The shift to life on land selected for planning

Ugurcan Mugan[1,2] & Malcolm A. MacIver[1,2,3,4]

[1] *The Neuroscience and Robotics Laboratory, Northwestern University, Evanston IL*

[2] *Department of Biomedical Engineering, Northwestern University, Evanston IL*

[3] *Department of Mechanical Engineering, Northwestern University, Evanston IL*

[4] *Department of Neurobiology, Northwestern University, Evanston IL*

It is uncontroversial that land animals have developed more elaborated cognitive abilities than aquatic animals, with the possible exception of formerly land-based mammals like whales and dolphins that have returned to an aquatic existence. Yet there is no apparent *a-priori* reason for aquatic and land animals to have different evolutionary capacities for the rise of cognition. One possibility is that investigators have been anthropocentric in their definition of cognition, rather than attuned to cognitive phenomena as appropriate to the broader ethological and ecological context of each species. However, this concern may not apply to the paradigmatically cognitive faculty of being able to imagine multiple complete sequences of actions to accomplish a goal and select one to enact, or planning. Although planning over space requires cognitive maps—which exist in many species including fish—behavioral and neural evidence for planning is presently restricted to birds and mammals. Here, we present evidence that a reason for the absence of planning in fish and many other aquatic animals is that in a key driver of natural selection, predator-prey interactions, there is no benefit to planning above habit-based action selection. In contrast, there is a significant benefit of planning under similar predator-prey scenarios for terrestrial conditions. This effect is dependent on a combination of increased visual range and the spatial complexity of terrestrial environments. The ability to plan in select terrestrial vertebrates may therefore be an adaptive response to the massive increase in visual sensory range (100x) occurring with the shift to life on land in combination with the complexity of terrestrial habitats and its affordance of strategic behavior during predator-prey interactions.

Introduction

The emergence of vertebrates on to land over 350 million years ago was preceded by a massive increase in visual range when their eyes moved to the top of the head to look over the water surface and tripled in size[1] (**Fig. 1a**). The consequent 100-fold increase in visual range (**Fig. 1b**) arose from a dramatic decrease in the attenuation of light from water to air[1,2], and was accompanied by the opportunity to see elements of spatial complexity, such as topography and vegetation (**Fig. 1c**). Animals looking through air were able to observe distant potential drivers of behavior, such as predators and prey, in a more complex space (**Fig. 1d**). It is likely that mobile threats and opportunities within a complex space increased the need for anticipating the consequences of actions based on either experience or imagination[3,4]. Here we address whether sensing the changing location of mobile threats and opportunities in the spatial environment, along with increased afforded delay between stimulus and response[5], would have advantaged mental evaluation of action sequences in animals emerging onto land[6].

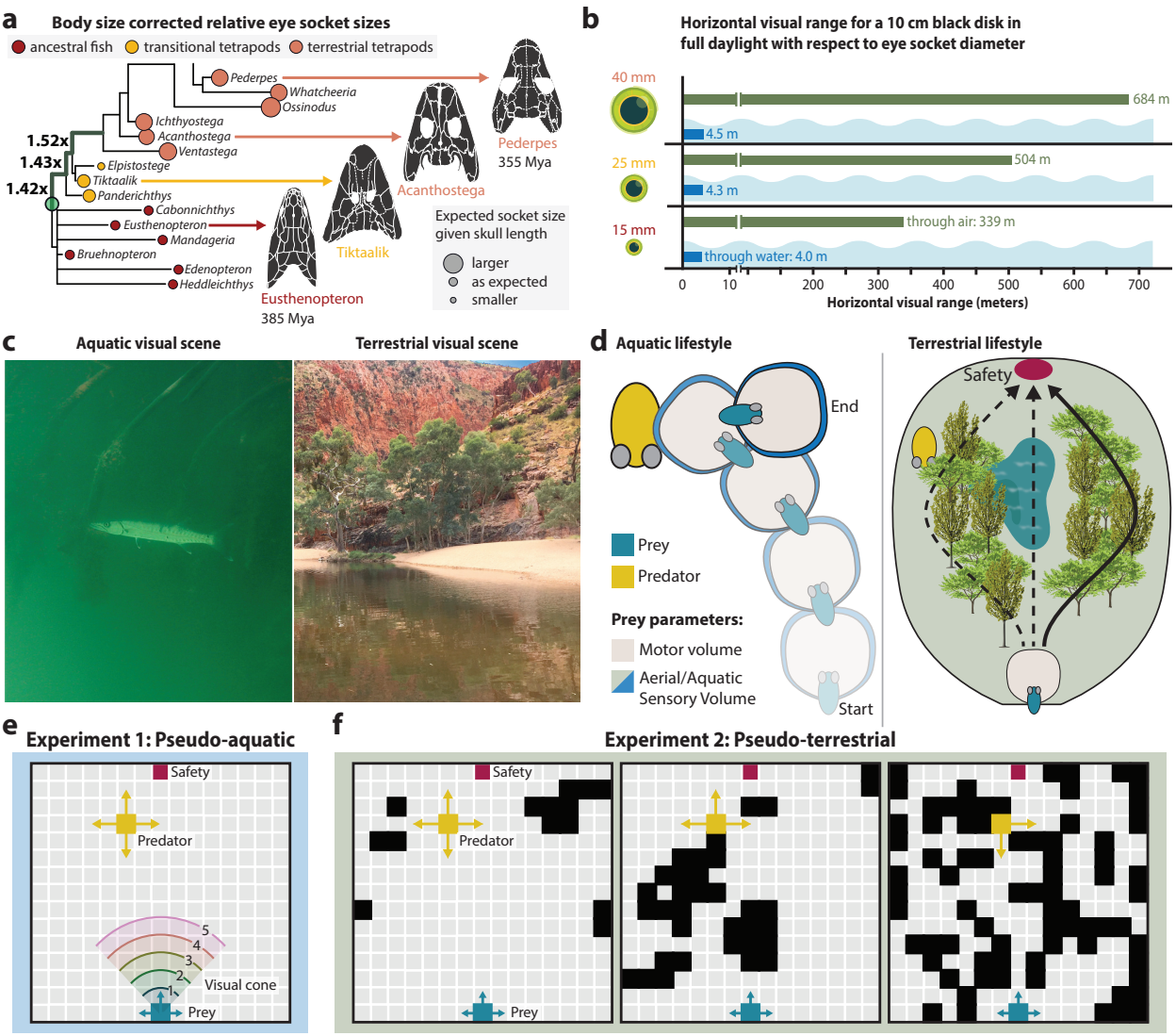

Figure 1: **Implications of the shift to life on land on eye size, visual range, and the affordance of planning versus reactive behavioral control. (a)** Data from a phylogenetic comparative study[1]. The thick branches indicate positions of selective shifts in eye size, with corresponding factor of increase in eye socket size compared with the ancestral regime (terminated by green circle). **(b)** Through computational visual ecology estimates visual range can be inferred from fossilized eye socket diameters[1]. Mean eye socket diameters are shown on left for ancestral fish, transitional tetrapods, and terrestrial tetrapods. **(c)** Examples of aquatic and terrestrial visual scenes[7]. **(d)** The impact of sensory ecology on affording reactive control versus planning. Aquatic schematic shows a prey's immediate "motor volume" (tan region, region being moved into in a short duration)[6,8] almost completely overlapping a diminutive visual "sensory volume" (blue region, where an animal detects objects). Short sensing range necessitates a high speed response. Terrestrial schematic shows similar motor volume, but visual sensory volume (green region) has massively increased. This affords more time for planning—here between multiple behavioral options (solid and dashed black arrows)—to choose the option with higher expected fitness value (solid black arrow). **(e–f)** Examples of the quantized environments used in computational experiments for this study; black squares represent occlusions. Examples of low, medium, and high clutter are shown for the pseudo-terrestrial experiment.

Much is now known about the neural circuitry underlying planning. The hippocampus and its functional homologues[9,10] in terrestrial animals has emerged as a key structure for evaluating the consequences of actions in imagination prior to execution[11]. Previous work has shown that mammals engage in hippocampal replay during sharp wave-ripple events to mentally represent action sequences before execution during avoid-approach conflicts in dynamic environments[12,13]. Studies that reconstruct mentally represented paths reveal that environmental topology and spatial complexity are accurately captured by hippocampal replay[14]. Birds have a hippocampus with the same developmental origin[15] and similar anatomy[10] as mammalian hippocampus, and there is accumulating behavioral evidence that it plays a similar role in planning for this group of animals[16]. These findings suggest that the planning abilities we see in mammals and birds may have evolved from a common toolkit of neural circuits in a stem amniote, perhaps over the 35 million years between the emergence of fully terrestrial forms (such as *Pederpes*, 355 million years ago (Mya); **Fig. 1a**) and the last common ancestor of birds and mammals (320 Mya[17]). Supporting this hypothesis, here we present evidence that after the shift to life on land, there was a higher selective benefit for the brain circuitry subserving planning because of the great increase in visual range and in-range spatial complexity this change in habitat afforded.

Results

We have selected the context of prey being pursued by a predator for this work, as it has both the needed features of being dynamic as well as having obvious evolutionary importance. Prior work has shown that hippocampal replay in spatial navigation tasks resembles model-based, sequential forward planning[18–20], which creates a tree-like structure of possible movements over the environment. The planning simulator developed for this study uses this same structure to evaluate both its moves ahead, and its adversary. Variation in the amount of planning—how many different possibilities for action are considered—was controlled by the number of forward states evaluated, hereafter referred to as "planning acuity" for brevity.

Two computational experiments (**Fig. 1e–f**) were designed to study the effects of how much planning an animal can do on performance and behavioral complexity, as a function of visual range and spatial complexity. For both of these computational experiments, the predator was designed as a model-based reflex agent, while the prey was designed as a planning agent (see Discussion), with a predetermined planning acuity and visual range.

In Computational Experiment 1 (pseudo-aquatic; **Fig. 1e**), the prey's visual range was varied in a simple (open) environment. In Computational Experiment 2 (pseudo-terrestrial; **Fig. 1f**), the environmental topology was varied by adding randomly generated clutter into the environment until a pre-determined level of environmental entropy was reached (see Methods). This caused the prey and predator's visual range to explicitly depend on the environment. Each experimental

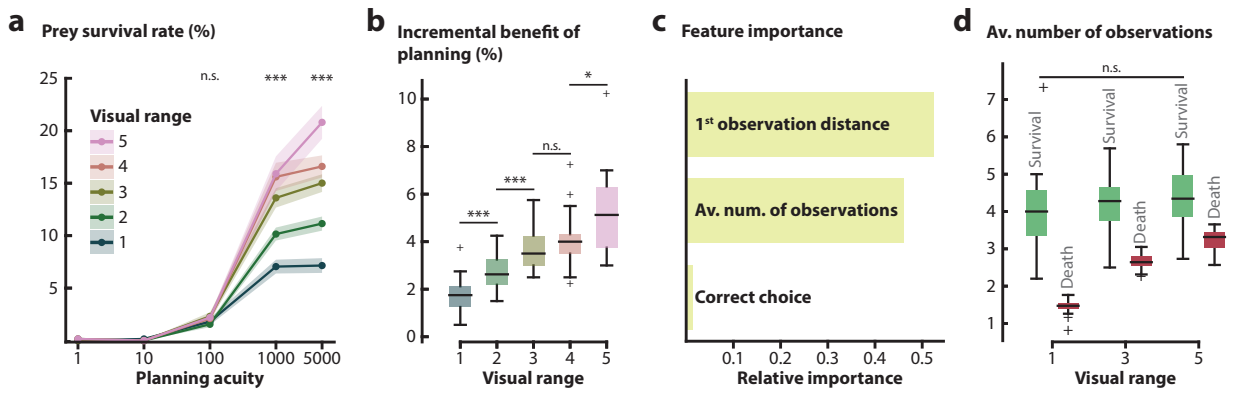

Figure 2: **The effect of visual range and planning acuity on survival.** **(a)** Prey survival rate in pseudo-aquatic environments (**Fig. 1e**) as a function of prey visual range and planning acuity across random initial predator locations ($n = 20$). The line plot shows the mean survival rate, and the surrounding fill indicates $\pm$ sem. **(b)** For each visual range, the mean gain of survival rate over all the planning acuities shown in panel (a)—1, 10, 100, 1000, and 5000. The outliers are indicated with a '+'. **(c)** The importance of the feature shown in C (average number of observations), as well as the 1st observation distance (typically, the visual range), and whether the prey moves away from the predator on its first action ("Correct choice"), in predicting survival. **(d)** The distribution of the average number of observations grouped by termination condition (survival, death) for a prey with high planning acuity (5000) and across visual range 1, 3, and 5.

trial had a fixed predator start locationand prey planning acuity, but both of these parameters were varied across trials.

In psuedo-aquatic conditions, the prey survival rate increases with respect to an increase in planning acuity, independent of visual range (**Fig. 2a**). However, survival rate quickly plateaus when visual range is short, indicating that that there is limited added gain in survival rate from increasing planning acuity (**Fig. 2b**; ***: $P < 0.001$, *: $P = 0.046$, One-way ANOVA $P < 0.001$). The two most important determinants of survival in open environments are observation distance—proportional to visual range—and the number of observations, which is the number of time steps when the predator is in view of the prey (**Fig. 2c**). Due to the randomness in both speed and direction of the predator, the prey is only certain about the predator location when it observes the predator. When the predator is out of view, the prey samples from a probability distribution of predator locations that is informed by knowledge of the predator's speed and the predator's desire to close the gap to the assumed location of the prey (the number of predator locations the prey samples is equal to the planning acuity; see Methods.)

A step increase in visual range results in an exponential increase in the amount of space monitored, resulting in a decrease in the uncertainty about the predator location in subsequent time steps. Frequent observations decrease the set of possible predator locations, making planned trajectories tailored to a particular area, rather than the entire space outside of the visual cone, thereby increasing the probability of survival (**Fig. 2d**). The proportional relationship between the

prey's visual range and the average number of time steps for the predator to reach the prey from the time it is observed[6] makes survival independent of the prey erroneously choosing to move toward the predator (**Fig. 2c; Correct choice**). Long visual ranges—resulting in lower predator location uncertainty—and increased predator-prey separation at observation, advantages higher planning acuity in spatially simple environments. Due to the near one-body length visual range afforded by typical ancestral aquatic habitats[1], regardless of the actual spatial complexity of their environment, animals within them would effectively reside in a similarly simple environment (**Fig. 1c**, aquatic visual scene).

In open dynamic environments planning has marginal utility with sensing ranges of near one body length (visual ranges 1 & 2) typical of the ancestral vertebrate aquatic visual range[1,21]. With the emergence of eyes above the water column, there was an increase in both the total volume of space monitored by vision (1-million fold)[1], and in-range spatial complexity due to vegetation[22] and other terrestrial barriers. Thus to model the selective benefit of planning in terrestrial environments, we varied the amount of clutter in the environment, and extended the visual range to span the entire environment, except as blocked by elements of clutter.

In pseudo-terrestrial environments (**Fig. 1e**), prey survival rate increases until midrange levels of environmental clutter (entropy 0.4–0.6), and then decreases in high clutter environments (entropy $>$ 0.6) (**Fig. 3a**). When the environment has very little clutter (low entropy), the predator speed and pursuit strategy (see Methods) restricts the prey's survival rate, limiting the gain in survival rate from increases in planning acuity (**Fig. 3b**, Mann-Whitney U test with Bonferroni correction Low–Mid: $P < 0.001$). In these relatively open environments the predator's visual range is rarely hampered, which results in longer periods of aggressive pursuit. When the environment has high levels of clutter (high entropy), the overall space in which the prey can move decreases, causing survival to be highly dependent on the predator's initial location and the distribution of occlusions. Interestingly, not only is planning of little benefit in this case, random action selection provides survival rates similar to those that occur with planning (planning acuity=1 approximates random action; **Fig. 3a–b**, Mann-Whitney U test with Bonferroni correction Mid–High: $P < 0.001$).

The amount of clutter along with the distribution of clutter in the environment changes how visible each location is with respect to all other locations. This causes different action sequences to be favored in relation to the predator's movement, changing behavioral variability with respect to clutter amount and distribution. As a result, in environments that have midrange clutter (entropy 0.4–0.6), the prey's predator avoidance behaviors are highly variable, even within a given environment.

To analyze the spatial distribution of clutter, we used a network generated from these location visibilities. This visibility network connects all the cells that do not have an occlusion along a connecting line of sight between cell centers (see Methods). The complexity of these networks,

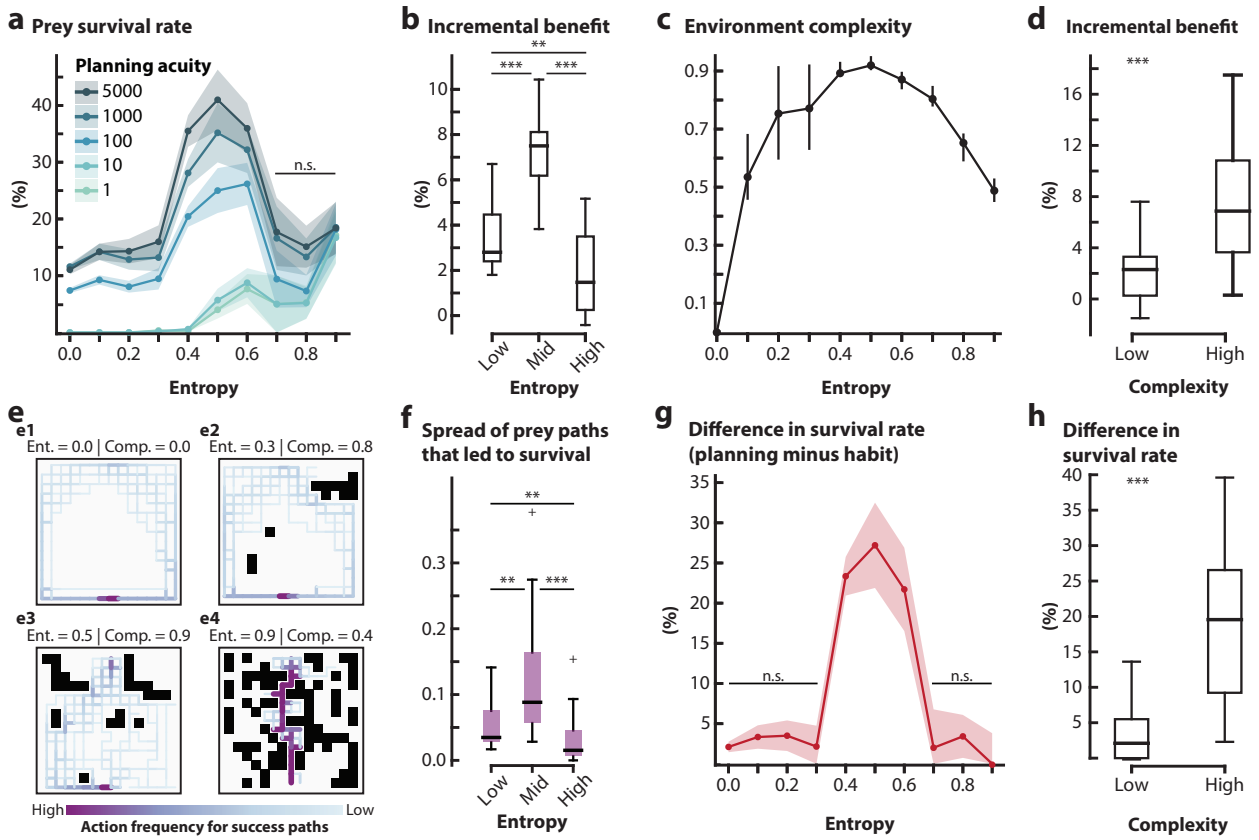

Figure 3: **The effect of clutter in the environment on survival and the emergence of multiple viable futures.** **(a)** Mean survival rate as a function of environment entropy for a given planning acuity, across different initial predator locations ($n = 5$). Fill indicates $\pm$ sem across randomly generated environments ($n = 20$). **(b)** The incremental benefit of planning—the mean difference in survival rate for each tested increase in planning acuity—grouped by environmental entropy range. Low: Environments of 0.0–0.3 entropy; Mid 0.4–0.6; High 0.7–0.9. **(c)** Network complexity with respect to entropy. The line plot shows the mean complexity and the interquartile range ($n = 20$). **(d)** The incremental benefit of planning (see **b**) when environments are grouped based on spatial complexity. **(e1–e4)** Heatmaps of all prey survival paths (from start to safety without being captured) that were taken by prey with high planning acuity (5000) 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. Black squares correspond to occlusions. **(f)** Path spread is defined as the variance of the paths employed by a prey with high planning acuity (5000) that result in survival (examples in **(e)**) across different trials within a fixed environment at a predetermined entropy level. High path spread indicates that the success paths are highly dissimilar. Low, mid, and high entropy ranges as in **(b)**. The outliers are indicated with a '+'. **(g)** Mean difference in survival rate between prey with high planning acuity, and prey that rely on habit-based action selection (see Methods). Fill indicates $\pm$ sem across the tested environments ($n = 20$). **(h)** Distribution of differences of survival rates between prey with high planning acuity, and prey that relies on habit-based action selection (full data in **(g)**) when environments are grouped based on spatial complexity.

defined in terms of equivalence and diversity of cell visibilities, has two boundary cases that are simple[23]: fully connected environments (open, entropy $= 0$), and unconnected environments (fully occluded). However, between these edge cases, complexity increase until midrange entropy values, and then decreases for high entropy values (**Fig. 3c**). It is in these midrange, spatially complex environments that planning becomes significantly more advantageous, and results in larger gains

in survival rate from increasing planning acuity (**Fig. 3d**, Mann-Whitney U test: $P < 0.001$) Interestingly, fractal analyses that are commonly used to quantify the detail in complex patterns found in natural environments[24] (**Fig. 5a**) indicates that environments that have high spatial complexity have terrestrial fractal geometries (**Fig. 5c**, Mann-Whitney U test with Bonferroni correction Open–Land: $P < 0.001$). Our results show that only terrestrial environments (fractal dimension corresponding to midrange entropy, 0.4–0.6, **Fig. 5b**) advantage higher degrees of planning (**Fig. 5d**, Mann-Whitney U test with Bonferroni correction Open–Land: $P < 0.001$).

To analyze the behavioral ramifications of increased planning in environments with variable complexity, we quantified prey behavior across episodes by counting the number of times an action between linked cells was taken when the prey succeeded in reaching safety. While this varied with the predator's initial position, all such action sequences are termed "success paths." The success paths in low entropy, low complexity environments reveal that planning generates action sequences that take the prey to the opposite wall from the believed predator location (**Fig. 3e1–e2**). While this causes the overall spread of success paths to be low (**Fig. 3f**, Mann-Whitney U test with Bonferroni correction Low–Mid: $P < 0.001$), it also shows that wall-following behavior—or thigmotaxis, commonly observed in rodents in open-field tests[25]—increases survival rate. This same behavior is seen in pseudo-aquatic environments independent of assigned visual range. Therefore, in simple environments success paths are easily learned, and are determined by environmental properties rather than prey sensory capabilities. In these environments, there is no statistically significant difference between survival rate with habit-based action selection[6,26], which uses previously employed success paths, and survival rate at high planning acuity (entropies 0.0–0.3, **Fig. 3g**, One-way ANOVA $P > 0.05$).

The wall-following behavior exhibited by the prey, which results in a low spread of success paths, suggests that the predator will follow a competing strategy that is similarly dependent on environmental properties. Complementary to our previous analysis, we quantified success paths for the predator by calculating the frequency of actions taken by the predator in episodes that resulted in predator success (capturing the prey) (**Fig. 4a**). The predator's trajectories are similar to those observed in pursuit tasks in open environments with primates[27], and seem to arise as a result of easy access to predicted prey locations. These findings suggest that in simple environments, the pursuer strategy is to go towards regions that are well connected (the middle of the environment), while the prey strategy is to go towards regions that have limited access.

The connectedness of cells was quantified by eigenvector centrality (eigencentrality), which represents the weighted sum of direct connections (actions to and from a cell), as well as indirect connections of every length[28] (see Methods) (**Fig. 4b**). High eigencentrality implies that the cell is connected to cells, which themselves are highly connected (have high eigencentrality values).

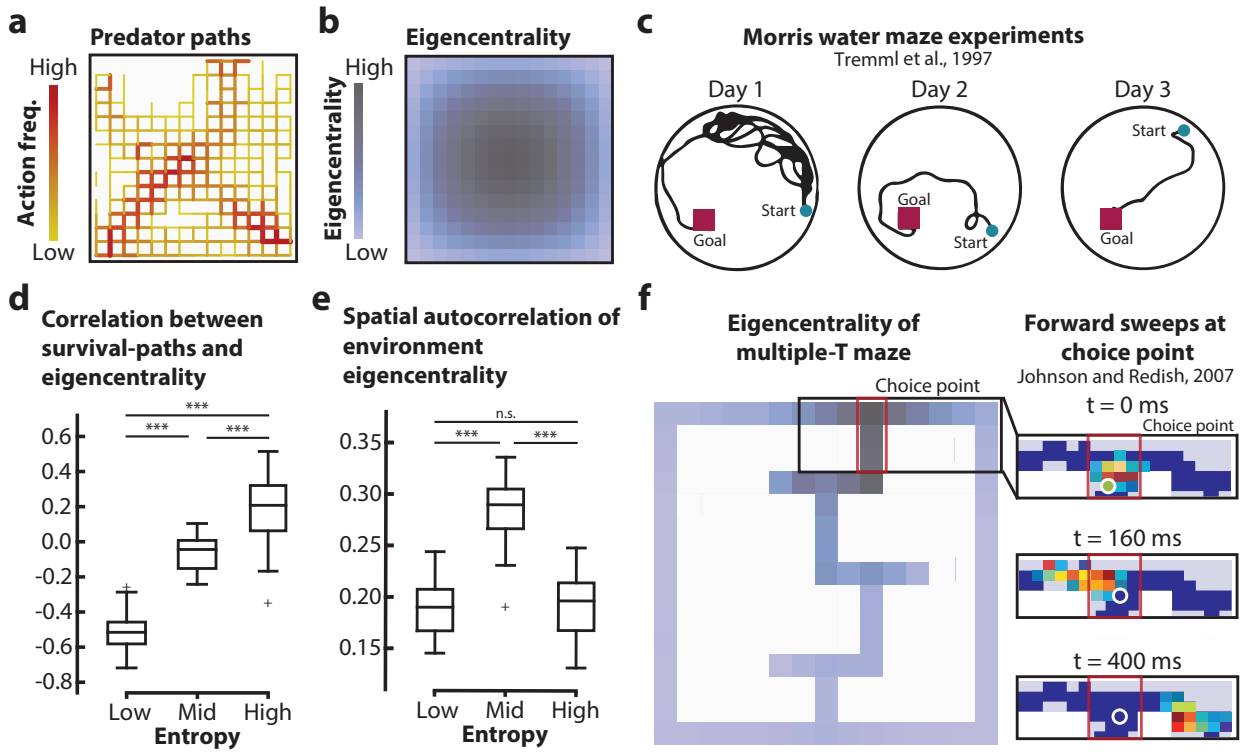

Figure 4: **Properties of the environment that influence behavioral strategy. (a)** Heatmap of all predator success paths (capturing the prey before the prey reaches safety), with color density proportional to frequency. Color bar: low: 1, high: 90. **(b)** Eigencentrality of an open environment. Color density proportional to the eigencentrality of each cell. Color bar: low: 0, high: 1. **(c)** Selected examples from swim paths of typical wildtype mice in a Morris water maze[29]. **(d)** Correlation between environment eigencentrality and prey success-paths (e.g. **Fig. 3a**). Low spread negative correlation indicates that the paths that led to success are similar and follow graph cells that have low eigencentrality. The outliers are indicated with a '+'. **(e)** Spatial autocorrelation (global Moran's I) of the environment eigencentrality. Higher spatial autocorrelation indicates that the low and high values of environment eigencentrality are more spatially clustered. The outliers are indicated with a '+'. **(f)** Multiple-T maze overlaid with eigencentrality. Color density proportional to the eigencentrality of each cell of the quantized maze. Color bar: low: 0, high: 1. Red box indicates the "choice" point where the rat pauses. Johnson and Redish[18] showed that the neural representation (reconstruction on the right) moved ahead of the animal while it paused at the choice point. A time stamped neural representation of the part of the maze highlighted by the black box is depicted on the right hand side. The intensity of the neural representation is shown in pseudocolor (red high probability; blue low probability). The actual position of the animal is shown with a white circle.

The few success paths that are seen in open (low entropy) environments, employed by prey with high planning acuity, are along cells of low eigencentrality (**Fig. 4d**, Spearman's rank correlation $P < 0.001$). Conversely, the predator success paths, independent of predator initial location, are along cells of high eigencentrality, which allows easy transitions to neighboring regions (**Fig. 4a,b**). Competitive sports analyses similarly point to the importance of reaching positions of higheigencentrality for rapid transitions to less connected regions[30].

This relationship between maneuverability within the environment and eigencentrality also has implications for exploration. It has previously been shown that anxiety in rodents (including in novel

environments) decreases exploration[25,31]. In open environments exploration requires animals to go towards regions of high eigencentrality, which results in high exposure to attack. As rodents become more comfortable, thigmotactic behavior decreases, and occupancy of the middle of the environment increases[29] (**Fig. 4c**) signaling an increase in exploration towards the goal point.

Contrary to low complexity environments, which allow simple strategies to succeed, in high complexity environments the prey's survival strategy becomes less stereotyped as the number and spread of viable paths increase (**Fig. 3e3,f**, Mann-Whitney U test with Bonferroni correction Low–Mid: $P = 0.003$, Mid–High: $P < 0.001$). The spread of trajectories that lead to success in midrange entropy environments—signaled by their diffuse occupancy frequency maps (**Fig. 3e3**)—suggests that prey success is dependent on the ability to rapidly update planned action sequences based on the predator's location with respect to occlusions. This implies that planning in dynamic environments is only beneficial for animals that can accurately localize the threat or opportunity within a limited region. Therefore, the sensory modality or combination of modalities used for such localization has to have high spatial and temporal resolution for rapid and precise updates. Vision through air is uniquely suited for this role (see Discussion). Consistent with the hypothesis that planning is needed for dynamic environments and requires spatially and temporally precise location updates, with habit-based action selection—no evaluation of future actions based on current observations—performance is much worse in spatially complex environments (**Fig. 3h** Mann-Whitney U test: $P < 0.001$). Unlike habit-based selection, re-planning in spatially complex environments generates complex behaviors that strategically deploy occlusions to escape from the predator, such as through hiding, or by engaging in diversionary tactics not unlike the broken-wing anti-predator tactic that is observed in birds[32] (see Supplementary Video S1).

Complex behaviors, exhibited in spatially complex environments, are a consequence of the distribution of eigencentralities. Unlike simple environments (**Fig. 4b**), which have a region of high eigencentrality that tapers in all directions away, spatially complex environments exhibit adjacent clusters of highly and poorly connected regions (**Fig. 4e**, Mann-Whitney U test with Bonferroni correction Low–Mid: $P < 0.001$,). Consequently, in such environments simple action sequences—such as following cells with low eigencentralities—fail (**Fig. 4d**, Mann-Whitney U test with Bonferroni correction Low–Mid: $P < 0.001$). Similarly, environments that fall within the terrestrial fractal dimension range are spatially complex, and exhibit higher divergence between success paths. This causes habit-based action selection to be unsuccessful (**Fig. 5e**, One-way ANOVA $P < 0.001$).

The clustered nature of eigencentralities in complex environments (**Fig. 4e**) forces the prey to move between regions of high and low eigencentrality. Transitioning from a poorly connected region to a highly connection region implies increased visibility and access from nearby cells. Planning at

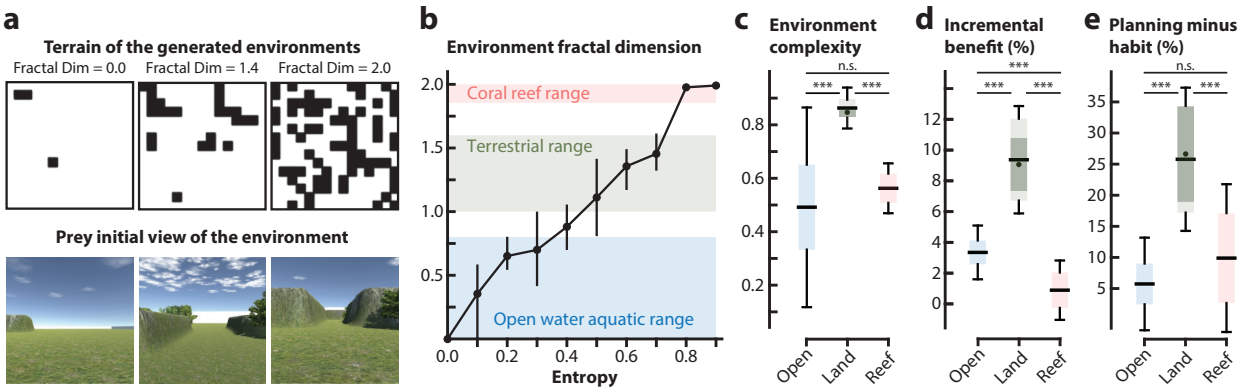

Figure 5: **Performance in typical example environments.** **(a)** Top-down view of example environments (black squares correspond to occlusions) and corresponding prey sideways view (Unity 3D reconstruction, occlusions are made into hills) from its start location. **(b)** Distribution of fractal dimensions of generated environments. The line plot shows the mean fractal dimension and the interquartile range ($n = 20$). The blue fill shows the range of fractal dimensions (0–0.8) that are observed in typical aquatic environments (see Methods), the green fill shows the range of fractal dimensions (1.0–1.6) that are observed in typical terrestrial environments[33,34], and the pink fill shows the range of fractal dimensions (1.9–2.0) that are observed in typical coral reefs[35,36]. **(c)** Environment network complexity (see **Fig. 3c**) of typical open water, terrestrial, coral reef environments. Dark green dot and band corresponds to mean environment complexity $\pm$ std, in environments that have a fractal dimension of $\approx$1.3 (peak human navigational performance[34]). **(d)** The incremental benefit of planning (see **Fig. 3b**) of typical open water, terrestrial, and coral reef environments. Dark green dot and band as in **(c)**. **(e)** The difference in survival rate between prey with high planning acuity, and prey that rely on habit-based action selection (see **Fig. 3g**) within typical open water, terrestrial, and coral reef environments. Dark green dot and band as in **(c)**.

these transition points is imperative for the prey to account for predator trajectories, and thereby safely navigate the exposed areas. Complex behaviors (detailed above) occur at these transition points, and result in highly variable action sequences contingent on the most recent observation of the predator and location of occlusions when transitioning to the highly connected regions.

Prior research in dynamic environments, where the threats and opportunities are always changing, has shown that animals rely on model-based planning methods to correctly react to the changes in the distribution of rewards[12,26]. Our results suggest that planning becomes more important, and results in higher behavioral variability at these transition regions form either high to low (**Fig. 4d**), or low to high eigencentrality dependent on the task. This framework provides a way to understand empirical phenomena influenced by environment topology, such as the rapid sweeping of hippocampal place cells to and from the maze junction that corresponds to a transition point from high to low eigencentrality in a multiple-T maze task[11,18].

Where environmental entropy is high and spatial complexity low, the amount of clutter in the environment constricts the profusion of success trajectories to one or two (**Fig. 3e4, Fig. 3f**, Mann-Whitney U test with Bonferroni correction Mid–High: $P < 0.001$). In these environments, the importance of planning is diminished (**Fig. 3b, d, h**), and we find no statistically significant differ-

ence between survival rate with habit-based strategies and survival rate at high planning acuity (**Fig. 3g**, One-way ANOVA $P > 0.05$). The high level of clutter causes a very local region of high eigencentrality that decreases sharply in the surrounding regions. Due to the reduced state-space, which causes there to be only a few routes to safety, the prey no longer needs a strategy that uses environment-dependent variables for success, eliminating the necessity to plan.

Discussion

Our results indicate that the complexity of the terrestrial world that greeted our fish ancestors[22], along with their massively increased visual range, advantaged planning in predator-prey interactions. The core basis for the apparent diminished elaboration of planning systems in most aquatic animals may therefore be low spatial complexity relative to the short range of aquatic vision in a behavior that is a key driver of evolutionary change. However, our open world (zero entropy) and short visual range approximation of the aquatic life of fish, while appropriate for both coastal fish[21] and for the turbid-water dwelling transitional tetrapods leading to land vertebrates[1], is less applicable to daylit coral reefs in clear water, due to a much higher visual range (eye parameters from a large set of diurnal reef fish; see Methods).

To examine the importance of planning in the coral reef environment, we used the fractal dimension of coral reefs (1.9–2.0[35,36]) to select from the range of environments we simulated. Our analysis revealed that coral reef environments are similar to high entropy environments (0.8–0.9; **Fig. 5b**). Thus coral reefs habitats plus their surrounding water can be thought of as a high entropy space with multiple safety points[37] surrounded by a zero entropy open space. In line with this idea, similar to other high entropy environments, there is little benefit to increasing the amount of planning due to the high level of clutter and few routes to multiple safety points (**Fig. 5d**). Moreover, based on our results, there is unlikely to be any advantage to planning in such highly cluttered but spatially simple environments (**Fig. 5c**). A simple habit-based action selection that relies on a cognitive map, using previously successful trajectories, would perform just as well as high planning acuity (**Fig. 5e**, One-way ANOVA $P > 0.05$).

Prior work on prospective coding in the hippocampus indicates that the forward sweeping of spatial representations occurs at important decision points when the reward contingencies are uncertain[18]. The predator-prey model we have used for this study is just a sub-category of this broader phenomena where the animal has uncertainty about where the reward is located. In our model, while the safety or goal position remains in the same location, the predator is an unpredictable, sometimes unobserved aversive stimulus that has to be avoided. The randomness in the predator model incentivizes the consideration of long action sequences, and therefore high planning acuities. However, our results indicate that even in tasks where there is a dynamic aversive stimulus, which causes the value function to change at each time step, after consolidating

long successful action-sequences, selecting actions by using a habit-based system would enable animals to succeed in getting to the goal position in spatially simple environments.

While the selective benefit of planning is dependent on habitat, we expect there is no dependence of our results on our choice to only model planning in the simulated prey, versus allowing the predator to plan. This choice was based on the high computational burden of having both the prey and predator engaged in planning[38]. In addition, while most animals are prey (including rodents, where the leading non-primate model of planning, vicarious trial and error[11], occurs), only a subset are predators, so this choice enables our results to have wider applicability. In the edge cases of open or highly cluttered habitats, our findings suggest that the failure of planning to provide advantage is unlikely to change as a result of our choice. We further chose to focus on the role of planning to the exclusion of learning by pre-equipping both predator and prey with a cognitive map of the boundaries, occlusions, and safety point of the habitats.

Despite our finding that there is no advantage to planning for most aquatic animals, it can still be highly useful for such animals to have a cognitive map. This is because there are many uses of a cognitive map that are independent of planning, such as present location on a map[39] and corrections to a dead reckoning system[40], with none of the prospective counterfactual aspects of checking multiple sequences of choices before enacting one as suggested by studies of planning in rodents[40,41]. Consistent with this hypothesis, it has been found that a likely homolog of the mammalian hippocampus in fish, the dorsolateral pallium, is hypertrophied in squirrelfish, a nocturnal reef fish[42], and evidence for how the pallium generates a cognitive map has been found in another animal that inhabits complex habitats with short sensing range, the electric fish[43]. With these animals and other fish that appear to need more spatial navigation abilities and have enlarged dorsolateral pallia[42], we may be observing an intermediate stage in evolution where cognitive maps are needed, but without the "simulation in the imagination" step needed for planning, a step that may require prefrontal cortex in mammals and its homolog in birds called the nidopallium caudolaterale[44].

Another potential constraint on the evolution of planning circuitry in vertebrates is its computational cost, which is exponential with the number of steps ahead being planned over. As a coarse indicator of this cost, 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 fully explored). This crude estimate does not take into account efficiencies related to various shortcuts such as hierarchical decomposition[45,46]. Nonetheless, even with such shortcuts planning is recognized to be demanding, and easily degrades into habit-based action selection under time pressure[47]. In contrast, although there is some cost associated with accumulating the statistics for habit-based

action selection, once learned the cost to select and release the action pattern is likely constant and does not increase with the number of steps within the habitized sequence.

In mammals planning engages prefrontal cortex in addition to the hippocampus[11]. Both of these forebrain structures approximately scale with overall brain size[48]. The evidence from brain development is that the most likely brain alteration for selection of a particular brain-size-dependent trait is a coordinated enlargement of the entire brain[48]. Brains of birds and mammals, where the existence of planning behaviors and its neuronal basis is most well documented, are greatly increased in size relative to fish—a factor of 10x larger at body sizes of 1 kg and 40x at large body sizes[49]. Interestingly, a third group of terrestrial animals, the reptiles, do not exhibit this increase in brain size; theirs is similar that of fish, which like them are also ectotherms.

The cognitive ability of reptiles is poorly understood, and there is little evidence of planning[50,51]. Yet, these animals have existed in the same complex terrestrial habitats, with aerial vision, as mammals and birds. There are several possibilities: the first is that these animals are able to plan and this is simply not known due to lack of study; the second is that the selective benefit of planning, while applicable to these animals, is not high enough to overcome an additional constraint; the third is that the selective benefit of planning arises from several other factors, either instead of or in addition to the ones we have uncovered, which apply to birds and mammals but not to terrestrial ectotherms. 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 previous point on the computational cost of planning. 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 avoidance[52]. A 10°C increase in temperature of muscle above ambient doubles the rate at which muscle can reach maximum power[53].

Endothermy likely played an important role in greatly increasing the size and computational power of the brain[49], and in increasing the sensitivity and temporal resolution of vision[54]. The flicker fusion frequency of the retina 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°C[54]. 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 x 30 = 24,000 times more mass flux assuming identical extraction efficien-

cies; even with the doubled efficiency of gills the mass flux is four orders of magnitude higher for respiration with water compared to air[1,55,56]. 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 planning is either unlikely to occur in underwater animals or occur in diminished form relative to birds and mammals.

Our work highlights an important role for one sensory modality above others for the aptly named capability of foresight: vision. Other than echolocation, a relatively recent mammalian innovation[57], there is no other sensory modality that provides the high temporal and spatial resolution best able to support planning over online dynamic scenes. In terms of our framework, one way to think of planning using olfactory, auditory, electrosensory, magnetosensory, or other teleceptive senses is that they reduce the entropy of the scene relative to what is obtained through high acuity vision. So a terrestrial scene that—when sensed with a high acuity visual system—is at midrange entropy where spatial complexity is maximized and the advantage of planning peaks, becomes a scene at lower entropy with lower spatial complexity where planning loses its advantage. In addition, for a modality like olfaction, the speed of update is much lower than with vision.

These considerations regarding the special importance of vision in the selective benefit of planning leads us to examine two invertebrates with exceptional visual acuity, one terrestrial and one aquatic. While the average acuity of insects is 0.25 cycles per degree (cpd)[58] (compared to 60 cpd for humans)—far below what we expect is needed for in-range spatial complexity to make strategic maneuvers useful—there is an insect with truly exceptional acuity, the jumping spider *Portia fimbriata* with acuity similar to cats and birds at 25 cpd[59]. These animals are known to stalk their prey, and to perform route detours to reduce the risks associated with attacking their quarry, which happen to be spiders that specialize on eating other spiders[60,61]. The octopus also has an exceptional acuity of 46 cpd and shows evidence of cognitive abilities[62,63]. Octopus researchers have noted that the best characterized species—due to simple considerations of accessibility to observation—inhabit shallow-water reef habitats, and further that the less accessible species seem to have more limited behavioral diversity[63, p. 286]. These points suggest that a comparative study of related species of octopus in reef environments versus open and deep water environments could be informative. Nonetheless, the prior points about planning in a reef environment apply here as well. The cleverness of the coleoid cephalopods may arise from a different source[64], such as predatory stress from animals that already have evolved planning circuits—namely birds and aquatic mammals[63]—in the context of having no armor to battle these predators with.

These findings suggest that with terrestrialization, there would be reduced need for specialized circuits in the vertebrate brain to generate habitual predator-avoidance maneuvers. The Mauthner system, which reduces the delay time between detection of a predator and the generation of an escape, is perhaps the best characterized of such circuits[65], and is only known to be present in vertebrates up through amphibians[66]. With the advent of vision on land and the planning it advantages, there is greater selection pressure for those circuits that support strategic behavior generation, such as the imagination of multiple futures prior to action selection. A key question for future research into the mechanisms of planning is what constrains the temporal and spatial limits of the futures we are able to imagine. Is it what was adaptive in the ancestral environment? Work on these issues—towards a neuroscience of sustainability—may become increasingly important as we face distant existential threats far beyond the range of our individual planning acuity.

Author Information The authors declare no competing financial interest. Correspondence should be addressed to M.A.M. (email: maciver@northwestern.edu).

Methods

This study concerns planning and as such, for both Computational Experiment 1 and 2, we isolate the role of planning by assuming that the prey and the predator have previously learned the environment (where the boundaries are, where the occlusions are, and the location of safety). The prey has also learned how the predator moves, and with what speed it moves. Some of the randomly generated worlds resulted in a situation where the prey and predator continued to move but the episode did not terminate. If a trial did not terminate within 200 time steps, it was excluded from further analysis. For Computational Experiment 1, the number of steps before termination was 19 $\pm$ 6 (mean $\pm$ std). For Computational Experiment 2, the number of steps before termination across all topologies was 30 $\pm$ 17. The higher standard deviation arises because the number of steps varies greatly with the degree of clutter.

Computational Experiment 1 (CE1): Pseudo-aquatic

A virtual prey and predator act in an empty 15×15 virtual discretized environment (gridworld) (**Fig. 1e**). The aim of the prey, which starts at a fixed initial position at the middle of the bottom row of cells, is to get to the fixed safety position while being pursued by a predator. The episode terminates with positive reward if the prey is able to reach the safety position, and negative reward if the predator "catches" the prey (meaning that the predator moves to the cell the prey is occupying). The prey is designed as a planning agent (see **CE1: Planning problem definition and evaluation of future states** for details) that forward simulates a specified number of states (termed planning acuity hereafter) before executing an action. As noted above, the forward simulation of states relies on an accurate environmental model (transitions between cells, and boundaries of the environment), and an accurate predator model (how the predator will move in space). The predator spawns at a random initial position exclusive of two cells: the cell occupied by the prey, and the safety cell. A total of $n = 20$ random predator locations were used for CE1 per visual range (1–5), and per planning acuity (1, 10, 100, 1000, 5000). Survival rate was calculated over 100 episodes at a given predator spawn location, prey visual range, and prey planning acuity (for a total of 50,000 trials). The predator is designed as a model-based reflex agent that selects actions based on the policy: aggressively pursue the prey with 75% probability, and act randomly with 25% probability. The predator is on average $1.5\times$ faster than the prey, which falls within typical terrestrial predator speeds relative to prey of 1.2–2[67,68].

The prey has a pre-specified visual cone that faces the direction of motion and extends outward 1–5 cells ahead (fixed at a single value per trial) (**Fig. 1e**). The prey always knows its own location within the gridworld, but only knows the predator location if the predator is inside of the prey's visual cone. If the predator is outside of the prey's visual cone, the prey samples a predator location from its belief state ($\mathcal{B}(., h_t)$) proportionate to the belief state distribution (for example, if

the predator is believed to be at cell (8,12) with 90% probability, then 9 out of 10 draws from the distribution will be (8,12)). The number of samples (predator locations) the prey draws from its belief state is equal to the prey's planning acuity. For each of these samples, the prey constructs a planning tree (see below) and evaluates to the termination condition.

Initially the prey's belief state is that the predator is located at all possible locations outside of its visual cone with equal probability. Until the first observation, the prey's belief state is propagated based on the prey's model of the predator movement; if the predator is then expected to be within the visual cone, but is not, then the set of belief states is correspondingly pruned and now the distribution reflects the new set of belief states. In between two observations, the prey's belief state consists of all the possible places the predator might have moved given the location of the predator at the time of observation and the prey's model of the predator action-selection policy.

The predator can observe the entire gridworld, and therefore knows the location of the prey at all time points. During aggressive pursuit, the predator chooses actions that minimizes the Euclidean distance between itself and the prey (if there is more than one cell the same distance to the prey, then the cell is randomly picked from this set). During random action selection, if the prey is within the reach of the predator, the predator chooses actions that terminates the episode with predator success. Otherwise, the predator chooses a random action that keeps it within the confines of the gridworld.

One time step consists of: 1) The prey greedily choosing an action $a \in \mathcal{A}$ based on its forward evaluation (see **CE1:Planning problem definition and evaluation of future states**). If that action brings the prey to the safety point then the episode is terminated (survival); 2) The predator choosing an action $a \in \mathcal{A}$ based on its pursuit policy (aggressive or random). If that action brings the predator to the prey position, then the episode is terminated (death); 3) The prey receiving an observation $o \in \mathcal{O}$ and reward $r = \mathcal{R}(s)$ from the environment; 4) The prey adding the action it chose and the corresponding observation to its history h; 5) The prey updating its belief state $\mathcal{B}(s, h)$ based on current history h; 6) The prey planning until a fixed planning acuity is reached.

Planning problem definition and evaluation of future states. We formulate planning as a partially observable Markov decision process (POMDP) consisting of the following variables[69]: a set of states $\mathcal{S}$ (prey and predator spatial location), a set of observations $\mathcal{O}$ (0 if the predator is not observed, cell number corresponding to predator location if the predator is observed), a set of actions $\mathcal{A}$ (cardinal directions), a list of action-observation pairs that constitutes the history h, a belief state $\mathcal{B}(s, h)$ specifying the probability distribution of the prey being in a state s given history h, a reward function $\mathcal{R}(s)$ defined as the expected immediate reward for a given state s, and discount factor $\gamma = 0.97$[70] that attenuates distal rewards.

The prey selects an action based on its policy $\pi(a|h)$, receives an observation, and collects rewards as it moves through states. Each state is associated with a value that denotes the expected total future reward starting from $s \in \mathcal{S}$ following a policy $\pi(a|h)$ ($\mathcal{V}^{\pi}(h)$). The maximum value that is achievable by any policy $\pi(a|h)$ gives the optimal value for that state ($\mathcal{V}^{*}(h)$). This value can be explicitly represented in terms of the reward function ($\mathcal{R}(s_t)$), and the belief state ($\mathcal{B}(s,h) = \mathbb{P}(s_t = s|h_t = h)$).

$$\mathcal{V}^{*}(h) = \max_{\pi} \left(\mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^{t} \mathcal{R}(s_t) \mathbb{P}(s_t = s|h_t = h) \right] \right)$$

Here, the prey's aim is to estimate $\mathcal{V}^{*}(h)$ by using its environmental model before taking an action (further disscussed in **CE1: Planning algorithm**). The prey internally simulates its own actions, the reactionary actions of the predator, and the corresponding observations and rewards. These internal simulations are used to approximate the value function without explicitly calculating it. The number of states ahead the prey evaluated was set by the planning acuity, which was one of [1, 10, 100, 1000, 5000].

Planning algorithm. Forward simulation of future states is implemented by a tree-like planning system (Monte-Carlo tree search adapted for POMDPs (POMCP)[69,70]), which relies on a previously learned model of the environment (boundaries and predator model). After an observation $o \in \mathcal{O}$ is received and a state sampled from the belief state $\mathcal{B}(., h_t)$, the prey begins planning from its current history h_t to estimate the optimal value function $\mathcal{V}^{*}(h)$. Each node in the search tree, denoted by $T(h)$, has three elements associated with it: $\mathcal{B}(h)$ specifying a set of possible predator locations (which converges to a single state when the prey observes the predator), number of times a specific history h has been visited ($N(h)$), and the expected value of an action and corresponding observation ($\mathcal{V}(h)$). $\mathcal{V}_{\text{init}}(h)$, and $N_{\text{init}}(h)$ are initialized to 0 for new nodes .

Planning tree construction and node value estimation in POMCP is divided into two stages: a tree-search policy that is on nodes with non-zero visit values (within-tree-search), and a rollout policy for nodes that have not previously been visited. After the evaluation of a state (predator and prey location), the node containing the first new history visited in the second stage is added to the search tree . The planner uses partially observable UCT (PO-UCT) during the first stage within-tree-search, which selects actions based on Upper Confidence Bound (UCB1)[71]; and a uniform random rollout policy during the second stage. PO-UCT has been proven to converge to the optimal value function[70],

$$\mathcal{V}(h) \rightarrow \mathcal{V}^{*}(h) \; \forall h \text{ as } N(h) \rightarrow \infty$$

which implies that when the planning acuity is high, an action that is greedily selected based on the search tree is the optimal action to perform.

After an action a_t is greedily selected (maximal expected reward) and an observation o_t is received from the environment, the planning agent's history is updated to reflect the new sample $\langle a_t, o_t \rangle$. The start node of the search tree and the associated belief state is updated to reflect the current history. The rest of the tree is pruned, since all other simulated histories are no longer representative of possible futures.

Statistics. Significance among groups were tested by using one-way ANOVA and Mann-Whitney U test with Bonferroni correction were applicable. All significance indicators follow: n.s.: $P \geq 0.05$, *: $0.01 \leq P < 0.05$, **: $0.001 \leq P < 0.01$, ***: $P < 0.001$

In **Fig. 2b**, the incremental benefit of planning is defined as the average difference in survival rate between tested planning acuities (1, 10, 100, 1000, 5000) (e.g. difference in survival rate between planning acuity 1000 and planning acuity 100) for a given visual range. Due to a non-uniform increase from planning acuity 1000 to planning acuity 5000 (difference is not 1 when converted to $\log$), a linear relationship was assumed, and the calculated difference was multiplied by 2.

In **Fig. 2c** relative feature importance on survival rate was calculated by using random forests. Feature importance was calculated by comparing the value lost or gained from the inclusion or exclusion of that particular feature. A total of 1000 random trees were created from a training set that was randomly chosen from the data.

Computational Experiment 2 (CE2): Pseudo-terrestrial

A virtual prey and predator act in a 15×15 virtual descretized environment (gridworld) (**Fig. 1f**), that features randomly added clutter with controlled density. A total of $n = 20$ random environments, with 10 levels of clutter—quantified by environmental entropy (see **CE2: Environment generation with randomized clutter** for details)—were generated. The predator and prey model used for this experiment are the same as **CE1**. The aim of the prey and the predator are kept the same as in **CE1: Task definition**. The predator spawns at a random initial position exclusive of prey start location, safety position, and occlusions. A total of $n = 5$ random predator locations were used for CE2 per environment randomization (20), per clutter level (10: entropy levels 0.0 through to 0.9 in steps of 0.1), and per planning acuity (5: 1, 10, 100, 1000, 5000). Survival rate was calculated over 50 episodes at a given predator spawn location, environment randomization, clutter level and planning acuity (for a total of 250,000 trials). A trial terminated if the prey reached the safety point, the predator moved the prey location, or if the episode reached cut-off point.

Unlike **CE1**, the prey can see the entire environment except where blocked by occlusions (**Fig. 1f**). If an occlusion exists on the ray[72] between the predator and the prey, the prey samples a state from

its belief state ($\mathcal{B}(., h_t)$). Initially, if the predator is not observed by the prey (behind an occlusion), the prey's belief state is all possible locations that are unobservable from the given prey location. Until the first observation, the prey's belief state is propagated then pruned based on the prey's model of the predator movement and locations that are hidden from the prey's position. In between observations, the prey's belief state constitutes all the possible places the predator might have moved (based on the prey's model of predator movement) given the location of the predator at the time of observation and all the locations that are not visible from the prey's position.

The existence of occlusions impedes both the prey's and the predator's line of sight. Therefore, the predator knows the exact location of the prey if an occlusion is not present on the ray between the predator and the prey. The predator keeps track of the prey location while the prey is in view. When the prey is hidden, the predator propagates the prey's last known location randomly within the gridworld (exclusive of occlusions) to form a belief state. During aggressive pursuit, if the prey is within view the predator uses the actual prey location to choose an action that minimizes the Euclidean distance. If the prey is hidden, the predator aims to minimize its Euclidean distance to a randomly sampled prey location from the belief state. During random action selection, the predator chooses a random action that does not move the predator to an occlusion or outside of the gridworld.

The sequences of computations that occur in one time step are the same as **CE1: Task definition**.

Environment generation with randomized clutter. The entropy of a general $m \times n$ descretized environment can be calculated by treating the descretized environment as a binary matrix, where 1's represent occlusions, and 0's represent unoccupied cells. The entropy of such an environment ($\mathrm{Ent}(g)$) can be written as:

$$\begin{aligned}\mathrm{Ent}(g) = &- \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{I}(g_{i,j}=0)}{mn}\log\left(\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{I}(g_{i,j}=0)}{mn}\right)\\ &- \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{I}(g_{i,j}=1)}{mn}\log\left(\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbb{I}(g_{i,j}=1)}{mn}\right)\end{aligned}$$

where $g_{i,j}$ refers to the value at the ith row and the jth column.

In generating the occlusions for the environment we assume a random walk policy of a random length that starts at an unoccupied random position. The number of random walks performed can at least be 1 and at most be the number of occlusions for a given entropy, here denoted as k. The total number of random walk lengths l must equal to k: $\sum_{1\leq l\leq k} l = k$. This process is repeated if a path from the fixed prey position to the fixed safety position does not exist (via the A* algorithm[73]). The algorithm for random environment generation with a fixed entropy level is described in the

supplement .

Environment complexity analysis. If an occlusion exists on the ray between the predator and the prey, both the prey and the predator are hidden from each other. By using the above principle we created a *visibility network* $G_v = (V_v, E_v)$ for all randomly generated environments. Vertices in this visibility network represent individual cells. An edge $e_{i,j}$ exists between two vertices $\{v_i, v_j\}$ if an occlusion does not exist on the line between the two vertices. This can formally be written as:

$$E_v = \{\{v_i, v_j\} \,|v_i \in V_v, v_j \in V_v, v_i \neq v_j \text{ and } l(v_i, v_j) \cap O = \emptyset\}$$

where $l(v_i, v_j)$ determines the vertices that fall on the line between v_i and v_j, and O refers to the set of occlusions specific to the environment.

Each vertex v_i has a degree $deg(v_i)$ that specifies the number of vertices that are connected to the vertex v_i. With such a network transformation, an environment with $\mathrm{Ent}(g) = 0$ is a complete graph with vertex degree $deg(v_i) = N - 1,\ \forall v_i \in V_v$. On the other hand, an environment that is only clutter is a disconnected graph with vertex degree $deg(v_i) = 0,\ \forall v_i \in V_v$. Therefore, the complexity of a graph passes through a maximum and goes down to zero for complete and disconnected graphs. An argument in support of this complexity definition arises from Shannon's information theory applied to random graphs[23]. Mathematically, the complexity of a network is defined as:

$$\alpha = \{deg(v_i)|v_i \in V_v\}$$
$$H(\alpha) = -\sum_{i=1}^{|\alpha|} \frac{\alpha_i}{|\alpha|} \log\left(\frac{\alpha_i}{|\alpha|}\right)$$

Environment fractal dimension analysis. The fractal dimension of the randomly generated environments (see **CE2: Environment generation with randomized clutter**) was calculated by using the box counting algorithm[74]. The basic procedure is to insert grids of decreasing box size over the environment, and count the number of cells that include occlusions for each box size. The fractal dimension of a gridworld is calculated by:

$$D = \frac{\log(N_r)}{\log(1/r)}$$

where N_r is the number of boxes that cover the pattern and r is the magnification, or the inverse of the box size. Therefore, the slope of the line when $\log N_r$ is plotted on the y-axis and $\log(1/r)$ is plotted on the x-axis equals the fractal dimension of the environment. A linear regression for this log-log plot was fitted to calculate the fractal dimension of each environment.

We used a similar approach on underwater images that ranged from murky to clear water to estimate the observed fractal dimension in aquatic environments. The colored aquatic images (I) of sizes $M \times M \times 3$ were converted to grayscale, and partioned into a grid with box sizes $s \times s$, where $M/2 \geq s > 1$. Based on the differential box-counting approach[75], $N_r = \sum_{i,j} n_r(i,j)$, where $n_r(i,j) = \max I(i,j)_k - \min I(i,j)_k + 1$ and k refers to the box number in the third dimension. N_r is counted for different values of box sizes s, which determines the scale r. By using the above equation, we estimated D from the least squares linear fit of $\log(N_r)$ against $\log(1/r)$. This analysis revealed that aquatic environments have fractal dimensions that range from 0–0.8.

Eigenvector centrality of environments. Environment quantization into a grid structure lends itself to a network representation based on how the system is connected together internally. If we again assume each cell is a vertex, in order to represent the environment dynamics, we can now define the edges in terms of actions. In such a network $G_w = (V_w, E_w)$, an edge $e_{i,j}$ between two vertices v_i, and v_j exists if there is an action connecting the two vertices. Similar to before we can formally write this as:

$$E_w = \left\{ \{v_i, v_j\} \,|v_i \in V_w, v_j \in V_w, v_i \neq v_j \text{ and } p(v_i) \xrightarrow{a} p(v_j) \; \forall a \in \mathcal{A} \right\}$$

where $p(v)$ returns the cell for vertex v.

Eigenvector centrality (eigencentrality) depends both on the vertex degree $\deg(v_i)$ and neighboring vertex centralities. The centrality score x of a vertex v_i is defined as[76]:

$$x_{v_i} = \frac{1}{\lambda} \sum_{j=1}^{n} A_{v_i,v_j} x_{v_j}$$

where λ is the largest eigenvalue of the adjacency matrix A_{v_i,v_j}.

Habit agent simulations implementing policy reuse. To model the consequences on performance for prey that acts based on policy reuse rather than planning (**Fig. 3g**), we implemented a variant of the PRQ-Learning algorithm[77]. A policy library $L = \{\Pi_1, ...\Pi_n\}$, specific to an environment, was created based on the prey success-paths (**Fig. 3e**) (prey going from start to safety without being captured). A policy $\Pi_k \in L$ was chosen by a softmax decision maker: $P(\Pi_k) = \frac{\exp^{\tau W_k}}{\sum_{p=0}^{n} \exp \tau W_p}$, where W_k is the reuse gain of implementing the chosen policy, and τ is the temperature parameter. After the implementation of the chosen policy Π_k, the reuse gain (W_k) is weighted by the total discounted reward R, and the number of times the policy Π_k has been chosen (N_k): $W_k = \frac{W_k N_k + R}{N_k + 1}$.

The predator action-selection policy was the same as the one implemented in the planning task (see **CE2: Task definition**). The full habit-based action selection algorithm is provided in the

supplement .

Statistics. For all environment groupings, environments with entropies below the 25th percentile were categorized as "low", and similarly environments with entropies above the 75th percentile were categorized as "high". Mid-level entropy was classified as entropies between "low" and "high".

For environment groupings in **Fig. 3d, h**), environments with spatial complexities below the 25th percentile were categorized as "low", and similarly environments with spatial complexities above the 75th percentile were categorized as "high".

For environment groupings in **Fig. 5c, d, e**), environments were grouped based on their fractal dimensions. Environments with fractal dimension below 0.8 were categorized as open water aquatic, environments with fractal dimension between 1.0–1.6 were categorized as terrestrial, and environments with fractal dimension between 1.9–2.0 were categorized as coral reef. Environments within the range 1.25–1.35 were categorized as environments in which peak human navigation performance occurs[34].

Significance among groups were tested by using one-way ANOVA and Mann-Whitney U test with Bonferroni correction were applicable. All significance indicators follow: n.s.: $P \geq 0.05$, *: $0.01 \leq P < 0.05$, **: $0.001 \leq P < 0.01$, ***: $P < 0.001$

In **Fig. 3b, d** and **Fig. 5d**, the incremental benefit of planning is defined as described above for **Fig. 2b**.

In **Fig. 4e**, the spatial autocorrelation of the environment eigencentrality was calculated by using global Moran's I. The weight matrix was set to be the inverse of the vertex distances (min. number of actions to get from v_i to v_j). This creates a weight matrix with greater values for vertices that are closer together. Global Moran's I evaluates whether a set of given values and their locations are clustered, dispersed, or random. For this statistic, the null hypothesis is that the spatial distribution of feature values (eigencentrality score of a vertex) is random. .

Coral Reef Fish Visual Range Calculations. In order to calculate the visual range for a typical coral reef fish we used a clear water model[1]. This water model has an attenuation length of 1.17 m at 575 nm and a Secchi depth of 5.92 m. Given the position of the eyes we simulated visual range for horizontal viewing, using horizontal radiance ($\theta = 90°$ relative to nadir) with solar photon travel path angle $\phi = 180°$. This causes the diffuse attenuation coefficent (K_d) to be 0 across all wavelengths and depths (for more detail see MacIver et al. supplementary materials: Aquatic Firing Threshold[1]). For a more realistic underwater range, we used the contrast threshold of goldfish (*Carassius auratus*) to account for object invisibility due to exponentially attenuated contrast (for more details see MacIver et al. supplementary materials: Contrast Threshold for Aquatic Vision[1])

We used a sample of 211 diurnal species of teleost reef fish in 43 families with a size range of 0.044–0.638 m across individual species (data obtained from Schmitz et al.[78]). Most species in this sample were reef inhabitants living in clear marine environments, with only a few entering murkier brackish and muddy coastal waters. For our simulations we did not differentiate between the two, and used a water model representing the clarity of the deepest oceanic water. Based on the data, pupil diameter increased with body mass. Therefore, we binned the body masses into 4 equal frequency bins and caluclated the average pupil diameter for each bin. This resulted in pupil diameters of 2.13 mm, 3.08 mm, 4.08 mm, and 5.17 mm. For each of these pupil diameters we calculated visual range using horizontal spectral radiance data for clear water at depths 5 m, 10 m, and 15 m. The calculated visual ranges at these depths for a 30 cm black disk "mock preator"[79] for fish with the above pupil diameters viewing the predator horizontally are: 5.02 m, 4.84 m, and 4.66 m for a fish with pupil diameter 2.13 mm; 5.04 m, 4.88 m, and 4.72 m for a fish with pupil diameter 3.08 mm; 5.05 m, 4.89 m, and 4.77 m for a fish with pupil diameter 4.08 mm; 5.06 m, 4.91 m, and 4.80 m for a fish with pupil diameter 5.17 mm. Given the size range of these fish, this causes the visual range to be $\approx$7–114 body lengths.

Computing Environment. The computational resources for this work were provided by the Quest high performance computing facility at Northwestern University which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. The cluster is composed of 244 nodes of Intel Haswell E5-2680 processors with 128 GB memory/node, 184 nodes of Intel Xeon E5-2680 processors with 128 GB memory/node, 72 nodes of Intel Xeon Gold 6132 processors with 96 GB memory/node. Approximate runtimes: **CE1** 2,000 total compute hours (20 hours on 100 Quest nodes); **CE2** 300,000 total compute hours (3,000 hours on 100 Quest nodes).

Data availability.

Simulation data will be available for download upon the completion of the peer review process.

References

1. MacIver, M. A., Schmitz, L., Mugan, U., Murphey, T. D. & Mobley, C. D. Massive increase in visual range preceded the origin of terrestrial vertebrates. *Proceedings of the National Academy of Sciences* **114**, E2375–E2384 (2017).

2. Balboa, R. M. & Grzywacz, N. M. Power spectra and distribution of contrasts of natural images from different habitats. *Vision Res.* **43**, 2527–2537 (2003).

3. Gupta, A. S., van der Meer, M. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705 (2010).

4. Doll, B. B., Duncan, K. D., Simon, D. A., Shohamy, D. & Daw, N. D. Model-based choices involve prospective neural activity. *Nature neuroscience* **18**, 767–772 (2015).

5. Keramati, M., Dezfouli, A. & Piray, P. Speed/accuracy trade-off between the habitual and the goal-directed processes. *PLoS computational biology* **7**, e1002055 (2011).

6. MacIver, M. A. Neuroethology: From morphological computation to planning. In Robbins, P. & Aydede, M. (eds.) *The Cambridge Handbook of Situated Cognition*, 480–504 (Cambridge University Press, New York, NY, 2009).

7. Nilsson, D. E. Evolution: An Irresistibly Clear View of Land. *Curr. Biol.* **27**, R715–R717 (2017).

8. Snyder, J. B., Nelson, M. E., Burdick, J. W. & MacIver, M. A. Omnidirectional sensory and motor volumes in an electric fish. *PLoS Biology* **5**, 2671–2683 (2007).

9. Colombo, M. & Broadbent, N. Is the avian hippocampus a functional homologue of the mammalian hippocampus? *Neuroscience & Biobehavioral Reviews* **24**, 465–484 (2000).

10. Striedter, G. F. Evolution of the hippocampus in reptiles and birds. *Journal of Comparative Neurology* **524**, 496–517 (2016).

11. Redish, A. D. Vicarious trial and error. *Nature Reviews Neuroscience* **17**, 147–159 (2016).

12. Choi, J.-S. & Kim, J. J. Amygdala regulates risk of predation in rats foraging in a dynamic fear environment. *Proceedings of the National Academy of Sciences* **107**, 21773–21777 (2010).

13. Walters, M., C J Adkins & Redish, A. D. Non-local representation of future decisions in dorsal hippocampal ca1 during anxiety- like hesitation behavior in a robotic predator-inhabited foraging task. In *Society for Neuroscience 2018* (San Diego, CA, 2018).

14. Wu, X. & Foster, D. J. Hippocampal replay captures the unique topological structure of a novel environment. *Journal of Neuroscience* **34**, 6459–6469 (2014).

15. Allen, T. A. & Fortin, N. J. The evolution of episodic memory. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 10379–10386 (2013).

16. Clayton, N. S. & Emery, N. J. Avian models for human cognitive neuroscience: A proposal. *Neuron* **86**, 1330–1342 (2015).

17. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* **34**, 1812–1819 (2017).

18. Johnson, A. & Redish, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189 (2007).

19. Mattar, M. G. & Daw, N. D. Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience* 1609–1617 (2018).

20. Pezzulo, G., van der Meer, M. A., Lansink, C. S. & Pennartz, C. M. Internally generated sequences in learning and executing goal-directed behavior. *Trends in cognitive sciences* **18**, 647–657 (2014).

21. Nilsson, D. E., Warrant, E. & Johnsen, S. Computational visual ecology in the pelagic realm. *Philosophical Transactions of the Royal Society B-Biological Sciences* **369** (2014).

22. Stein, W. E., Berry, C. M., Hernick, L. V. & Mannolini, F. Surprisingly complex community discovered in the mid-Devonian fossil forest at Gilboa. *Nature* **483**, 78–81 (2012).

23. Bonchev, D. & Buck, G. A. Quantitative measures of network complexity. In *Complexity in chemistry, biology, and ecology*, 191–235 (Springer, 2005).

24. Bolliger, J., Sprott, J. C. & Mladenoff, D. J. Self-organization and complexity in historical landscape patterns. *Oikos* **100**, 541–553 (2003).

25. Simon, P., Dupuis, R. & Costentin, J. Thigmotaxis as an index of anxiety in mice. Influence of dopaminergic transmissions. *Behav. Brain Res.* **61**, 59–64 (1994).

26. 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).

27. Yoo, S. B. M., Piantadosi, S. T. & Hayden, B. Y. Monkeys predict trajectories of virtual prey using basic variables from newtonian physics. *bioRxiv* 272260 (2018).

28. Bonacich, P. Some unique properties of eigenvector centrality. *Social networks* **29**, 555–564 (2007).

29. Tremml, P., Lipp, H.-P., Müller, U., Ricceri, L. & Wolfer, D. P. Neurobehavioral development, adult openfield exploration and swimming navigation learning in mice with a modified β-amyloid precursor protein gene. *Behavioural brain research* **95**, 65–76 (1998).

30. Vučković, G., Perš, J., James, N. & Hughes, M. Tactical use of the t area in squash by players of differing standard. *Journal of Sports Sciences* **27**, 863–871 (2009).

31. Zangrossi Jr, H. & File, S. E. Behavioral consequences in animal tests of anxiety and exploration of exposure to cat odor. *Brain research bulletin* **29**, 381–388 (1992).

32. Lima, S. L. Ecological and evolutionary perspectives on escape from predatory attack: a survey of north american birds. *The Wilson Bulletin* 1–47 (1993).

33. Mandelbrot, B. B. *The fractal geometry of nature*, vol. 1 (WH freeman New York, 1982).

34. Juliani, A. W., Bies, A. J., Boydston, C. R., Taylor, R. P. & Sereno, M. E. Navigation performance in virtual environments varies with fractal dimension of landscape. *Journal of environmental psychology* **47**, 155–165 (2016).

35. Bradbury, R. & Reichelt, R. Fractal Dimension of a Coral Reef at Ecological Scales. *Marine Ecology Progress Series* **10**, 169–171 (1983).

36. Young, G. C., Dey, S., Rogers, A. D. & Exton, D. Cost and time-effective method for multi-scale measures of rugosity, fractal dimension, and vector dispersion from coral reef 3d models. *PLOS ONE* **12**, 1–18 (2017).

37. Hixon, M. A. & Beets, J. P. Predation, prey refuges, and the structure of coral-reef fish assemblages. *Ecological Monographs* **63**, 77–101 (1993).

38. Hula, A., Montague, P. R. & Dayan, P. Monte carlo planning method estimates planning horizons during interactive social exchange. *PLoS computational biology* **11**, e1004254 (2015).

39. Foster, D., Morris, R. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10**, 1–16 (2000).

40. Redish, A. D. *Beyond the cognitive map from place cells to episodic memory* (MIT Press, Cambridge, Mass., 1999).

41. Miller, K. J., Botvinick, M. M. & Brody, C. D. Dorsal hippocampus contributes to model-based planning. *Nat. Neurosci.* **20**, 1269–1276 (2017).

42. Demski, L. S. The Pallium and Mind/Behavior Relationships in Teleost Fishes. *BRAIN BEHAVIOR AND EVOLUTION* **82**, 31–44 (2013).

43. 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).

44. Guentuerkuen, O. & Bugnyar, T. Cognition without Cortex. *TRENDS IN COGNITIVE SCIENCES* **20**, 291–303 (2016).

45. McNamee, D., Wolpert, D. M. & Lengyel, M. Efficient state-space modularization for planning: theory, behavioral and neural signatures. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I. & Garnett, R. (eds.) *Advances in Neural Information Processing Systems 29*, 4511–4519 (Curran Associates, Inc., 2016).

46. Huys, Q. J. M. *et al.* Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences* **112**, 3098–3103 (2015).

47. Keramati, M., Smittenaar, P., Dolan, R. J. & Dayan, P. Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences* **113**, 12868–12873 (2016).

48. Finlay, B. & Darlington, R. Linked regularities in the development and evolution of mammalian brains. *Science* **268**, 1578–1584 (1995).

49. 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 evolutionary biology* **14**, 178 (2014).

50. Wilkinson, A. & Huber, L. Cold-blooded cognition: Reptilian cognitive abilities. In Vonk, J. & Shackleford, K. (eds.) *Oxford handbook of comparative evolutionary psychology* (Oxford University Press, Oxford, UK, 2012).

51. Burghardt, G. M. Environmental enrichment and cognitive complexity in reptiles and amphibians: Concepts, review, and implications for captive populations. *Applied Animal Behaviour Science* **147**, 286 – 298 (2013). Welfare of Zoo Animals.

52. Ruben, J. The evolution of endothermy in mammals and birds: From physiology to fossils. *Annual Review of Physiology* **57**, 69–95 (1995).

53. Bennett, A. F. Thermal dependence of muscle function. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **247**, R217–R229 (1984).

54. Fritsches, K. A., Brill, R. W. & Warrant, E. J. Warm eyes provide superior vision in swordfishes. *Current Biology* **15**, 55 – 58 (2005).

55. 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).

56. Glass, M. L. *Physiological Evidence Indicates Lungfish as a Sister Group to the Land Vertebrates*, 161–177 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009).

57. Parker, J. *et al.* Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228 (2013).

58. Caves, E. M., Brandley, N. C. & Johnsen, S. Visual acuity and the evolution of signals. *Trends in Ecology & Evolution* **33**, 358–372 (2018).

59. Williams, D. S. & MeIntyre, P. The principal eyes of a jumping spider have a telephoto component. *Nature* **288**, 578–580 (1980).

60. 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).

61. Tarsitano, M. Route selection by a jumping spider (portia labiata) during the locomotory phase of a detour. *Animal Behaviour* **72**, 1437–1442 (2006).

62. Mather, J. A. & Dickel, L. Cephalopod complex cognition. *Current Opinion in Behavioral Sciences* **16**, 131 – 137 (2017). Comparative cognition.

63. Hanlon, R. T. & Messenger, J. B. Cephalopod behaviour (2018).

64. PACKARD, A. Cephalopods and fish: The limits of convergence. *Biological Reviews* **47**, 241–307 (1972).

65. Bhattacharyya, K., McLean, D. L. & MacIver, M. A. Visual Threat Assessment and Reticulospinal Encoding of Calibrated Responses in Larval Zebrafish. *Curr. Biol.* **27**, 2751–2762 (2017).

66. Bierman, H. S., Zottoli, S. J. & Hale, M. E. Evolution of the Mauthner Axon Cap. *BRAIN BEHAVIOR AND EVOLUTION* **73**, 174–187 (2009).

67. Elliott, J. P., Cowan, I. M. & Holling, C. Prey capture by the african lion. *Canadian Journal of Zoology* **55**, 1811–1828 (1977).

68. Hedenström, A. & Rosén, M. Predator versus prey: on aerial hunting and escape strategies in birds. *Behavioral Ecology* **12**, 150–156 (2001).

69. Sutton, R. & Barto, A. G. *Reinforcement Learning: An Introduction (2nd Edition)* (The MIT Press, Cambridge, Massachusetts, 2018).

70. Silver, D. & Veness, J. Monte-carlo planning in large POMDPs. In *Advances in neural information processing systems*, 2164–2172 (2010).

71. Kocsis, L. & Szepesvári, C. Bandit based monte-carlo planning. In *European conference on machine learning*, 282–293 (Springer, 2006).

72. Bresenham, J. E. Run length slice algorithm for incremental lines. In *Fundamental Algorithms for Computer Graphics*, 59–104 (Springer, 1985).

73. Hart, P. E., Nilsson, N. J. & Raphael, B. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics* **4**, 100–107 (1968).

74. Liebovitch, L. S. & Toth, T. A fast algorithm to determine fractal dimensions by box counting. *physics Letters A* **141**, 386–390 (1989).

75. Sarkar, N. & Chaudhuri, B. B. An efficient differential box-counting approach to compute fractal dimension of image. *IEEE Transactions on systems, man, and cybernetics* **24**, 115–120 (1994).

76. Newman, M. E. Mathematics of networks. *The new Palgrave dictionary of economics* 1–8 (2016).

77. Fernández, F. & Veloso, M. Probabilistic policy reuse in a reinforcement learning agent. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, 720–727 (ACM, 2006).

78. Schmitz, L. & Wainwright, P. C. Nocturnality constrains morphological and functional diversity in the eyes of reef fishes. *Bmc Evolutionary Biology* **11** (2011).

79. Holmes, T. H. & McCormick, M. I. Size-selectivity of predatory reef fish on juvenile prey. *Marine Ecology Progress Series* **399**, 273–283 (2010).