

Supplementary Information for

Spatial planning with long visual range benefits escape from visual predators in complex naturalistic environments

Mugan & MacIver

Supplementary Methods 1

Supplementary Notes 1–4

Supplementary Discussion 1–3

Supplementary Figures 1–11

Supplementary Tables 1–4

Supplementary References

Supplementary Methods

1 Lacunarity analysis

The effects of spatial distribution of ecological structures (e.g. vegetation) have been studied by ecologists to understand the differential dispersal ability of predators and prey¹, biodiversity, and intra- or interspecific interactions². One important variable that influences these phenomena is the heterogeneity in gaps between structures (e.g. the gaps that result from spatially discontinuous biogenic structures such as trees or shrubs). Lacunarity analysis has been used to characterize the pattern and spatial scaling of gap sizes in landscapes^{3,4}. Intuitively, it is a measure of ‘gapiness’ or ‘hole-iness’ of geometric structure⁵. Lacunarity was originally proposed by Mandelbrot⁶ as a measure to differentiate between objects with identical fractal dimensions but differing appearances. Despite lacunarity being developed for use with fractals, it also applies to non-fractal structure⁷.

Formally defined, lacunarity is a scale-dependent deviation of a geometrical object or pattern from invariance under translation, or from homogeneity^{8,9}, used in ecology^{10–12}. Objects with low lacunarity are finely textured and homogeneous with respect to the size distribution and spatial arrangement of gaps. Objects with high lacunarity are coarsely textured with high heterogeneity in size distribution and spatial arrangement of gaps¹³. Estimates of lacunarity are therefore dependent on both the spatial scale (i.e. resolution), and the spatial extent of the region to be characterized^{3,14}. Given that we are using environments that vary from open (no clutter) to highly cluttered, the calculated lacunarities are between ∞ and near 0, respectively.

A common algorithm that is used to calculate lacunarity is the ‘gliding-box algorithm’⁷, which is similar to the box-counting method commonly used in fractal dimension analyses. Information on the algorithm is provided in the methods. Lacunarity mathematically is defined as: $\Lambda(r) = \frac{\text{Var}(S)}{\mathbb{E}[S]^2} + 1$, where S is the occupied sites by the variable of interest (such as occlusions in our case), r is the box size. Thus, it is a variance to mean squared ratio of the number of non-occupied cells. Some properties that arise are³:

1. As box size r increases, the average box mass ($n(S, r)$) also increases, causing a decrease in variance and therefore, a decrease in lacunarity at that box size $\Lambda(r)$.
2. As the mean number of occlusions goes to zero ($\rightarrow H(S) = 0.0$), $\text{Var}(S) \rightarrow 0$, $\mathbb{E}[S] \rightarrow 0 \implies \Lambda(r) \rightarrow \infty$.
3. For a given fraction of cell occupancy, higher lacunarity represents fewer but overall larger gaps.

These properties imply that there is a large amount of information in looking at the distribution of lacunarities as a function of a wide range of box sizes. Variance ($\text{Var}(S)$) will approach to 0 ($\implies \Lambda(r) \rightarrow 1$) for a given box size if the map has random structure at that spatial scale. Self-similar fractals produce lacunarity curves with a slope equal to the fractal dimension D minus the Euclidean dimension E ^{4,7}. In other cases, the rate of decline in lacunarity curves (lacunarity vs box size) depends on the presence and heterogeneity of coarser-scale patterns. Curves

that decline with less slope as box size is increased indicates heterogeneity at coarser-scale patterns^{3,4}. Conversely, curves that decline with more slope as box size increases indicates spatial randomness or uniformity at coarser-scale patterns^{3,4}. In our lacunarity analysis, instead of plotting individual curves, we provide the natural log of mean lacunarity over various box sizes ($\Lambda_{\text{avg}} = \frac{\sum_{g=1}^{N_r} \Lambda(r)}{N_r}$).

In the lacunarity applications of interest (e.g. ^{15–17}), the top down view of the environment is being used. We are also assuming that the non-occupied regions do not obstruct the vision of the agents in the space. This may not be true depending on the height of the agent’s visual system: for example, a region with trees and intervening grassland taller than the agent’s body. It is important to note that within our study, while occlusions blocked the line of sight, neither the prey nor the predator relied on vision to locate the occlusions—this is only relied on for each agent’s detection of the other agent in the arena. The acting agent’s having a perfect map of the environment means that they were navigating and (for the prey) imagining trajectories in the environment from a perspective-independent representation. Therefore, in conducting our lacunarity analysis, we use a the perspective independent map of the environment.

Supplementary Note

1 Cognitive map assumptions for both the predator and the prey

In the case of our assumption that the predator and prey have a perfect cognitive map (in the predator's case, excluding the location of the safety point that the prey is trying to reach), our modeling choice is driven by a desire to surgically isolate the relative properties of planning versus habit. First, the assumption that the predator has a perfect cognitive map (minus the prey's goal) is the most detrimental to the prey's survival. Here we consider three different cases and their possible outcomes under different levels of cognitive map fidelity.

Case 1: No cognitive map. Neither the predator nor the prey has a cognitive map, and instead, the prey learns a cognitive map by acting and interacting with its environment. Without a model to begin with, the prey agent can not use its internal model to reason about the future and to seek positive outcomes while avoiding the adverse consequences of trial-and-error in the real environment. During the building of the model, we would have to make a large number of empirically poorly-justified modeling choices on learning rates, and the number of exposures to a given random environment to enable the planning algorithm to use a high fidelity model to reason with. The predator in such a case could not be a simple reactive agent, since similar to the prey, the predator would have to learn the environment before choosing its actions. This would cause both the prey and the predator to be learning agents and would necessitate the predator to be more than just a reflex agent.

Case 2: Inaccurate cognitive map updated with learning. Similar to *Case 1*, this scenario would add many more modeling choices. Methods for planning or model-based learning that are robust against model imperfections is an active research area¹⁸. In such cases, the prey would still use its own model to make decisions but would also have to reason about its imperfect predictions. Given the way in which we have implemented planning, this would cause many actions to be based on a possibly incorrect value function that corresponds to this inaccurate cognitive model. Therefore the decisions would, in the limit of number states forward simulated, not be optimal. If the decisions fail to be optimal, in the context of this scenario (irreversibility since death occurs upon capture), there may not be much to learn; but inasmuch as learning can occur, the issues raised about the complexity of our study and poorly justified parameters enter in. If there is no learning within this case, the decisions of the prey will not reach optimality over time. The predator may not be as affected by the imperfect cognitive map. However, the amount of imperfect information about the location of the occlusions could cause it to get stuck in regions, decreasing its threat to the prey.

Case 3: Perfect cognitive map. The attraction of this choice for our study is avoiding uncertain parameterization of learning the model, and avoiding uncertain parameterization of having an inaccurate model. We believe this enables us to surgically isolate the impact of planning in the context of a generous assumption of learning having previously been completed, and completed with perfect fidelity to the underlying structure (not counting the predator) where that structure is necessarily static (since learning cannot occur to accommodate a dynamic

environment, and a dynamic environment without learning would quickly lead to an inaccurate model—case 2).

2 Visual conditions of the prey and predator

The simulations in which the prey and predator’s visual range differ are 1) for the empty world simulations (pseudo-aquatic) (Fig. 1a), and 2) simulations conducted in a set of mid-entropy environments (pseudo-terrestrial) where the prey’s visual range was explicitly controlled (1, 3, or 5 cells ahead) (Supplementary Figure 6). In all simulations that included occlusions (pseudo-terrestrial conditions), independent of the prey’s visual range (whether full or limited), the prey and predator could not observe one another if there was an occlusion on the line between them. This disconnect between the predator’s and the prey’s visual range in the aforementioned simulations ensured that there would be observation events between the predator and the prey.

The predator policy in most time steps was aggressive pursuit towards the prey’s location. In prior simulations (not shown in the paper), the predator, similar to the prey, had a visual cone in the direction of travel. When the prey was not in view, the predator created a belief distribution of all possible prey locations that were not inside its visual cone (as is done for the prey with respect to the predator). In order to reason about the value of taking an action in these cases, during planning the prey sampled as many predator locations as the number of states forward simulated (100–5000) with probability proportionate to the belief state distribution. For computational expedience, the predator only sampled one prey location proportionate to the belief state distribution, and chose an action to minimize its distance to that sampled prey location. Furthermore, because the predator did not have access to a model for prey action selection (as the prey had for the predator) nor access to the goal location, the pruning and propagation of the predator’s belief about the prey location was based on the assumption that the prey acted randomly. This in turn resulted in most episodes to terminate without any observation events (i.e. neither the prey nor the predator ever observed each other) since the predator’s actions were effectively random. Given that we did not provide the predator with the location of the goal, or the overall aim of the prey, we could not use a perhaps more natural prey model (e.g. the prey is choosing actions that minimizes its distance to the goal).

Even in the case of a more reasonable prey action selection model for the propagation of the predator’s belief distribution (such as using the goal location), we expect that the predator and the prey would only observe each other when the prey was near to the goal position. Therefore, unless the predator was positioned near the prey at the start of the simulation, it would not have enough time steps (actions left) to reach the prey before the prey reached the goal position. The only way to enable the predator to make good decisions, as well as maintain a well supported belief distribution about the prey’s location, is to simulate both the prey and the predator as planners. This would allow the predator to choose actions based on more than one sample of the prey location, and model the prey action selection as an independent planner. While this is clearly ideal, it is not computationally tractable given the way planning is implemented (as recognized in the literature¹⁹): Both the prey and the predator would have to plan an action at each imagined step for each other. However, given the number of action

choices, the state-space of the simulated environment, and how temporally extended the task is, such an implementation is not feasible.

In sum, we have made the predator's visual range larger than the prey's because 1) it is not feasible to simulate both agents as planners; and 2) in the absence of that framework, most of the trials ended with no observation events between the predator and prey. Extending the predator's visual range avoids this problem; it provides some of the benefits of simulating the predator as a planner without the computational intractability of so doing.

How does this impact our results? First, note that the predator having larger visual range makes it more challenging for the prey to survive. As such, it is a harsher test of the efficacy of either habit- or plan-based action selection in the prey. Our expectation is that as we reduce the predator's visual range, survival rate would increase, but our key findings—a) thigmotaxis and b) there is no advantage of planning in low visibility open environments—would not be affected.

3 Arena size

One aim of this study was to examine whether the increased visual range resulting from aerial vision advantages plan-based action selection. As can be seen from Supplementary Figure 1, after the vertebrate invasion of land, visual range increased by a factor of roughly 100. Given our arena size, we have simulated a visual range increase of, at most, a factor of 15. While ideally we would have a larger arena to encompass the actual range increases from aquatic to aerial vision, as a consequence of the planning algorithm used along with other computational limitations, we chose to use a 15 x 15 environment. This arena size, goal location (7, 14), and the prey start location (7, 0) ensured that the space being planned over during plan-based action control was as large as feasible; if the goal, for example, was closer, the environment would effectively be reduced in size. The infeasibility of larger environments is due to the state-space of the problem. For example, by going from a 15 x 15 environment to a 20 x 20 environment, the number of states increases from approximately 10^{36} to 10^{52} . A trial in a pseudo-terrestrial environment (15 x 15) requires roughly 2 hours of compute time (there are 250,000 such trials in total in the study).

4 Feasibility of the planning algorithm

The planning algorithm implemented in this contribution is not biologically feasible. As a coarse indicator of this, in the worst case for our gridworlds, where there are only four possible actions at each cell, the number of states that need to be estimated for N steps ahead is 4^N —4,096 states for only six steps ahead (reached by planning acuity 5,000 when all the branches are full explored). This crude estimate does not take into account efficiencies related to the use of Upper Confidence Bound (UCB1)²⁰ however, it does illustrate that in general this approach is highly costly and unlikely to be biologically instantiated. Recent research suggests that animals organize their behavior with respect to goals and subgoals^{21,22}, which falls within the framework of hierarchical reinforcement learning²³. Hierarchical reinforcement learning expands the definition of a single action to an action sequence called an option. One hypothesis, based on

this idea of hierarchical organization of choices at different levels of temporal abstraction, postulates that planning is used to select subgoals²⁴. These selected subgoals act as an internal stimulus that triggers a motor routine or policy. On the other hand, another possibility is that the overall goal acts as a stimulus that triggers a cognitive response of selecting subgoals under the habit system²⁵. The actions or policies to reach these subgoals are then chosen through a planning algorithm. There has been a significant amount research on abstract state-space representations that would reduce the size of the original problem (e.g.²⁶). Nevertheless, even with such shortcuts, planning is recognized to be demanding and easily degrades into habit-based action selection under time pressure²⁷.

While the way in which we have implemented plan-based action selection is likely not biologically feasible, it has numerous advantages (e.g. limited number of parameter initializations, theoretical convergence to optimal value function). With such advantages, we expect the approach to provide an upper bound on biological performance that is approximated by more efficient subgoal methods closer to biological feasibility, without altering our key conclusions with respect to visual range and environmental complexity. Finally, tree search model-based methods for Markov decision processes are widely used to model planning in animals^{28–32}.

Another approach could be to implement simpler planning methods on the value function. It is possible that in environments where we do not see behavioral variability (low and high entropy environments), despite the task featuring a dynamic threat that would conventionally necessitate the relearning of the value function at each time step (essentially, to plan), there is in practice very little change in the value function. Therefore in some sense, while our task is dynamic, it may not be fully dynamic in all environments (which is further evidenced by the success of blind habit in low and high entropy environments). In such rather stable environments, it may be possible to perform entropy minimization over the likelihood of survival, as with infotaxis³³, and simplify the planning process.

Infotaxis chooses actions by minimizing the entropy of a belief distribution representing, say, target location. The planning algorithm presented here optimizes a policy using Bellman backups of a terminal reward. A theorem³⁴ shows that with sufficient samples, the method converges to the optimal policy. As a result, although these algorithms are distinct, their results could potentially agree if the chosen reward were entropy of a belief. To use infotaxis-like methods here, one would need to use entropy minimization on the belief distribution representing likelihood of survival. However, the complexity of infotaxis (the need to grid the space and compute entropy for all possible actions) makes it impractical to use directly on survival beliefs. Alternatively, bio-inspired algorithms^{35–37} that approximate infotaxis in a physiologically plausible manner might be sufficient to show that the rough framework provided by infotaxis can be used for the problems pursued here. More broadly, both infotaxis and the algorithms used here are examples of the more general problem of optimal planning under uncertainty, so one should expect that there will ultimately be many algorithms that lead to similar near-optimal solutions.

These methods can be used in low or high entropy environments where we expect a stable distribution of survival probabilities. However, in mid-entropy environments that feature adjacent

clusters of high and low eigenvector centralities, a possible reason for why we see behavioral variability and the need to plan in regions of high eigenvector centrality is that there is a significant amount of change in the value function after each step of the interaction—analogous to sensorimotor latency or planning cycle duration in animals that plan. This makes the task fully dynamic, and necessitates the re-learning of the value function via planning at each step. Unless short update horizons are used, and robustness to multi-peaked distributions is available (as in ergodic information harvesting^{38,39}), infotaxis-like methods may become infeasible.

Supplementary Discussion

1 Ectothermy and the computational complexity of planning

The cognitive ability of reptiles is poorly understood, and while there's evidence of cognitive maps⁴⁰⁻⁴², there is less evidence of planning. Yet, these animals have existed in the same complex terrestrial habitats, with aerial vision, as mammals and birds. We consider three possibilities to resolve this conflict. The first is that these animals are able to plan, and this is simply not known due to lack of study^{40,43}. While plausible, it seems likely that even if planning is present in reptiles, it exists in a relatively diminished form. The second is related to the splayed, close-to-the-ground posture of reptiles. Our lacunarity analysis and simulations all assume that open areas are transparent to vision. This is only true if the visual system is high enough to view over shrubs and grassland that are typically constitutive of open terrestrial domains. Nonetheless, while our results suggest this may mean that the selective benefit of planning in dynamic scenarios may not be present, planning over static rewards should be fully functional. The third possibility is that the selective benefit of planning is present in reptiles, but is not large enough to overcome an additional constraint.

There is certainly not enough evidence to settle this issue presently, but there are a number of considerations in favor of the “additional constraint” hypothesis related to the computational cost of planning. While computers process information differently from neural circuits, the differences in the algorithmic complexity of habit versus planning as it is formalized by reinforcement learning theory is instructive. With habit-based action selection, policy lookup time is independent of the number of actions within the action sequence, but with plan-based action selection, the time needed for an action choice scales exponentially with the number of steps forward simulated.

Both mammals and birds are endotherms. In mammals, endothermy appears to have come about due to selection for increased aerobic capacity in Permian theriodont therapsids, many of which were active predators, resulting in extended capacity for prey pursuit and predator avoidance⁴⁴. A 10°C increase in temperature of muscle above ambient doubles the rate at which muscle can reach maximum power⁴⁵ and can triple rate constants in neural dynamics (modeled elsewhere⁴⁶ with references therein). For body sizes over 1 kg, endotherms have relative brain masses that are a factor of 10x to 40x larger than ectotherms^{47,48}. Thus endotherms have brains that are an order of magnitude larger than ectotherms like fish and reptiles, and those much larger brains run faster due to warming.

Warming also allows animals to acquire sensory information more rapidly. It been shown that the flicker fusion frequency (the rate at which strobed images can no longer be temporally discriminated) of swordfish, a highly active underwater predator that rapidly transitions between warmer surface waters and cold deep waters, rises from 5 Hz in 10°C water to over 40 Hz at 20°C⁴⁹. This is one explanation for why swordfish is one of only around 30 out of 30,000 species of bony fish that has a heating mechanism, in this case for its eyes and brain only. However, endothermy is rare even among active predators underwater, likely because heat is

dissipated 3,000 times faster within water than in air. In addition to this barrier to endothermy underwater, any increase in aerobic capacity of animals that gill rather than breathe has to contend with water having only 1/30th the oxygen of air while being 800 times denser. Gill ventilation therefore requires $800 \times 30 = 24,000$ times more mass flux assuming identical extraction efficiencies; even with the doubled efficiency of gills, the mass flux is four orders of magnitude higher for respiration with water compared to air^{50–52}.

Given the disadvantageous energy load due to rapid cooling and the difficulty of recouping this energetic cost through additional oxygen expenditure, for most underwater animals, the heightened velocity of movement, brain power, and sensory performance of endothermy is out of reach. Yet these factors seem to be important for realizing the selective benefit of planning, which also suggests that even if the spatial complexity of underwater environments were sufficient to favor planning in predator-prey contexts, planning may be difficult to evolve in underwater animals or likely to occur in diminished form relative to birds and mammals. A possible exception, discussed above in the section on invertebrate planning, is cephalopods, which may have unusual adaptations for high metabolic rate, including the use of diffusion of oxygen to supply the highly energetic skin display and reception system in addition to gill ventilation. There is evidence, within species of squid, of temperature-corrected rates higher than mammals of similar size⁵³.

2 The sensory ecology of planning

The specific regime of our study is non-habitable planning—environments so dynamic that plan-based action selection, with its unique ability to rapidly propagate the value of future actions back to inform what the next action should be, is favored over habit-based action selection. In our study, the key factor advantaging plan-based action selection is the presence of a mid-range level of occlusions. Such occlusions, such as terrestrial features like hillocks⁵⁴ and biogenic structure such as shrubs, trees, and grasses, do not typically actively emit signals that can be strategically exploited, but rather are signal reflectors. This makes their use contingent on one of two possibilities: 1) the perceiver has an imaging modality—vision or echolocation; 2) the agent has an accurate and detailed cognitive map in combination with cues from a non-imaging modality (or several in combination) that provides sufficiently precise localization of a moving threat or opportunity. Regarding the first possibility, vision is clearly used by many animals for this purpose (see Discussion). Echolocation, a more recent innovation in mammals⁵⁵, provides good resolution and fast update rates but is range-limited by power requirements⁵⁶. Nonetheless, sound propagation through water is as favorable as light propagation through air, so underwater echolocators can gain substantial range advantage over aquatic vision⁵⁶. We would expect our results to generalize to terrestrial mammals engaged in echolocation-guided behaviors provided range is sufficient, and the environment has the appropriate level of lacunarity. It does not appear that aquatic environments meet this criterion. Regarding the second possibility—the use of a cognitive map in combination with non-imaging modalities—our study provisioned an accurate cognitive map to both predator and prey (Methods), so vision was only needed for predator localization. However, predators emit olfactory,

mechanical (substrate vibration), and auditory cues. Given the olfactory dominance of many mammals, including rodents and elephants^{57,58}, with evidence of cognitive maps^{59,60}, the use of a cognitive map in combination with a non-imaging modality seems quite likely to occur for survival in non-habitizable contexts. In the absence of a cognitive map, or given an incomplete or inaccurate cognitive map, an animal may need to rely on habit-based action selection or on imaging modalities for planning.

In habitizable planning, where planning generates action sequences initially used for advantage over relatively stable features prior to being shifted over to the habit system, olfaction, audition, path integration, and somatosense are all relevant. Hearing provides range and fast updates, but spatial resolution of sound emitters is generally poor outside of owls that hunt by precise sound localization. Olfaction provides long range, a slow update rate, and variable resolution depending on whether the odor is on a substrate or dispersed in air. It may be immensely important for the generation and indexing of a cognitive map, particularly in many olfaction-dominant mammals such as rodents and elephants^{57–61}. Cognitive maps have many uses that are independent of planning, such as present location on a map⁶² and corrections to a dead reckoning system⁶³. This may result in broader phylogenetic distribution of this trait than planning, with evidence for its presence in fish^{41,64,65}, in turtles⁴², and in insects⁶⁶.

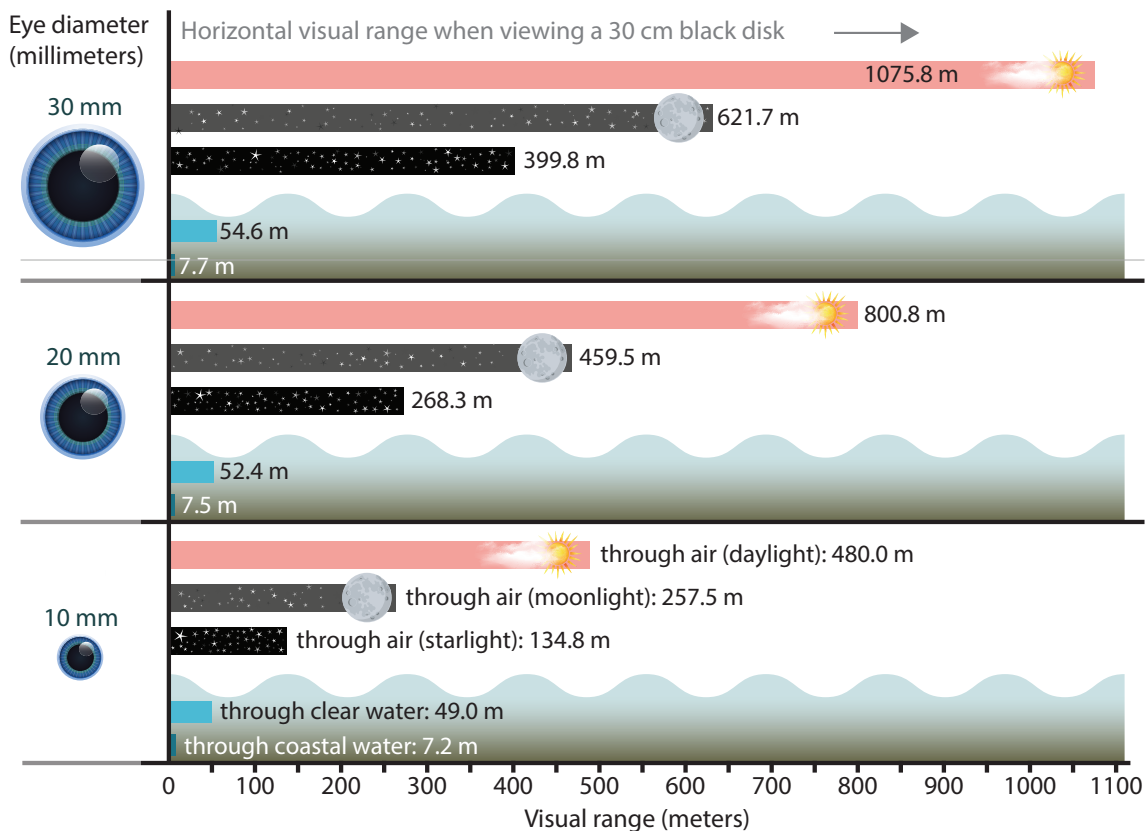
3 Planning in invertebrates

The selection and maintenance of behavior is a need of all mobile animals. In terms of habits, the basal ganglia is a crucial structure in vertebrates, and genetic and molecular evidence points to the occurrence of a preceding structure in the last common ancestor of arthropods and vertebrates⁶⁷. Our study suggests that with respect to a visually guided behavior, predator-prey encounters, certain terrestrial environments—those gappy enough to reveal and hide adversaries at long distances (necessitating vision) so that strategizing pays off—confer a selective benefit to planning. A major obstacle for the same set of factors to select for planning in similar conditions for invertebrates is their very low visual acuity, with a median of 0.25 cycles per degree (cpd) for insects and only up to 0.4 cpd for crustaceans⁶⁸. Here, a fascinating exception is provided by two invertebrates, one terrestrial and one aquatic: the jumping spider, at 13 cpd (similar to passerine birds), and cephalopods, at 46 cpd—only lower than primates and raptors. The jumping spider shows evidence of planning while stalking its typically stationary prey, orb-weaver spiders^{69,70}. Cephalopods have exceptionally large brains and unusually high metabolic rates (partly accomplished by combining diffusion of oxygen through the skin in addition to gills; one species of squid measured at a temperature-corrected metabolic rate higher than mammals of the same size⁵³). The planning abilities of cephalopods have yet to be measured, although it is certainly clear that they have unusually complex cognition⁷¹. Some have pointed to the complexity of controlling what is, essentially, an infinite-degree of freedom body, a muscular hydrostat with eight arms. A more interesting possibility, relating back to their exceptional visual acuity—which, given the evidence shown in Supplementary Figure 1, cannot be for long range vision—is provided by the fact that their skin is akin to a very high resolution video display, containing millions of neurally controlled pigment cells⁷² as well as

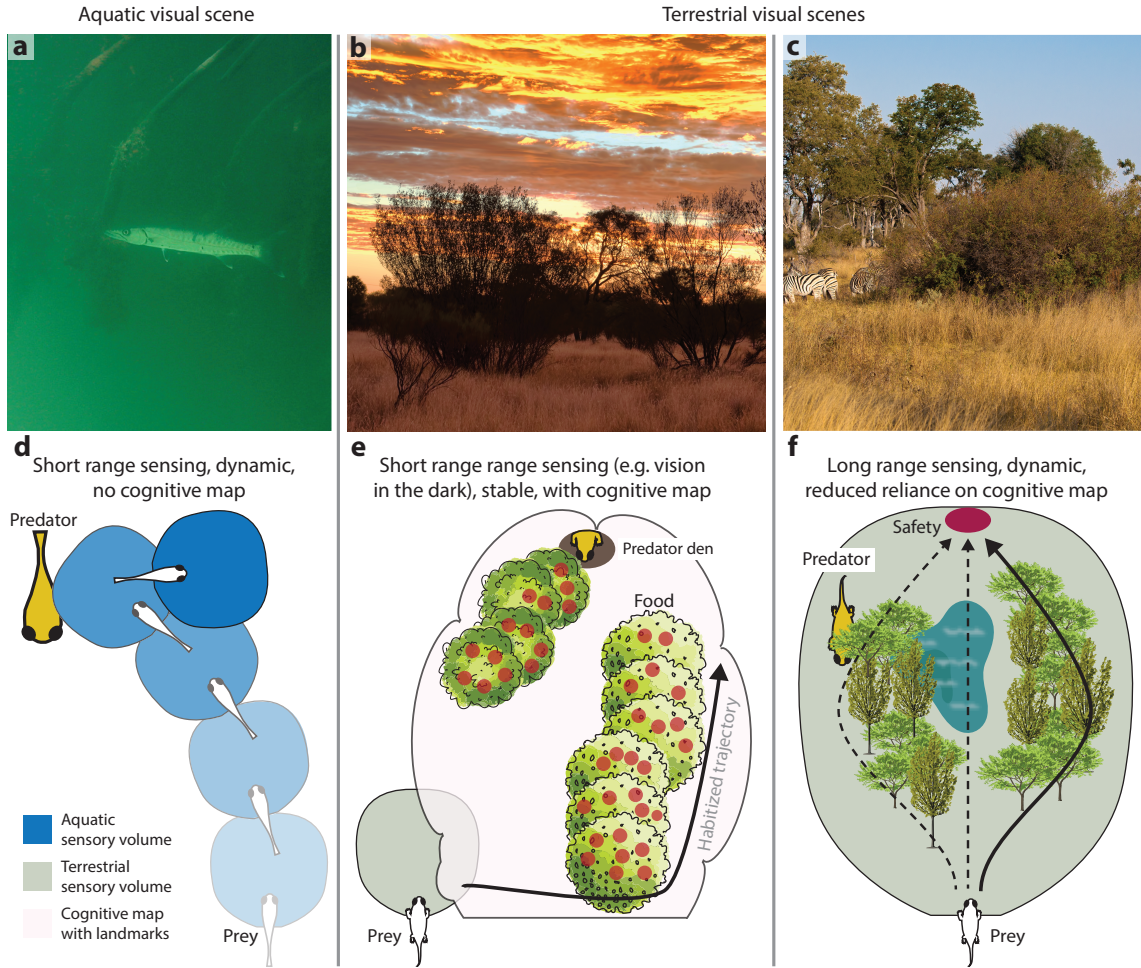
phototransduction elements^{73,74}. This allows projecting detailed and rapidly changing patterns for camouflage around predators and for deception during sexual competitions^{71,75}. During sexual competitions, a male will get between a female and rival, simultaneously displaying the female pattern to the rival on one side of its body and the male pattern to the female on the other side of its body⁷⁵. This is a possible context for a selective benefit of non-habitizable planning without extended sensory range, as it occurs at close quarters with slow changes in position.

In both the camouflage and sexual competition contexts, we can observe a logic similar to that underlying the selective benefit of long-range vision interacting with patchy structure on land: “I see the predator, and if I move there, it will also see me unless I rapidly alter my skin pattern” in the case of camouflage; and “the male thinks I’m a female, and if I move there, it will see that I’m a male unless I rapidly alter my skin pattern” in the case of sexual competitions. The possibility of task logic driving convergence of neural function has been more thoroughly discussed elsewhere⁷⁶.

Supplementary Figures

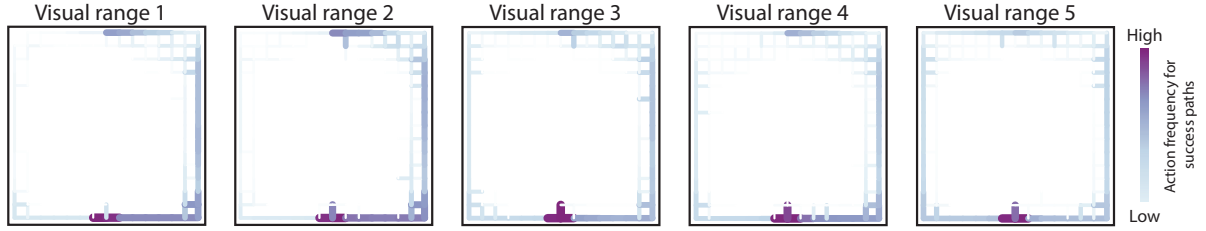


Supplementary Figure 1: **Visual range data for a 30 cm black disk versus eye diameter.** Visual range was calculated⁵⁰ for bright daylight (noon, no clouds), moonlight, starlight, clear water, and coastal water. For reference, typical human eye diameter is 24 mm. The water clarity used for the clear water range estimate (in full sun, at a depth of 10 m) is based on a very clear sample taken in the Bahamas. The water clarity used for the coastal water estimate (also full sun) is based on water typical of freshwater bodies and saltwater near coastlines (Baseline model in our prior work⁵⁰). These values are an upper bound on possible visual range for a black disk (as detailed previously⁵⁰). The fundamental reason for the increase in visual range in air is the level of light attenuation in water compared to air. Light intensity, at its most penetrating (blueish) wavelengths, decreases to 37% of its original value at distances of about 20 m (the attenuation length) in perfectly clear water—only approximated at the deepest reaches of oceans⁵⁰. In comparison, air has an attenuation length of 10,000–100,000 m⁵⁰. For less than full sun, or for more naturalistic contrast ratios than a black disk provides, range decreases rapidly⁵⁰. This results in aquatic visual ranges on the order of one body length^{50,77} for coastal and freshwater biomes where much of aquatic life is concentrated. Given body sizes in the range of meters, and typical locomotion speeds on the order of a body length per second, these range estimates show that visually-guided decision making often has to occur on the order of a second underwater. Above water, the same visual system (after corneal shape changes to correct for the different index of refraction) provides ≈ 100 times more time for these decisions to be made. The absorption and beam attenuation data used for the clear water range simulation were measured near Lee Stocking Island, Bahamas on May 17, 2000 during the Coastal Benthic Optical Properties (CoBOP) program sponsored by the Office of Naval Research, and the data is courtesy of Curtis Mobley, Sequoia Scientific, Inc. Measurements were with a WETLabs ac-9, and made according to the manufacturer's protocols for instrument operation and processed according to NASA protocols for these measurements⁷⁸. Eye design courtesy of Vecteezy.com.

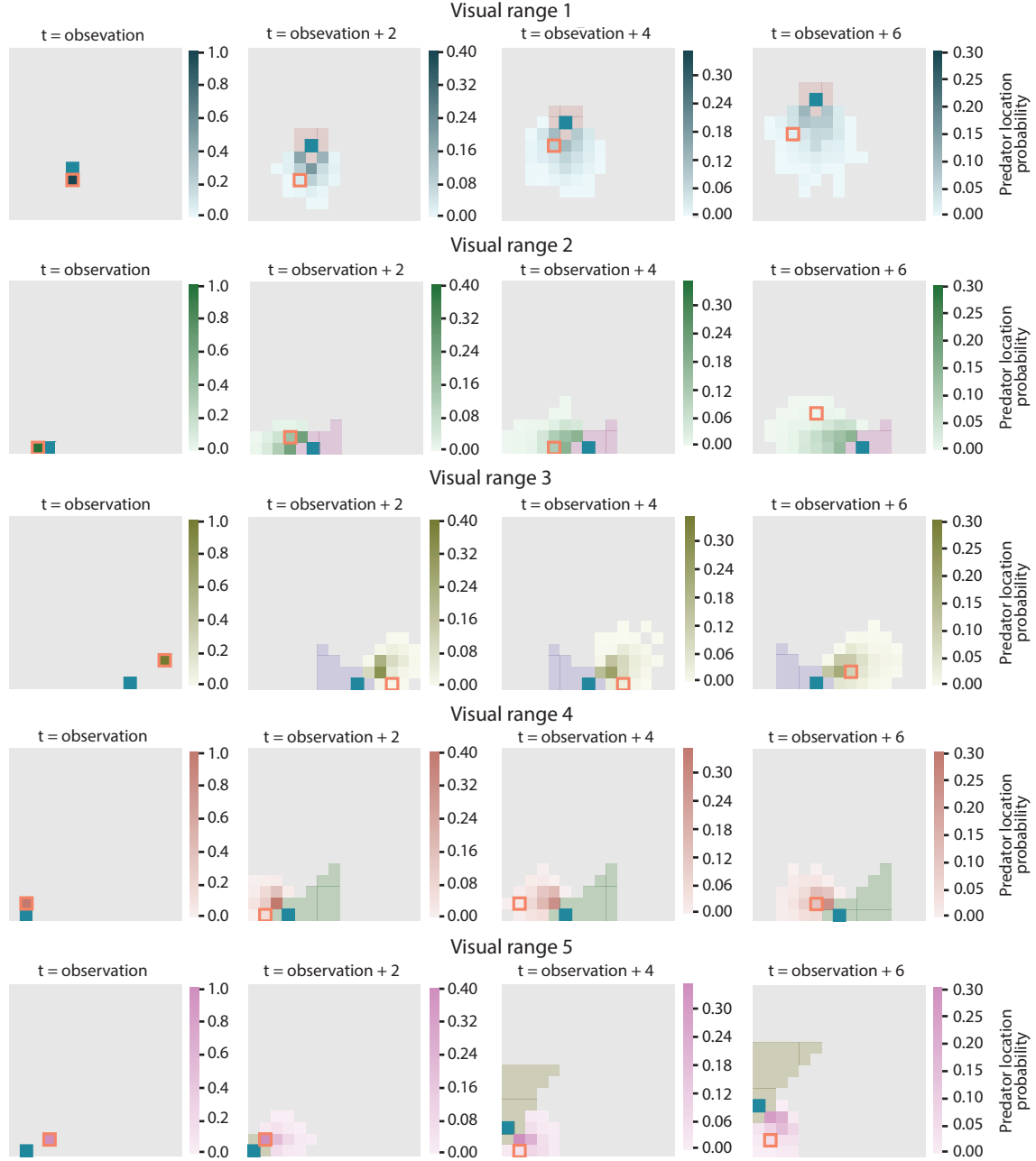


Supplementary Figure 2: **Aquatic versus aerial visual scenes, and the hypothesis under test of how these habitats affect the utility of habit- versus plan-based action selection in a critical visually-guided behavior.**

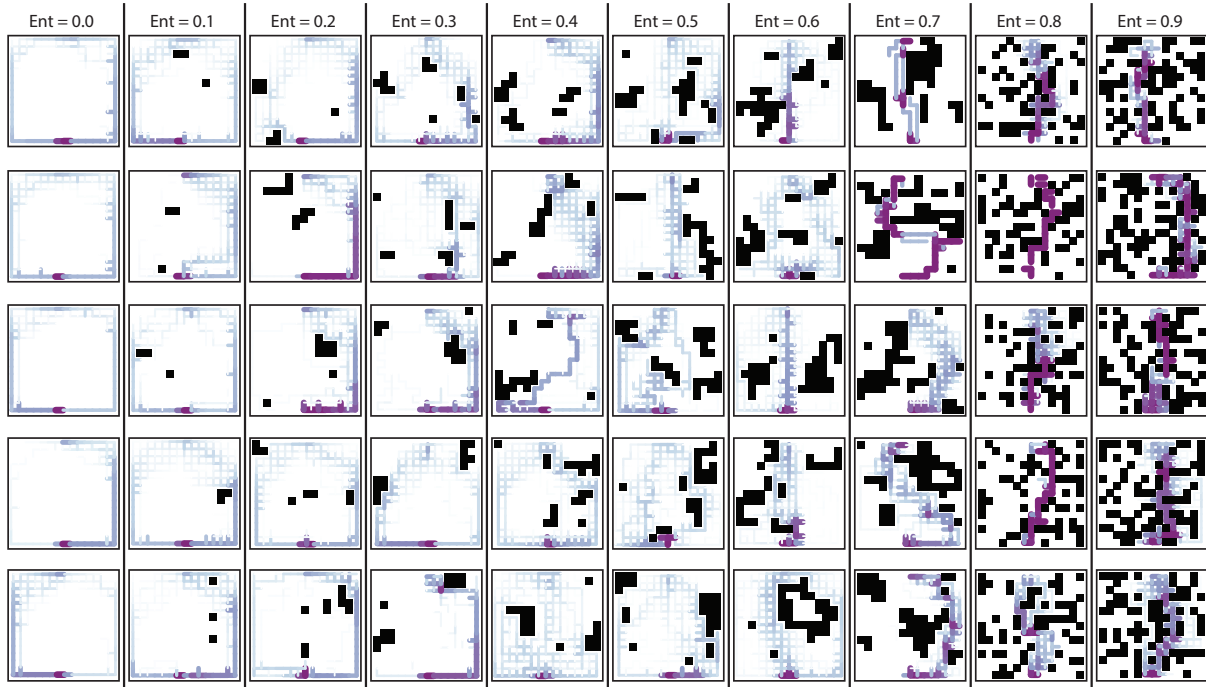
a Example of an aquatic visual scene⁷⁹ from Current Biology, Vol. 27 Issue 14, Dan E. Nilsson, Evolution: An Irresistibly Clear View of Land, R716, Copyright (2017), with permission from Elsevier. **b** Example of a terrestrial visual scene that depicts a shrubland at dusk ("Outback sunset" by Kathleen, used under CC BY 2.0 / Cropped from original). **c** Example of a terrestrial visual scene in which long action sequences such as stalking takes place ("Zebras and giraffe" by Cathy T, used under CC BY 2.0 / Cropped from original). **d** Dynamic scenes at short sensory ranges, typical of the sensory ecology of aquatic environments⁵⁰ (e.g. (a)), affords habit-based action selection. Aquatic behaviors are typically short range, and predator-prey interactions are at close quarters, requiring rapid and simple responses. **e** Planning under situations that are not continuously dynamic. Even with short sensory range (e.g. olfaction), after the prey learns and updates its cognitive map about the location of static food sources (bushes, and berries) and possible threats (predator den), it can use its cognitive map to devise paths to avoid threats⁸⁰ or get to food⁸¹. After a period of time these devised paths can be compiled into habit (black arrow). **f** Planning in dynamic situations with long visual range. The long visual ranges afforded by terrestriality (≈ 100 times larger than aquatic visual ranges (Supplementary Figure 1)) could possibly enable animals with appropriate circuitry to plan over the contents of the sensorium, rather than over the cognitive map. This in turn would allow for the imagining of multiple possible futures (solid and dashed black arrows) followed by selection of the option with higher expected value (solid black arrow).



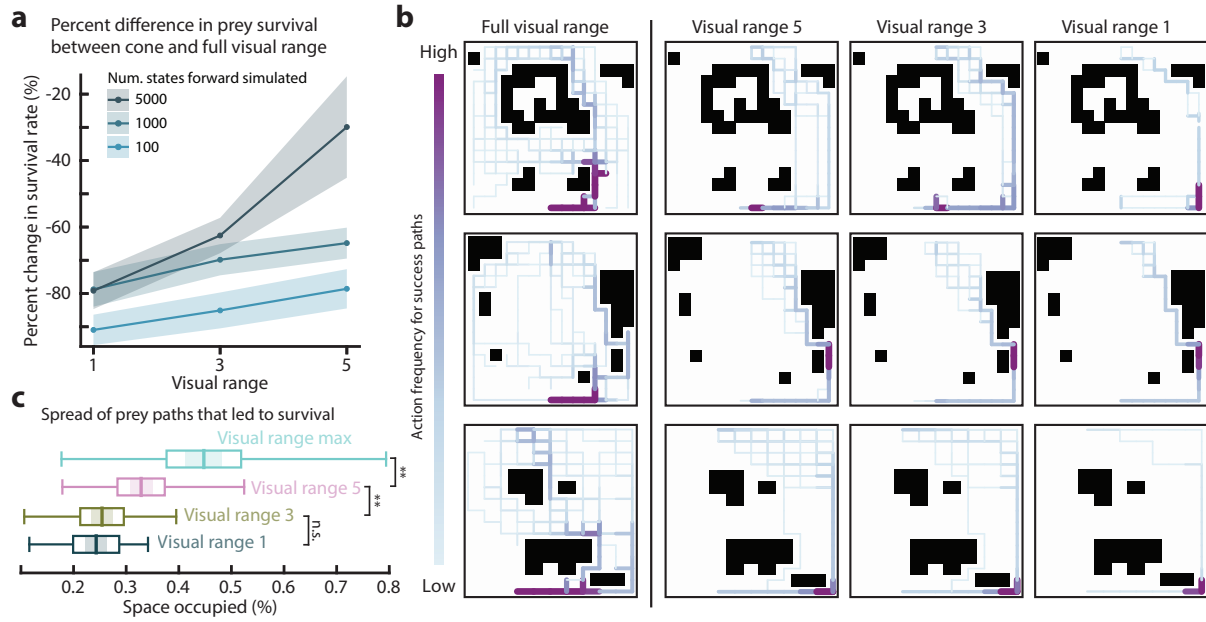
Supplementary Figure 3: **Prey success trajectories with respect to visual range.** Heatmaps of all action sequences taken by the prey that resulted in prey survival at maximum planning level (5000 states forward simulated), with color density proportional to frequency. Color bar action frequencies range from 1–138, dependent on visual range. This representation shows that across visual ranges, the prey strategy remains the same. Longer visual ranges enable the prey to turn away from a sensed predator at a greater distance, allowing it to escape. Survival paths are aggregated across all tested predator locations ($n = 20$), and total number of episodes ($n = 100$). Source data are provided as a Source Data file.



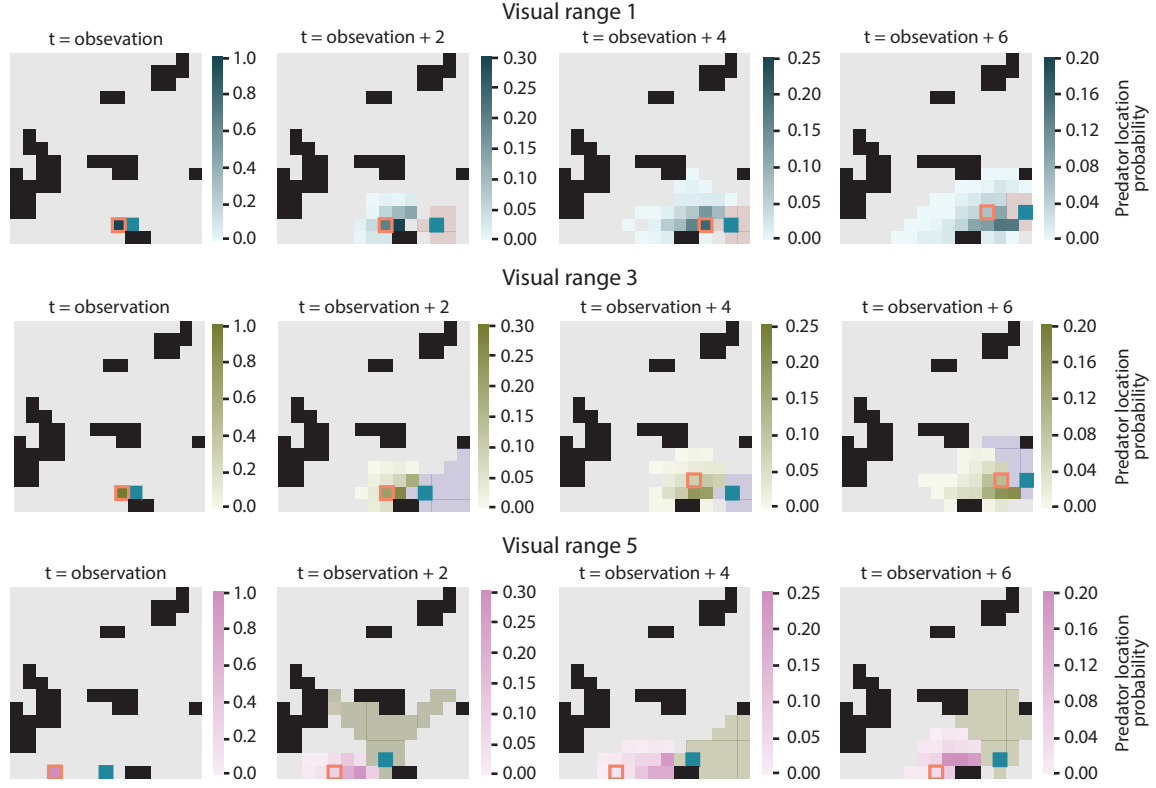
Supplementary Figure 4: **Uncertainty growth as a function of visual range.** Heatmaps depict the spread of the prey's belief state based on its knowledge about how the predator moves (both speed and randomness). The variance of the prey's belief state changes with the number of time steps that the predator is out of the prey's visual cone, and the prey's visual range. A random episode with at least 6 unobserved time steps was chosen, and the policy implemented by the prey from the onset of observation to 6 time steps ahead was implemented. During the prey's implementation of this action sequence, based on the initial predator observation, the predator location was randomly sampled and moved according to the predator model ($N = 1000$). The location of the prey is depicted in teal, the actual location of the predator is depicted with an orange boundary, and the extending cone is the prey's visual cone (1 & 2: pink, 3: purple, 4 & 5: green). Source data are provided as a Source Data file.



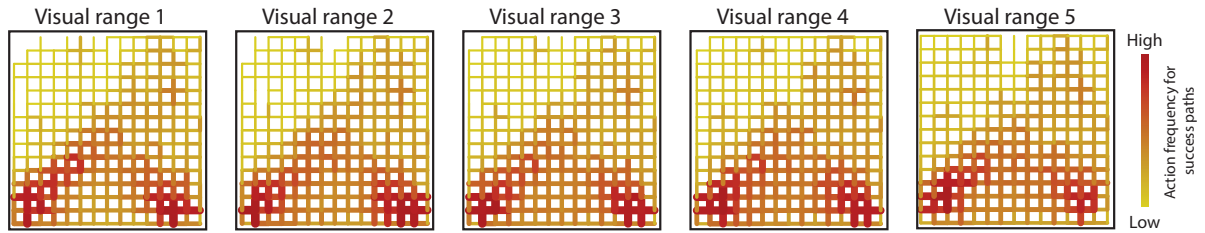
Supplementary Figure 5: **Prey success trajectories with respect to environment entropy.** Heatmaps of all action sequences taken by the prey that resulted in prey survival at maximum planning level (5000 states forward simulated), with color density proportional to frequency. Color bar action frequencies range from 1–115, dependent on environment entropy. This representation shows that at low and high entropy success paths are fairly stereotypical. In contrast, in midentropy environments, with the emergence of multiple viable futures, success paths are dissimilar. Survival paths are aggregated across all randomly generated environments per entropy level ($n = 20$), random initial predator location ($n = 5$), and total number of episodes ($n = 50$). Source data are provided as a Source Data file.



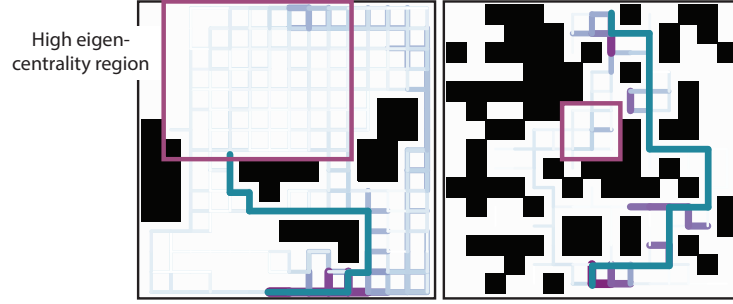
Supplementary Figure 6: **Survival rate and prey success paths with limited visual range.** **a** Mean percent change in survival rate (computed as [survival rate with limited visual range minus survival rate with unlimited visual range] divided by [survival rate with unlimited visual range]) versus visual range in environments that had prey success path spreads above the 75th percentile for prey with ‘unlimited’ visual range ($n = 30$) (see Methods). The fill indicates $\pm$ s.e.m. of percent change in survival rate across selected environments. With both full visual range **b** For three different environments (rows), we illustrate how success path diversity varies with visual range (columns). Heatmaps are of all action sequences taken by the prey that resulted in prey survival at the maximum planning level (5000 states forward simulated), with color density proportional to frequency. Success paths for full visual range are given for the same predator start location. Color bar action frequencies range from 1–62, dependent on environment entropy and visual range. Survival paths are aggregated across all randomly generated environments per entropy level ($n = 20$), and total number of episodes ($n = 50$). **c** High path spread indicates that the success paths are highly dissimilar. Within the boxes, the vertical line corresponds to the mean, the shaded region corresponds to the s.e.m., and the box extent corresponds to the 95% confidence interval of the mean. The lines extending from the boxes show the range of the data ($n_1 = 19$, $n_3 = 26$, $n_5 = 26$, and $n_{full} = 26$). MWU tests (two-tailed) across visual ranges: $U_{1,3} = 238.5$, $U_{3,5} = 201.5$, $U_{5,full} = 231.0$ (** $p < 0.008$; n.s. is not significant $p = 0.43$). Source data are provided as a Source Data file.



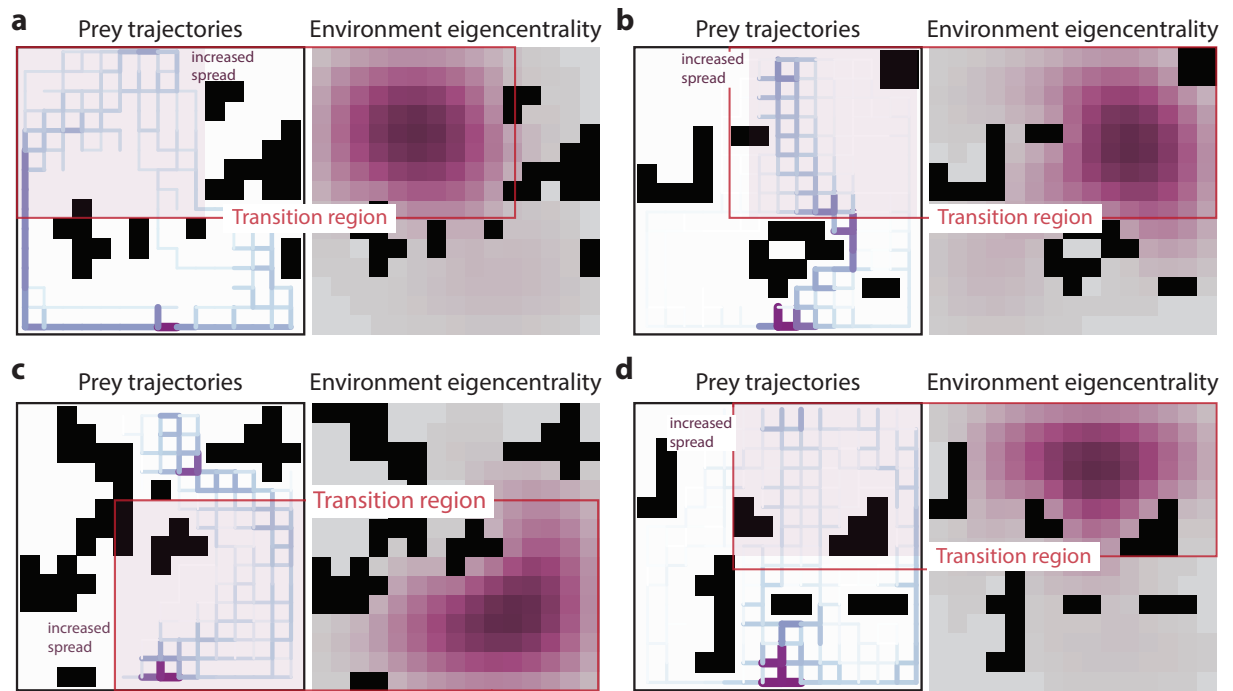
Supplementary Figure 7: **Uncertainty growth as a function of visual range.** Heatmaps depict the spread of prey's belief state based on its knowledge about how the predator moves (both speed and randomness). The variance of the prey's belief state changes with the number of time steps that the predator is out of the prey's visual cone, and the prey's visual range. A random episode with at least 6 unobserved time steps was chosen, and the policy implemented by the prey from the onset of observation to 6 time steps ahead was implemented. During the prey's implementation of this action sequence, based on the initial predator observation, the predator location was randomly sampled and moved according to the predator model ($N = 1000$). The location of the prey is depicted in teal, the actual location of the predator is depicted with an orange boundary, and the extending cone is the prey's visual cone (1: pink, 3: purple, 5: green). Source data are provided as a Source Data file.



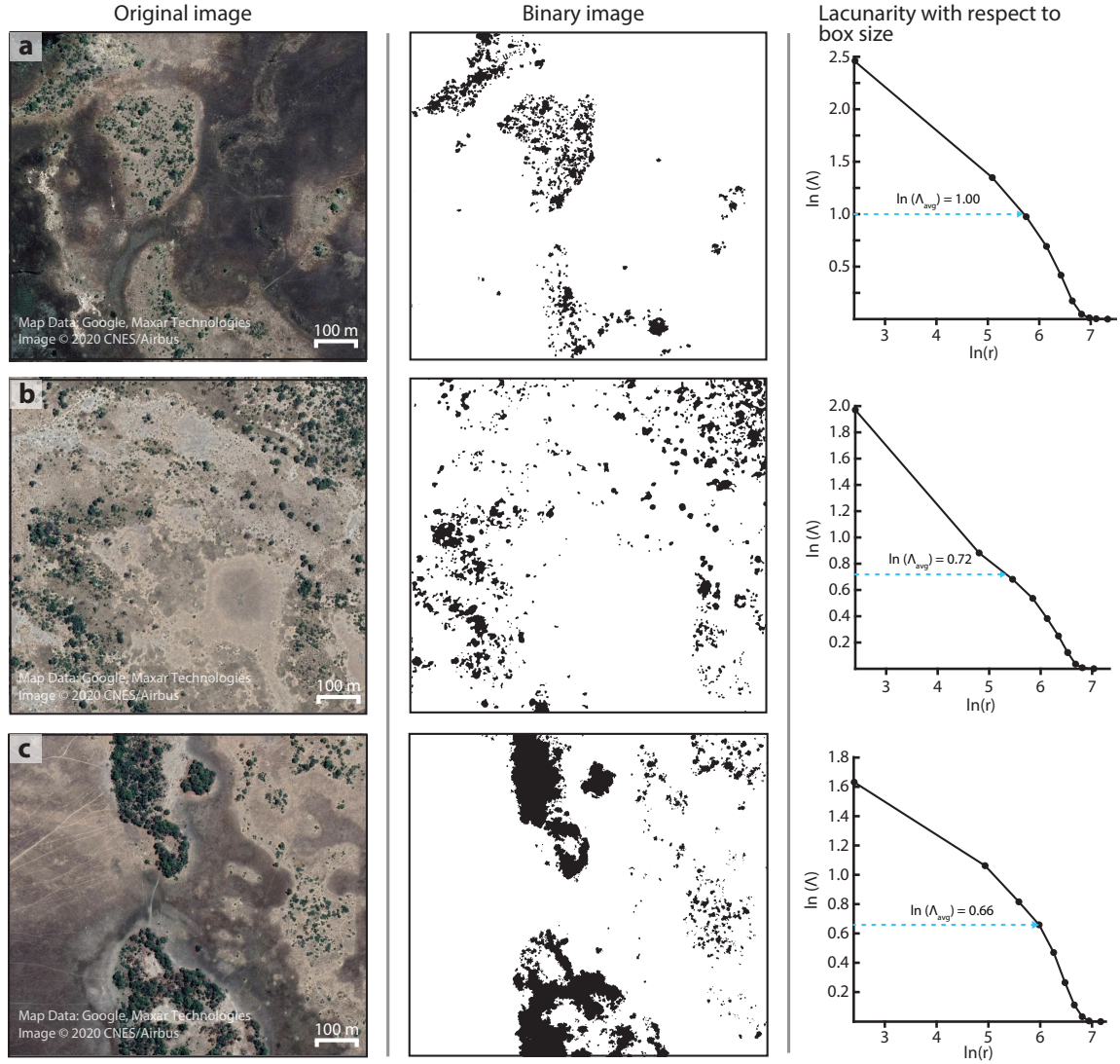
Supplementary Figure 8: **Predator success trajectories with respect to prey visual range.** Heatmaps of all action sequences taken by the predator that resulted in predator capture of the prey, with color density proportional to frequency. Color bar action frequencies range from 1–579, dependent on visual range. This representation shows that across visual ranges, the predator strategy remains the same. Predator success paths are aggregated across all tested predator locations ($n = 20$), and total number of episodes ($n = 100$). Source data are provided as a Source Data file.



Supplementary Figure 9: **Example prey success trajectories and high eigencentrality regions.** Heatmaps of all action sequences taken by the prey that resulted in prey survival at the maximum planning level (5000 states), with color density proportional to frequency. Color bar action frequencies range from 1 for low, to 92 for the most saturated color, dependent on entropy level. The pink box indicates the high eigencentrality region. Survival paths are aggregated across all randomly generated environments per entropy level ($n = 20$), random initial predator location ($n = 5$), and total number of episodes ($n = 50$). Source data are provided as a Source Data file.



Supplementary Figure 10: **Example prey success trajectories and corresponding environment eigencentality.** Example environments and their eigenvector centralities. Color densities of each cell are proportional to the metric. Transition regions from low to high eigencentality—based on a change in gradient and value of eigencentality—are shown by the red box. The pink box highlights the increase in path spread that corresponds to regions of high eigencentality. Survival paths are aggregated across all randomly generated environments per entropy level ($n = 20$), random initial predator location ($n = 5$), and total number of episodes ($n = 50$). Source data are provided as a Source Data file.



Supplementary Figure 11: **Sample images from the Okavango Delta, Botswana and their corresponding binary images and lacunarity plots.** Images from the Okavango Delta in Botswana providing example landscapes across the range of mean lacunarity values in which planning results in higher survival rates when compared to habit than habit ($\approx 0.7\text{--}1.0$). Each spans ≈ 16 ha ($800\text{ m} \times 800\text{ m}$). This geographic region is considered to be a modern analogue of the early hominin paleoenvironment^{82,83}. Research suggests that the environments that hominins lived in after branching from chimpanzees 6–7 million years ago were composed of open areas of grassland interspersed with dense and moderately dense woodlands with trees and bushes⁸³. Prior studies hypothesize a connection between habitat patchiness and the evolution of bipedalism, which facilitates easier navigation between patches and exploitation of food in open regions given its more favorable locomotion energetics^{83,84}. Images are from Google Earth, imagery date November 1 2018 (copyright 2020 CNES/Airbus). Binary images were generated manually from the source images using Adobe Photoshop. Lacunarity analyses of the binary images were performed using the FracLac plug-in⁸⁵ for ImageJ⁸⁶. Box size is in pixels. Coordinates of the image locations (decimal degrees) and image sizes: **(a)** Coordinate: $-19.229956, 22.986520$. Image size: 1507×1442 pixels. **(b)** Coordinate: $-19.216485, 22.925382$. Image size: 1208×1134 pixels. **(c)** Coordinate: $-19.335282, 22.914340$. Image size: 1462×1312 pixels. Image originals and source data are provided within Source Data files.

Supplementary Tables

Supplementary Table 1: Lacunarity values for different environments

Environment	Value ($\ln(\Lambda_{avg})$)	Reference
Coastal Aquatic Environment Lacunarity Values ($\ln(\Lambda_{avg})$)		
Fine and coarse sand flat seabed	asymptotic - unary image	Data from ⁸⁷ Synthetic aperture sonar (SAS) image near La Spezia Italy, and SAS image Riga, Latvia
Seagrass with small patches covering <7% and ~13% of sampled area	1.16–1.26	Data from ⁸⁸ Geographical information system (GIS) coverage samples from Owen Anchorage and Cockburn Sound region south of Fremantle, Western Australia. Additional reference ⁸⁹
Structured Aquatic Environment Lacunarity Values ($\ln(\Lambda_{avg})$)		
Seagrass with small patches covering ~37% of sampled area	0.18	Data from ⁸⁸ Geographical information system (GIS) coverage samples from Owen Anchorage and Cockburn Sound region south of Fremantle, Western Australia. Additional reference ⁸⁹
Seagrass with medium patches covering 16%–25% of sampled area	0.40–0.45	
Seagrass with medium patches covering 31%–37% of sampled area	0.25–0.27	
Seagrass with large patches covering 17%–21% of sampled area	0.42–0.46	
Seagrass with large patches covering 32%–45% of sampled area	0.12–0.13	
Salt marsh spatial distribution of plant species	0.03–0.49	Data from ⁹⁰ Skallingen salt marsh in southwestern Jutland, Denmark

Continued on next page

Supplementary Table 1 – continued from previous page

Environment	Value ($\ln(\Lambda_{\text{avg}})$)	Reference
Coral reefs in the Upper Palaeozoic Era	0.41	Data from ¹⁷ Norwegian Barents Sea
Modern day coral reefs	0.23–0.41	Data from ¹⁷ Alacranes Reef and data from ⁹¹ Hawaii and the Carribean
Terrestrial Environment Lacunarity Values ($\ln(\Lambda_{\text{avg}})$)		
Herbaceous rangeland (ecosystems dominated by grasses and forbs ⁹²)	1.34	
Shrub rangeland (ecosystems dominated by woody vegetation ⁹²)	1.19	Data from ¹⁶ central coast region of California; land use land cover data developed by the US Geological Survey
Mixed rangeland (ecosystems where more than one-third of the land is a mixture of herbaceous and shrub or bush ⁹²)	0.26	
Typical grazing land (shrub rangeland)	1.2	
Acre with groves	1.35	
Meadow	0.63	Data from ⁹⁴ normalized digital surface model (NDSM) from German low-range mountain area in Baden-Wurttemberg
Meadow with fruit	0.56	
Orchard	0.47	
<i>Terra firme</i> forest	0.23	Data from ⁹⁵ rainforest at Caxiuana, Para, Brazil. Additional references ^{12,96}

Supplementary Table 2: General parameters used for all simulations

Parameter	Value	Note
Environment size	15×15	The environment size was chosen to have both a large state-space and be feasible in terms of duration of simulations given our compute time on the cluster used for this study (Methods). While the advent of terrestriality increased visual range by a factor of 100, within this environment we were only able to simulate an increase of at most a factor of 5 due to this size limitation for computational expediency. The x - and y -coordinate values for location are $\in [0, 14]$, with the origin at the bottom left.
Prey initial location	(7, 0)	Prey initial location was chosen arbitrarily and was fixed across all computational experiments
Predator initial location		Predator initial location was chosen randomly from unoccupied cells.
Safety location	(7, 14)	Safety location was chosen arbitrarily and was fixed across all computational experiments
Predator pursuit strategy	75% aggressive, 25% random	Pursuit probabilities were chosen arbitrarily, however the pac-man example provided in Silver et al.'s work ³⁴ also uses this same pursuit strategy. The randomness places importance on the prey observing the predator. After each observation the randomness in predator pursuit strategy causes belief state diffusion (Supplementary Figure 4 & Supplementary Figure 7). In the case of 100% aggressive pursuit, after one observation the prey would know exactly where the predator is after each subsequent time step.

Continued on next page

Supplementary Table 2 – continued from previous page

Parameter	Value	Note
Predator:Prey speed ratio	1.5	With 50% probability (drawn from a Bernoulli distribution) the predator moves 2 cells ahead. Therefore, on average the predator is $1.5\times$ faster than the prey. This speed falls within typical terrestrial predator pursuit speeds relative to prey of 1.2–2 ^{97,98} .
Rewards	-1, -25, -100, +1000	At each time step the prey is required to move; movement of one square received a reward of -1. If the prey hit the wall or an occlusion it received a reward of -25. If the predator occupied the same cell at the same time step as the prey the episode terminates and the prey receives a reward of -100 (death). If the prey is able to reach the safety position the episode terminates and the prey receives a reward of +1000 (survival). These values are from Silver et al. ³⁴ .
Tested environment entropies		Entropy 0.0: 0 ± 0 occlusions ; Entropy 0.1: 3 ± 0 ; Entropy 0.2: 9 ± 1 ; Entropy 0.3: 13 ± 1 ; Entropy 0.4: 20 ± 2 Entropy 0.5: 28 ± 3 ; Entropy 0.6: 35 ± 3 ; Entropy 0.7: 45 ± 3 ; Entropy 0.8: 59 ± 4 ; Entropy 0.9: 75 ± 3
Number of states the prey forward simulates		Fixed within trial, but varying across trials, this parameter took these values: 1, 10, 100, 1000, 5000 for Simulation 1, and 100, 1000, 5000 for Simulation 2.

Supplementary Table 3: Parameters for POMCP (Plan-based action selection)

Parameter	Symbol	Value	Note
Exploration constant for PO-UCT	c	100	During the within tree search (over nodes that have been visited) the value of an action is augmented by an exploration bonus that is highest for rarely tried actions $\tilde{Q}(s, a) = Q(s, a) + c\sqrt{\frac{\log N(s)}{N(s, a)}}$. This scalar constant determines the relative ratio of exploration to exploitation. Thus, when $c = 0$ the explored nodes are selected greedily. The value chosen was set to the value used by Silver et al.³⁴. Changing the value of c, especially lowering it to close to 0, could possibly change the survival rates, but is unlikely to change the overall trend we see.
Softmax inverse temperature parameter	β	3	Experimental work on human and animal decision making has found that the stochasticity in the observed behavior can be captured by softmax policies^{99,100}. The parameter β controls how diffuse the probability of choosing an action is. As $\beta \rightarrow \infty$ the actions are chosen greedily, and at $\beta = 0$ actions are chosen randomly. This seemingly low β value is a consequence of the sizes of the rewards: very negative at death, and very positive at survival. This value ensures that most action are chosen greedily from the constructed tree. Increasing β would cause more actions to be chosen greedily, and though unlikely, could increase survival rate, but similar to before, would not change the overall observed trend, since the same β value is used in all simulations.
Continued on next page			

Supplementary Table 3 – continued from previous page

Parameter	Symbol	Value	Note
Discount factor	γ	0.95	The discount factor affects the weight of the future rewards. Therefore, $\gamma = 0$ will result in state or state-action values to represent only the immediate reward. On the other hand, $\gamma = 1$ will make the agent consider the cumulative effects of long term rewards. Lowering the discount value will decrease the agents cumulative reward at survival, and will increase the impact of the negative reward given at each time step. Thus, this might cause behaviors such as the agent going towards the predator, since the overall imagined cumulative reward might end up being lower. Increasing the γ won't have any foreseeable effect, although $\gamma = 1$ could have divergence problems. The value used was taken from the experiments done by Silver et al. ³⁴ .
Rollout count		$200 \times 1.25^{\text{Env. entropy} \times 10}$	In POMCP if a node has not been previously visited in order to determine the value of the node, random action rollouts are carried out. The rollout, similar to the actual simulation, terminates if the prey reaches the safety position or if the predator catches the prey. Therefore, the only stipulation is for the rollout count to be high enough so that the prey encounters either of the termination conditions. In environments where there is not a lot of clutter around 200 was found to be more than enough. When clutter is added to the environment, the nature of random action rollout, causes the agent to need more time steps. The above modulation to the base line of 200 random rollouts was found to be sufficient.

Supplementary Table 4: Parameters for PRQL (Habit-based action selection)

Parameter	Symbol	Value	Note
Softmax temperature parameter	τ	0	This parameter controls how diffuse the probability of choosing a policy from the policy library is. Initially τ is set to 0, therefore the first chosen policy from the policy library is always random. In consequent trials τ is updated by: $\tau = \tau + \Delta\tau$. This value was set to the value used by Fernandez et al. ¹⁰¹ .
Incremental increase of τ	$\Delta\tau$	0.001	This parameter controls how fast the softmax temperature parameter increases. As $\tau \rightarrow \infty$ policies from the policy library are chosen greedily based on their weight. Given reward values, this increment allows for continual exploration after one implementation. If this value was increased, it would inhibit the agent from exploring other policies which might also result in favorable outcomes. Decreasing this value would result in policies to be chosen randomly over more trials. This value was set to the value used by Fernandez et al. ¹⁰¹ .
Discount factor	γ	0.95	Was kept the same as the discount factor that was used in plan-based action selection (Supplementary Table 3).

Supplementary References

1. Barbosa, P. & Castellanos, I. *Ecology of predator-prey interactions* (Oxford University Press, 2005).
2. King, A. W. *et al.* Dispersal success on fractal landscapes: a consequence of lacunarity thresholds. *Landsc. Ecol.* **14**, 73–82 (1999).
3. Plotnick, R. E., Gardner, R. H. & O'Neill, R. V. Lacunarity indices as measures of landscape texture. *Landsc. Ecol.* **8**, 201–211 (1993).
4. Plotnick, R. E., Gardner, R. H., Hargrove, W. W., Prestegard, K. & Perlmutter, M. Lacunarity analysis: a general technique for the analysis of spatial patterns. *Phys. Rev. E* **53**, 5461 (1996).
5. Kaye, B. H. *A random walk through fractal dimensions* (John Wiley & Sons, 2008).
6. Mandelbrot, B. B. *The Fractal Geometry of Nature*, vol. 1 (WH freeman New York, 1982).
7. Allain, C. & Cloitre, M. Characterizing the lacunarity of random and deterministic fractal sets. *Phys. Rev. A* **44**, 3552 (1991).
8. Gefen, Y., Meir, Y., Mandelbrot, B. B. & Aharony, A. Geometric implementation of hypercubic lattices with noninteger dimensionality by use of low lacunarity fractal lattices. *Phys. Rev. Lett.* **50**, 145 (1983).
9. Cheng, Q. Multifractal modeling and lacunarity analysis. *Math. Geol.* **29**, 919–932 (1997).
10. Sun, G. & Ranson, K. Radar modelling of forest spatial patterns. *Int. J. Remote Sens.* **19**, 1769–1791 (1998).
11. Weishampel, J. F., Blair, J. B., Knox, R. G., Dubayah, R. & Clark, D. B. Volumetric lidar return patterns from an old-growth tropical rainforest canopy. *Int. J. Remote Sens.* **21**, 409–415 (2000).
12. Frazer, G. W., Wulder, M. A. & Niemann, K. O. Simulation and quantification of the fine-scale spatial pattern and heterogeneity of forest canopy structure: A lacunarity-based method designed for analysis of continuous canopy heights. *Forest Ecol. Manag.* **214**, 65–90 (2005).
13. Blumenfeld, R. & Mandelbrot, B. B. Lévy dusts, mittag-leffler statistics, mass fractal lacunarity, and perceived dimension. *Phys. Rev. E* **56**, 112 (1997).
14. Dale, M. Lacunarity analysis of spatial pattern: a comparison. *Landsc. Ecol.* **15**, 467–478 (2000).
15. Tellez, O. L. & Papili, S. An introductory study of the impact of environmental parameters in the performances of imaging sonar systems. In *OCEANS 2016 MTS/IEEE Monterey*, 1–6 (IEEE, 2016).

16. Perry, J. *et al.* Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. *Ecography* **25**, 578–600 (2002).
17. Purkis, S., Casini, G., Hunt, D. & Colpaert, A. Morphometric patterns in Modern carbonate platforms can be applied to the ancient rock record: Similarities between Modern Alacranes Reef and Upper Palaeozoic platforms of the Barents Sea. *Sediment. Geol.* **321**, 49–69 (2015).
18. Racanière, S. *et al.* Imagination-augmented agents for deep reinforcement learning. In *Adv. Neur. In.*, 5690–5701 (2017).
19. Hula, A., Montague, P. R. & Dayan, P. Monte Carlo planning method estimates planning horizons during interactive social exchange. *PLoS Comput. Biol.* **11**, e1004254 (2015).
20. Kocsis, L. & Szepesvári, C. Bandit based Monte-Carlo Planning. In *ECML*, 282–293 (Springer, 2006).
21. Botvinick, M. M. Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* **12**, 201–208 (2008).
22. Ito, M. & Doya, K. Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. *Curr. Opin. Neurobiol.* **21**, 368–373 (2011).
23. Sutton, R. S., Precup, D. & Singh, S. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artif. Intell.* **112**, 181–211 (1999).
24. Dezfouli, A. & Balleine, B. W. Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* **35**, 1036–1051 (2012).
25. Cushman, F. & Morris, A. Habitual control of goal selection in humans. *Proc. Natl. Acad. Sci. U.S.A* **112**, 13817–13822 (2015).
26. McNamee, D., Wolpert, D. M. & Lengyel, M. Efficient state-space modularization for planning: theory, behavioral and neural signatures. *Adv. Neur. In.* 4511–4519 (2016).
27. Keramati, M., Smittenaar, P., Dolan, R. J. & Dayan, P. Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proc. Natl. Acad. Sci. U.S.A* **113**, 12868–12873 (2016).
28. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–325 (2013).
29. Daw, N. D. Model-based reinforcement learning as cognitive search: neurocomputational theories. *Cognitive search: Evolution, algorithms and the brain* 195–208 (2012).
30. Huys, Q. J. *et al.* Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.* **8** (2012).
31. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* **8**, 1704–1711 (2005).

32. Solway, A., Lohrenz, T. & Montague, P. R. Simulating future value in intertemporal choice. *Sci. Rep.* **7**, 43119 (2017).
33. Vergassola, M., Villermaux, E. & Shraiman, B. I. 'Infotaxis' as a strategy for searching without gradients. *Nature* **445**, 406–409 (2007).
34. Silver, D. & Veness, J. Monte-Carlo planning in large POMDPs. In *Adv. Neur. In.*, 2164–2172 (2010).
35. Álvarez-Salvado, E. *et al.* Elementary sensory-motor transformations underlying olfactory navigation in walking fruit-flies. *Elife* **7**, e37815 (2018).
36. Calhoun, A. J., Chalasani, S. H. & Sharpee, T. O. Maximally informative foraging by *Caenorhabditis elegans*. *Elife* **3**, e04220 (2014).
37. Masson, J.-B. Olfactory searches with limited space perception. *Proc. Natl. Acad. Sci. U.S.A* **110**, 11261–11266 (2013).
38. Miller, L. M., Silverman, Y., MacIver, M. A. & Murphey, T. D. Ergodic exploration of distributed information. *IEEE Transactions on Robotics* **32**, 36–52 (2016).
39. Chen, C., Murphey, T. & MacIver, M. Sense organ control in moths to moles is a gamble on information through motion. *bioRxiv* (2019).
40. Wilkinson, A. & Huber, L. Cold-blooded cognition: Reptilian cognitive abilities. In Vonk, J. & Shackelford, K. (eds.) *Oxford handbook of comparative evolutionary psychology* (Oxford University Press, Oxford, UK, 2012).
41. Broglio, C. *et al.* Hippocampal pallium and map-like memories through vertebrate evolution. *J. Behav. Brain Sci.* **5**, 109 (2015).
42. López, J., Vargas, J., Gómez, Y. & Salas, C. Spatial and non-spatial learning in turtles: the role of medial cortex. *Behav. Brain Res.* **143**, 109 – 120 (2003).
43. Burghardt, G. M. Environmental enrichment and cognitive complexity in reptiles and amphibians: Concepts, review, and implications for captive populations. *Appl. Anim. Behav. Sci.* **147**, 286 – 298 (2013). Welfare of Zoo Animals.
44. Ruben, J. The evolution of endothermy in mammals and birds: From physiology to fossils. *Annu. Rev. Physiol.* **57**, 69–95 (1995).
45. Bennett, A. F. Thermal dependence of muscle function. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **247**, R217–R229 (1984).
46. Goldin, M. A. & Mindlin, G. B. Temperature manipulation of neuronal dynamics in a forebrain motor control nucleus. *PLoS Comput. Biol.* **13**, e1005699 (2017).
47. Yu, Y., Karbowski, J., Sachdev, R. N. & Feng, J. Effect of temperature and glia in brain size enlargement and origin of allometric body-brain size scaling in vertebrates. *BMC Evol. Bio* **14**, 178 (2014).

48. Gillooly, J. F. & McCoy, M. W. Brain size varies with temperature in vertebrates. *PeerJ* **2**, e301 (2014).
49. Fritsches, K. A., Brill, R. W. & Warrant, E. J. Warm eyes provide superior vision in swordfishes. *Curr. Biol.* **15**, 55 – 58 (2005).
50. MacIver, M. A., Schmitz, L., Mugan, U., Murphey, T. D. & Mobley, C. D. Massive increase in visual range preceded the origin of terrestrial vertebrates. *Proc. Natl. Acad. Sci. U.S.A* **114**, E2375–E2384 (2017).
51. Glass, M. L. & Rantin, F. T. *Gas Exchange and Control of Respiration in Air-Breathing Teleost Fish*, 99–119 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009).
52. Glass, M. L. *Physiological Evidence Indicates Lungfish as a Sister Group to the Land Vertebrates*, 161–177 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009).
53. Seibel, B. A. On the depth and scale of metabolic rate variation: scaling of oxygen consumption rates and enzymatic activity in the Class Cephalopoda (Mollusca). *J. Exp. Biol.* **210**, 1–11 (2007).
54. Mech, L. D. Possible use of foresight, understanding, and planning by wolves hunting muskoxen. *Arctic* **60**, 145–149 (2007).
55. Parker, J. *et al.* Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228 (2013).
56. Nelson, M. E. & MacIver, M. A. Sensory acquisition in active sensing systems. *J. Comp. Physiol. A* **192**, 573–586 (2006).
57. Niimura, Y., Matsui, A. & Touhara, K. Extreme expansion of the olfactory receptor gene repertoire in african elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res.* **24**, 1485–1496 (2014).
58. Schmitt, M. H., Shuttleworth, A., Ward, D. & Shrader, A. M. African elephants use plant odours to make foraging decisions across multiple spatial scales. *Anim. Behav.* **141**, 17 – 27 (2018).
59. Redish, A. D. Vicarious trial and error. *Nat. Rev. Neurosci.* **17**, 147–159 (2016).
60. Viljoen, P. J. Spatial distribution and movements of elephants (*loxodonta africana*) in the northern namib desert region of the kaokoveld, south west africa/namibia. *J. Zool.* **219**, 1–19 (1989).
61. Jacobs, L. F. From chemotaxis to the cognitive map: The function of olfaction. *Proc. Natl. Acad. Sci. U.S.A* **109**, 10693–10700 (2012).
62. Foster, D., Morris, R. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10**, 1–16 (2000).

63. Redish, A. D. *Beyond the cognitive map from place cells to episodic memory* (MIT Press, Cambridge, Mass., 1999).
64. Wallach, A., Harvey-Girard, E., Jun, J. J., Longtin, A. & Maler, L. A time-stamp mechanism may provide temporal information necessary for egocentric to allocentric spatial transformations. *ELife* **7** (2018).
65. Trinh, A.-T., Clarke, S. E., Harvey-Girard, E. & Maler, L. Cellular and Network Mechanisms May Generate Sparse Coding of Sequential Object Encounters in Hippocampal-Like Circuits. *eNeuro* **6** (2019).
66. Menzel, R. *et al.* Honey bees navigate according to a map-like spatial memory. *Proc. Natl. Acad. Sci. U.S.A* **102**, 3040–3045 (2005).
67. Strausfeld, N. J. & Hirth, F. Deep homology of arthropod central complex and vertebrate basal ganglia. *Science* **340**, 157–161 (2013).
68. Caves, E. M., Brandley, N. C. & Johnsen, S. Visual acuity and the evolution of signals. *Trends Ecol. Evol.* **33**, 358–372 (2018).
69. Tarsitano, M., Jackson, R. R. & Kirchner, W. H. Signals and signal choices made by the araneophagic jumping spider *Portia fimbriata* while hunting the orb-weaving web spiders *Zygiella x-notata* and *Zosis geniculatus*. *Ethology* **106**, 595–615 (2000).
70. Tarsitano, M. Route selection by a jumping spider (*Portia labiata*) during the locomotory phase of a detour. *Anim. Behav.* **72**, 1437–1442 (2006).
71. Mather, J. A. & Dickel, L. Cephalopod complex cognition. *Curr. Opin. Behav. Sci* **16**, 131 – 137 (2017). Comparative cognition.
72. Messenger, J. B. Cephalopod chromatophores: neurobiology and natural history. *Biological Reviews of the Cambridge Philosophical Society* **76**, 473–528 (2001).
73. Ramirez, M. D. & Oakley, T. H. Eye-independent, light-activated chromatophore expansion (LACE) and expression of phototransduction genes in the skin of *Octopus bimaculoides*. *J. Exp. Biol.* **218**, 1513–1520 (2015).
74. Kingston, A. C. N., Kuzirian, A. M., Hanlon, R. T. & Cronin, T. W. Visual phototransduction components in cephalopod chromatophores suggest dermal photoreception. *J. Exp. Biol.* **218**, 1596–1602 (2015).
75. Brown, C., Garwood, M. P. & Williamson, J. E. It pays to cheat: tactical deception in a cephalopod social signalling system. *Biol. Letters* **8**, 729–732 (2012).
76. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93**, 480–490 (2017).
77. Nilsson, D. E., Warrant, E. & Johnsen, S. Computational visual ecology in the pelagic realm. *Phil. Trans. R. Soc. B* **369** (2014).

78. Mueller, J. *et al.* *Ocean Optics Protocols for Satellite Ocean Color Sensor Validation, Revision 4, Volume III: Radiometric Measurements and Data Analysis Protocols* (National Aeronautics and Space Administration, 2003).
79. Nilsson, D. E. Evolution: An Irresistibly Clear View of Land. *Curr. Biol.* **27**, R715–R717 (2017).
80. Wu, C.-T., Haggerty, D., Kemere, C. & Ji, D. Hippocampal awake replay in fear memory retrieval. *Nat. Neurosci.* **20**, 571 (2017).
81. Miller, K. J., Botvinick, M. M. & Brody, C. D. Dorsal hippocampus contributes to model-based planning. *Nat. Neurosci.* **20**, 1269–1276 (2017).
82. Brunet, M. *et al.* New material of the earliest hominid from the upper miocene of chad. *Nature* **434**, 752–755 (2005).
83. Le Fur, S., Fara, E., Mackaye, H. T., Vignaud, P. & Brunet, M. The mammal assemblage of the hominid site TM266 (Late Miocene, Chad Basin): ecological structure and paleoenvironmental implications. *Naturwissenschaften* **96**, 565–574 (2009).
84. Rodman, P. S. & McHenry, H. M. Bioenergetics and the origin of hominid bipedalism. *Am. J. Phys. Anthropol.* **52**, 103–106 (1980).
85. Karperien, A. Fraclac for imagej. *Charles Sturt University* (2013).
86. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. Nih image to imagej: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
87. Williams, D. P. Fast unsupervised seafloor characterization in sonar imagery using lacunarity. *IEEE Trans. Geosci. Remote Sens.* **53**, 6022–6034 (2015).
88. Sleeman, J. C., Kendrick, G. A., Boggs, G. S. & Hegge, B. J. Measuring fragmentation of seagrass landscapes: which indices are most appropriate for detecting change? *Mar. Freshw. Res.* **56**, 851–864 (2005).
89. Fonseca, M. S. & Bell, S. S. Influence of physical setting on seagrass landscapes near Beaufort, North Carolina, USA. *Mar. Ecol. Prog. Ser.* **171**, 109–121 (1998).
90. Kim, D., Cairns, D. & Bartholdy, J. Spatial heterogeneity and domain of scale on the Skallingen salt marsh, Denmark. *Dan. J. Geogr.* **109**, 95–104 (2009).
91. Donovan, M. K. *A Synthesis of Coral Reef Community Structure in Hawaii and the Caribbean*. Ph.D. thesis (2017).
92. Anderson, J. R. *A land use and land cover classification system for use with remote sensor data*, vol. 964 (US Government Printing Office, 1976).
93. Seifan, M. & Kadmon, R. Indirect effects of cattle grazing on shrub spatial pattern in a mediterranean scrub community. *Basic Appl. Ecol.* **7**, 496–506 (2006).

94. Hoechstetter, S., Walz, U. & Thinh, N. X. Adapting lacunarity techniques for gradient-based analyses of landscape surfaces. *Ecol. Complex.* **8**, 229–238 (2011).
95. Malhi, Y. & Román-Cuesta, R. M. Analysis of lacunarity and scales of spatial homogeneity in IKONOS images of Amazonian tropical forest canopies. *Remote Sens. Environ.* **112**, 2074–2087 (2008).
96. Li, X. *et al.* Evaluating the effectiveness of neutral landscape models to represent a real landscape. *Landsc. Urban Plan.* **69**, 137–148 (2004).
97. Elliott, J. P., Cowan, I. M. & Holling, C. Prey capture by the African lion. *Can. J. Zool.* **55**, 1811–1828 (1977).
98. Hedenström, A. & Rosén, M. Predator versus prey: on aerial hunting and escape strategies in birds. *Behav. Ecol.* **12**, 150–156 (2001).
99. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585–595 (2010).
100. Yoshida, W., Dolan, R. J. & Friston, K. J. Game theory of mind. *PLoS Comput. Biol.* **4**, e1000254 (2008).
101. Fernández, F. & Veloso, M. Probabilistic policy reuse in a reinforcement learning agent. In *AAMAS '06*, 720–727 (ACM, 2006).