is given by

$$\pi_0(g) = H(g - \theta_c),$$

where $H(g)$ is the Heaviside function, equal to 1 if $g \geq 0$ and 0 otherwise. It follows that

$$\pi(g) = NH(g - \theta_c) + \dots,$$

which yields $\pi(g) \sim NH(g - \theta_c)$, therefore, all individuals should commit when $g \geq \theta_c$.

Upper bound on the optimal policy when $g < \theta_c$. We now establish that the optimal policy $\pi(g) \leq \Pi_N(g) = \mathcal{O}(\ln N)$ when $g < \theta_c$. The proof proceeds in four steps: (i) establishing pointwise convergence of the posterior update, (ii) bounding the difference between the finite and limiting operators, (iii) showing the relevant constants scale linearly in N , and (iv) deriving the upper bound on the optimal action.

Pointwise convergence of the posterior update. Define the rescaled posterior update $\tilde{\mathcal{G}}(\xi|g, a) = \mathcal{G}(\xi a|g, a)$ for $\xi \in [0, 1]$. The pointwise limit as $a \rightarrow \infty$ is given by

$$\tilde{\mathcal{G}}_\infty(\xi|g) := \lim_{a \rightarrow \infty} \tilde{\mathcal{G}}(\xi|g, a) = \begin{cases} \epsilon & , \xi < 1/2, \\ (1 - \epsilon)g + \epsilon(1 - g) & , \xi = 1/2, \\ 1 - \epsilon & , \xi > 1/2. \end{cases} \quad [9]$$

We now quantify the rate of convergence. For $\xi < 1/2$, we have

$$\tilde{\mathcal{G}}(\xi|g, a) - \tilde{\mathcal{G}}_\infty(\xi|g) = (1 - 2\epsilon)\tilde{\mathcal{U}}(\xi|g, a),$$

where $\tilde{\mathcal{U}}(\xi|g, a) = \mathcal{U}(\xi a|g, a)$ satisfies

$$\tilde{\mathcal{U}}(\xi|g, a) = \frac{g\mathcal{Q}_\xi^{-a}}{g\mathcal{Q}_\xi^{-a} + (1 - g)}, \quad \mathcal{Q}_\xi := \left(\frac{\gamma}{1 - \gamma} \right)^{|2\xi - 1|}.$$

Here we assume $\gamma > 1/2$, so that $\mathcal{Q}_\xi > 1$ for $\xi \neq 1/2$. Since $g \leq \theta_c < 1$, we obtain the bound

$$\tilde{\mathcal{U}}(\xi|g, a) \leq \frac{g}{1 - g} \mathcal{Q}_\xi^{-a} \leq \frac{\theta_c}{1 - \theta_c} \exp(-a \ln \mathcal{Q}_\xi).$$

Therefore,

$$|\tilde{\mathcal{G}}(\xi|g, a) - \tilde{\mathcal{G}}_\infty(\xi|g)| \leq Z \exp(-a \ln \mathcal{Q}_\xi), \quad Z := \frac{(1 - 2\epsilon)\theta_c}{1 - \theta_c}. \quad [10]$$

A symmetric argument establishes the same bound for $\xi > 1/2$. Note that the decay rate $\ln \mathcal{Q}_\xi \rightarrow 0$ as $\xi \rightarrow 1/2$, reflecting slower convergence near $x = a/2$.

Asymptotic analysis of the operator $\mathcal{P}$. Define the limiting operator $\mathcal{P}_\infty V(g) := \lim_{a \rightarrow \infty} \mathcal{P}V(g|a)$ and the difference

$$\Delta \mathcal{P}V(g|a) := \mathcal{P}V(g|a) - \mathcal{P}_\infty V(g).$$

Fix a parameter $\zeta \in (0, \gamma - 1/2)$ and introduce the set

$$\Lambda(a, \zeta) := \left\{ x \in \{0, 1, \dots, a\} \mid \left| \frac{x}{a} - \frac{1}{2} \right| \leq \zeta \right\},$$

which captures observations near the boundary $x/a = 1/2$. We decompose the difference as

$$\begin{aligned} \Delta \mathcal{P}V(g|a) &= \sum_{x=0}^a \left[(V \circ \tilde{\mathcal{G}}) \left(\frac{x}{a} \middle| g, a \right) - (V \circ \tilde{\mathcal{G}}_\infty) \left(\frac{x}{a} \middle| g \right) \right] p(x|g, a) \\ &= \Sigma_1 + \Sigma_2, \end{aligned}$$

where Σ_1 sums over $x \in \Lambda(a, \zeta)$ and Σ_2 sums over $x \notin \Lambda(a, \zeta)$.

The distribution $p(x|g, a)$ is a mixture of two binomials:

$$p(x|g, a) = g B(x|a, \gamma) + (1 - g) B(x|a, 1 - \gamma),$$

where $B(x|a, p) = \binom{a}{x} p^x (1 - p)^{a-x}$. Since $\gamma > 1/2$, the first component $B(x|a, \gamma)$ has mean $\gamma a > a/2$, while the second component $B(x|a, 1 - \gamma)$ has mean $(1 - \gamma)a < a/2$. Both components concentrate away from $x = a/2$. By Cramér's theorem applied to each component separately:

$$\begin{aligned} \sum_{x \in \Lambda(a, \zeta)} B(x|a, \gamma) &\leq B_1 \exp \left(-a \mathcal{I} \left(\frac{1}{2} - \zeta \middle| 1 - \gamma \right) \right), \\ \sum_{x \in \Lambda(a, \zeta)} B(x|a, 1 - \gamma) &\leq B_2 \exp \left(-a \mathcal{I} \left(\frac{1}{2} + \zeta \middle| \gamma \right) \right), \end{aligned}$$

where $\mathcal{I}(y|p) = y \ln \frac{y}{p} + (1 - y) \ln \frac{1-y}{1-p}$ is the Kullback-Leibler divergence (rate function for binomial large deviations), and B_1, B_2 are constants. Define

$$I_0 := \min \left\{ \mathcal{I} \left(\frac{1}{2} - \zeta \middle| 1 - \gamma \right), \mathcal{I} \left(\frac{1}{2} + \zeta \middle| \gamma \right) \right\} > 0.$$

Note that $I_0 > 0$ according to Eq. (6). Therefore,

$$\begin{aligned} \sum_{x \in \Lambda(a, \zeta)} p(x|g, a) &= g \sum_{x \in \Lambda(a, \zeta)} B(x|a, \gamma) + (1 - g) \sum_{x \in \Lambda(a, \zeta)} B(x|a, 1 - \gamma) \\ &\leq g B_1 \exp(-a I_0) + (1 - g) B_2 \exp(-a I_0) \\ &\leq B \exp(-a I_0), \end{aligned}$$

where $B = \max\{B_1, B_2\}$. Thus,

$$|\Sigma_1| \leq 2 \|V\|_\infty \sum_{x \in \Lambda(a, \zeta)} p(x|g, a) \leq 2B \|V\|_\infty \exp(-a I_0). \quad [11]$$

For $x \notin \Lambda(a, \zeta)$, we have $|x/a - 1/2| > \zeta$, so the bound Eq. (10) applies with $\ln \mathcal{Q}_{x/a} \geq \ln \mathcal{Q}_\zeta > 0$, where

$$\mathcal{Q}_\zeta = \left(\frac{\gamma}{1 - \gamma} \right)^{2\zeta}.$$

Assuming V is Lipschitz continuous with constant $\|V'\|_\infty$, we have

$$\begin{aligned} \left| (V \circ \tilde{\mathcal{G}}) \left(\frac{x}{a} \middle| g, a \right) - (V \circ \tilde{\mathcal{G}}_\infty) \left(\frac{x}{a} \middle| g \right) \right| &\leq \|V'\|_\infty \left| \tilde{\mathcal{G}} \left(\frac{x}{a} \middle| g, a \right) - \tilde{\mathcal{G}}_\infty \left(\frac{x}{a} \middle| g \right) \right| \\ &\leq \|V'\|_\infty \cdot Z \exp(-a \ln \mathcal{Q}_\zeta). \end{aligned}$$

Summing over $x \notin \Lambda(a, \zeta)$ and using $\sum_x p(x|g, a) = 1$,

$$|\Sigma_2| \leq \|V'\|_\infty \cdot Z \exp(-a \ln \mathcal{Q}_\zeta). \quad [12]$$

Combining Eq. (11) and Eq. (12), we obtain

$$|\Delta \mathcal{P}V(g|a)| \leq D(V) \exp(-a \tilde{I}), \quad [13]$$

where

$$\tilde{I} := \min\{I_0, \ln \mathcal{Q}_\zeta\} > 0, \quad D(V) := 2B \|V\|_\infty + Z \|V'\|_\infty.$$

Scaling of $D(V)$ with N . Consider the ergodic Bellman equation

$$V(g) + \rho = \max_{a \in \{0,1,\dots,N\}} \{-a\delta(g) + \mathcal{P}V(g|a)\}, \quad [14]$$

where $\delta(g) := -r(g) > 0$ for $g < \theta_c$. Introduce the scaled quantities

$$U(g) := \frac{V(g)}{N}, \quad \rho' := \frac{\rho}{N}.$$

Then U satisfies

$$U(g) + \rho' = \max_{a \in \{0,1,\dots,N\}} \left\{ -\frac{a\delta(g)}{N} + \mathcal{P}U(g|a) \right\}.$$

The immediate reward is uniformly bounded:

$$\left| \frac{a\delta(g)}{N} \right| \leq \sup_g |\delta(g)| =: C_0 < \infty.$$

By standard results on the span of the bias function in ergodic MDPs (see, e.g., (5)), we have

$$\text{span}(U) := \max_g U(g) - \min_g U(g) \leq \frac{2C_0}{1-\beta},$$

where $\beta < 1$ is a contraction coefficient depending on the mixing properties of the chain. Normalizing U so that $\min_g U(g) = 0$, we obtain $\|U\|_\infty \leq 2C_0/(1-\beta)$. For the derivative bound, differentiating the Bellman equation and using regularity of $\mathcal{P}$ with respect to g , one can show (under appropriate smoothness assumptions on the model) that

$$\|U'\|_\infty \leq C_1,$$

for some constant C_1 independent of N . Rescaling back to V :

$$\|V\|_\infty \leq C_2 N, \quad \|V'\|_\infty \leq C_1 N,$$

where $C_2 = 2C_0/(1-\beta)$. Therefore,

$$D(V) = 2B\|V\|_\infty + Z\|V'\|_\infty \leq (2BC_2 + ZC_1)N =: \tilde{Z}N. \quad [15]$$

Upper bound on the optimal policy. Define the Bellman operator for a fixed action:

$$\mathcal{T}V(g|a) := -a\delta(g) + \mathcal{P}V(g|a).$$

The value difference from increasing the commitment level is

$$\begin{aligned} \mathcal{T}V(g|a+1) - \mathcal{T}V(g|a) &= -\delta(g) + \mathcal{P}V(g|a+1) - \mathcal{P}V(g|a) \\ &= -\delta(g) + (\Delta\mathcal{P}V(g|a+1) - \Delta\mathcal{P}V(g|a)), \end{aligned}$$

where we used $\mathcal{P}_\infty V(g|a+1) = \mathcal{P}_\infty V(g|a) = \mathcal{P}_\infty V(g)$. By the triangle inequality and Eq. (13)–Eq. (15),

$$|\Delta\mathcal{P}V(g|a+1) - \Delta\mathcal{P}V(g|a)| \leq 2D(V) \exp(-a\tilde{I}) \leq 2\tilde{Z}N \exp(-a\tilde{I}).$$

Define the threshold

$$\Pi_N(g) := \frac{1}{\tilde{I}} \left(\ln N + \ln \frac{4\tilde{Z}}{\delta(g)} \right).$$

For any $a > \Pi_N(g)$, we have

$$2\tilde{Z}N \exp(-a\tilde{I}) < \frac{\delta(g)}{2},$$

and therefore

$$\mathcal{T}V(g|a+1) - \mathcal{T}V(g|a) \leq -\delta(g) + \frac{\delta(g)}{2} = -\frac{\delta(g)}{2} < 0.$$

This shows that increasing a beyond $\Pi_N(g)$ strictly decreases the value, so the optimal policy satisfies

$$\pi^*(g) \leq \Pi_N(g) = \frac{\ln N}{\tilde{I}} + \mathcal{O}(1).$$

We conclude that $\pi^*(g) \leq \mathcal{O}(\ln N)$ for all $g < \theta_c$.

3. Probabilistic decision-making in homogeneous collectives

Consider a collective of N individuals who independently follow the same probabilistic strategy $\varphi(g)$, defined as the probability of committing given belief g . The corresponding Bellman equation is given by

$$V(g) + \rho = \sum_{a=0}^N \mathcal{B}(a|N, \varphi(g)) (r(g|a) + \mathcal{P}V(g|a)), \quad [16]$$

where a is the number of committed individuals and $\mathcal{B}(a|n, p)$ is the Binomial distribution with n trials and success probability p . As before, $r(g|a)$ is the expected reward and $\mathcal{P}V(g|a)$ the future expected return given belief g and action (or committed individuals) a . Note that the Bellman optimality equation can be written as

$$V(g) + \rho = \sum_{a=0}^N 1_{a=\pi(g)} (r(g|a) + \mathcal{P}V(g|a)), \quad [17]$$

where $1_{a=\pi(g)}$ represents the indicator function, equal to 1 if and only if $a = \pi(g)$. Even when $\varphi(g) = \pi(g)$, randomness in individual decisions leads to random over- or under-commitment relative to the optimal decision at the group level. However, for the homogeneous collective, the induced binomial distribution satisfies

$$\mathcal{B}(a|N, \pi(g)) \xrightarrow{p} 1_{a=\pi(g)} \quad [18]$$

as $N \rightarrow \infty$, by the law of large numbers. In other words, despite individual-level randomness, the collective behavior concentrates on the optimal group action and becomes asymptotically near-optimal as group size N increases.

References

1. M DeGroot, *Optimal Statistical Decisions*, Wiley Classics Library. (Wiley), (2005).
2. Kv Frisch, *The Dance Language and Orientation of Bees*. (Harvard University Press), (1993).
3. JBS Haldane, H Spurway, A statistical analysis of communication in “apis mellifera” and a comparison with communication in other animals. *Insectes Sociaux* **1**, 247–283 (1954).
4. R Schürch, FLW Ratnieks, The spatial information content of the honey bee waggle dance. *Front. Hum. Neurosci.* **3** (2015).
5. ML Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. (Wiley), (1994).
6. S Mahadevan, Average reward reinforcement learning: Foundations, algorithms, and empirical results. *Mach. Learn.* **22**, 159–195 (1996).
7. A Dembo, O Zeitouni, *Large Deviations Techniques and Applications*. (Springer Berlin Heidelberg), (2010).
8. PC Bressloff, *Probability theory and martingales*. (Springer International Publishing), p. 681–703 (2021).

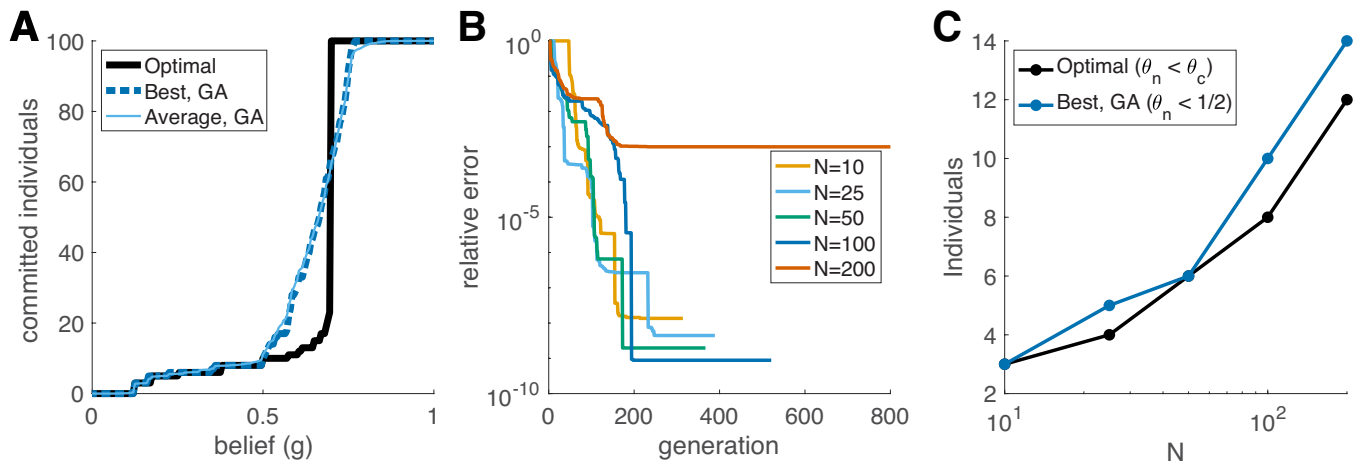


Fig. S1. Natural selection gives rise to the optimal division of labor. A genetic algorithm (GA) is chosen as a model of natural selection. A population of 200 groups of size N is initialized with all individuals greedy ($\theta_n = \theta_c$ for all $n = 1, \dots, N$), and groups are selected and evolved based on long-run average foraging return ρ . The algorithm terminates when no improvement is observed over 100 consecutive generations. **(A)** Collective policy as a function of belief g for $N = 100$, comparing the optimal policy obtained by solving the Bellman equation (black), the best GA solution (blue dashed), and the GA population average (blue solid). The GA recovers the optimal policy for $g < 1/2$ and for large g , but deviates near $g \approx \theta_c$. However, since beliefs near θ_c are visited very rarely, the best GA policy yields effectively the same foraging performance as the optimal policy. **(B)** Relative error $(\rho_{\text{opt}} - \rho_{\text{best}})/\rho_{\text{opt}}$ as a function of generation for $N = 10, 25, 50, 100, 200$. Smaller groups converge to near-optimal solutions in fewer generations, while larger groups require more generations, consistent with a larger policy search space. **(C)** Number of scouts as a function of group size N , comparing the best GA solution (individuals with $\theta < 1/2$, blue) with the optimal collective (individuals with $\theta < \theta_c$, black). The cutoff $\theta < 1/2$ serves as a conservative, model-free proxy for scouts in the GA solution.

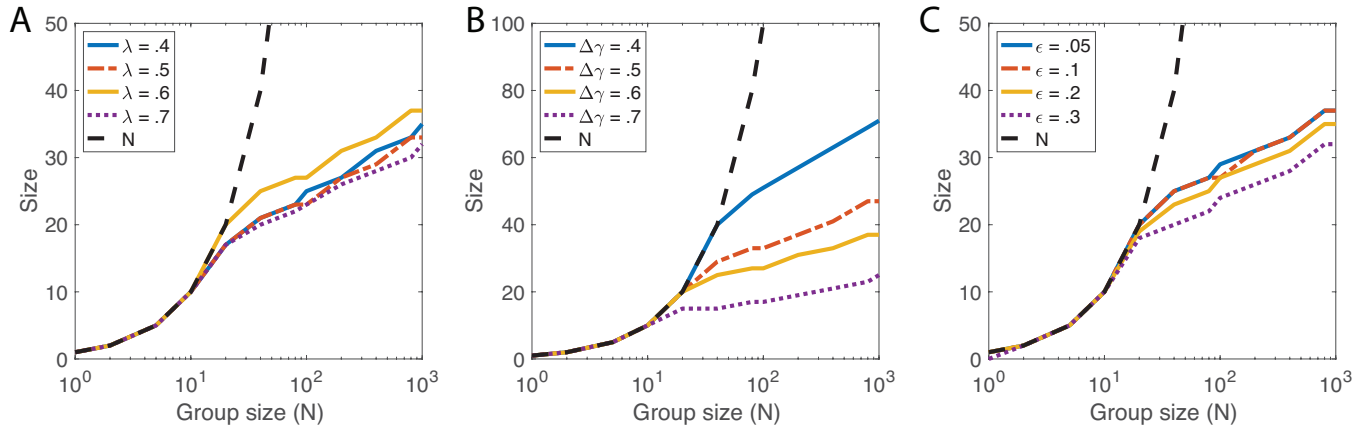


Fig. S2. Sublinear scaling of risk-tolerant individuals across ecological conditions. (A) The number of risk-tolerant individuals as a function of group size N for different commitment costs, λ . (B) Scaling of risk-tolerant individuals with group size under varying reward conditions, $\Delta\gamma = \gamma_+ - \gamma_-$ (difference in high- and low-reward probabilities). (C) Scaling under different environmental switching rates, ϵ . Across all conditions, most individuals are risk-tolerant in small groups, whereas in larger groups the number of risk-takers increases sublinearly and logarithmically with group size.

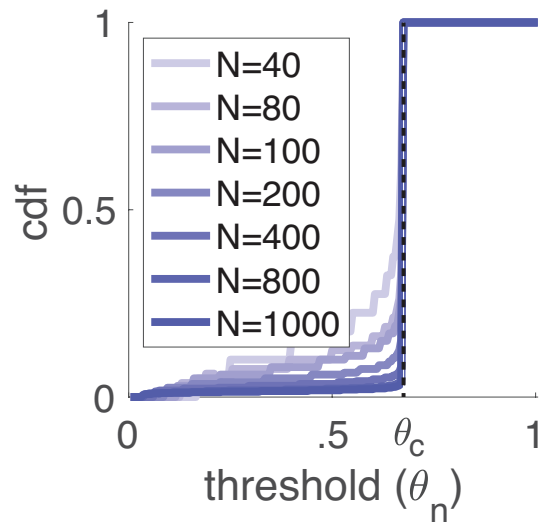


Fig. S3. Large collectives fully commit once expected rewards become positive. The cumulative distribution of optimal thresholds indicates that all individuals commit when $g \geq \theta_c$, i.e., when the expected reward is positive, in sufficiently large collectives.

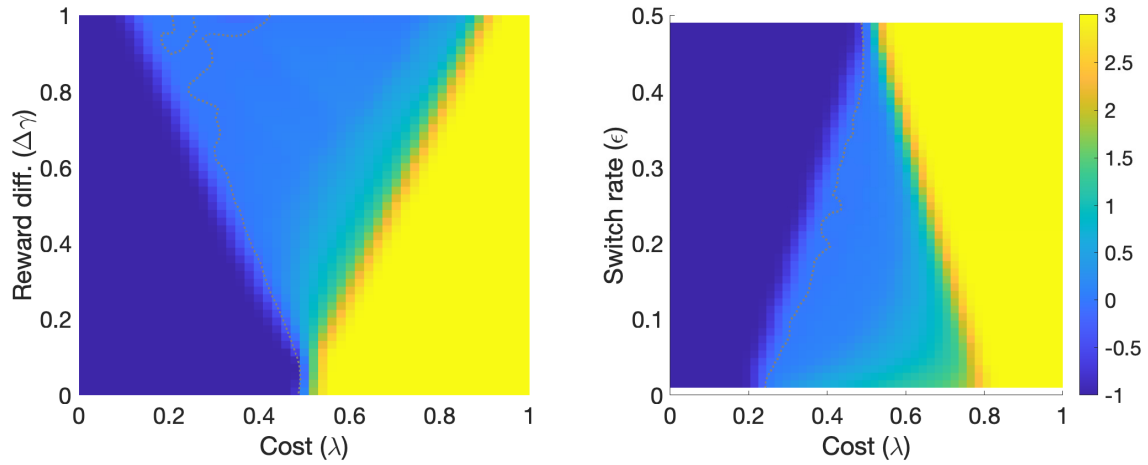


Fig. S4. Asymmetry in collective response times. Collective responses are slower for low-to-high than high-to-low environmental switches. Response times for low-to-high and high-to-low transitions are denoted by $\tau_{-,+}$ and $\tau_{+,-}$, respectively, and are estimated as mean response times from Monte Carlo simulations (with unsuccessful responses counted as inter-switch intervals). Shown is the response-time bias $\beta = \log(\tau_{-,+}/\tau_{+,-})$ across variation in environmental parameters: **(A)** Commitment cost λ vs. reward difference $\Delta\gamma$; **(B)** Commitment cost λ vs. environmental switching rate ϵ . The gray-dashed curve shows the contour $\beta = 0$, corresponding to equal response times for low-to-high and high-to-low environmental switches. At low commitment cost, strong commitment (in low reward state) enables rapid responses to low-to-high switches, whereas at high cost weak commitment impedes timely responses to high-to-low switches.

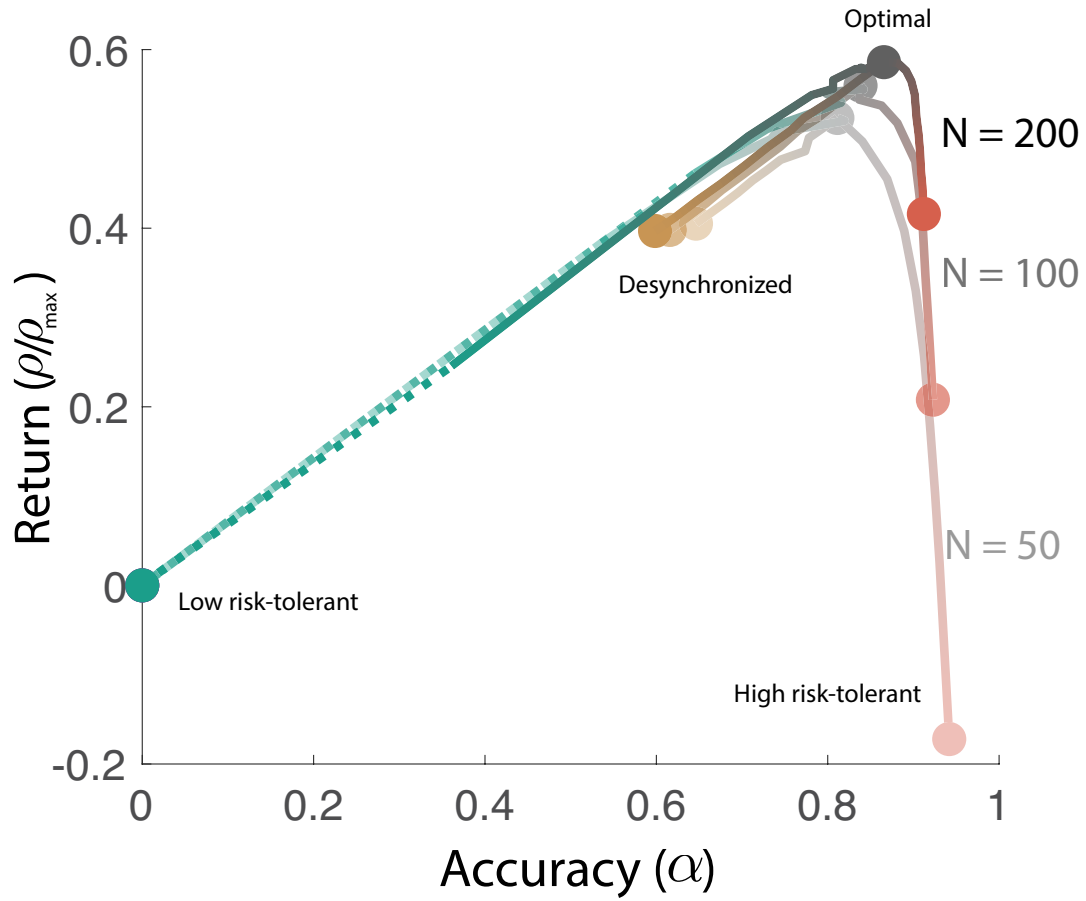


Fig. S5. Group size effect on collective performance under various perturbations to the optimal threshold structure. Optimal performance increases with group size N ($N = 50, 100, 200$, shown by increasing color saturation). As group size increases, the normalized return increases under the same perturbations. Since the optimal fraction of risk-tolerant individuals decreases in larger groups, assigning maximal risk to the risk-tolerant individuals takes a lower relative cost than in small groups. All other parameters are identical to those in Fig. 3.

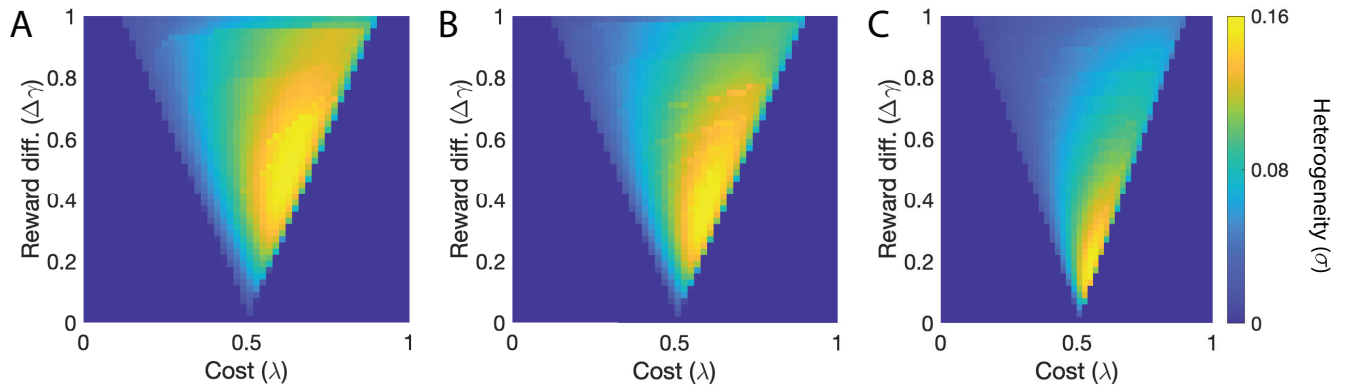


Fig. S6. Group size effect on optimal heterogeneity. Heterogeneity of the optimal threshold structure in decentralized collectives decreases with group size N : **(A)** $N = 50$, **(B)** $N = 100$, **(C)** $N = 200$. All other parameters are identical to those in Fig. 4.

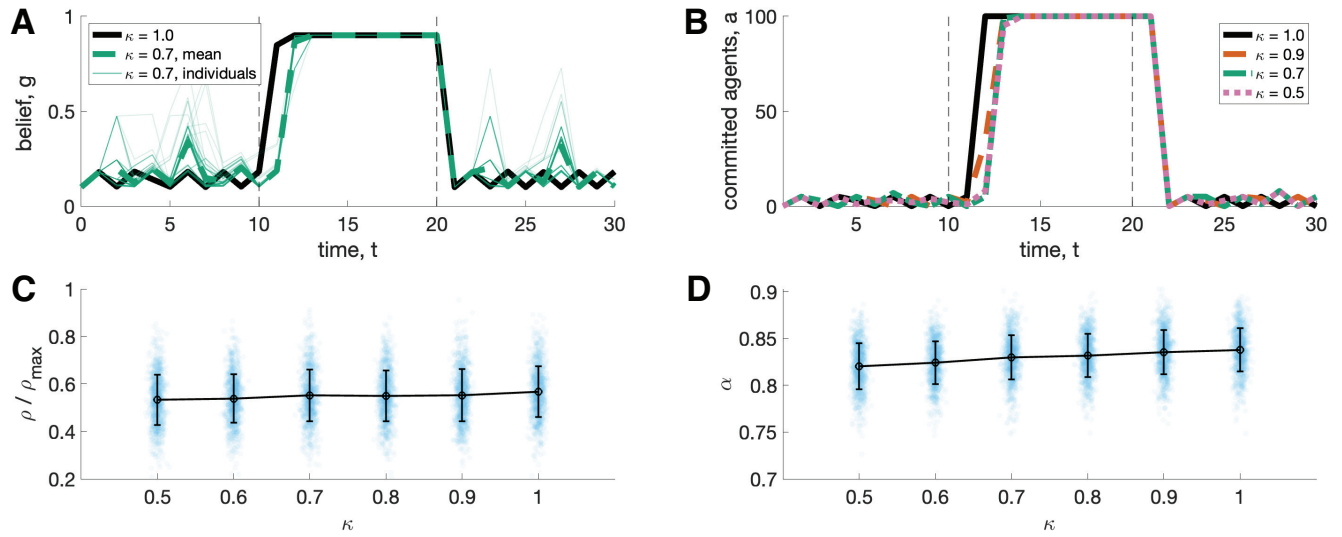


Fig. S7. Effect of partial communication on collective foraging. Instead of sharing a common belief, each individual maintains and updates its own belief as an independent Bayesian observer. Each individual randomly samples the foraging outcomes (success or failure) of committed foragers with probability κ , so that $\kappa = 1$ corresponds to full communication and smaller κ to increasingly sparse information sharing. We assume that individuals know the number of committed foragers. Individuals act according to the threshold distribution optimized under full communication ($\kappa = 1$), which is held fixed as κ varies, so that any difference in collective performance reflects the effect of information quality alone. **(A-B)** Collective foraging dynamics with partial communication. Environmental state switches at $t = 10, 20$ (vertical dashed lines) and is the high-reward state between them. **(A)** Belief trajectories of individual agents (light traces) and population mean (dashed green) under partial communication ($\kappa = 0.7$), compared to the population mean under full communication ($\kappa = 1$, solid black). **(B)** Number of committed foragers over time for various κ . Partial communication delays collective adoption during the favorable environmental phase, while abandonment upon environmental deterioration is largely unaffected. **(C-D)** Monte Carlo estimates of normalized return $\rho / \rho_{\max}$ and decision accuracy α during the high-reward phase as functions of κ . Individual simulation runs (blue dots) and error bars of ± 1 standard deviation are depicted alongside mean values (black circles). Collective foraging performance degrades gracefully as communication becomes sparser, remaining near-optimal across a wide range of κ .