



# Sector search strategies for odor trail tracking

Gautam Reddy<sup>a,b</sup>, Boris I. Shraiman<sup>c,1</sup>, and Massimo Vergassola<sup>b,d,1</sup>

<sup>a</sup>NSF–Simons Center for Mathematical & Statistical Analysis of Biology, Harvard University, Cambridge, MA 02138; <sup>b</sup>Department of Physics, University of California San Diego, La Jolla, CA 92093; <sup>c</sup>Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106; and <sup>d</sup>Laboratoire de physique de l'École Normale Supérieure, CNRS, Paris Sciences et Lettres Research University, Sorbonne Université, Paris 75005, France

Edited by Eric Siggia, Center for Studies in Physics and Biology, The Rockefeller University, New York, NY; received April 27, 2021; accepted October 23, 2021

**Ants, mice, and dogs often use surface-bound scent trails to establish navigation routes or to find food and mates, yet their tracking strategies remain poorly understood. Chemotaxis-based strategies cannot explain casting, a characteristic sequence of wide oscillations with increasing amplitude performed upon sustained loss of contact with the trail. We propose that tracking animals have an intrinsic, geometric notion of continuity, allowing them to exploit past contacts with the trail to form an estimate of where it is headed. This estimate and its uncertainty form an angular sector, and the emergent search patterns resemble a “sector search.” Reinforcement learning agents trained to execute a sector search recapitulate the various phases of experimentally observed tracking behavior. We use ideas from polymer physics to formulate a statistical description of trails and show that search geometry imposes basic limits on how quickly animals can track trails. By formulating trail tracking as a Bellman-type sequential optimization problem, we quantify the geometric elements of optimal sector search strategy, effectively explaining why and when casting is necessary. We propose a set of experiments to infer how tracking animals acquire, integrate, and respond to past information on the tracked trail. More generally, we define navigational strategies relevant for animals and biomimetic robots and formulate trail tracking as a behavioral paradigm for learning, memory, and planning.**

tracking | algorithm | behavior | optimization

Experimental studies demonstrate the ability of ants, dogs, humans, and rodents to track odor trails (1–6). Rodents accurately track trails in the dark, remaining close to the trail and casting when contact is lost (Fig. 1A) (5). Carpenter ants closely follow a trail while sampling it using a “crisscross” pattern with their two antennae (Fig. 1B) (1). Current models of this behavior rely on variants of chemotaxis (7) based on continuous estimates of the rising and falling odor gradients as the trail is crossed. One such strategy compares simultaneous odor concentrations detected by two spatially separated sensors (8). Yet, rats with a blocked nostril (5) and ants with a single antenna (1) are still able to track trails, although less accurately. An alternative chemotaxis strategy has the animal measuring odor gradients along its trajectory and turning when a significant decrease is perceived (5).

While chemotaxis-based strategies can allow for trail tracking when trails are continuous, they fail when trails are broken and gradients are absent, which is certainly relevant for animals tracking trails in the wild. In experiments with broken trails (1, 5), the absence of signal triggers casting, which is a fundamental feature shared with olfactory searches in a turbulent medium (9–11). Even though turbulent searches also feature sporadic cues, airborne odor signals tend to be localized in a cone, and even within the cone, the signal is highly fluctuating (12, 13). Therefore, beyond qualitative similarities between terrestrial trail tracking and airborne olfactory searches, the specific statistics of detections, geometric constraints, and behavioral patterns are distinct.

In contrast with chemotaxis-based algorithms, we propose an alternative framework built on the searcher exploiting past contacts with the trail to maintain an estimate of the trail’s local heading and its uncertainty. A minimal memory of the

approximate locations of the two most recent contacts suffices to delineate an angular sector of probable trail headings that radiates from the most recent detection point. The resulting “sector search” provides a quantitative description of trail-tracking behavior that unifies its various phases and yields specific experimental predictions.

## Results

We first show that reinforcement learning (RL) based on the sector search idea can recapitulate natural behavior. An RL agent in this scheme learns to traverse the trail as quickly as possible while minimizing the probability of losing it (*Materials and Methods* has details). Our in silico RL experiments show that general aspects of animal tracking behavior naturally emerge (Fig. 1C and D). Specifically, casts are observed around the most likely heading of the trail, and their amplitude is within the angular sector defined by the initial uncertainty  $\sigma$  of the trail’s heading  $\phi$ . The reason for the oscillatory pattern of casting is intuitive. Indeed, while moving along a path  $\mathcal{C}$  without detecting the trail, the estimated heading’s probability distribution  $P(\phi)$  (Fig. 1E) is updated into  $P_C(\phi)$  as

$$P(\phi) \rightarrow P_C(\phi) \propto \Gamma_C(\phi)P(\phi) \quad [1]$$

where  $\Gamma_C(\phi)$  is the probability of not detecting the trail headed along  $\phi$ . Irrespective of the explicit form of  $\Gamma_C(\phi)$ , the depletion of headings already explored generally leads to a bimodal posterior distribution, with the two modes at the edges of the angular sector (Fig. 1E). The search process is analogous to an agent “foraging” for the trail at two spatially separated patches. The

## Significance

Surface-bound odor trail tracking is critical for the survival of terrestrial animals dependent on olfaction. Little is known about how animals track trails at the algorithmic level. In the present study, we propose that a tracking animal maintains a noisy estimate of where the trail is headed based on its past contacts with the trail. We show that virtual agents trained to exploit this strategy reproduce the tracking patterns of ants and rodents. The observed patterns emerge simply as a consequence of common geometric constraints, which also impose fundamental limits on how quickly an animal can track trails. A series of experiments is proposed to quantify how past experience and trail statistics shape tracking behavior.

Author contributions: G.R., B.I.S., and M.V. designed research, performed research, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

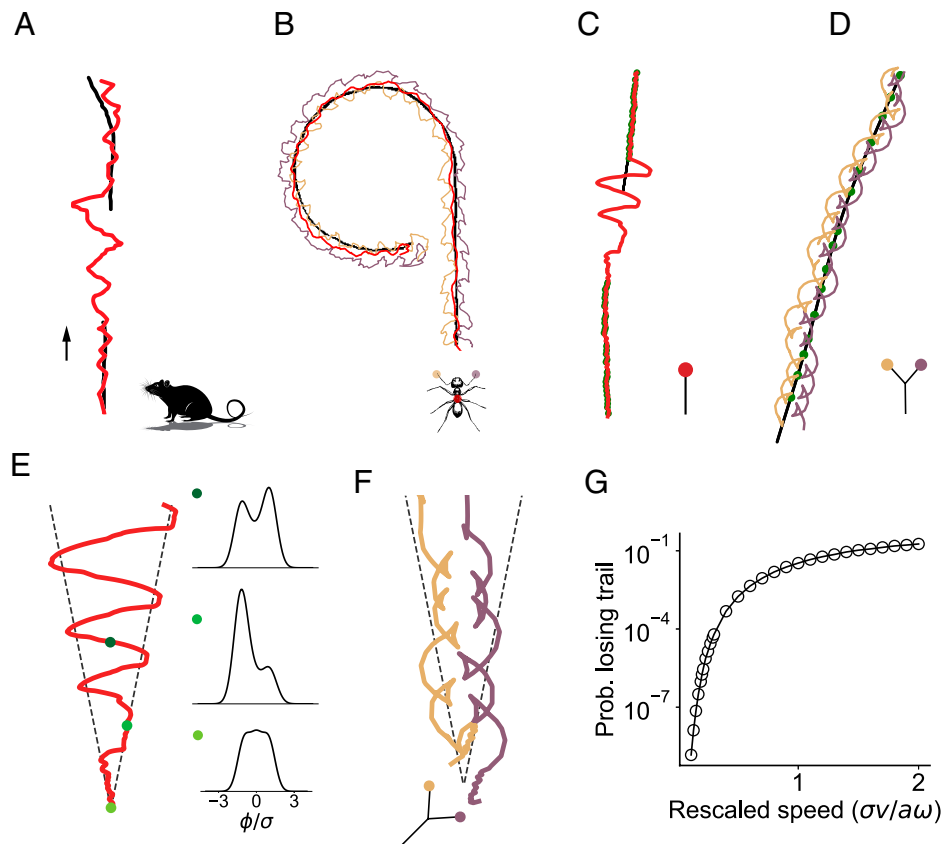
This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

See online for related content such as Commentaries.

<sup>1</sup>To whom correspondence may be addressed. Email: shraiman@ucsb.edu or massimo.vergassola@phys.ens.fr.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2107431118/-DCSupplemental>.

Published December 30, 2021.



**Fig. 1.** Sample trail-tracking trajectories from previous experiments and our RL simulations. (A) A rat (head position in red) tracking a trail (in black). Note the wide casts on extended loss of contact with the trail. Data reproduced from ref. 5. (B) A carpenter ant tracking an odor trail (black) using a stereotyped crisscrossing strategy (1). (C and D) Sample trajectories obtained from RL for agents with one sensor (C) and two sensors (D) recapitulate experimentally observed tracking patterns in A and B. (E, Left) Search paths executed by RL agents with a single sensor upon loss of contact with the trail. (E, Right) The initial prior distribution (E, Bottom Right) over trail headings transforms into a bimodal posterior distribution (E, Top Right and E, Middle Right), which drives the oscillatory pattern of casting. (F) RL agents with two sensors show a characteristic crisscrossing pattern close to the last detection point. The search path is similar to the single-sensor agent at long distances (SI Appendix, Fig. S1A). (G) RL agents show a trade-off between tracking speed (rescaled by the sector angle  $\sigma$ , sensor size  $a$ , and sampling frequency  $\omega$ ) and the probability of losing the trail entirely.

emergence of oscillations is then understood in terms of marginal value theory (14, 15); we show using a minimal model of casting (Materials and Methods) that the turning point of a cast occurs when the marginal value of continuing on one side of the sector (i.e., without paying the cost of traveling) is outweighed by the probability of finding the trail on the opposite side.

We proceed now by establishing geometric limits on tracking speed. A typical RL curve for the probability of losing the trail vs. speed is shown in Fig. 1G. Its monotonicity epitomizes universal limits that “staying on the trail” imposes on tracking speed. Intuitively, searching slowly reduces the distance between detections (the interdetection interval [IDI]), decreasing the uncertainty in the estimate of the trail’s heading and thus, the probability of losing the trail. However, these benefits come at the cost of slow forward progression along the trail. In contrast, moving quickly reduces the detection rate, leading to longer IDIs, increased uncertainty, and loss probability.

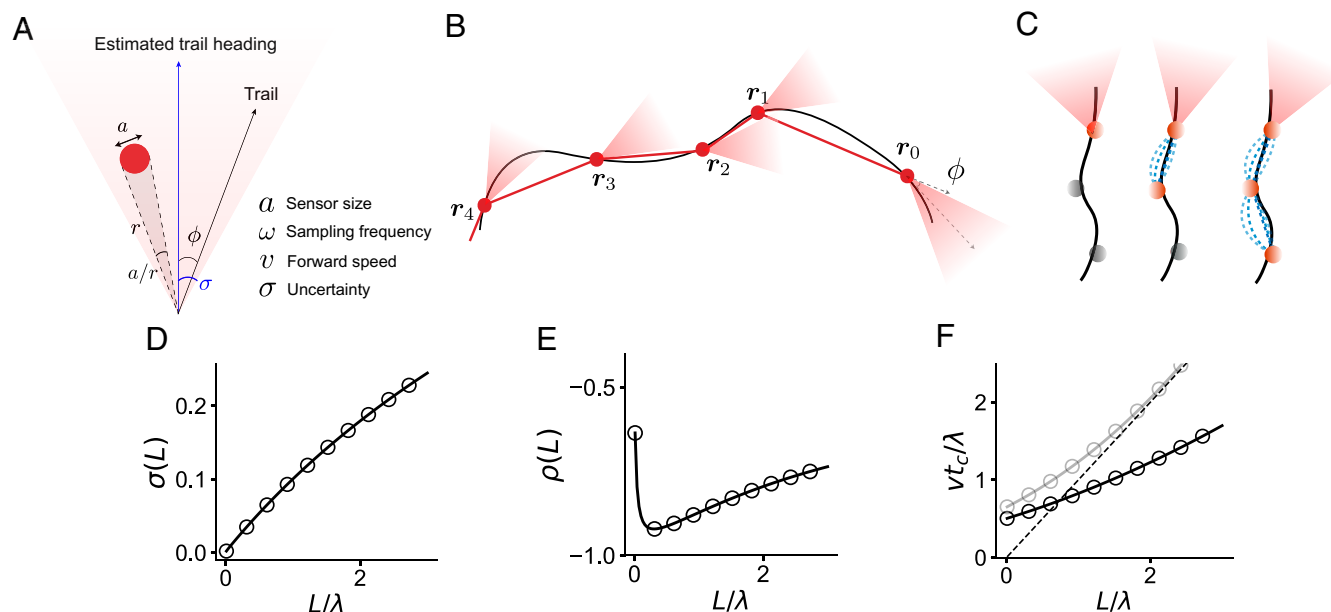
We quantify the above trade-off using simple scaling arguments. Suppose the tracking agent has a sensor of size  $a$ , samples at a frequency  $\omega$ , and moves with a fixed forward speed  $v$ . As shown in Fig. 2A, the angle subtended by the detector at distance  $r$  from the last contact is  $a/r$ . The agent searching over an angular sector then scans at a rate  $d\phi/dt = \omega a/r \simeq \omega a/vt$ . Integrating the above expression for the angular rate,  $\frac{v}{a\omega} \int d\phi = \int dt/t$ , we obtain the typical time for searching over a sector angle  $\sigma$ :  $t_c \sim \omega^{-1} e^{\sigma v/a\omega}$ . The corresponding distance  $L \sim vt_c$  is obtained using  $r \sim vt$ . The heading of the trail is known with uncertainty  $\sigma$ ,

which is the opening angle of the conical sector shown in Fig. 2A. Uncertainty is expected to depend on the distance  $L'$  from the previous detection as  $\sigma(L') = (L'/\ell)^\gamma$ , where  $\ell$  and  $\gamma$  characterize the statistics of trails (below and Fig. 2D). Importantly, a stable strategy for long-term tracking requires that successive IDIs should on average be equal (i.e.,  $L = L'$ ). Combining  $L = vt_c$  with  $L' = \sigma^{1/\gamma} \ell$  and the expression for  $t_c$ , we finally obtain an upper bound on the tracking speed  $v$ :

$$\frac{v^{1+\gamma}}{a\ell^\gamma \omega^{1+\gamma}} = (\omega t_c)^{-\gamma} \log(\omega t_c) \leq \gamma^{-1} e^{-1}. \quad [2]$$

Its maximum  $v_{\max} \sim \omega (a\ell^\gamma)^{\frac{1}{1+\gamma}}$  defines the optimal stable tracking speed in terms of the tracker’s sensory parameters and trail statistics. The basic element that leads to this bound is the geometric factor  $1/r$  that underlies searching over an angular sector. The result from Eq. 2 that  $\omega t_c$  is of order one ( $e^{1/\gamma}$ ) explains experimental observations (Fig. 1) that tracking animals typically take only a few samples to reestablish contact with the trail.

The above argument implies that tracking speed depends on the trail statistics via the relation between uncertainty and the distance between points of contact. We use ideas from polymer physics to quantify how this relationship depends on geometric properties of the trails. Specifically, we ask how detecting the trail at a set of points  $r_0, r_1, r_2, \dots$  (Fig. 2B) constrains the searcher’s estimate of its future heading. We consider the case when the



**Fig. 2.** History dependence and trail models. (A and B) Trail tracking naturally splits into distinct episodes punctuated by trail detections by the searcher. In each episode, we propose that the tracker searches for the trail using an estimate of the trail’s heading updated based on the past points of contact with the trail and a model of trail statistics. We affix a polar coordinate system with the origin at the most recent contact point and the azimuthal angle defined relative to the estimated trail heading. The uncertainty  $\sigma$  fixes the angular width of the search. The searcher moves forward with a speed  $v$  while sampling at a frequency  $\omega$ . A sensor of size  $a$  spans  $a/r$  radians at distance  $r$ , which determines the rate at which the angular space is searched. (C) To estimate where the trail is headed and its uncertainty from past contacts, the tracker can either use local anisotropy estimated from a single contact (C, Left) or extrapolate from previous points of contact using a model of trail statistics (C, Center and C, Right). In the latter case, the most likely trail paths (dashed blue lines) are similar to interpolated splines, which capture basic geometric notions of persistence in heading and curvature. (D) The uncertainty in trail heading (in radians) as a function of the distance,  $L$ , between points of contact for the GWLC model of trails discussed in the text.  $\lambda$  is the correlation length scale of the trail’s curvature. *SI Appendix, Fig. S3C* illustrates the various scaling regimes exhibited by the GWLC model. (E) The correlation between trail heading at the most recent and second most recent points of contact for the GWLC model changes with the distance between these points, yet it is generally expected to be negative. (F) The expected search distance,  $vt_c$ , against the distance,  $L$ , between the previous two points of contact for  $a/\lambda = 0.1$  and  $v/a\omega = 5, 6.25$  (black and gray, respectively) (discussion above Eq. 2). The point of intersection with the 45° dashed line is the condition for a stable tracking strategy. The gray curve corresponds to  $v_{\max}$  beyond which tracking is unstable.

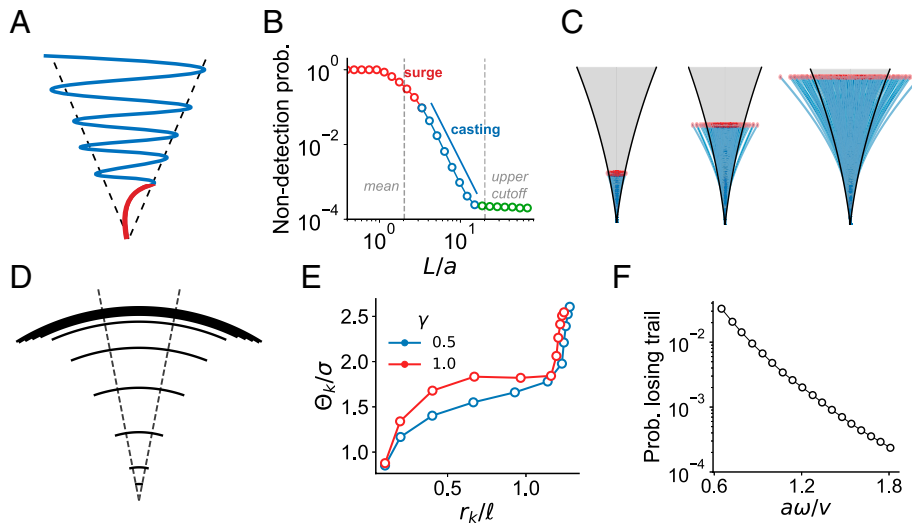
searcher keeps track of the two most recent points of contact with perfect memory of their location. A more extended memory is discussed further below; an imperfect memory adds to the uncertainty and can be easily accommodated within the framework developed below. Intuition for the two-point case is provided by the familiar “curve” tool in graphical design software, which draws a cubic spline through a set of prescribed points (Fig. 2C). The tool captures the simple intuition that tangents to a curve are continuous (i.e., the trail’s heading has local persistence), which is a plausible, minimal assumption about trails. We show in *SI Appendix* that cubic spline interpolation corresponds to the most likely path (through a fixed set of points) in the so-called worm-like chain (WLC) ensemble (originally introduced for polymers) (16, 17). In this ensemble, the tangent direction undergoes diffusion with rate  $\kappa$ , and the uncertainty is then  $\sigma = (\kappa L/3)^{1/2}$ , which determines the two parameters: the scaling law,  $\gamma = 1/2$ , and the correlation length scale,  $\ell = 3\kappa^{-1}$ , in Eq. 2. Actual trails could be smoother and have a well-defined curvature (the rate of change of heading) that persists on a characteristic length scale  $\lambda$ . We capture this ensemble of curves by introducing two additional parameters: persistence length  $\lambda$  and typical radius of curvature  $\xi$  (*Materials and Methods*). Uncertainty is then given by  $\sigma \approx L/2\xi$  (hence,  $\gamma = 1$  and  $\ell = 2\xi$ ) at distances  $L < \lambda$ , while diffusive scaling is recovered at larger distances with an effective diffusivity  $\kappa = 2\lambda\xi^{-2}$ . This extended model defines a generalized worm-like chain (GWLC) ensemble with crossovers across the various regimes (*Materials and Methods*). In summary, the model leads to a “propagator,” which encodes how information about past contacts is integrated to form an estimate of the trail’s heading while taking into account geometric aspects

of trails. A general feature is that the headings at two consecutive contacts are anticorrelated (Fig. 2E), which reflects the bending of the spline relative to the chord seen in Fig. 2C. We emphasize that although the general strategy of the agent depends on the statistical properties of the trail ensemble, the specific actions taken by the tracking agent along a particular trail, such as reorientation based on the most likely trail heading, will depend on the history of contact points via the propagator for the WLC (or GWLC) model.

Why and when do searchers need to cast? The question stems from our previous result that a few samples are typically sufficient to reestablish contact with the trail. To address it quantitatively, we consider again the setup of Eqs. 1 and 2. The nondetection probability averaged over the ensemble of trails that pass through past contact points is

$$\Gamma_C = \langle e^{-\omega \int_C \frac{ds}{v} I_a(r(s), y(s))} \rangle_y, \quad [3]$$

where  $s$  parametrizes the searcher’s path  $C$  and the Boolean indicator function  $I_a$  measures if the agent at  $r(s)$  is within sensing range  $a$  of the trail at  $y$  (i.e., the integral is the time spent in contact with the trail). Numerical simulations of the search show a power law scaling regime for  $\Gamma_C$ , which is cut off at short distances by the initial surge and at long distances by trails escaping out of the casting envelope (Fig. 3A–C). We proceed to explain these three regimes shown in Fig. 3B. Intuitively, at short radial distances  $r < a/\sigma \sim v/\omega$  (the latter from Eq. 2), the sensor covers the entire sector of likely headings, the searcher can just move forward, and  $\Gamma_C \propto e^{-\omega r/v}$  (Fig. 3B). Casting sets in if the searcher reaches, without detection, a distance  $r \gtrsim$



**Fig. 3.** The role of casting in trail tracking. (A–C) The probability of not detecting the trail at distance  $L$  decays exponentially during the initial forward surge (red) and as a power law during casting (blue) over a conical envelope. The searcher often finds the trail within the initial surge (mean in B), whereas casting determines the probability of losing the trail (determined by the upper cutoff) if the trail is not found during the surge. Beyond a characteristic length scale  $\ell$ , the detection rate becomes negligible (green region in B) as trails initially well inside the casting envelope escape out of the envelope (C). (D and E) The casting policy obtained from Bellman optimization. The specific casting strategy depends on trail statistics, shown here for diffusive and curvature-dominated trail ensembles (scaling exponent  $\gamma = 0.5, 1.0$ ) in blue and red, respectively. Note the increasing casting angle and the slowing down of the agent at  $r \approx \ell$  (E). (F) The trade-off between the probability of missing the trail for fixed  $\langle L \rangle$  and searcher speed  $v$ .

$a/\sigma$  (i.e., when the sector is not fully covered any longer by the sensor of size  $a$ ). The sector geometry in Fig. 3A implies that the length of a single casting sweep is proportional to  $r$ . A fixed forward speed then implies that the distance between successive encounters with the trail also scales with  $r$ . Hence, the number of times the searcher crosses the trail (and thus, the time spent on the trail) per unit radial distance decreases as  $1/r$ . This  $1/r$  scaling in the overlap then leads to a logarithmic integral in the exponent of Eq. 3 and thus, a power law regime  $\Gamma_C \propto r^{-\beta}$  during casting. The optimal exponent  $\beta^*$  depends on the statistics of the trails, yet it generally satisfies  $\beta^* > 1$  ( $\beta^* = 1.63$  for the curvature-dominated GWLC model) (Eq. 19). Since  $\Gamma_C$  is a cumulative distribution,  $\beta > 1$  implies that the mean distance is indeed determined by the lower cutoff (i.e., the trail is typically found in a few samples, as estimated in Eq. 2). However, the power law decay implies that casting phases are frequent and can span up to the upper cutoff where all aspects of trail statistics and sector geometry come into play, as discussed below.

How should an agent perform sustained casts so as to minimize the probability of losing the trail while maximizing tracking speed? To go beyond the above scaling arguments, we now consider the geometry of sector search in detail. The searcher's path is parameterized by the sequence of turning points of its casting trajectory  $\{r_k, \theta_k\}$  (in polar coordinates with respect to the most recent point of contact), which are to be optimized. We maximize the average tracking speed  $L/T$ , where  $L$  and  $T$  are the distance and duration between the most recent and the subsequent point of contact with the trail, respectively. As discussed previously, the uncertainty estimated for a bout of sector search depends on the IDI. To constrain the uncertainty, we therefore constrain the average,  $\langle L \rangle$ . Hence, we consider the following optimization problem:

$$V_{v,\Lambda} = \max_{\{r_k, \theta_k\}} \left[ \left\langle \frac{L}{T} \right\rangle - \Lambda \langle L \rangle \right], \quad [4]$$

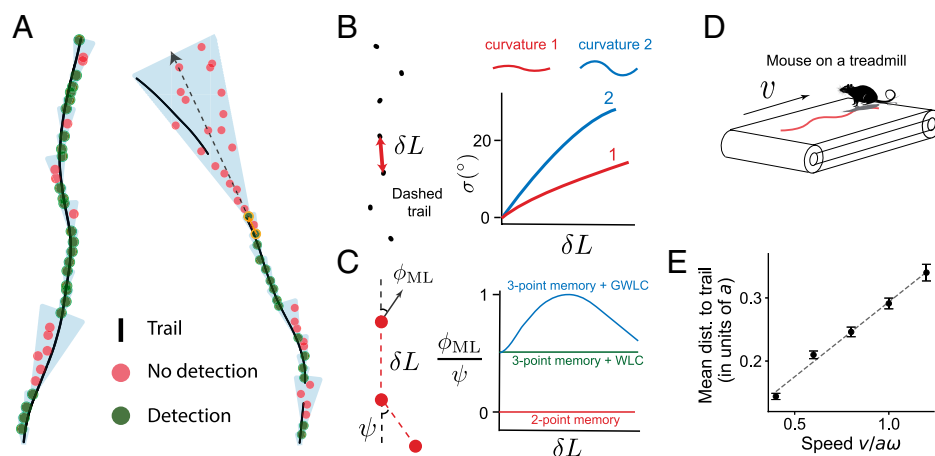
where  $\Lambda$  is the Lagrange multiplier enforcing the  $\langle L \rangle$  constraint. The turning points and the searcher's speed  $v$  affect the probability of detecting the trail in a single cast, which is implicit in the expectation in Eq. 4. We solve the Bellman equation corresponding to the above optimization problem using dynamic programming

(Materials and Methods), which sequentially optimizes for the turning points by considering at each step the two possibilities that either the trail is detected during the cast or the agent advances to the next cast. The probabilities for these two events are controlled by the nondetection probability given by Eq. 3. In the event of no detection, the estimate of the trail's heading is updated according to Eq. 1. The resulting optimization yields a search strategy with an increasing sequence of casting angles (Fig. 3D). The specific casting strategy depends on how trails meander and curve (Fig. 3E). The choice of  $v$  controls the trade-off between the tracking speed and the probability of losing the trail entirely (Fig. 3F). Independent of the choice of  $v$  and  $\langle L \rangle$ , azimuthal excursions are by and large conical but extend dramatically (Fig. 3D and E) when trails that were initially inside the cone escape from it with high probability (Fig. 3C). This happens at a distance that scales with  $\ell$  but also depends on the sector envelope, which in turn, depends on  $\sigma$  (recall that  $\ell$  is the correlation length of the trail heading). Intuitively, at length scales  $\sim \ell$ , the trail's heading decorrelates from its initial value, the relevance of information on past detections has expired, the trail is effectively lost, and it is optimal to stop progressing forward.

### Experimental Tests

A number of transformative experimental assays are suggested by our theoretical framework. The broad theme is whether and how animals adapt their behavior to the statistics of trails. For field experiments, it would be informative to measure the statistics of natural trails, analogous to the statistics of natural images that has brought insight into the adaptation of visual responses (18–21). Specifically, one can measure the autocorrelation of local heading and curvature of natural trails, which would test the validity and fix the parameters of our WLC-type models. In laboratory settings, the statistics of trails can be controlled by varying persistence or curvature or using broken trails (Fig. 4A). The general issue of adaptation is articulated in the following four specific questions that stem from our work.

First, how long after the loss of contact do animals “give up” tracking? Our prediction is that they should when they get beyond the characteristic correlation length of the trails. At this point, the value of past information has expired, and it is best to



**Fig. 4.** Behavioral assays to dissect trail-tracking strategies. (A) A tracker executing a sector search often maintains continuous contact with the trail (A, Left). Extended sector searches can instead be systematically elicited using broken trails (A, Right). The subsequent search envelope's orientation and width relative to past contact points inform the tracker's internal estimate and uncertainty  $\sigma$  of the trail's heading. (B) We propose dashed trails as an assay to infer how a tracking animal integrates past information. The distance between dashes forces contact points to be separated by at least  $\delta L$ , and the subsequent search sector yields an estimate of  $\sigma$ . (C) Our theory predicts that the most likely trail heading  $\phi_{ML}$  is proportional to the angle  $\psi$  between the line segments joining points of contact (red dots), with a prefactor that depends on the distance  $\delta L$  between the two most recent contact points, the trail model and memory. (D) Automated behavioral tracking of rodents on a treadmill allows control of tracking speed and trail statistics (5, 22). (E) The mean distance to the trail with speed in simulations of sector search (as in A, Left), which recapitulates the linear relationship (dashed line) observed in experiments with rats (5). We use a sector search strategy, where the longitudinal speed  $v$  is fixed and the tracker rapidly casts within a conical envelope (SI Appendix, Fig. S4A shows a sample trajectory in a single bout). For each  $v$ , we simulate 100 trials, where each trial consists of successive 10 successful contacts with the trail. Trails are generated from the GWLC ensemble with  $\kappa = 0$ ,  $\xi/\lambda = 3$ ,  $a/\lambda = 0.1$ . Error bars are one SEM.

turn back or start a new search. This prediction can be tested by varying trail statistics, interrupting the trails, and measuring when animals give up.

Second, does the amplitude of casting depend on the statistics of trails and the IDI? We predict that it should, and the specific quantitative relationship is a signature of the underlying predictive model employed by the animal (Fig. 4B). Our prediction should be contrasted with the nonadaptive casting envelope assumed in ref. 5. The IDI can be experimentally manipulated by generating dashed trails as shown in Fig. 4B, which forces the animal to detect the trail sporadically yet at controllable intervals. It would be particularly informative to verify whether or not animals include curvature in their estimates of the trails' future heading or limit to persistence.

Third, what is the memory of past trail contacts? Experiments with forked trails (5) show that rats exhibit a predictive component, suggesting a memory that extends over the recent past. Our theory posits that the tracker remembers (at least) the two most recent detection points. For the case of two-point memory, we predict the sector search is oriented along the line connecting those two points. If more than two points are remembered, the expected heading deviates from this line (by an angle that we calculate explicitly in SI Appendix) as illustrated in Fig. 4C. Note that the heading is not an average of the past headings, as assumed in ref. 5, and actually depends on the IDI between recent contacts; this prediction could again be tested by using curved, dashed trails as in Fig. 4B.

Fourth, does the tracking speed vary with the typical IDI, reducing with increased uncertainty as predicted by the speed-accuracy trade-off Eq. 2? This can be tested by varying the speed, for instance, of the treadmill in ref. 5 (Fig. 4D) and measuring tracking accuracy. Available data for three speeds in ref. 5 are captured by our theory (Fig. 4E), which highlights the importance of revisiting those pioneering experiments and measuring additional quantities, namely the explicit prediction in Fig. 3F.

## Conclusion

In conclusion, we show that an optimized sector search strategy based on the memory of two or more recent detection events

yields an oscillatory search path with increasing amplitude that naturally unifies the observed low-amplitude “zigzagging” and larger-amplitude “casting” behaviors into the same quantitative framework. This framework elucidates the geometric and computational constraints faced by tracking animals and identifies general features of the algorithms that efficiently solve the task, which can also be implemented for robotic applications. Insights and predictions developed here impact and should inform the design and analysis of future animal behavior experiments.

While the computational constraints discussed in this work apply generally, animals may face an additional set of species-specific physical and physiological constraints that influence aspects of their trail-tracking strategies. For example, inertial effects and gait constraints will affect the rate at which an animal sweeps across a sector and consequently, its speed selection. Further flexibility is offered by extra degrees of freedom in the sensorimotor apparatus, such as modulation of the inhalation rate or independent control of segments of an ant's antennae. In our model, the sensorimotor mechanism impacts the strategy via the probability of detection of the trail from the animal's sensors, which we have assumed to be fixed. Indeed, without additional trade-offs, our model implies that the optimal strategy is to simply pick the parameters that maximize the probability of detection (i.e., the maximal sensor size and sampling frequency). Physiological constraints may impose trade-offs (e.g., between the strength and rate of inhalation), which then lead to an altogether new dimension in the animal's decision making. We postpone a detailed examination of these trade-offs and how they influence the trail-tracking strategy to future work.

We have assumed that the searcher's behavior is guided by a model of continuous trails and investigated how an animal should behave in the presence of a sporadic break (a common perturbation in experiments). Recurrent breaks in the trails can be incorporated into the decision-making framework by introducing an additional term in the exponent of Eq. 3 for the presence or absence of a trail. When averaged over the trail ensemble and the statistics of brokenness, the effect of broken trails appears as a reduced detection rate if the breaks are much shorter than the correlation length scale  $\ell$ , and the general conclusions remain

unaffected. However, if the breaks last longer than  $\ell$ , the trail is “patchy,” with each patch containing little information about the location of the next patch. In such a scenario, a strategy that employs biased random exploration or foraging may prove optimal.

## Materials and Methods

**An RL Framework for Trail Tracking.** Trail tracking naturally splits into discrete episodes where after each loss of contact, the tracker searches and attempts to reestablish contact with the trail. We use RL to identify optimal search strategies for each episode and explore how factors such as sensory configuration or movement constraints influence the strategy. The task is a one-dimensional search over angular space  $\theta$ , the geometry of which is illustrated in Fig. 2A. The tracker controls its tangential speed  $u(t) \equiv r\dot{\theta}$ , while its radial speed  $v$  is kept constant. For simplicity, we focus here on the configuration featuring a single sensor of size  $a$  sampling with a Poisson frequency  $\omega$ . The generalization to two sensors is found in *SI Appendix*.

In each episode, the (Bayesian) agent maintains a posterior probability distribution function (PDF)  $P(\phi)$  over possible trail headings, which is continuously updated based on the locations already visited until the trail is recontacted. The agent’s strategy of decisions about its future trajectory a priori depends on the full high-dimensional distribution  $P(\phi)$ , which is difficult to learn. To circumvent this issue, we formulate the search task using a tractable parametrization of the posterior as a mixture of  $K$  Gaussian basis functions:

$$P(\phi) = \sum_{i=1}^K q_i B_i(\phi), \quad B_i(\phi) = s^{-1} \varphi\left(\frac{|\phi - \mu_i|}{s}\right), \quad [5]$$

where  $\varphi$  is the standard normal PDF and the  $q_i$ ’s are normalized weights. The posterior is encoded by the weights  $q$ , the posterior probabilities of the latent states given the agent’s history. The corresponding vector is lower dimensional and yields to standard RL methods. For simulations in the text, we used  $K = 3$ ; equal initial weights;  $\mu_1, \mu_2, \mu_3 = -\sigma, 0, \sigma$ ; and  $s = 0.5\sigma$ . These values were chosen so that the prior has mean zero and SD  $\approx \sigma$ . Using  $K > 3$  led to similar strategies of search, but training was slower. We define the detection probability given latent state  $i$  for an agent at  $r$  as

$$\bar{\gamma}_i(r) = \int \gamma(r, \phi) B_i(\phi) d\phi, \quad [6]$$

where  $\gamma(r, \phi)$  is the detection probability of finding a trail headed along  $\phi$  if the searcher is at  $r$ . We assume a Gaussian detection kernel of size  $a$ , with distance measured to the closest point on the trail:  $\gamma(r, \phi) = e^{-r^2 \sin^2(\theta - \phi)/2a^2} \approx e^{-r^2(\theta - \phi)^2/2a^2}$ , where we have used the small-angle approximation. Conditional on no detection at  $r$ , Bayes’ rule yields  $\hat{q}_i = -\omega q_i(\bar{\gamma}_i(r) - \bar{\gamma}(r))$ , where  $\bar{\gamma} = \sum_i q_i \bar{\gamma}_i$  is the total probability of detection. From Eq. 6, we have

$$\bar{\gamma}_i(r) = \sqrt{\frac{\pi}{2}} \frac{\sigma_r}{\sigma'_r} \varphi\left(\frac{\theta - \mu_i}{\sigma'_r}\right), \quad [7]$$

where  $\sigma_r = a/r$  and  $\sigma'_r = \sqrt{\sigma_r^2 + s^2}$ .

We use a discount rate  $\lambda$  and provide a reward as discussed below, after Eq. 8. Training is performed in an episodic fashion with each episode lasting time  $T$ . The kinematic variables are updated with time step  $dt$ , and actions are taken with time step  $dt_{\text{act}} \geq dt$ . Movement constraints are imposed by restricting the set of actions to three values,  $u/a\omega \in \{-\alpha, 0, \alpha\}$ . The state space has four dimensions:  $r, \theta, -\ln q_1/q_2$ , and  $-\ln q_3/q_2$ . We discretize our state space using a nonoverlapping tile coding scheme (23). We refer to *SI Appendix* for hyperparameter values, details about the state space architecture, and the case of two sensors.

We use the SARSA (state-action-reward-state-action)  $Q$ -learning algorithm (23), which learns the so-called  $Q$  function, that is the value function for each action in a given state:

$$Q_\pi(r, \mathbf{q}, u) = \bar{\gamma}(r)dt + [1 - (\lambda + \bar{\gamma}(r))dt] V_\pi(r + vdt, \mathbf{q} + \dot{\mathbf{q}}dt) \quad [8]$$

where  $V_\pi(r, \mathbf{q}) = \sum_u Q_\pi(r, \mathbf{q}, u)\pi(u|r, \mathbf{q})$ , and the index  $\pi$  highlights the dependence on the probabilistic policy  $\pi(u|r, \mathbf{q})$ . The above equation differs from the standard SARSA update by the addition of  $\bar{\gamma}dt$  in the discount term, which is due to our formulation of the search as a continuing process conditional on no contact with the trail. Alternatively, one may provide a unit reward when the trail is found, stop the episode, and start over.

However, the credit assignment problem in goal-oriented tasks makes the training (Eq. 9 hereafter) problematic, even though the final optimal policy is equivalent (e.g., ref. 24). Our formulation circumvents both issues by 1) giving a local reward  $\bar{\gamma}dt$  rather than a final one, which addresses the credit assignment, and 2) including the detection probability  $\bar{\gamma}dt$  into the discount rate to account for the condition of no contact.

To learn the  $Q$  function as defined in Eq. 8, we use a “softmax” training policy: that is,  $\ln \pi(u|r, \mathbf{q}) \propto \hat{Q}(r, \mathbf{q}, u)/T_{\text{explor}}$ , where  $\hat{Q}$  is the current estimate of the  $Q$  function and  $T_{\text{explor}}$  is a “temperature” parameter that is annealed as training progresses to allow for sufficient exploration of actions. Given an action  $u$  at state  $(r, \mathbf{q})$  and a subsequent action  $u'$  at state  $(r', \mathbf{q}')$ ,  $\hat{Q}$  is updated during training as

$$\hat{Q}_\pi(r, \mathbf{q}, u) \rightarrow \hat{Q}_\pi(r, \mathbf{q}, u)(1 - \eta) + \eta(\bar{\gamma}(r)dt + (1 - (\lambda + \bar{\gamma}(r))dt)\hat{Q}_\pi(r', \mathbf{q}', u')) \quad [9]$$

where  $\eta$  is the learning rate. The function  $\hat{Q}$  obtained at the end of the training period yields a search strategy as  $\pi^*(r, \mathbf{q}) = \arg \max_u \hat{Q}(r, \mathbf{q}, u)$ .

We applied the algorithm just described for a range of values of  $v, \alpha$ , and  $d$  (half the distance between sensors for the two-sensor case). For each parameter set, we obtain a search strategy, the corresponding probability  $\Gamma_C(T)$  of missing the trail in time  $T$ , and the expected number of samples to find the trail. Comparing  $\Gamma_C(T)$  on a test set for different numbers of training episodes (*SI Appendix, Fig. S1B*) shows that nontrivial learning takes place, typically saturating at  $\sim 10^5$  training episodes.

**Casting in a Minimal Model of Sector Search.** In order to establish the relation between casting and marginal value theory, here we propose a minimal model of sector search. The model lends to an analytical solution, which allows us to quantify how the frequency of casting and the efficiency of search depend on the movement and computational constraints imposed upon the tracker.

We consider the same setting as the above episodic RL framework, where the tracker is searching for the trail over a sector after losing contact with it. To focus on casting, we analyze the behavior of the searcher after an initial forward excursion along the most likely heading. This surge decreases the probability weight  $q$  of the mode at  $\phi = 0$  and yields a symmetric bimodal posterior distribution concentrated at the two modes,  $\pm\phi_0$  (typically  $\phi_0 \sim \sigma$ ). The resulting model is equivalent to Eq. 5 with  $K = 2$ ,  $\mu_1 = -\phi_0, \mu_2 = \phi_0$ , and  $s$  small. The searcher moves radially as  $r(t)$  (fixed), controls its tangential speed  $u = r\dot{\theta}$ , receives a unit reward when it finds the trail, and incurs a movement cost per unit time  $\mu u^2/2$ , where  $\mu$  sets the movement constraint. The reward and cost are discounted at a rate  $\lambda$ . The two-dimensional state space of the agent consists of  $\theta$  and the probability  $q$  of finding the target at  $\phi_0$  (the probability is  $1 - q$  at  $-\phi_0$ ). For full details, we refer to *SI Appendix*.

The above model is exactly solvable. An optimal searcher exhibits oscillations between  $-\phi_0$  and  $\phi_0$  (casting) until it finds the trail (*SI Appendix, Fig. S2A*). After an initial transient, the searcher traverses a loop in state space  $(\theta, q)$ , alternating between sampling at  $-\phi_0$  or  $\phi_0$  and casting to the other side:

$$(\phi_0, q_s^*) \xrightarrow{\text{cast}} (-\phi_0, q_s^*) \xrightarrow{\text{sample}} (-\phi_0, 1 - q_s^*) \xrightarrow{\text{cast}} (\phi_0, 1 - q_s^*) \xrightarrow{\text{sample}} (\phi_0, q_s^*), \quad [10]$$

where  $q_s^*$  is the optimal switching probability at which the searcher stops sampling at  $\phi_0$  and traverses to  $-\phi_0$ . The speed of traversal from  $\phi_0$  to  $-\phi_0$  is determined by balancing the cost of traversing quickly and the potential value at  $-\phi_0$  discounted due to the limited time horizon.

Intuitively, the searcher casts when the marginal value of continuing to sample at  $\phi_0$  is just outweighed by the marginal value that the searcher receives if it stops sampling, traverses from  $\phi_0$  to  $-\phi_0$ , and samples at  $-\phi_0$ . Balancing the marginal value of these two possibilities then yields the optimal switching probability  $q_s^*$ . For small  $q_s^*$ , we derive (*SI Appendix*)

$$q_s^* \approx \frac{\lambda}{1 + \lambda} \frac{\zeta^2}{1 + \zeta + (1 + \lambda)\zeta^2 - \zeta^3}, \quad [11]$$

where  $\zeta \equiv 1 - \sqrt{2\phi_0^2 \mu \lambda (1 + \lambda)} > 0$ . When  $\zeta < 0$ , the movement cost outweighs the value the agent may receive, and the optimal strategy is to simply not move.  $q_s^*$  get smaller, and thus, the searcher casts less frequently with increasing time horizon ( $\lambda \ll 1$ ) or movement costs ( $0 < \zeta \ll 1$ ). The probability of not detecting the trail after time  $t$  is given by

$$\Gamma_C(t) \simeq e^{-\int_0^t \frac{dq}{r(s)}}, \quad [12]$$

where  $I$  is interpreted as the rate of information acquisition. Its dependence on  $\mu$ ,  $\lambda$ , and  $\phi_0$  is shown in *SI Appendix, Fig. S2B*. As expected, increasing movement cost (decreasing  $\zeta$ ) decreases how quickly information is acquired. Similarly, a large time horizon ( $\lambda$  small) makes the agent sample at  $\pm\phi_0$  longer and cast slower, decreasing the rate of information acquisition. For a constant radial speed  $v$ , the probability  $\Gamma_C(t) \simeq t^{-1/\nu}$  decreases as a power law, which arises quite generally from the sector search geometry as discussed in the text.

**Trail Statistics.** If an agent makes contact with the trail at two points separated by distance  $L$ , statistical and geometric information is encoded in the propagator  $P(\phi_L, \phi_0)$ , where  $\phi_L, \phi_0$  are the trail headings at the two points (measured relative to the line joining them). If  $H$  points of contact are remembered, then the propagator can be used to compute the posterior distribution of the heading:  $P(\phi) = \int \mathcal{D}\phi_i P(\phi|\phi_1)P(\phi_1|\phi_2) \dots P(\phi_{H-2}, \phi_{H-1})$ . We introduce an ensemble of trails (which we call the GWLC ensemble) that have persistence in heading and curvature, quantified respectively by the parameters  $\kappa$  and  $\lambda$  (distinct from the discount rate used for RL) and a typical radius of curvature,  $\xi$ . Typical samples of trails from this ensemble are presented in *SI Appendix, Fig. S3A*. The WLC ensemble (16, 17) previously introduced for polymers is a special case with  $\xi = \infty$ .

The propagator  $P(\phi_L, \phi_0)$  takes into account all the trails constrained to pass through two contact points, weighted by their probability. We affix a coordinate system, where the line joining the two contact points is along the  $x$  axis. Specifically, the trail's  $y$  coordinate,  $y_x$ , satisfies the constraint  $y_0 = y_L = 0$ ; the small-angle approximation is used so that  $\dot{y}_x = \phi_x$ ; and the trail has end-point headings  $\phi_0, \phi_L$ . We define

$$P(\phi_L, \phi_0) = Z^{-1} \int \mathcal{D}\phi_x \mathcal{D}\chi_x e^{-\mathcal{E}(\{\phi_x, \chi_x\})}, \quad [13]$$

where  $Z$  is a normalization constant, and the integral is over all possible headings and curvatures  $\phi_x, \chi_x$  at the various positions  $x$ . The action  $\mathcal{E}$  of a path is given by

$$\mathcal{E}(\{\phi_x, \chi_x\}) = \frac{1}{2\kappa} \int_0^L dx (\dot{\phi}_x - \chi_x)^2 + \frac{\lambda \xi^2}{4} \int_0^L dx \left( \dot{\chi}_x + \frac{\chi_x}{\lambda} \right)^2 + \frac{\xi^2 \lambda_0^2}{2}. \quad [14]$$

The model Eq. 14 is a Gaussian process. Since symmetry dictates that  $\langle \phi_L \rangle = \langle \phi_0 \rangle = 0$ , it follows that  $P(\phi_L, \phi_0)$  is defined entirely by the variance  $\sigma^2(L) \equiv \langle \phi_L^2 \rangle = \langle \phi_0^2 \rangle$  and by the correlation  $\rho(L) \equiv \langle \phi_0 \phi_L \rangle / \sigma_L^2$ .

We present the full calculation of  $\sigma(L)$  and  $\rho(L)$  in *SI Appendix*. In summary, the first integral over  $\chi_x$  can be performed using the Gaussian integral formula and leads to an effective action in  $\phi_x$ . The Euler-Lagrange equation of this effective action then yields extremal paths ("splines") that minimize the effective action (*SI Appendix* has details). The splines have the form

$$y_x = a_0(L-x) + a_1 x + c_0 \frac{(L-x)^3}{6} + c_1 \frac{x^3}{6} + \lambda^2 d_0 e^{-(L-x)/\lambda} + \lambda^2 d_L e^{-x/\lambda}, \quad [15]$$

where the constants in the above equation are set so that  $y_0 = y_L = 0$  and can be expressed in terms of  $\phi_0, \phi_L$ . *SI Appendix, Fig. S3B* shows the splines between contact points spaced at increasing intervals. Plugging Eq. 15 in the effective action, we obtain

$$\sigma^2(L) = \frac{\kappa \Omega^2 \lambda^2}{L} \frac{b_1 + b_2}{2}, \quad \rho(L) = \frac{b_2 - b_1}{b_2 + b_1}, \quad [16]$$

where the two functions  $b_1$  and  $b_2$  are

$$b_1 = \frac{L^2}{2} - \frac{\lambda L}{2\Omega^2 V^2} \left( 1 - e^{-\frac{L}{\lambda}} \right) \quad [17]$$

$$b_2 = \frac{L^2}{6} - \frac{\lambda(L+2\lambda)}{2\Omega^2 V^2} \left[ (L-2\lambda) + e^{-\frac{L}{\lambda}} (L+2\lambda) \right],$$

and  $V = \sqrt{\kappa \xi^2 \lambda / 2}$ ,  $\Omega^2 = \lambda^{-2} + V^{-2}$ . The variance  $\sigma(L)$  and the correlation  $\rho(L)$  are plotted in *SI Appendix, Fig. S3C and D*. Three distinct regimes are apparent. For  $L/\lambda \ll 1$ , we can approximate  $\sigma^2(L) \approx \kappa L/3 + L^2/4\xi^2$ , which reflects diffusive  $\propto L$  and curvature-dominated  $\propto L^2$  scalings. When diffusion dominates,  $\rho(L) = -1/2$ , whereas  $\rho(L) \approx -1$  when curvature dominates. The perfect anticorrelation in the curvature-dominated regime is intuitive as the line joining the two points of contact can be viewed as the

chord of a circle with radius  $\xi$ , whereas  $\sigma(L) = L/2\xi$  is the angular deviation of the trail around this chord. When  $L/\lambda \gg 1$ , the heading is randomized over many correlation lengths, and the diffusive scaling  $\sigma^2(L) \approx 2\lambda L/3\xi^2$  is recovered.

While  $P(\phi_L|\phi_0)$  (the interpolation model) is required to integrate past information, search strategies also require the forward propagator  $P(\phi_L, y_L|\phi_0)$ , which keeps track of trail headings and locations while the agent searches for the trail. The methods described above can be used again (*SI Appendix* has details) to yield extremal paths as Eq. 15 and the three quantities  $\langle \phi_L^2 \rangle_{\text{fwd}}$ ,  $\langle y_L^2 \rangle_{\text{fwd}}$ ,  $\langle \phi_L y_L \rangle_{\text{fwd}}$ . These quantities fully describe  $P(\phi_L, y_L|\phi_0)$ . We validated our interpolation model Eq. 16 and the forward model using numerical simulations (*SI Appendix, Fig. S3E and F*).

**The Nondetection Probability during Surge and Cast.** We introduce a sector search strategy that allows us to quantify the nondetection probability taking into account the full dynamics of the trails and yields intuition on the factors that contribute toward losing the trail entirely. We suppose the radial speed  $v$  is fixed and the tangential speed  $u \gg a\omega$ . In other words, the agent casts rapidly within a conical envelope of semiaperture angle  $\sigma\Theta_0$ , where  $\sigma$  is the prior uncertainty of the trail's heading and  $\Theta_0 \gtrsim 1$  (*SI Appendix, Fig. S4A*). To simplify the presentation, we assume curvature-dominated trails [i.e.,  $\sigma(L) = L/2\xi$ ], although the arguments below are general.

The rapid casting limit  $u/a\omega \gg 1$  allows us to compute the nondetection probability at distance  $r$  from the most recent contact point defined by Eq. 3:

$$\Gamma_C(r) = \left\langle e^{-\frac{\omega}{v} \int_0^r dx \gamma(x, y(x))} \right\rangle_{y(x)}, \quad [18]$$

where the expectation is over the full ensemble of trails  $\{y(x)\}$  that pass through past detection points. Here, we provide intuition for the detection rate  $\omega\gamma(x, y(x))$  and refer to *SI Appendix* for full details. For small  $x$  (i.e.,  $x \lesssim a/2\sigma\Theta_0$ ),  $\gamma(x, y(x)) = 1$  for most trails as the sensor size  $a$  spans the entire casting envelope  $2\sigma x\Theta_0$ . In the casting regime ( $x \gtrsim a/2\sigma\Theta_0$ ), since the time spent on the trail in a single cast is  $a/u$ , the probability of nondetection per crossing is  $e^{-a\omega/u}$ . After  $n$  crossings, the nondetection probability is then  $e^{-na\omega/u}$ . As the agent moves a distance  $dx$  in the radial direction, it crosses the trail  $n = udx/2v\sigma x\Theta_0$  times [i.e.,  $\gamma(x, y(x)) = a/2\sigma x\Theta_0$ ], which is independent of  $u$ . From Eq. 18, these relations yield an exponential  $\Gamma_C(r)$  during the initial surge followed by a power law,  $r^{-\beta}$  with exponent  $\beta \equiv a\omega/2\sigma v\Theta_0$ . This heuristic argument aligns well with the results of numerical simulations in *SI Appendix, Fig. S4B*.

The spline formulation of the GWLC yields a geometric picture of the dynamics. Intuitively, the nondetection probability (and thus, the posterior probability from Eq. 1) is large if the trail has a large probability of escaping the casting envelope before it is found by the tracker. In order to "escape" the casting envelope, the trails that are initially well within the casting envelope have to bend significantly. This bending incurs a cost in the action Eq. 14, which reduces with increasing  $r$ . The nondetection probability conditional on initial trail heading,  $\Gamma_C(r|\phi)$ , flattens out (i.e., the detection rate vanishes) when a significant fraction of trails escapes the casting envelope, including trails with initial heading  $\phi = 0$  (Fig. 3C).

This geometric picture yields a length scale  $L_{\text{in}}$ , which is the distance at which the trails initially along the most likely heading  $\phi = 0$  escape the casting envelope. For curvature-dominated trails, escaping trails should deviate by an angle  $\sigma\Theta_0 \simeq L_{\text{in}}/2\xi$ , which gives  $L_{\text{in}} \simeq 2\sigma\xi\Theta_0$ . The probability of losing the trail can then be estimated as  $\epsilon_{\text{in}} = \Gamma_C(L_{\text{in}})$ . An additional contribution to the flattening of  $\Gamma_C(r)$  comes from the trails that are headed outside of the cone  $|\phi| > \Theta_0$  even before reaching  $L_{\text{in}}$ . The probability  $\Gamma_C(r)$  would then flatten out at a different length scale  $L_{\text{out}}$ , where  $\Gamma_C(L_{\text{out}}) = \epsilon_{\text{out}}$ . The relative contributions of the two mechanisms are discussed in detail in *SI Appendix*, and they are validated by using numerical simulations as shown in *SI Appendix, Figs. S4C and S5B*.

Due to the power law tail for  $\Gamma_C(r)$  with exponent  $\beta$ , the mean distance to find the trail,  $\bar{r}$ , is determined by the upper or lower cutoff for  $\beta < 1$  or  $\beta > 1$ , respectively. Since the upper cutoff is much larger than the lower cutoff, it is more convenient for the tracker to use  $\beta > 1$ . By using the self-consistency condition  $\sigma = \bar{r}/2\xi$  in the curvature-dominated regime and the relation  $\bar{r} \simeq v\omega^{-1} (1 + e^{-\beta}/(\beta - 1))$ , we obtain for the maximum speed

$$v^* = \omega \sqrt{a\xi/\Theta_0} \max_{\beta} \left( \beta + \frac{\beta e^{-\beta}}{\beta - 1} \right)^{-1} \simeq 0.47\omega \sqrt{a\xi/\Theta_0}, \quad [19]$$

where the maximum occurs at  $\beta^* = 1.63$  (*SI Appendix, Fig. S5C*). The mean number of samples to find the trail at this optimum is  $\approx 1.3$ , indicating that the optimal strategy is for the agent to move slow enough that it typically finds the trail within one or two samples.

Intuition for the above results holds quite generally, including for other statistical ensembles of trails and for nonconical casting envelopes. The latter is relevant because we expect that a slowdown in the radial direction and a widening casting envelope may increase the likelihood of finding the trail. Relaxing the constraints of fixed radial speed and conical envelopes constitutes the aim of the next section.

**Bellman Optimization of the Casting Strategy.** Our final step is to formulate the problem of optimizing the casting strategy. The geometry of the search is shown in *SI Appendix, Fig. S6A*. To simplify, we parametrize the path by the set of turning points  $\{r_i, \theta_i\}$ . We assume a fixed speed  $v$  (note that the average radial speed depends on the strategy), which is set later based on the probability of losing the trail. As discussed in the text, we optimize the average tracking speed,  $\langle L/T \rangle$ , after constraining the IDI,  $\langle L \rangle$ , using a Lagrange multiplier  $\Lambda$  (Eq. 4). This optimization can be recast as a Bellman-type dynamic programming problem by breaking the problem up into discrete steps, each corresponding to a cast from  $\{r_i, \theta_i\} \rightarrow \{r_{i+1}, \theta_{i+1}\}$  (denoted  $\{r', \theta'\} \rightarrow \{r, \theta\}$  below for conciseness):

$$\mathcal{V}(r', t', P'(\phi)) = \max_{r, \theta} \left[ \left( \frac{r}{t} - \Lambda r \right) (1 - \bar{\Gamma}_{r, \theta}) + \bar{\Gamma}_{r, \theta} \mathcal{V}(r, t, P_{r, \theta}(\phi)) \right], \quad [20]$$

where the required reward function in Eq. 4 is  $V_{v, \Lambda} = \mathcal{V}(0, 0, P_0(\phi))$  with the Gaussian normal prior  $P_0(\phi) = \mathcal{N}(0, \sigma^2)$ . The first term in the square bracket returns the reward if the trail is found (weighted by the detection probability  $1 - \bar{\Gamma}_{r, \theta}$ ), while the second term advances to the next cast if detection failed. The Bellman equation thus relates  $\mathcal{V}$  before and after a cast by updating the current state variables,  $r', t', P'(\phi)$ . The time elapsed is updated as  $t \rightarrow t' + (r - r' + 2r|\theta|)/v$ , where we approximate  $|\theta' - \theta| \sim 2|\theta|$  for simplicity. The prior  $P'(\phi)$  is updated using Bayes' rule  $P_{r, \theta}(\phi) = \bar{\Gamma}_{r, \theta}(\phi)P'(\phi)/\bar{\Gamma}_{r, \theta}$ , where  $\bar{\Gamma}_{r, \theta}(\phi)$  is the nondetection probability given the initial trail heading  $\phi$  and  $\bar{\Gamma}_{r, \theta} = \langle \bar{\Gamma}_{r, \theta}(\phi) \rangle_\phi$  is the normalization.

It remains to calculate  $\bar{\Gamma}_{r, \theta}(\phi)$ , which depends on the forward model of the trail. We consider generally that the trail's azimuthal position expands as  $\sigma_{\text{fwd}}(r)$  [i.e., a trail initially headed along  $\phi$  is located at the

azimuthal position,  $\phi + \Delta\phi$ , where  $\Delta\phi \sim \mathcal{N}(0, \sigma_{\text{fwd}}^2(r))$ ]. For the GWLC model, we show in *SI Appendix* that  $\sigma_{\text{fwd}}(r) = \sigma(r)$ , where  $\sigma(r)$  is given in Eq. 16. Since the time spent on the trail within the casting envelope in a single cast is  $a/v$ , we have  $\bar{\Gamma}_{r, \theta}(\phi) = e^{-a\omega/v}$  if  $|\phi + \Delta\phi| < \theta$  and equal to unity otherwise. Since  $\Delta\phi$  is normal distributed, we have  $\langle \mathbb{I}(|\phi + \Delta\phi| < |\theta|) \rangle_{\Delta\phi} = \Phi\left(\frac{\theta - \phi}{\sigma_{\text{fwd}}(r)}\right) - \Phi\left(\frac{-\theta - \phi}{\sigma_{\text{fwd}}(r)}\right)$ , where  $\Phi$  is the normal cumulative distribution function. We finally approximate  $\bar{\Gamma}_{r, \theta}(\phi) \approx e^{-\frac{a\omega}{v} \langle \mathbb{I}(|\phi + \Delta\phi| < |\theta|) \rangle_{\Delta\phi}}$  and compute the expectation over  $\phi$  in  $\bar{\Gamma}_{r, \theta}$  numerically. The previous approximation greatly simplifies the optimization as  $P(\phi)$  is then a sufficient statistic for past measurements (thus allowing the decomposition in Eq. 20) and yet, captures the effect of the trail's widening trajectory on the optimized casting strategy.

At each casting step, we optimize (using standard black box optimization methods using the SciPy library in Python) for  $\Delta r = r - r' > 0$  and  $\theta$  by expanding Eq. 20 two steps further into the future. The optimization then involves six variables, the immediate pair  $\Delta r, \theta$  and the subsequent two pairs. The optimized immediate pair is then used for updating as detailed above, and the process is repeated. Optimizing more than two steps did not yield different results as the  $\bar{\Gamma}_{r, \theta}$  factors introduce an effective planning horizon, effectively suppressing contributions from future rewards beyond two steps. For less than two steps, optimization is "greedy," and the corresponding landscape was found to have a qualitatively different, shallow landscape around the optimum. After completing the optimization procedure, the constraint of fixed  $\langle L \rangle$  is imposed by varying  $v$  and the Lagrange multiplier  $\Lambda$  (*SI Appendix, Fig. S6B*).

**Data Availability.** There are no data underlying this work.

**ACKNOWLEDGMENTS.** We thank Venkatesh Murthy for stimulating discussions throughout our work. This research was supported in part by NSF Grant PHY-1748958, NIH Grant R25GM067110, and Gordon and Betty Moore Foundation Grant 2919.02. G.R. was partially supported by NSF-Simons Center for Mathematical & Statistical Analysis of Biology at Harvard Award 1764269 and the Harvard Quantitative Biology Initiative. B.I.S. was supported by NSF Grant PHY-1707973.

1. R. W. Draft, M. R. McGill, V. Kapoor, V. N. Murthy, Carpenter ants use diverse antennae sampling strategies to track odor trails. *J. Exp. Biol.* **221**, jeb185124 (2018).
2. A. Thesen, J. B. Steen, K. B. Døving, Behaviour of dogs during olfactory tracking. *J. Exp. Biol.* **180**, 247–251 (1993).
3. P. G. Hepper, D. L. Wells, How many footsteps do dogs need to determine the direction of an odour trail? *Chem. Senses* **30**, 291–298 (2005).
4. J. Porter et al., Mechanisms of scent-tracking in humans. *Nat. Neurosci.* **10**, 27–29 (2007).
5. A. G. Khan, M. Sarangi, U. S. Bhalla, Rats track odour trails accurately using a multi-layered strategy with near-optimal sampling. *Nat. Commun.* **3**, 703 (2012).
6. U. Bhattacharyya, U. S. Bhalla, Robust and rapid air-borne odor tracking without casting. *eNeuro* **2**, ENEURO.0102-15.2015 (2015).
7. H. C. Berg, *E. coli in Motion* (Springer Science & Business Media, 2008).
8. N. Riman, J. D. Victor, S. D. Boie, B. Ermentrout, The dynamics of bilateral olfactory search and navigation. *SIAM Rev.* **63**, 100–120 (2021).
9. C. David, J. Kennedy, A. Ludlow, Finding of a sex pheromone source by gypsy moths released in the field. *Nature* **303**, 804–806 (1983).
10. J. Murlis, J. S. Elkinton, R. T. Carde, Odor plumes and how insects use them. *Annu. Rev. Entomol.* **37**, 505–532 (1992).
11. A. Mafra-Neto, R. T. Cardé, Fine-scale structure of pheromone plumes modulates upwind orientation of flying moths. *Nature* **369**, 142–144 (1994).
12. M. Vergassola, E. Villermaux, B. I. Shraiman, 'Infotaxis' as a strategy for searching without gradients. *Nature* **445**, 406–409 (2007).
13. A. Celani, E. Villermaux, M. Vergassola, Odor landscapes in turbulent environments. *Phys. Rev. X* **4**, 041015 (2014).
14. E. L. Charnov, Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* **9**, 129–136 (1976).
15. D. W. Stephens, J. R. Krebs, *Foraging Theory* (Princeton University Press, 1986), vol. 1.
16. M. Doi, S. F. Edwards, S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, 1988), vol. 73.
17. J. F. Marko, E. D. Siggia, Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
18. D. J. Field, Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
19. D. L. Ruderman, W. Bialek, Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.* **73**, 814–817 (1994).
20. E. P. Simoncelli, B. A. Olshausen, Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
21. B. Wark, B. N. Lundstrom, A. Fairhall, Sensory adaptation. *Curr. Opin. Neurobiol.* **17**, 423–429 (2007).
22. A. Mathis et al., DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
23. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 2018).
24. G. Reddy, J. Wong-Ng, A. Celani, T. J. Sejnowski, M. Vergassola, Glider soaring via reinforcement learning in the field. *Nature* **562**, 236–239 (2018).