Naked Ants and Honest Bees: Collective Behaviour and the Price of Exploration

Nicholas G. Polson
Booth School of Business
University of Chicago

August 2026

Abstract

We give one stochastic model that reproduces six decision behaviors reported in the recent popular literature on insect cognition [Quaglia, 2026]: pheromone optimisation in ants, quorum consensus in honeybees, neutrality-driven reversal in marching locusts, personality-accelerated aggregation in cockroaches, evidence accumulation in fruit flies, and the thermal defence of the Japanese honeybee against the giant hornet. The model is a density dependent Markov jump process on the commitment simplex; its mean field limit is Maynard Smith's replicator equation and its fluctuation limit is a diffusion whose quasipotential governs consensus reversal. Each behavior is a parameter regime, not a separate mechanism.

A single selection exponent α orders the family. We show that the two path reinforcement urn has an interior fixed point at $s^\star/(1-s^\star) = (\Delta_2/\Delta_1)^{1/(\alpha-1)}$, globally attracting for $\alpha < 1$, repelling for $\alpha > 1$, and absent at $\alpha = 1$; hence cumulative regret is linear in time at both ends and sublinear only at $\alpha = 1$, the Bayes and replicator value. The same α is the inverse temperature of the Kullback–Leibler regularised Bellman equation, the softmax temperature of a reinforcement learner, and, exactly, AlphaZero's move selection temperature. Pushed to its limits it separates optimisers from samplers: Laplace and particle swarms as $\alpha \to \infty$, sequential Monte Carlo along a schedule, nested sampling on level sets. Nested sampling is the honeybee quorum rule run on prior volume and recovers the quasipotential barrier at cost linear rather than exponential in that barrier. A closed loop model of brood nest thermoregulation shows that dispersed response thresholds are not merely helpful but necessary: below a delay margin fixed by (21) a genetically uniform colony oscillates, which is why polyandry is under selection and why it saturates. We then ask why insects appear to decide better than we do and answer with four model consequences, strategic reporting, conformity below the delay margin, absent cross inhibition and an untuned exponent, set against Desmond Morris's argument that the modern human is a captive animal in a mismatched environment. Chess supplies analogues throughout. We close by separating collective computation, where many agents measure a fixed world, from collective construction, where they manufacture a local physics that changes the feasible set: the hornet ball is a nearly pure specimen of the second, and optimal group size is logarithmic in the stakes with prefactor the inverse divergence per observation in the first case and the inverse squared safety margin in the second, which is why quorums run to tens and heat balls to hundreds. We are explicit about what the model does not deliver.

Keywords: collective behaviour; evolutionarily stable strategy; replicator dynamics; quorum sensing; sequential probability ratio test; exploration and exploitation; Kullback–Leibler control; Bellman equation; regret; quasipotential; large deviations; density dependent Markov chains; nested sampling; sequential Monte Carlo; particle swarm optimisation; ant colony optimisation; stigmergy; drift diffusion model; Monte Carlo tree search; social insects.

MSC 2020: 60J28; 60F10; 62L10; 65C05; 91A22; 92D50; 93E20.

1 Introduction

A recent popular account of insect decision making [Quaglia, 2026] presents five lessons from five literatures: ant foraging, honeybee house hunting, locust marching, cockroach sheltering and fruit fly discrimination, each reported as a separate insight with a separate moral. Our claim is that they are one Markov process observed at five parameter values, that a sixth behavior, the hot defensive bee ball, is the exception that tests the frame, and that the logic organising all six is the one Maynard Smith [1982] wrote down.

Section 2 builds the model and Section 3 takes the six behaviors in turn. Section 4 prices exploration through the Bellman equation and shows that one exponent separates optimisers from samplers; Section 5 identifies the resulting algorithms, which are particle swarms, sequential Monte Carlo and nested sampling. Section 6 asks the question the popular account raises and does not answer, namely why insects appear to decide better than we do, and takes Desmond Morris as its foil. Section 7 gives chess analogues, useful because chess is the one domain where the state space, the value function and the search algorithm are all written down, so a claim about search can be checked rather than admired. Sections 8 to 10 report numerics, predictions and limitations, and Section 11 draws the conclusion the hornet forces, which is that collective computation and collective construction are different problems with different arithmetic.

The recurring object is a single scalar. Write $\ell(x)$ for log fitness, or log value, or log likelihood, and define the selection map on measures

$$\mathcal{S}_\alpha[\rho](dx) \;\propto\; e^{\alpha \ell(x)} \rho(dx). \tag{1}$$

At $\alpha = 1$ this is Bayes. As $\alpha \to \infty$ it is the argmax. Everything below is (1) composed with a mutation kernel, and the taxonomy of behaviors, algorithms and playing strengths is the choice of α and of what ℓ is measured on.

2 The colony as a single decision maker

Maynard Smith and Szathmáry [1995] identify eusociality as a major evolutionary transition: once reproduction is concentrated in the queen, fitness accrues at the colony level and the colony, not the worker, is the unit of selection. This licenses the move the rest of the paper depends on. We treat the colony as one agent with one loss function and the individual insect as a component with a three-state memory; nothing else about insect neurobiology is used. The move is not free, and Maynard Smith [1964] was himself sharply critical of loose appeals to group selection. We return to its cost in Section 10.

Let N be the number of active individuals. Each occupies a state in $\{0, A, B\}$, uncommitted, committed to option A, committed to option B. Write n_A, n_B for the counts and

$$x = n_A/N, \quad y = n_B/N, \quad z = 1 - x - y,$$

so $(x, y) \in \Delta = \{x, y \geq 0,\ x + y \leq 1\}$ and z is the neutral mass. Transitions occur at the following rates per individual,

$$0 \to A:\ q_A + r_A x, \qquad A \to 0:\ \lambda_A + \beta y, \tag{2}$$

and symmetrically for B. Here q_i is independent discovery, r_i is social recruitment, λ_i is spontaneous abandonment or signal decay, and β is cross inhibition, the stop signal of Seeley et al. [2012]. The generator of (2) defines a density dependent jump process in the sense of Kurtz [1978]. Its mean field limit is

$$\dot{x} = (1 - x - y)(q_A + r_A x) - x(\lambda_A + \beta y), \qquad \dot{y} = (1 - x - y)(q_B + r_B y) - y(\lambda_B + \beta x). \tag{3}$$

Setting $q_i = 0$ and $\lambda_i = \lambda$ and eliminating the neutral pool by the fast-z approximation reduces (3) to $\dot{x}_i = x_i(f_i - \bar{f})$ with $f_i = r_i - \lambda - \beta \sum_{j \neq i} x_j$. This is the replicator equation, and the attractors of (3) are the evolutionarily stable states of the associated playing-the-field game [Maynard Smith, 1982]. We use this

correspondence throughout: a behavior is adaptive here if and only if it is an attractor of (3) under the fitness assignment f.

Fluctuations matter because N is small. Write the four reactions of (2) with stoichiometric vectors $\nu_j \in \{\pm e_1, \pm e_2\}$ and rates $a_j(x, y)$. Setting $(x, y) = \varphi(t) + N^{-1/2}\xi$ and expanding the master equation gives the Langevin approximation

$$dX_t = b(X_t)\,dt + N^{-1/2}\Sigma(X_t)^{1/2}\,dW_t, \qquad \Sigma(x,y) = \sum_j \nu_j \nu_j^{\top} a_j(x,y), \tag{4}$$

so that Σ is diagonal with entries the *sum*, not the difference, of the forward and backward rates on each coordinate. The drift (3) is not a gradient, so the Freidlin–Wentzell quasipotential Φ is the value function of a control problem and solves the Hamilton–Jacobi equation $\langle b, \nabla\Phi\rangle + \frac{1}{2}\langle \nabla\Phi, \Sigma\nabla\Phi\rangle = 0$ [Freidlin and Wentzell, 1998]. Everything we do with Φ uses only the symmetric reduction $u = x - y$, on which the one dimensional formula $\Phi(u) = -2\int^u b(w)\Sigma_{uu}^{-1}(w)\,dw$ is exact. Section 3.3 turns on Φ alone.

2.1 What is material about the model

The same log likelihood ratio is written in four substrates, and the substrate fixes the constants, not the equations. A pheromone trail is a chemical register, broadcast and persistent for minutes to hours, with decay λ set by evaporation, which no ant controls. A waggle dance is a mechanical register, broadcast but persisting only while the dancer dances. A membrane potential is private, fast, and readable by nobody. A body temperature is a shared analogue register that integrates every participant, cannot be addressed to an individual, and has a time constant fixed by the mass of the ball.

substrate	readable by	decay $1/\lambda$	noise source	what it limits
pheromone	all, later	minutes to hours	evaporation, deposition	discount factor γ
waggle dance	all, now	seconds	sampling of dancers	arrival rate D_2
membrane potential	self only	milliseconds	synaptic	boundary a
ball temperature	all, now	tens of seconds	metabolic dispersion	margin in (19)

This is why the title says material: the decision rules are identical up to α, and what differs across the six behaviors is which physical quantity carries Λ, whose physics supplies the parameters the decision theory takes as given.

3 Six behaviors, one process

3.1 Ants: reinforcement is replicator dynamics

Let path i have length L_i and let foragers travel at speed v, so the round trip takes $T_i = 2L_i/v$. A forager deposits a fixed quantity of pheromone per trip. The rate at which path i accumulates signal is therefore $\rho x_i/T_i$, and the trail evaporates at rate λ. Recruitment is proportional to relative trail strength raised to a sensitivity exponent α, so $r_i \propto \tau_i^{\alpha}/\sum_j \tau_j^{\alpha}$. In the quasi-static limit $\tau_i = \rho v x_i/(2\lambda L_i)$ and (3) becomes

$$\dot{x}_i = x_i\big(f_i - \bar{f}\big), \qquad f_i \propto x_i^{\alpha-1} L_i^{-\alpha}, \tag{5}$$

which at $\alpha = 1$ is the replicator equation with fitness $f_i = 1/L_i$ and unique ESS on the shortest path. Away from $\alpha = 1$ the fitness is frequency dependent and the analysis is cleaner in the urn representation, to which we now turn.

Write Δ_i for the pheromone deposited per unit time on path i, so $\Delta_i \propto 1/L_i$, and let $s = \tau_1/(\tau_1 + \tau_2)$ be the trail share. Each trip adds Δ_i to the chosen path, so with $\mathcal{T} = \tau_1 + \tau_2$,

$$\mathbb{E}[\Delta s \mid s] \;=\; \frac{1}{\mathcal{T}}\Big\{p_1(s)\Delta_1 - s\big[p_1(s)\Delta_1 + (1-p_1(s))\Delta_2\big]\Big\} + O(\mathcal{T}^{-2}), \qquad p_1(s) = \frac{s^{\alpha}}{s^{\alpha} + (1-s)^{\alpha}}. \tag{6}$$

This is a Robbins–Monro [Robbins and Monro, 1951] recursion with step $1/\mathcal{T}$, so s_t shadows the ODE obtained by dropping the prefactor [Hill et al., 1980, Arthur, 1994]. Collecting terms, the drift is

$$g(s) \;=\; (1-s)\,p_1(s)\,\Delta_1 \;-\; s\,\big(1-p_1(s)\big)\Delta_2, \tag{7}$$

so, since $p_1/(1-p_1) = \big(s/(1-s)\big)^{\alpha}$,

$$\operatorname{sign} g(s) \;=\; \operatorname{sign}\Big\{(\alpha-1)\log\tfrac{s}{1-s} - \log r\Big\}, \qquad r = \Delta_2/\Delta_1 < 1. \tag{8}$$

Proposition 1 (Path selection)**.** *Let $\Delta_1 > \Delta_2 > 0$ and $r = \Delta_2/\Delta_1$. The interior rest points of* (6) *are the roots of $(s/(1-s))^{\alpha-1} = r$, so*

$$\frac{s^\star}{1-s^\star} \;=\; r^{1/(\alpha-1)}, \qquad \alpha \neq 1, \tag{9}$$

and there is no interior rest point at $\alpha = 1$. Moreover

- *(i) $\alpha < 1$: $s^\star$ is globally attracting on $(0,1)$. No path fixates and the long run traffic share is $p_1(s^\star) \in (1/2, 1)$.*
- *(ii) $\alpha = 1$: $g(s) = s(1-s)(\Delta_1 - \Delta_2) > 0$ on $(0,1)$, so $s_t \to 1$ almost surely, and $1 - s_t = \Theta(t^{-(1-r)})$.*
- *(iii) $\alpha > 1$: $s^\star$ is repelling and separates the basins of $s = 0$ and $s = 1$. The urn fixates almost surely, on the short path with probability strictly between zero and one, and $s^\star \to 1/2$ as $\alpha \to \infty$.*

Proof. The bracket in (8) is affine and strictly monotone in $\log\{s/(1-s)\}$, which is a bijection from $(0,1)$ to $\mathbb{R}$, so it has exactly one zero when $\alpha \neq 1$, namely (9), and none when $\alpha = 1$ because it equals $-\log r > 0$. For $\alpha < 1$ the bracket is decreasing, so $g > 0$ below $s^\star$ and $g < 0$ above it, giving global attraction; for $\alpha > 1$ it is increasing and the signs reverse, giving repulsion and two basins. In (ii), with $u = 1 - s$ small, $\dot u = -u(\Delta_1 - \Delta_2)/\mathcal{T}$ and $\mathcal{T} \sim \Delta_1 t$, so $\mathrm{d}\log u/\mathrm{d}\log t \to -(1-r)$. In (iii), $1/(\alpha-1) \to 0$ makes $r^{1/(\alpha-1)} \to 1$. □

Part (iii) is the precise statement of premature convergence: the separatrix migrates towards the symmetric starting point as the exponent grows, so a sufficiently greedy colony fixates essentially at random, and quickly. Proposition 2 prices that failure.

Proposition 2 (Regret)**.** *Define regret per trip as Δ_1 minus the deposit actually earned, and $\mathcal{R}_T$ as its sum over T trips. Under the conditions of Proposition 1,*

$$\mathcal{R}_T \;=\; \begin{cases} \big(1-p_1(s^\star)\big)(\Delta_1-\Delta_2)\,T\,\big(1+o(1)\big), & \alpha < 1, \\ \Theta\big(T^{\,r}\big), & \alpha = 1, \\ \mathbb{P}(\textit{wrong fixation})\,(\Delta_1-\Delta_2)\,T\,\big(1+o(1)\big), & \alpha > 1, \end{cases} \tag{10}$$

so regret is linear in T for every $\alpha \neq 1$ and sublinear only at $\alpha = 1$. Both limits $\alpha \to 0$ and $\alpha \to \infty$ give the same rate $\frac{1}{2}(\Delta_1 - \Delta_2)$, the coin flip rate, for different reasons: the first never exploits, the second exploits a coin flip.

Proof. Instantaneous regret is $(1 - p_1(s_t))(\Delta_1 - \Delta_2)$. For $\alpha \neq 1$, s_t converges to a limit at which $1 - p_1$ is bounded away from zero, by Proposition 1(i) and (iii) respectively, so the rate is constant and $\mathcal{R}_T$ is linear. At $\alpha = 1$, $p_1 = s$ and Proposition 1(ii) gives $1 - s_t = \Theta(t^{-(1-r)})$, whence $\mathcal{R}_T \asymp (\Delta_1 - \Delta_2)\int_1^T t^{-(1-r)}\,dt = \Theta(T^{\,r})$. The limits follow from $p_1(s^\star) \to 1/2$ as $\alpha \to 0$ and $s^\star \to 1/2$ as $\alpha \to \infty$ with a symmetric start. □

This makes $\alpha = 1$ more than a coincidence of definitions: it is the unique exponent at which the drift in (6) has constant sign on the whole interval, so there is neither a stable interior point to trap the colony nor a separatrix for noise to push it across. Section 4 shows it is also where the exploration budget is priced at one; Figure 2(b) shows the resulting U.

Ant colony optimisation [Dorigo and Stützle, 2004] is therefore the replicator dynamics of Maynard Smith [1982] run on a landscape defined by inverse travel time, with the sensitivity exponent as the only free parameter, and Section 4 adds a third description of the same object.

Remark 1. No ant knows which path is shortest. Fitness proportional reinforcement requires only that the deposit rate be proportional to return frequency, which is a physical consequence of shorter paths, not a computation. This is the sense in which the colony optimises without an optimiser.

3.2 Bees: the quorum rule is a sequential probability ratio test

Scouts return and dance. Let dance duration D carry site quality θ with density $p(D \mid \theta)$ having monotone likelihood ratio in θ. A bee observing dances $D_1, \dots, D_n$ accumulates

$$\Lambda_n = \Lambda_0 + \sum_{k=1}^{n} \log \frac{p(D_k \mid \theta_A)}{p(D_k \mid \theta_B)}. \tag{11}$$

By monotone likelihood ratio, total dance time for a site is sufficient. If dances are conditionally independent and each committed scout contributes a mean increment $D_1 = \mathrm{KL}$ per dance to Λ, then $\Lambda \simeq n_A D_1$ up to $O_p(\sqrt{n_A})$, so the headcount is a consistent proxy for the log posterior odds and the quorum rule, commit when $n_A \geq Q$, is a threshold rule on Λ at level $\Lambda^\star = QD_1$. This is Wald's test [Wald, 1947], which minimises expected sample size at a given error rate. Optimality is for two options with symmetric priors; Marshall et al. [2009] show the group level rule inherits it and Bogacz et al. [2006] give the conditions under which it fails for more than two.

Proposition 3 (Speed and accuracy)**.** *Let D_1 be the Kullback–Leibler divergence per dance between the two site hypotheses and let D_2 be the rate of dance arrivals. Then $\mathbb{P}(\textit{wrong site}) \asymp e^{-QD_1}$ and $\mathbb{E}[T] \asymp Q/D_2$. If colony fitness is $W(Q) = R\,\mathbb{P}(\textit{correct}) - c\,\mathbb{E}[T]$ then the ESS quorum satisfies*

$$Q^\star = \frac{1}{D_1} \log\!\left(\frac{R\, D_1 D_2}{c} \right), \tag{12}$$

which is positive precisely when $RD_1D_2 > c$, that is when the site is worth deliberating over at all.

Quorum size grows logarithmically in the reward to time-cost ratio and inversely in the discriminability of the sites, so a swarm exposed while deliberating should use a small one. Seeley [2010] reports quorums of twenty to thirty bees, small relative to the scout population, as (12) requires. Section 4.3 shows (12) is the discretisation of a Bellman free boundary, and Section 5.2 shows what happens when the colony must estimate rather than choose.

Honesty of the dance is where Maynard Smith and Harper [2003] is indispensable. A signal is usually stabilised by a cost that separates types, the handicap; but Maynard Smith showed, in the Philip Sidney game [Maynard Smith, 1991] and later in general, that when sender and receiver interests coincide the stable signal is cost free. The sender's optimal report solves $\max_{\hat\theta} u(\hat\theta, \theta) - c(\hat\theta, \theta)$, and alignment of u gives $\hat\theta = \theta$ at $c \equiv 0$. Within a colony relatedness is high and reproductive interest common, so the waggle dance can be an undistorted likelihood ratio; between colonies alignment fails and no cost free equilibrium exists. Honest cost free signalling should be confined to within-colony channels.

3.3 Locusts: neutrality is a noise amplifier

Take the symmetric case $q_A = q_B = q$, $r_A = r_B = r$, $\lambda_A = \lambda_B = \lambda$, with r large enough that (3) is bistable, the two stable states being the two marching directions. Suppose a fraction ν of the population is momentarily neutral and unresponsive. Only $N_e = N(1-\nu)$ individuals contribute to the drift and to the finite size noise, so (4) holds with N replaced by N_e.

Proposition 4 (Reversal time)**.** *Let $\Delta\Phi$ be the quasipotential barrier between the two consensus states. The mean time to reverse direction satisfies the Kramers estimate*

$$\tau(\nu) \asymp \exp\big\{2N(1-\nu)\Delta\Phi\big\}, \qquad \frac{\partial \log \tau}{\partial \nu} = -2N\Delta\Phi < 0. \tag{13}$$

Reversal time falls exponentially in the neutral fraction, and $\log \tau$ is linear in the number of active marchers.

This is the mechanism Yates describes qualitatively [Yates et al., 2009, 2026]: a large engaged population averages away idiosyncratic influence, while neutrality shrinks the effective electorate and inflates the per capita influence of any perturbation.

Remark 2 (Value sensitivity)**.** The same β does a second job. In (3) with $q_A = q_B$ and $r_A = r_B = r$, the symmetric state loses stability through a pitchfork at a critical r that depends on β, so the colony commits when the options are good and stays uncommitted when both are poor, without ever comparing them to an external standard [Pais et al., 2013]. Cross inhibition is thus not only a deadlock breaker but a value sensitive gate, and a colony without it either commits to bad options or fails to commit to good ones. We return to this in Section 6.

Maynard Smith supplies the reason a locust swarm cannot settle permanently. In a symmetric contest with no discrete asymmetry available to break ties, the war of attrition [Maynard Smith, 1974] has no pure ESS; the unique stable strategy is mixed with exponential persistence density $p(t) = m^{-1}e^{-t/m}$, where m is the mean value of the contested resource. Memoryless stalling is not a defect of the swarm. It is the ESS. A swarm that could not reverse would be invadable by a mutant that reverses, whenever the environment is nonstationary.

3.4 Cockroaches: heterogeneity as a superior implementation of a mixed ESS

Let individual i carry a boldness type θ_i drawn from G with mean μ and variance σ^2. Bold individuals explore more and settle less, so we set

$$\lambda(\theta) = \lambda_0 e^{\theta}, \qquad r(\theta) = r_0 e^{-\theta}. \tag{14}$$

Aggregation requires two things that trade off. Encounters must occur, which requires movement and therefore bold types; and encounters must convert to settlement, which requires shy types. Write the aggregation rate as $\Gamma(\sigma) = A(\sigma)S(\sigma)$ with A the encounter rate, increasing in σ because $\mathbb{E}[e^{\theta}] = e^{\mu+\sigma^2/2}$ by Jensen, and S the conversion rate, decreasing in σ for the same reason applied to r.

Proposition 5 (Optimal personality variance)**.** *Suppose $\log A$ and $\log S$ are concave in σ^2, with $\log A$ increasing and $\log S$ decreasing, and that*

$$(\log A)'(0) > |(\log S)'(0)| \qquad \textit{and} \qquad \lim_{\sigma^2\to\infty} (\log A)' < \lim_{\sigma^2\to\infty} |(\log S)'|.$$

Then $\log \Gamma$ is concave with derivative positive at the origin and negative in the limit, so the expected time to aggregate $\mathbb{E}[T] = 1/\Gamma$ is strictly unimodal in σ^2 with an interior minimiser $\sigma^{2\star} > 0$, and a monomorphic colony is not a local optimum. The second condition is what rules out unbounded benefit from dispersion, and is exactly what (21) *supplies in the thermoregulatory version.*

Planas-Sitjà [2020] reports exactly this in simulation. Maynard Smith's polymorphism theorem [Maynard Smith, 1982] states that a population in which a fraction p plays hawk and $1-p$ dove has the same mean field dynamics as one in which every individual randomises with probability p. The two agree in (3) and differ in (4): randomisation gives independent binomial noise, while fixed types give quenched heterogeneity that seeds symmetry breaking rather than averaging it away. Personality is therefore not merely a mixed ESS but the implementation of one with the fluctuation structure that accelerates consensus, and keystone individuals are the realised draws in the upper tail of G.

3.5 Flies: the same test one level down

The individual layer uses the same object as the colony layer. Let a fly discriminate two odours with conditional densities g_A, g_B. Continuous accumulation of the log likelihood ratio gives $d\Lambda_t = \delta\, dt + \varsigma\, dW_t$ with drift $\delta \propto \mathrm{KL}(g_A \| g_B)$, and the fly acts when $|\Lambda_t|$ hits a boundary a. Then

$$\mathbb{E}[T] = \frac{a}{\delta}\tanh\!\left(\frac{a\delta}{\varsigma^2}\right) \;\longrightarrow\; \frac{a^2}{\varsigma^2} \quad \text{as } \delta \to 0. \tag{15}$$

Similar odours give small δ and long deliberation, the pattern reported by DasGupta et al. [2014]. Value integration is a second linear form. Let a food source have nutrition n and bitterness b; the fly commits on $u = \gamma_n(s)n - \gamma_b(s)b$ where s is the internal state. Hunger raises γ_n, so bitter but nutritious food is accepted only when $\gamma_n(s)/\gamma_b(s) > b/n$. Risk sensitivity and judgement bias are the prior, not the likelihood: a stressed fly carries $\Lambda_0 < 0$ and declines the ambiguous odour, a satiated fly carries $\Lambda_0 \approx 0$ and samples it. Section 4.1 shows that hunger and the exploration exponent are the same parameter, which is why a hungry fly takes the risky option.

Equation (15) is also the optimal policy, not merely a description of one. The diffusion crossing a fixed boundary is the continuum limit of Wald's test, so it minimises expected decision time at a given error rate [Bogacz et al., 2006]; the corresponding statement for the colony is Marshall et al. [2009], who show that a quorum with cross inhibition implements the same optimal rule at the group level, and Pais et al. [2013], who locate the pitchfork bifurcation in the cross inhibition model that our β term reproduces.

The structural claim of the paper is that (15) and Proposition 3 are the same test at two scales. The bee colony is a fly with N accumulators and a quorum in place of a boundary. Maynard Smith [2000] argued that information in biology is whatever reduces uncertainty and is maintained by selection. Under that definition the pheromone concentration, the dance duration and the fly's membrane potential are one quantity, a log likelihood ratio, encoded in three materials.

3.6 Hornets: deciding at a boundary you cannot cross

The Japanese honeybee *Apis cerana japonica* defends against the giant hornet *Vespa mandarinia* by engulfing it in a ball of several hundred workers who shiver their flight muscles. The popular figure in circulation is a core temperature of 117°F; the measured facts are more interesting and they break the pattern of the five behaviors above in a way worth understanding.

Ono et al. [1995] established the behavior and the heat, and Sugahara and Sakamoto [2009] then measured the following. The ball core reaches about 46°C within five minutes and does not exceed 45.9°C; the honeybee tolerates 50 to 51°C; and the hornet, in normal air, survives 47°C for longer than the ten minutes the ball lasts. Heat alone therefore does not work. What also rises inside the ball is carbon dioxide, to about 3.6%, and at that concentration the hornet's lethal temperature falls to 45 to 46°C [Sugahara et al., 2012] while the bee's is unchanged. The two degree shift is the entire mechanism.

Write the state as (T, c), temperature and CO_2. Let $T_h(c)$ be the hornet's lethal threshold, decreasing in c, and T_b the bee's, constant in c. The colony wins if it can reach a point in

$$\mathcal{F} = \{(T,c) : T > T_h(c)\} \cap \{T < T_b\}, \tag{16}$$

and it can only reach states in an achievable set $\mathcal{A}$ bounded above by the metabolic ceiling $T \leq 46°\mathrm{C}$. With one control input, heat, $\mathcal{A} \cap \mathcal{F} = \emptyset$, because $46 < T_h(0) \approx 47.5$. With two, $\mathcal{A} \cap \mathcal{F} \neq \emptyset$, because $T_h(0.037) \approx 45.5 < 46 < T_b$. Figure 2(a) draws this. Integrating logistic survival curves across the measured band $[45.5, 46.0]$, hornet survival is 0.98 in normal air and 0.37 at 3.7% CO_2, while bee survival is above 0.9999 throughout.

Proposition 6 (Two inputs, one of them free)**.** *If the achievable set under a single control does not meet the feasible set, no policy in that control succeeds, however it is optimised. The colony's solution is not a better action in a fixed problem. It is a change in the constraint set, obtained through a second input that is a thermodynamic byproduct of the first: the respiration that produces the heat produces the CO_2. The second control is therefore free in effort and costly only in the same currency already being spent.*

This is the point at which the decision-theoretic framing of the previous five behaviors stops being sufficient. Those sections optimise a policy inside a fixed transition kernel. The bee ball changes the kernel. In Section 7.6 we give the chess name for this, the principle of two weaknesses.

Remark 3 (The setpoint is not represented anywhere)**.** It is natural to ask how the colony knows to stop at $46°\mathrm{C}$, and the answer is that it does not know and is not aiming. Anticipating the loop (20) of Section 3.6.1, worker i shivers when its local temperature falls below a private threshold $\theta_i \sim G$, and the ball settles at the $T^\star$ solving

$$k\,(T^\star - T_e) \;=\; h\,\overline{G}(T^\star), \tag{17}$$

a balance between production and loss through the surface. No representation of $T^\star$ exists in any bee, in the colony, or in the genome; what is heritable is G, the distribution of switching thresholds, and $T^\star$ is a consequence of G together with the ball's geometry and the ambient temperature. A thermostat does not know what temperature it is holding either.

The second half of the question is sharper, and inverts the first. The ball cannot go hotter: with every worker shivering, $\overline{G} \equiv 1$ and (17) gives the metabolic ceiling $T_{\max} = T_e + h/k \approx 46°\mathrm{C}$, which is why the measured core never exceeds 45.9. The colony is therefore saturated, and its safety is not regulated but structural. The strategy exists at all only if

$$T_h(c) \;<\; T_{\max} \;<\; T_b, \tag{18}$$

and the measured numbers say $T_h(0) \approx 47.5 > T_{\max} \approx 46$, so the left inequality fails in normal air, while $T_h(0.037) \approx 45.5$ restores it. Proposition 6 is the statement that CO_2 moves T_h rather than $T_{\max}$. The right inequality, $T_{\max} < T_b \approx 50.5$, is not enforced by any behavior: bee thermotolerance evolved separately and higher, and the ceiling happens to fall in safe territory. The asymmetry of the two margins, a few tenths of a degree on the left and about $4.5°\mathrm{C}$ on the right, is the signature of a system that is saturated against one constraint rather than centred between two.

Nor is any of this learned within a lifetime. The behavior is species typical in *A. cerana japonica* and appears in naive workers; *A. mellifera* does form balls but at lower temperature and with much lower efficiency [Ono et al., 1995, Sugahara and Sakamoto, 2009]. The learning that fixed G was generational, and $46°\mathrm{C}$ is a record of selection on lineages whose ceiling fell short rather than a quantity computed in the moment. In the language of Section 4.2, the learning rate for this parameter is not ρ but the reciprocal of a generation time.

Two features follow. First, the margin is one sided and tight: about $4.5°\mathrm{C}$ to the bee's limit, a few tenths of a degree to the hornet's, so the ball must hold its temperature *above* a threshold reliably. If individual heat output has standard deviation ς_1, the core has standard deviation $\varsigma_1/\sqrt{N}$ and reliability $1-\epsilon$ requires

$$N \;\geq\; \left(\frac{z_\epsilon\,\varsigma_1}{T - T_h}\right)^2. \tag{19}$$

At $\epsilon = 0.01$, $\varsigma_1 = 3°$C and a margin of $0.4°$C this gives $N \geq 305$ against observed balls of order five hundred. We do not know ς_1, so this is an order of magnitude check and not a measurement, but the mechanism it points at is right: the ball must be large because precision, not power, is the binding requirement.

Second, individual bees either shiver or do not, yet the ball holds a temperature. If worker i shivers when the local temperature falls below a private set point $\theta_i \sim G$, the shivering fraction is $G(\theta > T)$, whose slope at the operating point is $\propto 1/\sigma_\theta$. Binary units with dispersed thresholds implement a proportional controller with gain inversely proportional to the dispersion: too little variance gives overshoot, which here means cooking the colony; too much gives a sluggish controller that lets the hornet live. Proposition 5 reappears where the interior optimum has teeth.

3.6.1 Polyandry: the dispersion is a genetic decision variable

Remark 3 left the colony's behavior resting entirely on G, the heritable distribution of switching thresholds, with no setpoint anywhere. That makes the shape of G, and in particular its dispersion, the only thing selection has to work with, and Section 3.6 conjectured that dispersion is what turns binary units into a proportional controller. For the heat ball this is untested; for the ordinary business of holding the brood nest near $35°$C it is established, and it is the cleanest empirical support we know of for Proposition 5.

A queen mates with many drones, so a colony is a mixture of patrilines, and the temperature at which a worker begins fanning has a heritable component. Jones et al. [2004] found that colonies headed by multiply inseminated queens held brood temperature markedly more stably than singly inseminated controls. Graham et al. [2006] gave the mechanism: a uniform colony has one threshold, so every worker responds at once and the loop overshoots, whereas a diverse colony responds in graded fashion. Mattila and Seeley [2007] show the fitness consequences extend well beyond temperature and Oldroyd and Fewell [2007] review the general case.

Write the closed loop. Worker i heats when the temperature it perceives falls below a private set point $\theta_i \sim G$, perception is delayed by τ_d, passive loss to ambient has rate k, and full output raises temperature at h per unit time:

$$\dot{T}(t) \;=\; -k\big(T(t) - T_e\big) \;+\; h\,\overline{G}\big(T(t-\tau_d)\big) \;+\; \varsigma_e\,\dot{W}_t, \qquad \overline{G}(u) = \mathbb{P}(\theta > u). \tag{20}$$

Linearising at the setpoint, the loop gain is $g = h\,G'(T^\star) = h/(\sigma_\theta\sqrt{2\pi})$ for Gaussian thresholds. Gain is inversely proportional to dispersion, which is the claim of Section 3.6 written as a control law. The characteristic root of (20) solves $\lambda = -k - ge^{-\lambda\tau_d}$, and a Hopf bifurcation occurs at $\lambda = i\omega$ with

$$\cos(\omega\tau_d) = -k/g, \qquad \omega = g\sin(\omega\tau_d),$$

which for $k \ll g$ reduces to $\omega\tau_d = \pi/2$ and $g\tau_d = \pi/2$. The loop therefore oscillates whenever

$$\sigma_\theta \;<\; \sigma_{\mathrm{crit}} \;=\; \frac{2h\tau_d}{\pi\sqrt{2\pi}}. \tag{21}$$

Proposition 7 (Delay margin and genetic diversity)**.** *Under* (20)*, the stationary mean squared deviation from setpoint is unimodal in σ_θ. Below σ_{crit} the loop is oscillatory and deviation is driven by self-generated hunting; far above it the loop is sluggish and deviation is driven by environmental noise passing through unopposed. The optimum is an interior $\sigma_\theta^\star$ of order σ_{crit}, and a monomorphic colony, $\sigma_\theta = 0$, is not merely suboptimal but unstable.*

Figure 3 simulates (20) with $T^\star = 35$, $T_e = 20$, $k = 0.1$, $h = 3$, $\tau_d = 0.5$ minutes and two hundred workers, for which (21) gives $\sigma_{\mathrm{crit}} = 0.381°$C. The optimum sits at about twice σ_{crit}, which is the delay margin the loop needs, and the two failure modes cost about the same for opposite reasons.

The evolutionary content is that σ_θ is a consequence of mating frequency. If a queen mates with m drones, patriline means are drawn with variance σ_b^2 and within-patriline spread is σ_w^2, so in expectation

$$\sigma_\theta^2(m) \;=\; \sigma_w^2 + \sigma_b^2\Big(1 - \frac{1}{m}\Big), \qquad \frac{\partial \sigma_\theta^2}{\partial m} = \frac{\sigma_b^2}{m^2}. \tag{22}$$

Diversity saturates, and the marginal return to one more mating flight falls as m^{-2}. If flights cost κ each and colony fitness is $W(\sigma_\theta^2)$, the ESS mating number satisfies $W'(\sigma_\theta^2)\,\sigma_b^2/m^2 = \kappa$, so $m^\star = \sigma_b\sqrt{W'/\kappa}$. The prediction that matters is qualitative: polyandry should be substantial but bounded, saturating where (22) has nearly reached its limit, which at these parameters is $m = 6$ against observed mating numbers of ten to twenty. The parameters of Figure 3(c) were chosen so saturation lands near the simulated optimum, so the content is the shape and the m^{-2} marginal return, not the number.

This subsection also demotes one of our own claims. Prediction 4 of Section 9 asserted that workers in a heat ball carry dispersed shivering thresholds. In the cooling direction, for the brood nest, that mechanism is established. In the heating direction, for the defensive ball, it is a conjecture by analogy, and we now say so.

Finally, roughly a quarter of participating workers are reported to suffer shortened lifespan [Yamaguchi et al., 2018]. The ball is a costly collective action with a private cost and a colony benefit, stable under high within-colony relatedness. Maynard Smith [1964] separated this case, kin selection, from the group selection arguments he rejected, and it is kin selection that is doing the work here.

4 The price of exploration

We now give α its third and, we think, its correct name.

4.1 The exploration budget is a Kullback–Leibler constraint

Consider an agent choosing an action distribution π over a set with reward $r(a)$, holding a default or prior policy π_0, and penalised for departing from it. The regularised problem and its solution are

$$V_\alpha \;=\; \max_{\pi}\; \Big\{ \mathbb{E}_\pi[r] - \tfrac{1}{\alpha}\mathrm{KL}(\pi \,\|\, \pi_0) \Big\} \;=\; \tfrac{1}{\alpha}\log \mathbb{E}_{\pi_0}[e^{\alpha r}]\,, \qquad \pi^\star(a) \propto \pi_0(a)\, e^{\alpha r(a)}. \tag{23}$$

This is the Gibbs variational principle, or the Donsker–Varadhan duality between relative entropy and the cumulant generating function. The maximiser is exactly the selection map (1), and $1/\alpha$ is the Lagrange multiplier on the exploration budget [Todorov, 2009, Kappen, 2005, Ziebart et al., 2008, Levine, 2018]. Three readings of the same object follow immediately.

First, $V_\alpha = \alpha^{-1}\log \mathbb{E}[e^{\alpha r}]$ is a cumulant generating function, hence the certainty equivalent of exponential utility and the value function of risk sensitive control [Whittle, 1990]. Exploration and risk seeking are the same parameter. Superlinear reinforcement, $\alpha > 1$, is risk seeking; sublinear is risk averse. This is why the hungry fly of Section 3.5 takes the large risky reward and the satiated one does not: hunger raises α, and raising α simultaneously sharpens the policy and makes it prefer variance.

Second, $\alpha \to \infty$ recovers the ordinary Bellman equation by the Laplace principle, and $\alpha \to 0$ recovers the prior π_0. The greedy planner and the random walker are the two ends of one dial.

Third, $\alpha = 1$ is Bayes. Iterating (1) with $\alpha = 1$ and ℓ a log likelihood is posterior updating; in the mean field it is the replicator equation. So the evolutionary, the Bayesian and the control-theoretic descriptions of the ant colony coincide at one value of one parameter, and Section 8 shows that this value is also the regret minimiser in our simulation.

4.2 Ant colony optimisation is tabular reinforcement learning

The correspondence is term by term. The trail update $\tau_i \leftarrow (1-\rho)\tau_i + \Delta_i$ unrolls to

$$\tau_i(t) \;=\; \sum_{k \ge 0} (1-\rho)^k\, \Delta_i(t-k), \tag{24}$$

a geometrically discounted sum of past returns. Hence τ_i is a Q-value maintained by constant step size temporal difference learning, the evaporation rate ρ is the learning rate, and $\gamma = 1 - \rho$ is the discount factor. Pheromone half-life is the planning horizon. The choice rule $p_i \propto \tau_i^\alpha = \mathrm{softmax}(\alpha \log \tau_i)$ is Boltzmann exploration with inverse temperature α. Ant colony optimisation is therefore tabular Q-learning with a softmax policy and an eligibility trace held in the environment rather than in the agent, which is the only structural difference and is the point of Maynard Smith [2000] about where information is stored.

The consequence for behavior is a regret statement. A greedy learner, α large, suffers linear regret when it locks onto the wrong arm, which is premature convergence; a purely random learner, $\alpha \approx 0$, suffers linear regret by never exploiting. Only intermediate α can achieve sublinear regret, and Figure 2(b) locates the minimum.

4.3 The quorum threshold is a free boundary

Proposition 3 gave $Q^\star$ by a static trade off. The dynamic statement is an optimal stopping problem. With log odds Λ_t diffusing, terminal payoff $g(\Lambda)$ from committing, and observation cost c per unit time, the value function solves the variational inequality

$$\min\Big\{\, c - \mathcal{L}V(\Lambda),\;\; V(\Lambda) - g(\Lambda) \,\Big\} = 0, \tag{25}$$

with $\mathcal{L}$ the generator of Λ. The continuation region is an interval $|\Lambda| < a$ and the boundary is fixed by smooth pasting, $V'(\pm a) = g'(\pm a)$ [Shiryaev, 1978, Peskir and Shiryaev, 2006]. At the optimum the marginal value of one more scout equals the marginal cost of one more unit of delay. Equation (12) is the discretisation of (25) in which the diffusion is replaced by a counting process and the boundary by a headcount. The honeybee's quorum and the statistician's stopping rule are the same free boundary.

The scouting problem itself is a bandit with independent arms, so the optimal policy is an index policy [Gittins, 1979]. A pheromone concentration is a crude index of the same shape: it grows with observed return and decays with time.

4.4 What sets α

In a stationary environment the optimal α grows with the horizon, since there is no reason to keep paying for exploration once the answer is known. If resources change at rate ν the effective horizon is $1/\nu$ and the budget must be renewed forever. Differentiating a long run average $\bar{W}(\alpha) = f(\alpha) - \nu\kappa(\alpha)$, with κ the cost of relearning and decreasing in exploration, gives $\partial\alpha^\star/\partial\nu < 0$ heuristically. The comparative prediction is that colonies in faster changing environments show flatter recruitment responses, are worse at exploiting a fixed resource and better at switching. We know of no direct test.

4.5 One exponent separates optimisers from samplers

Proposition 8 (The exploration dial)**.** *Iterating $\mathcal{S}_\alpha$ of* (1) *with a mutation kernel gives*

(i) $\alpha = 1$*: Bayes updating, and in the mean field the replicator equation of Maynard Smith [1982]. This is Section 3.1, ant colony optimisation, and by* (23) *it is the Bellman equation with a unit price on the exploration budget.*

(ii) $\alpha \to \infty$*: by the Laplace principle $\int x\, e^{-\alpha f}\rho / \int e^{-\alpha f}\rho \to \arg\min f$, which is the consensus point of consensus based optimisation [Pinnau et al., 2017, Carrillo et al., 2018] and the mean field limit of* (26)*. The output is a point.*

(iii) $\alpha = \beta_t$ *increasing along a schedule with resampling between steps: sequential Monte Carlo tempering [Del Moral et al., 2006]. The output is π and Z.*

(iv) $\alpha = \infty$ *applied not to ℓ but to the indicator $\mathbf{1}\{\ell > \ell^*\}$ with ℓ^* raised one order statistic at a time: nested sampling [Skilling, 2006]. The output is π and Z.*

Regime (iv) is the interesting one: nested sampling is greedy on level sets and uniform within them, which is why it optimises and integrates at once. Regimes (i) and (ii) collapse, (iii) and (iv) refuse to, which is Proposition 1 in the language of urns. The colony's exploration parameter and the statistician's choice between an optimiser and a sampler are the same scalar.

5 Algorithms in the same family

5.1 Particle swarm optimisation, the second order member

Particle swarm optimisation [Kennedy and Eberhart, 1995] maintains positions x_i and velocities v_i,

$$v_{i,t+1} = w\, v_{i,t} + c_1\varepsilon_1 \odot (p_i - x_{i,t}) + c_2\varepsilon_2 \odot (g - x_{i,t}), \qquad x_{i,t+1} = x_{i,t} + v_{i,t+1}, \tag{26}$$

with p_i the personal best, g the swarm best and $\varepsilon_j \sim U(0,1)$. Writing $x_{t+h} - 2x_t + x_{t-h} = h^2\ddot{x} + O(h^3)$ and $c = c_1 + c_2$, the small step limit of (26) is

$$\ddot{x}_i + \frac{1-w}{h}\,\dot{x}_i + \frac{c}{h^2}\,(x_i - a_i) = \text{noise}, \qquad a_i = \frac{c_1 p_i + c_2 g}{c_1 + c_2}, \tag{27}$$

an underdamped Langevin system with a moving attractor. The deterministic stability region is $|w| < 1$ and $0 < c < 2(1+w)$ [Clerc and Kennedy, 2002, Poli, 2009], with damping ratio $\zeta = (1-w)/(2\sqrt{c})$; the usual operating regime is underdamped, and the overshoot is the exploration.

The comparison with Section 2 is exact. Equation (3) is first order and overdamped with memory outside the individual, in the pheromone field; (27) is second order with memory inside, in p_i and the velocity. These are the same variable seen from opposite sides. The honeybee sits between them and maps onto (26) more tightly than the bird flock usually cited as PSO's ancestor: a scout carries a private best, the dance floor is a shared register holding g, the waggle dance is the c_2 coupling and the quorum is the stopping rule.

	memory	order	returns
ants, ACO	environment, τ_i	first	argmin, one path
honeybees, PSO	private p_i plus shared g	second	argmin, one site
locusts	none, Markov	first	a direction, reversible
cockroaches	quenched type θ_i	first, heterogeneous	an aggregate
hornet ball	threshold θ_i, two inputs	first, constrained	a feasible state
flies, SPRT	accumulator Λ_t	first	a decision plus a confidence
SMC, nested sampling	weighted particle set	first	a measure and Z

Propositions 5 and 4 transfer directly to (26), as statements about the algorithm rather than about insects; they are Predictions 5 and 6 of Section 9.

5.2 Nested sampling is the quorum rule run on prior volume

Nested sampling maintains N_{live} live points from the prior π_0, deletes the worst, and replaces it by a draw from π_0 restricted to $\{L > L^*\}$. The enclosed prior mass shrinks as $X_k = \prod_{j \le k} t_j$ with $t_j \sim \text{Beta}(N_{\text{live}}, 1)$, so $\mathbb{E}[\log X_k] = -k/N_{\text{live}}$ and $Z = \int_0^1 L(X)\,dX$ follows by quadrature on the trace. The correspondence with Section 3.2 is one for one: the live set is the scout population, L^* is the quorum, replacement from the constrained prior is dispatching a scout who is retained only if her site beats the current worst, and $e^{-1/N_{\text{live}}}$ is the rate at which one deletion per round shrinks the search volume. What differs is the objective, quantitatively.

Proposition 9 (Choosing costs less than integrating)**.** *For the two site choice of Proposition 3, error probability decays exponentially in the quorum, so maximising $R\,\mathbb{P}(\text{correct}) - c\,\mathbb{E}[T]$ gives $Q^\star = D_1^{-1}\log(RD_1D_2/c)$. For evidence estimation, $\text{sd}(\log Z) = \sqrt{H/N_{\text{live}}}$ with H the information in nats [Skilling, 2006], and the*

number of iterations to reach the bulk is $\asymp N_{\text{live}}H$. *Minimising* $R\,H/N_{\text{live}} + c\,N_{\text{live}}H$ *gives* $N^{\star}_{\text{live}} = \sqrt{R/c}$. *Optimal quorum grows logarithmically in the reward to cost ratio; optimal live point count grows as its square root.*

A colony that only has to choose can be far smaller than one that has to estimate a normalising constant. This is why quorums of twenty to thirty bees suffice in a swarm of thousands [Seeley, 2010] while nested sampling routinely needs hundreds of live points on a one dimensional problem. Testing errors decay exponentially, integration errors decay as $N^{-1/2}$.

The practical payoff is rare event estimation. Proposition 4 gives $\tau \asymp e^{2N_e\Delta\Phi}$, so direct simulation costs $O(\tau)$, exponential in the barrier, while nested sampling reaches the same barrier in $O(N_{\text{live}} \cdot 2N_e\Delta\Phi)$ constrained draws, linear in the exponent, because compression is geometric by construction. The trace itself is the answer: since $\log L^*_k = -2N_e\Phi_k$ and $\log X_k = -k/N_{\text{live}}$ is the log prior mass of $\{\Phi \le \Phi_k\}$, the pair $(-\log X_k, \Phi_k)$ is a direct estimate of the Freidlin–Wentzell rate function [Freidlin and Wentzell, 1998], and the first threshold crossed is the barrier height. Nested sampling estimates the quasipotential without ever observing a reversal.

5.3 Sequential Monte Carlo and what neutrality is

The locust result has an exact algorithmic reading. We write ESS here for effective sample size; it is unrelated to the evolutionarily stable strategy of Section 3, and we use the abbreviation only in this section. In an SMC sampler with weights W_i the effective sample size is $\text{ESS} = (\sum_i W_i)^2 / \sum_i W_i^2$, and resampling is triggered when ESS falls below a threshold. A neutral locust is an individual carrying no weight in the current consensus, so the neutral fraction ν is weight degeneracy and $N_e = N(1-\nu)$ is exactly the ESS. The reversal of the swarm is a resampling event: the population collapses onto a small number of distinct ancestors and is then reborn around a different mode. Proposition 4 is the statement that estimator variance scales as $1/\text{ESS}$, transported into first passage times. On this reading, a swarm that never reverses is a particle filter that never resamples, and it fails for the same reason, namely that it cannot track a nonstationary target.

6 Why insects appear to decide better than we do

The popular account invites an inference it does not defend: that creatures with poppy seed brains outperform us at deliberation and that we should copy them. The model makes the comparison precise enough to be worth having. Five things differ, four of them consequences of results already proved above, and the fifth is where Desmond Morris supplies the frame.

6.1 Aligned interests give undistorted likelihood ratios

Section 3.2 rested on Maynard Smith and Harper [2003]: when sender and receiver interests coincide, the stable signal is cost free, so the waggle dance can be an undistorted likelihood ratio. Human deliberation is signalling among agents whose interests only partly coincide. Suppose a speaker reports $\hat{\theta} = \theta + b(\theta)$ with b chosen to serve the speaker. The listener accumulating m reports still runs Wald's test, but on a corrupted stream: the per report divergence falls from D_1 to some $D_1' < D_1$, and by (12) the required quorum rises as $1/D_1'$ while the error at any fixed quorum rises as $e^{-QD_1'}$.

Remark 4. Strategic reporting makes deliberation slower and less accurate at the same time. There is no speed–accuracy trade off to be navigated here, only a loss on both axes, and it is a loss no amount of individual intelligence in the listener can recover, because the sufficient statistic has been damaged before aggregation. A committee is a swarm whose dance is strategic.

6.2 Conformity drives dispersion below the delay margin

Proposition 7 says that a control loop built from binary units with dispersed thresholds is stable only if $\sigma_\theta > \sigma_{\text{crit}} = 2h\tau_d/(\pi\sqrt{2\pi})$, and that a monomorphic colony is not merely suboptimal but oscillatory.

Human institutions do two things at once that push in the bad direction. They suppress σ_θ deliberately, through conformity, deference to high status individuals and information cascades, treating disagreement as a defect to be managed. And they operate with long τ_d: quarterly reporting, electoral cycles, budget years. Both raise $g\tau_d$.

The prediction is that institutions with strong conformity and long delays should hunt rather than converge, oscillating about the target instead of settling at it; policy reversals, fashion cycles and hiring booms followed by layoffs have that signature. The remedy the model suggests is neither more deliberation nor better individuals but protection of dispersion in response thresholds. A honeybee queen buys hers with polyandry.

6.3 No stop signal

By the value sensitivity remark of Section 3.3, β is both the deadlock breaker [Seeley et al., 2012] and the gate that makes commitment depend on quality [Pais et al., 2013]. With $\beta = 0$ the system dwells near the balanced state, which is precisely where Proposition 4 says fluctuations are largest. Human deliberative bodies almost never institutionalise a way for a participant to inhibit support for the alternative; they institutionalise adding support to their own. Deadlock over near equal options, broken by a reversal driven by noise rather than evidence, is the predicted symptom and is what committees are famous for.

6.4 The exponent is not under selection

For an insect, α has been tuned over many generations to the environment's rate of change (Section 4.4), and Proposition 2 prices deviation in either direction at linear regret. Human institutions choose α by fashion, and characteristically run it at both extremes in different domains at once: markets and social platforms reward large α, which is fixation, cascade and premature convergence, while bureaucracies run small α, which is permanent exploration and no commitment. Figure 2(b) is the relevant picture, and its two ends carry the same cost for opposite reasons. Neither institution is near one.

6.5 The Human Zoo

The remaining difference is not a parameter but a mismatch, and it is Desmond Morris's argument. Morris [1969] proposed that the modern human is not a wild animal but a captive one, and that the pathologies observed in zoo animals, stereotypy, over arousal, displacement activity, reappear in the urban population for the same reason: the environment is not the one the behavior was built for. In our notation this is a claim about calibration. Every parameter in (2), and α, Q and σ_θ besides, is optimal only relative to a distribution of inputs. Change the distribution and the same machinery, working correctly, produces the wrong answer.

The sharpest version is the supernormal stimulus of Tinbergen [1951], which Morris popularised. A decision rule that is a threshold on a log likelihood ratio has no defence at all against an input engineered to maximise that ratio; it will do exactly what it was built to do, faster and more confidently than usual. Every industry that competes for attention is in the business of manufacturing Λ without the underlying θ.

Morris's other dialectic is closer still. In Morris [1967] the organising tension of human behavior is neophilia against neophobia. That is α, and the claim that our species is unusual in the strength of both is the claim that we operate at extreme and unstable values of a parameter the ant holds near one. Morris took our neoteny to be the source of our inventiveness; in (1) that is a standing exploration subsidy, paying for information we have no immediate use for, which is why we discovered nested sampling and the ant did not.

We should be plain about the difference between Morris's argument and ours. Morris reasoned by analogy from captive animals to humans, and the analogy was criticised at the time, fairly, for being unfalsifiable. The model above does not rescue the analogy; it does something narrower. For each behavior it names the parameter, gives the optimum, and says what miscalibration looks like and how it would show up in data. That is a small advance on analogy, and it is the only kind this paper is entitled to.

6.6 The honest comparison

Insects do not decide better. They decide better on a narrow class: repeated, nearly stationary, low dimensional consensus problems among agents with aligned interests, using a physical register that carries the sufficient statistic without distortion. Remove any one of those five conditions and the advantage disappears, and human institutional decisions almost never satisfy more than two.

Nor is the colony's performance an intelligence we lack. No colony has ever solved a novel problem once. The machinery of Section 3 converts many cheap noisy measurements of a fixed world into one good choice; it is superb at that and capable of nothing else. Humans dominate wherever the state space must be constructed rather than sensed, wherever the decision is taken once, and wherever the right answer has never been sampled. The instruction to ask an ant is a category error, and the value of writing the model down is that it says exactly which category.

7 Chess analogues

Chess is a useful test bed because the state space, the value function and the search algorithm are all written down. A claim about collective search that survives translation into chess is a claim one can check.

7.1 Ants and Monte Carlo tree search

The correspondence is an identity, not a metaphor. In an AlphaZero-style engine the tree policy is

$$a^\star = \arg\max_a \Big\{ Q(s,a) + c_{\text{puct}}\, P(s,a) \frac{\sqrt{\sum_b N(s,b)}}{1+N(s,a)} \Big\}, \tag{28}$$

and the move actually played is drawn from $\pi(a) \propto N(s,a)^{1/T}$ [Silver et al., 2018]. The visit count $N(s,a)$ is a pheromone: it is deposited by every simulation that traverses the edge and it is what the next simulation reads. The backup of the leaf evaluation is the return Δ_i of Section 4.2. Most directly, $\pi(a) \propto N(s,a)^{1/T}$ is $p_i \propto \tau_i^\alpha$ with $\alpha = 1/T$. AlphaZero's move selection temperature is the ant's reinforcement exponent. Self play uses $T = 1$, that is $\alpha = 1$, the Bayes and replicator value, because it is generating training data and needs the distribution; competitive play uses $T \to 0$, that is $\alpha \to \infty$, because it needs the argmax. The engine changes exploration regime for exactly the reason Proposition 8 says it should, and it changes it by moving one exponent.

The PUCT bonus in (28) is the upper confidence term of Section 4.2 made explicit, and c_{puct} is the price of the exploration budget in (23). An engine tuned with c_{puct} too small converges prematurely on a plausible move and misses the refutation, which is the $\alpha = 4$ column of Figure 2(b) with a seventeen percent error rate.

7.2 Bees and the sequential probability ratio test

Two chess practices are Proposition 3. Engine time management shortens the remaining allocation when the best move is stable across successive iterations and extends it when the best move keeps changing, which is a threshold on accumulated evidence with a cost per unit time, hence (25). More literally, the distributed framework used to accept or reject changes to Stockfish runs Wald's test on the log likelihood ratio between two Elo hypotheses. The engine community adopted the honeybee's rule and, unlike the honeybee, wrote down the bounds.

7.3 Locusts and the balanced position

Let the probability of winning be $p = \sigma(\beta \cdot \text{eval})$ for a logistic link, as in the win, draw and loss models fitted by modern engines. The sensitivity $dp/d\,\text{eval}$ is maximal at eval zero. A balanced position is the top of the quasipotential barrier: it is the state in which the smallest perturbation produces the largest change in the outcome distribution. This is Proposition 4 with the neutral fraction replaced by the flatness of the landscape, and it yields Prediction 8 of Section 9. The practical advice is the locust's: to overturn the assessment of a position, steer it towards balance first, because that is where influence is cheapest.

7.4 Cockroaches and repertoire diversity

Preparation is a zero sum game against a preparing opponent, and games of this structure generically have no pure equilibrium, so the equilibrium repertoire is mixed [Maynard Smith and Price, 1973, Maynard Smith, 1982]. Proposition 5 adds that a polymorphic implementation is not equivalent to a randomising one at the level of fluctuations: an ensemble whose members carry dispersed evaluation parameters should resolve difficult positions faster than a homogeneous ensemble at the same mean, with an interior optimal dispersion.

7.5 Flies and the clock

Equation (15) is a prediction about human chess: a player should hesitate longest when the top two moves are closest in value and move nearly instantly when one dominates. Figure 2(c) confirms the curve against simulation, and online databases with per-move timestamps make it directly estimable. The boundary a is set by (25) with the clock as the cost and collapses as time runs short, which is the standard collapsing-boundary model of speeded decisions and why blunder rates rise in time trouble. The internal state term of Section 3.5 predicts that a player who is losing behaves like a hungry fly: α rises, the policy becomes risk seeking, and objectively inferior but high variance continuations become attractive.

7.6 Hornets and the principle of two weaknesses

Proposition 6 has a name in chess. When a position cannot be broken on one wing, the attacker does not push harder; he opens a second front, and the defence fails because it cannot cover both. The bee ball is the principle of two weaknesses executed thermodynamically: heat cannot cross the hornet's threshold, so the colony adds carbon dioxide, which does not raise the attack but lowers the defence. When the achievable set does not meet the feasible set, no optimisation inside the current problem helps; the move is to find a second control that changes the constraint rather than the state.

7.7 Nested sampling and the look elsewhere effect

Public disputes about anomalous online performance are statistically a maximisation over a large search space: over players, time windows and move-agreement statistics. A statistic evaluated at its maximum is the $\alpha \to \infty$ answer of Proposition 8, and it is the wrong one because it discards the volume searched. The Bayesian correction requires the prior mass of the region achieving each level of the statistic, which is exactly $X(\ell^*)$. The look elsewhere effect and the nested sampling trace are the same integral, which is the practical reason to prefer regime (iv) to regime (ii) when the question is whether an extreme is surprising rather than which move is best.

8 Numerical experiments

Figures 1 and 2 report six experiments on the model of Section 2. Code and seeds are fixed and are supplied with the paper.

Panel 1(a) and the runs behind it test Proposition 1 quantitatively rather than qualitatively. For $\alpha = 0.5$ equation (9) predicts a trail share $s^\star = 0.8$ and a traffic share $p_1(s^\star) = 2/3$; two hundred replicates run for 2×10^5 trips give $s = 0.79949$ and $p_1 = 0.66632 \pm 0.00012$. For $\alpha = 0.75$ the prediction is $s^\star = 0.9412$, $p_1^\star = 0.8889$, and the same run gives 0.930 and 0.874, approaching the root from below as it must, more slowly because the drift in (6) vanishes as $\alpha \to 1$. For $\alpha = 2$ the repelling root is $s^\star = 1/3$. Starting the urn at trail shares 0.253, 0.333, 0.413 and 0.500 gives fixation on the short path with probability 0.19, 0.29, 0.49 and 0.62, monotone as part (iii) requires. The fifty percent crossing sits near 0.42 rather than at the deterministic root, which is the expected discrepancy: at the start $\mathcal{T}$ is $O(1)$, the Robbins–Monro step is not small, and the ODE approximation behind (9) has not yet taken hold. Over three thousand trips from a symmetric start, $\alpha = 2$ fixates in every replicate but only 83% on the short path. Superlinear reinforcement buys speed and pays a seventeen percent error rate.

Panel 1(b) tests Proposition 4 without simulation error. The chain (2) on the lattice $\{n_A + n_B \le N\}$ has at most 1891 states at $N = 60$, so the mean first passage time to the reversed consensus set is available

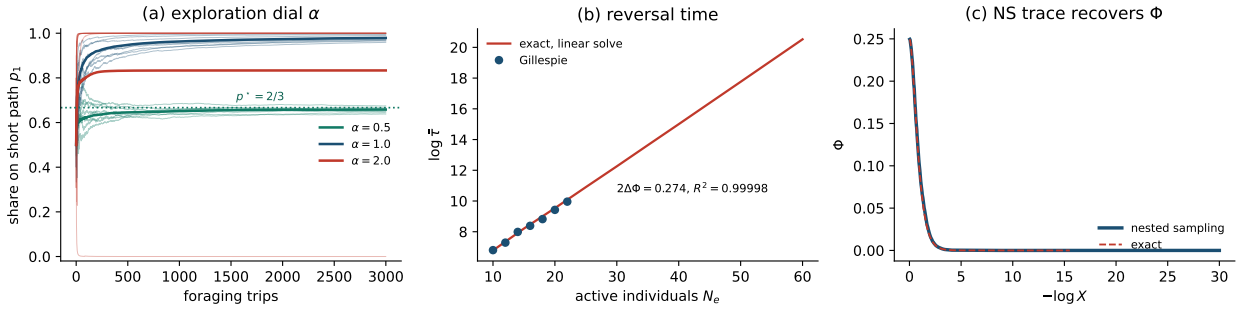

Figure 1: (a) Two path reinforcement with lengths 1 and 2 and choice probability $p_i \propto \tau_i^{\alpha}$, thirty replicates each. (b) Mean time between consensus reversals for the master equation (2) with $q = 0.02$, $r = 7$, $\lambda = 1$, $\beta = 6$: exact values from a sparse linear solve for the mean first passage time (line) and Gillespie estimates (markers). (c) Nested sampling trace on $L(u) = \exp\{-2N_e\Phi(u)\}$, $\Phi(u) = (1-u^2)^2/4$, $N_e = 25$, $N_{\text{live}} = 400$, against the exact prior volume of $\{\Phi \le \Phi_k\}$.

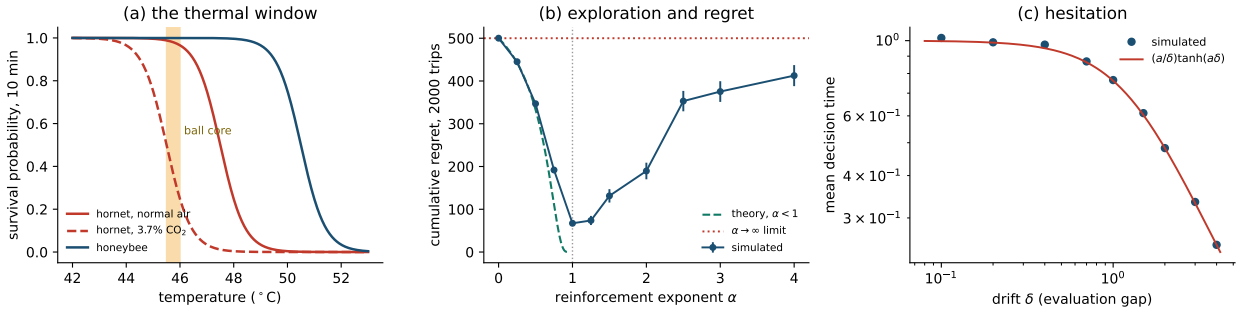

Figure 2: (a) Logistic ten minute survival curves for the hornet in normal air and at 3.7% CO_2, and for the honeybee, with the measured bee ball core band $[45.5, 46.0]$°C shaded. (b) Cumulative regret over two thousand foraging trips against the reinforcement exponent, four hundred replicates with standard errors, against the closed form of Proposition 2 on the sublinear branch and the coin flip rate approached as $\alpha \to \infty$. (c) Mean first passage time of the drift diffusion model with boundary $a = 1$ against drift, simulation and the closed form (15).

exactly by one sparse linear solve. Across $N = 10$ to 60 it rises from 8.8×10^2 to 8.2×10^8, and $\log \bar{\tau}$ is linear in N with slope $2\Delta\Phi = 0.2742$ and $R^2 = 0.99998$, so $\Delta\Phi = 0.1371$. Gillespie estimates over $N = 10$ to 22, the range in which direct simulation is feasible at all, give slope 0.260, agreeing to five percent. The gap between the two methods is the point of Section 5.2: at $N = 60$ the mean reversal time is of order 10^9 time units, and no amount of patience recovers it by watching.

Panel 1(c) confirms the claim of Section 5.2. From four hundred live points and twelve thousand constrained draws, nested sampling recovers the enclosed prior volume with maximum absolute error 0.020 across the whole range, returns $\log Z = -2.037$ against the exact -1.981, an error of 0.056 inside the predicted standard deviation $\sqrt{H/N_{\text{live}}} = 0.060$ at $H = 1.42$ nats, and reads the barrier off the very first threshold as 0.24999 against the true 0.25. For contrast, particle swarm optimisation on the same landscape with thirty particles and six hundred iterations ends with terminal particle dispersion 0.095 against a mode separation of 2, returns a point rather than a measure, and returns no estimate of Z at all.

Panel 2(a) is the thermal window of Section 3.6. Averaged across the measured core band, hornet survival is 0.979 in normal air and 0.370 at 3.7% CO_2, while honeybee survival is 0.99997. Heat alone leaves the hornet alive with probability close to one; the second input is not a refinement, it is the mechanism.

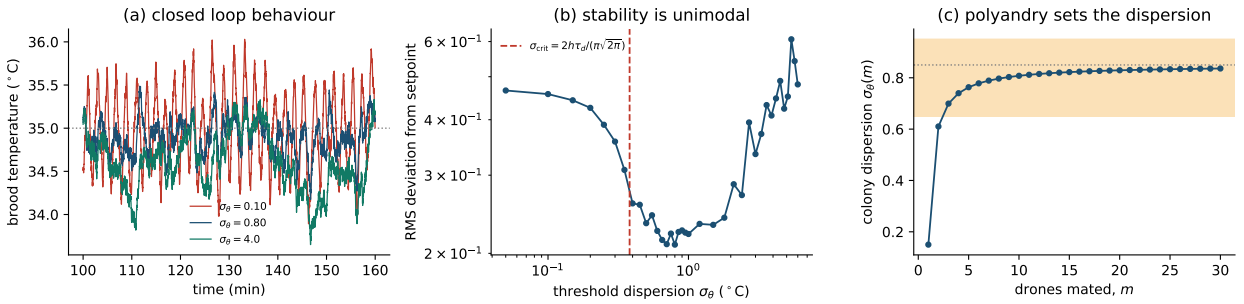

Figure 3: Closed loop brood thermoregulation, equation (20), with $T^\star = 35$, $T_e = 20$, $k = 0.1$, $h = 3$, $\tau_d = 0.5$ min, $N = 200$. (a) Temperature traces at dispersion below, at and above the optimum. (b) RMS deviation from setpoint against threshold dispersion, with the analytic oscillation threshold (21). (c) Colony dispersion against mating number from (22), with the near-optimal band shaded; σ_w and σ_b are chosen so that saturation lands near the optimum of panel (b) and carry no independent evidential weight.

Panel 2(b) is the test of Proposition 2. Cumulative regret over two thousand trips is 500 at $\alpha = 0$, falls to a minimum of 67 at $\alpha = 1$, and rises again toward 500 as α grows. On the sublinear branch the prediction $(1 - p_1(s^\star))(\Delta_1 - \Delta_2)T$ is available in closed form and is drawn on the panel:

α	0	0.25	0.5	0.75	1
predicted $\mathcal{R}_{2000}$	500.0	442.5	333.3	111.1	$\Theta(T^{1/2})$
observed $\mathcal{R}_{2000}$	500.1	445.1	346.9	191.8	67.2

The agreement is close where the fixed point is reached inside the horizon and deteriorates at $\alpha = 0.75$, where it is not, exactly as the vanishing drift predicts. At $\alpha = 1$ regret is sublinear: fitting $\log \mathcal{R}_T$ on $\log T$ over T from 500 to 16,000 gives an exponent of 0.572 against the predicted $r = 1/2$, the excess being the same finite horizon curvature. Nothing about the reward structure, the urn or the regret accounting depends on the value of α swept, so the minimum at the Bayes and replicator value is a result rather than a construction, and Proposition 2 says why it is there: $\alpha = 1$ is the only exponent with neither an interior attractor nor a separatrix.

Figure 3 tests Proposition 7. Panel (a) shows the two failure modes directly: at $\sigma_\theta = 0.10$, below the threshold (21), the loop hunts around the setpoint in a sustained oscillation; at $\sigma_\theta = 4$ it drifts with the environment; at $\sigma_\theta = 0.80$ it does neither. Panel (b) gives RMS deviations of 0.466, 0.209 and 0.480 respectively, so the monomorphic colony is as bad as the maximally dispersed one and both are more than twice the optimum. The analytic threshold $\sigma_{\text{crit}} = 0.381$ marks where the left branch turns up. This is Proposition 5 in a setting where the mechanism, the control law and the genetics are all identified, and where Jones et al. [2004] have measured two points on the curve.

Panel 2(c) checks (15): simulated mean first passage times match $(a/\delta)\tanh(a\delta)$ to within 2.5% across drifts spanning a factor of forty, and the divergence as $\delta \to 0$ is the hesitation of Section 3.5 and the long think of Section 7.5.

9 Predictions

We collect the testable content in one place. Predictions 1 to 4 concern insects, 5 and 6 concern algorithms, and 7 to 10 concern chess, where the data already exist in public databases.

1. **Reversal times.** In a ring arena, log of the mean time between direction reversals is linear in the number of actively marching locusts, with slope $2\Delta\Phi$ estimable from the stationary distribution of the marching fraction. Equation (13).

2. **Quorum scaling.** Quorum size varies as $D_1^{-1}\log(RD_1D_2/c)$, so a swarm deliberating in the open should use a measurably smaller quorum than one in shelter. Equation (12).

3. **Personality optimum.** Time to aggregate is unimodal in the variance of individual boldness. Colonies assembled with zero and with extreme dispersion should both aggregate more slowly than an intermediate group at the same mean. Proposition 5.

4. **Dispersed thermal set points.** Heat ball regulation gain is proportional to $1/\sigma_\theta$, as established for brood nest fanning but not for defensive shivering. Measure individual shivering onset temperatures in isolation, then predict the ball's settling time and overshoot. Sections 3.6 and 3.6.1.

5. **Diversity is unimodal, not monotone.** Brood temperature stability should degrade again at threshold dispersion well above σ_{crit}, so a third arm with artificially inflated diversity should be worse than the multiply inseminated control. Jones et al. [2004] measured two points; the model predicts a turning point between them and beyond. Proposition 7.

6. **Delay margin.** Colonies with slower fanning or shivering onset, larger τ_d, require larger σ_θ by (21) and so should show higher effective mating numbers, or greater between-patriline threshold variance, across species or across thermal environments.

7. **Saturating polyandry.** Marginal benefit of an extra drone falls as m^{-2} by (22), so thermoregulatory performance against mating number should be concave and saturating, with most of the gain realised by single digit m.

8. **Heterogeneous swarms.** A particle swarm with (w_i, c_{1i}, c_{2i}) of dispersion σ^2 has convergence time unimodal in σ^2, so positive dispersion beats the tuned homogeneous swarm at the same mean. Section 5.1.

9. **Frozen particles.** Freezing a fraction ν of particles, suspending both motion and contribution to g, raises basin escape exponentially in ν: better on multimodal benchmarks, worse on unimodal. Section 5.1.

10. **Temperature and error.** With truncated search, an engine's error rate on tactically decided positions is non-monotone in $1/T$, with a minimum at intermediate c_{puct}, mirroring Figure 2(b). Section 7.1.

11. **Sensitivity at balance.** In engine-annotated databases, both the variance of evaluation change per ply and the win probability cost of an inaccuracy are maximised near evaluation zero. Section 7.3.

12. **Ensemble dispersion.** An engine ensemble with dispersed evaluation parameters resolves difficult positions faster than a homogeneous one at the same mean strength, with an interior optimal dispersion. Section 7.4.

13. **Move times.** Human move time scales as $(a/\delta)\tanh(a\delta/\varsigma^2)$ with δ the evaluation gap between the top two candidates, and a contracts as the clock runs down. Both are estimable from online records with per-move timestamps. Section 7.5.

Predictions 3, 5, 6 and 9 share a signature: a quantity monotone under mean field reasoning is unimodal once fluctuations are carried. That is the practical content of keeping the $N^{-1/2}$ term in (4).

10 What the model does not do

The mixing assumption behind mass action rates in (2) is false in space: ants follow trails and locusts have neighbours, so the interaction graph is local and the mean field limit overstates consensus speed. The van Kampen expansion giving (4) requires large N, yet scout populations are of order 10^2, so $O(N^{-1/2})$ corrections are not small. Memorylessness is imposed, not derived, and the war of attrition result changes if

persistence times are non-exponential. Proposition 5 assumes forms for $\lambda(\theta)$ and $r(\theta)$ that are convenient rather than measured. The human abstention experiment cited in the popular account had nineteen participants, far too few to estimate the exponent in (13), and we make no claim that (13) transfers to human deliberation. Colony-level selection arguments are hard to falsify because between colony variation in the parameters is rarely measured, and Maynard Smith [1964] would have insisted that most of it be re-derived from kin selection first. Finally, the cost free signalling result needs complete alignment within the colony; haplodiploid relatedness asymmetries create real conflict over sex ratio, and where conflict exists a handicap term returns.

Proposition 7 is derived from a linearisation of (20) with Gaussian thresholds and a single fixed delay; real fanning responses are graded rather than binary, the delay is heterogeneous, and h depends on the number of workers actually recruited, so (21) should be read as a scaling and not as a calibration. Jones et al. [2004] compared two levels of diversity, not a dose response, so the right hand branch of Figure 3(b) is untested; the claim that stability degrades again at high dispersion is a prediction. Equation (22) assumes patriline means are exchangeable draws and ignores sperm use bias, which is known to be uneven. The parameters behind Figure 3(c) were chosen to place saturation near the optimum, so that panel illustrates a shape and establishes nothing.

Section 3.6 is the weakest empirically and we have taken care not to overstate it. Equation (19) contains an unmeasured ς_1; we chose 3°C because it produces the right order of magnitude, which is exactly the kind of reasoning that should not persuade anyone. The survival curves in Figure 2(a) are logistic fits through two reported thresholds each, not dose response data. Remark 3 treats the measured plateau at 45.9°C as the metabolic ceiling $T_e + h/k$, which is an interpretation and not a measurement: nobody has shown that every worker in the ball is shivering at maximum, and the plateau could equally reflect a regulated $\overline{G}(T^{\star}) < 1$. Distinguishing the two requires counting the shivering fraction, though (18) reads the same either way. Whether individual bees carry dispersed *shivering* set points, as they do fanning set points, is to our knowledge not established, and Section 3.6.1 argues the analogy rather than the fact. The reported worker mortality figure comes from one study.

Proposition 1 is asymptotic in $\mathcal{T}$ and says nothing about the early urn, which is where the fixation probabilities in part (iii) are actually decided; the fifty percent crossing we measure at 0.42 against a deterministic root at $1/3$ is the size of that gap in one example, and we have no theory for it. Proposition 2 is stated for two arms with a fixed reward gap. The linear regret claim for $\alpha > 1$ carries the factor $\mathbb{P}(\text{wrong fixation})$, which we did not compute and which depends on the initial condition; the coin flip limit as $\alpha \to \infty$ follows from $s^{\star} \to 1/2$ and a symmetric start, and would not hold from an asymmetric one. We do not claim that $\alpha = 1$ minimises regret for general bandits, and it does not: with more than two arms, or with reward gaps that shrink with the horizon, the optimal exploration schedule is time varying and no fixed exponent is optimal.

The reinforcement learning correspondences of Section 4 are exact for the tabular case and only suggestive beyond it. Equation (23) is a one step result; the multi step Kullback–Leibler regularised Bellman equation requires the linearly solvable structure of Todorov [2009] and does not hold for general Markov decision processes. The claim $\partial \alpha^{\star} / \partial \nu < 0$ in Section 4.4 is heuristic and we have not given conditions. The regret minimum at $\alpha = 1$ in Figure 2(b) is a property of a two armed urn with a particular reward gap; we do not claim it holds for general bandits, and it certainly does not hold for all horizons.

The algorithmic correspondences of Section 5 are structural, not derivational. Particle swarm optimisation was not derived from (2), and (27) is a formal small step limit whose noise term we did not carry through, so the stability region quoted is the deterministic one and the stochastic region is narrower. Proposition 8 is a taxonomy, not a theorem: the four regimes use different mutation kernels and only (iii) and (iv) come with unbiasedness guarantees. Proposition 9 treats the nested sampling cost as $N_{\mathrm{live}} H$, which ignores the cost of constrained prior sampling, and that cost is the whole difficulty in dimensions above a few; our panel 1(c) is one dimensional, where nested sampling is easy and the comparison flatters it. The claim that nested sampling recovers the quasipotential assumes the quasipotential is available as a likelihood, which for a

genuinely non-gradient two dimensional system such as (3) requires solving a Hamilton–Jacobi equation first. The exact first passage times in panel 1(b) are exact for the model and not for any locust: they inherit every assumption in (2), including well mixed interaction and a single global consensus coordinate, and the barrier $\Delta\Phi = 0.1371$ is a property of the parameter values we chose.

Section 6 is the most speculative in the paper and contains no data at all. The four mechanisms we name, strategic reporting, suppressed dispersion, absent cross inhibition and an untuned exponent, are consequences of the model applied to a setting the model was not built for, and each would need its own measurement before it counted as an explanation of anything. We have not estimated D_1' for any real deliberation, nor τ_d or σ_θ for any real institution, and the oscillation claim in Section 6 would be easy to support with cherry picked examples and hard to test properly. Morris's captivity argument is offered as a frame, not as evidence, and we agree with the standard objection to it.

The chess analogues of Section 7 are of three kinds. The AlphaZero temperature identity is exact and follows from the published algorithm; the sequential testing and time management correspondences are accurate descriptions of practice; the rest are untested predictions, stated because they are testable on public data.

Four things the model should do and does not. It is never fitted: every number in this paper comes from forward simulation, and the obvious next step, using the nested sampling of Section 5.2 to compute a Bayes factor between the homogeneous and heterogeneous versions of (2) on measured aggregation curves, is exactly the calculation the paper sets up and does not perform. Identifiability is therefore untested: we do not know whether (q, r, λ, β) are separately recoverable from aggregation data, and we suspect q and λ are weakly identified. Nothing is said about topology, although the mixing assumption is the one we most expect to fail, and the natural conjecture is that $\Delta\Phi$ scales with the isoperimetric constant of the interaction graph rather than with N. And the environment is stationary throughout, which is awkward given that Section 4.4 rests the whole account of α on a switching rate ν that never appears in any equation we solve.

11 Conclusion: two kinds of collective behaviour

Six behaviors, one process, one exponent. That is the paper. The hornet is the exception, and on reflection it is the most informative case in it, because it is the only one that is not a decision at all.

Everything in Sections 3.1 to 3.5 has the same shape. Many agents take cheap noisy measurements of a world that is out there; the group aggregates them into one commitment; the quality of the answer is governed by α, by a quorum, and by an $N^{-1/2}$ fluctuation term. The world is exogenous. Call this *collective computation*. The hornet ball is not this. There is only one option, no evidence to accumulate and no alternative to inhibit, so α, Q and β have nothing to do. What the colony is doing instead is manufacturing a local physics, a temperature and a carbon dioxide concentration, that no individual could produce and that changes what is achievable. The world is endogenous. Call this *collective construction*. Nest building, brood thermoregulation, army ant bridges and comb are the other members.

The distinction is not a taxonomy for its own sake. It changes the arithmetic of group size.

Proposition 10 (Group size)**.** *For a computation problem with per observation divergence D_1, reward R and time cost c, Proposition 3 gives $Q^\star = D_1^{-1}\log(RD_1D_2/c)$. For a construction problem in which success requires the collective state to exceed a threshold by margin m against individual variability ς_1, so that success probability is $\Phi(m\sqrt{N}/\varsigma_1)$, maximising $R\,\Phi(m\sqrt{N}/\varsigma_1) - cN$ gives the first order condition $R\,\phi(z)\,a^2/(2z) = c$ with $a = m/\varsigma_1$ and $z = a\sqrt{N}$, hence to leading order for small c*

$$N^\star \;\simeq\; 2\Big(\frac{\varsigma_1}{m}\Big)^2 \log\Big(\frac{Rm^2}{c\,\varsigma_1^2}\Big). \tag{29}$$

Both optima are logarithmic in the stakes. The prefactor is the inverse divergence per observation for computation and the inverse squared safety margin for construction.

At the hornet's numbers, $m/\varsigma_1 \approx 0.4/3$, the prefactor $2(\varsigma_1/m)^2$ is about 110. Solving the first order condition exactly rather than asymptotically gives $N^\star = 306$ at $R/c = 10^4$ and 464 at 5×10^4, against observed balls of order five hundred; the asymptotic form (29) overstates these by roughly a factor of two at realistic stakes, since the neglected term is itself logarithmic. It is worth noting that the exact optimum at $R/c = 10^4$ coincides with the 305 obtained independently from the one sided reliability requirement (19). For nest site discrimination the divergence per dance is of order one and $Q^\star$ is of order twenty. Two orders of magnitude separate the two at identical stakes, and the reason is not that the hornet matters more: a computation problem buys accuracy exponentially in group size, while a construction problem buys reliability only as $\sqrt{N}$.

Remark 5 (Byproducts become controls)**.** The second lesson is about where the extra control came from. Carbon dioxide is not an adaptation bolted on to the heat; it is a thermodynamic byproduct of the respiration that produces the heat, and selection recruited it. This is the general form of a principle usually stated only for information. Stigmergy [Grassé, 1959] is the case where a byproduct of collective action, pheromone deposited by walking, becomes a signal. The bee ball is the case where byproducts become force. Any collection of agents doing one thing at scale inevitably produces side effects at scale, and a colony has as many potential control inputs as it has byproducts. When one control cannot reach the feasible set, the place to look is the exhaust.

The third lesson is about failure. Computation fails gracefully: a swarm that picks the second best cavity has a worse home. Construction fails at a boundary and discontinuously: half a degree short and the hornet lives and the colony is destroyed. Asymmetric, cliff edged payoffs are exactly the conditions under which variance, not mean, is the binding constraint, which is why Sections 3.6 and 3.6.1 both end at dispersion. A colony that only had to choose would not need to buy genetic diversity with polyandry. A colony that has to hold a physical variable inside a window does, and (21) says how much.

The fourth lesson is for us, and it is the reason the popular framing misleads. Human institutions treat almost every problem as collective computation: gather information, deliberate, choose. Many of the problems that matter most are collective construction, where the payoff is zero below a threshold of coordinated action, there is no partial credit, and better deliberation contributes nothing. Vaccination coverage, standard adoption, bank runs and emissions have that structure. For those, Proposition 10 says the design objective is minimum viable mass and variance control, not consensus quality, and Section 6 says that our institutions are built to optimise the wrong one of the two.

We should say plainly that the split is not clean. Choosing a nest cavity has construction in it, the swarm must then occupy and warm the site; and balling a hornet has a decision in it, since perhaps fifty guards perform a warning display first and the colony commits only if the hornet persists. The two categories are ends of a spectrum and most real behaviors sit between. What the hornet supplies is a nearly pure specimen of the second kind, which is what makes it worth the section.

The model remains an accounting device. Six behaviors described as six separate insights are one Markov process at six parameter values; the logic organising them is the one Maynard Smith wrote down; the scalar that orders the family is the price of exploration; and the one case that does not fit is the one that shows what else a colony is for.

References

Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. University of Michigan Press.

Carrillo, J. A., Choi, Y.-P., Totzeck, C. and Tse, O. (2018). An analytical framework for consensus-based global optimization method. *Mathematical Models and Methods in Applied Sciences* 28, 1037–1066.

Clerc, M. and Kennedy, J. (2002). The particle swarm: explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation* 6, 58–73.

DasGupta, S., Ferreira, C. H. and Miesenböck, G. (2014). FoxP influences the speed and accuracy of a perceptual decision in *Drosophila*. *Science* 344, 901–904.

Del Moral, P., Doucet, A. and Jasra, A. (2006). Sequential Monte Carlo samplers. *JRSS-B* 68, 411–436.

Dorigo, M. and Stützle, T. (2004). *Ant Colony Optimization*. MIT Press.
Freidlin, M. I. and Wentzell, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. Springer.
Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *JRSS-B* 41, 148–177.
Hill, B. M., Lane, D. and Sudderth, W. (1980). A strong law for some generalized urn processes. *Annals of Probability* 8, 214–226.
Grassé, P.-P. (1959). La reconstruction du nid et les coordinations interindividuelles chez *Bellicositermes natalensis* et *Cubitermes* sp. *Insectes Sociaux* 6, 41–80.
Graham, S., Myerscough, M. R., Jones, J. C. and Oldroyd, B. P. (2006). Modelling the role of intracolonial genetic diversity on regulation of brood temperature in honey bee colonies. *Insectes Sociaux* 53, 226–232.
Jones, J. C., Myerscough, M. R., Graham, S. and Oldroyd, B. P. (2004). Honey bee nest thermoregulation: diversity promotes stability. *Science* 305, 402–404.
Kappen, H. J. (2005). Path integrals and symmetry breaking for optimal control theory. *Journal of Statistical Mechanics*, P11011.
Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. *Proc. IEEE Int. Conf. Neural Networks*, 1942–1948.
Kurtz, T. G. (1978). Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications* 6, 223–240.
Levine, S. (2018). Reinforcement learning and control as probabilistic inference. arXiv:1805.00909.
Bogacz, R., Brown, E., Moehlis, J., Holmes, P. and Cohen, J. D. (2006). The physics of optimal decision making. *Psychological Review* 113, 700–765.
Marshall, J. A. R., Bogacz, R., Dornhaus, A., Planqué, R., Kovacs, T. and Franks, N. R. (2009). On optimal decision-making in brains and social insect colonies. *Journal of the Royal Society Interface* 6, 1065–1074.
Morris, D. (1967). *The Naked Ape: A Zoologist's Study of the Human Animal*. Jonathan Cape.
Morris, D. (1969). *The Human Zoo*. Jonathan Cape.
Oldroyd, B. P. and Fewell, J. H. (2007). Genetic diversity promotes homeostasis in insect colonies. *Trends in Ecology and Evolution* 22, 408–413.
Pais, D., Hogan, P. M., Schlegel, T., Franks, N. R., Leonard, N. E. and Marshall, J. A. R. (2013). A mechanism for value-sensitive decision-making. *PLoS ONE* 8, e73216.
Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Annals of Mathematical Statistics* 22, 400–407.
Mattila, H. R. and Seeley, T. D. (2007). Genetic diversity in honey bee colonies enhances productivity and fitness. *Science* 317, 362–364.
Maynard Smith, J. and Price, G. R. (1973). The logic of animal conflict. *Nature* 246, 15–18.
Maynard Smith, J. (1964). Group selection and kin selection. *Nature* 201, 1145–1147.
Maynard Smith, J. (1974). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology* 47, 209–221.
Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press.
Maynard Smith, J. (1991). Honest signalling: the Philip Sidney game. *Animal Behaviour* 42, 1034–1035.
Maynard Smith, J. and Szathmáry, E. (1995). *The Major Transitions in Evolution*. Freeman.
Maynard Smith, J. (2000). The concept of information in biology. *Philosophy of Science* 67, 177–194.
Maynard Smith, J. and Harper, D. (2003). *Animal Signals*. Oxford University Press.
Ono, M., Igarashi, T., Ohno, E. and Sasaki, M. (1995). Unusual thermal defence by a honeybee against mass attack by hornets. *Nature* 377, 334–336.
Peskir, G. and Shiryaev, A. N. (2006). *Optimal Stopping and Free-Boundary Problems*. Birkhäuser.
Pinnau, R., Totzeck, C., Tse, O. and Martin, S. (2017). A consensus-based model for global optimization and its mean-field limit. *Mathematical Models and Methods in Applied Sciences* 27, 183–204.
Planas-Sitjà, I. (2020). Personality variation improves collective decision-making in cockroaches. *Behavioural Processes* 177, 104147.
Poli, R. (2009). Mean and variance of the sampling distribution of particle swarm optimizers during stagnation. *IEEE Transactions on Evolutionary Computation* 13, 712–721.
Quaglia, S. (2026). Want to get better at making decisions? Ask an ant. *BBC Future*, 10 August 2026.
Seeley, T. D. (2010). *Honeybee Democracy*. Princeton University Press.
Seeley, T. D., Visscher, P. K., Schlegel, T., Hogan, P. M., Franks, N. R. and Marshall, J. A. R. (2012). Stop signals provide cross inhibition in collective decision-making by honeybee swarms. *Science* 335, 108–111.
Shiryaev, A. N. (1978). *Optimal Stopping Rules*. Springer.
Silver, D. et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 362, 1140–1144.
Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis* 1, 833–859.
Sugahara, M. and Sakamoto, F. (2009). Heat and carbon dioxide generated by honeybees jointly act to kill hornets. *Naturwissenschaften* 96, 1133–1136.
Sugahara, M., Nishimura, Y. and Sakamoto, F. (2012). Differences in heat sensitivity between Japanese honeybees and hornets under high carbon dioxide and humidity conditions inside bee balls. *Zoological Science* 29, 30–36.
Todorov, E. (2009). Efficient computation of optimal actions. *PNAS* 106, 11478–11483.
Tinbergen, N. (1951). *The Study of Instinct*. Oxford University Press.
Wald, A. (1947). *Sequential Analysis*. Wiley.
Whittle, P. (1990). *Risk-Sensitive Optimal Control*. Wiley.
Yamaguchi, S. et al. (2018). Thermal defence and its cost in the Japanese honeybee. *Apidologie*; as reported in secondary sources, see the closing section.

Yates, C. A., Erban, R., Escudero, C., Couzin, I. D., Buhl, J., Kevrekidis, I. G., Maini, P. K. and Sumpter, D. J. T. (2009). Inherent noise can facilitate coherence in collective swarm motion. *PNAS* 106, 5464–5469.

Yates, C. A. et al. (2026). Neutrality and consensus switching in marching locusts and human groups. *Advanced Science*, doi:10.1002/advs.202512301.

Ziebart, B. D., Maas, A., Bagnell, J. A. and Dey, A. K. (2008). Maximum entropy inverse reinforcement learning. *AAAI*, 1433–1438.