



## Survival in a world of complex dangers

Dean Mobbs<sup>a,b,\*</sup>, Toby Wise<sup>c</sup>, Sarah Tashjian<sup>a</sup>, JiaJin Zhang<sup>a</sup>, Karl Friston<sup>d</sup>, Drew Headley<sup>e</sup>

<sup>a</sup> Department of Humanities and Social Sciences, USA

<sup>b</sup> Computation and Neural Systems Program at the California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA

<sup>c</sup> Department of Neuroimaging, King's College London, London, UK

<sup>d</sup> Institute of Neurology, and The Wellcome Centre for Human Imaging, University College London, London WC1N 3AR, UK

<sup>e</sup> Center for Molecular and Behavioral Neuroscience, Rutgers University–Newark, 197 University Avenue, Newark, NJ 07102, USA

### ARTICLE INFO

#### Keywords:

Action construction  
Dynamic survival coding  
Generalization  
Multidimensionality  
Specialization  
State construction  
Survival

### ABSTRACT

How did our nomadic ancestors continually adapt to the seemingly limitless and unpredictable number of dangers in the natural world? We argue that human defensive behaviors are dynamically constructed to facilitate survival in capricious and itinerant environments. We first hypothesize that internal and external states result in state constructions that combine to form a meta-representation. When a threat is detected, it triggers the action construction. Action constructions are formed through two contiguous survival strategies: generalization strategies, which are used when encountering new threats and ecologies. Generalization strategies are associated with cognitive representations that have high dimensionality and which furnish flexible psychological constructs, including relations between threats, and imagination, and which converge through the construction of defensive states. We posit that generalization strategies drive ‘explorative’ behaviors including information seeking, where the goal is to increase knowledge that can be used to mitigate current and future threats. Conversely, specialization strategies entail lower dimensional representations, which underpin specialized, sometimes reflexive, or habitual survival behaviors that are ‘exploitative’. Together, these strategies capture a central adaptive feature of human survival systems: self-preservation in response to a myriad of threats.

### 1. Introduction

Darwin’s observations across the archipelagos of the Galapagos Islands revealed that when organisms are faced with different ecological obstacles, they genetically adapt their physical phenotypes and behaviors to survive and flourish. Latent within the Darwinian framework is the notion that when the ecological dimensionality is low, species evolve to specialize and possess a genetic predisposition that tailors them to specific habitats (Darwin, 1859; Futuyma and Moreno, 1988; Finlay et al., 2001; Poisot et al., 2011). Such genetic profiles can result in biological stasis and a constrained yet effective range of survival strategies, particularly if the environment remains stable. While evolution often favors specialization, unpredictable or volatile ecological pressures—such as invasive predators, food depletion, or meteorological disasters can render the organism prone to extinction (Futuyma and Moreno, 1988; Gurevitch and Padilla, 2004; Carthey et al., 2017). As the renowned evolutionary biologist Ernst Mayr pointed out:

“Whenever a species acquires a new capacity it acquires, so to speak, the key to a different niche or adaptive zone in nature.” (Mayr, 2001; p 208).

These new capacities form in several ways: (i) over evolution, where the organism evolves physical phenotypes including biological machinery that promotes survival. This includes ‘fitter phenotypes’ including sensory ‘threat detection’ capabilities, reproductive adaptations (e.g., fecundity or where the organism produces more offspring than can survive), and body plans (e.g., shells, spikes); (ii) over the organism’s lifetime, where survival pivots on the organism’s learning and behavioral adaptability. The organism’s ability to integrate and oscillate between these strategies reflects different approaches to survival in ecologies that can be conceived from low to multidimensional (Hutchinson, 1957; Ingram et al., 2018). Generalizers respond to novel and high-dimensional environments by flexibly linking a large array of concepts and behaviors in an experience-dependent manner (Futuyma and Moreno, 1988), but at the cost of a relatively high sampling requirement. Theoretically, multidimensional environments can refine

\* Correspondence to: Department of Humanities and Social Sciences and Computation and Neural Systems Program at the California Institute of Technology, USA.  
E-mail address: [dmobbs@caltech.edu](mailto:dmobbs@caltech.edu) (D. Mobbs).

generalized models of the world and foster specialized repertoires, thereby providing a mixed profile of specialized and generalized strategies and a way for organisms to survive across multiple landscapes.

In this article, we put forward the Dynamic Survival Coding (DSC) framework, where human defensive behavior involves a collection of internal and external states that are dynamically constructed to facilitate survival across multidimensional environments. We examine specialized and generalized defensive strategies through the lens of survival decisions in humans, yet often discuss them in the context of research on other species to enrich our discussion. We aim to solve the disparity between ecological and contemporary affective models of defensive states by synthesizing models that extend from hardwired (e.g., Jaak Panksepp; LeDoux, 2012) to constructed emotions and behaviors (e.g., Feldman-Barrett et al., 2016). Our perspective reshapes the prevailing contemporary neuroscience framework by challenging the conventional dichotomy between reactionary and reflection systems by suggesting that specialization provides a successful solution to recurrent threats, and generalization strategies allow for ways that prey can survive an almost infinite number of dangers.

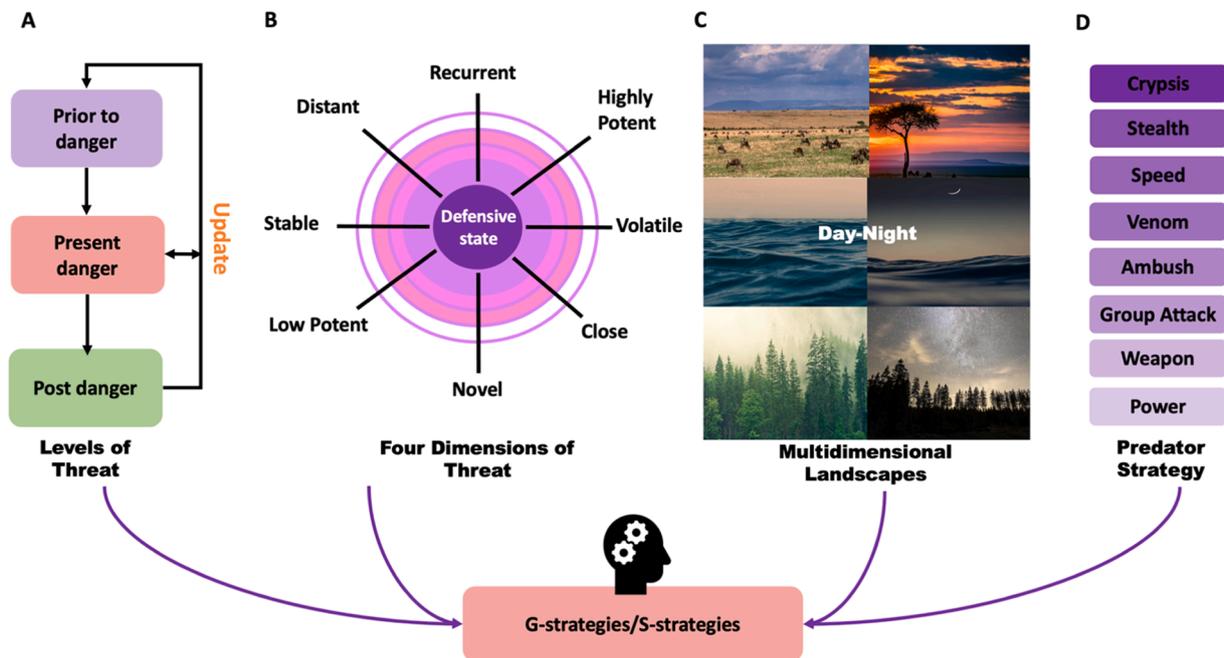
### 1.1. Limitless hazards and survival

Any serious theoretical synthesis of how the organism computes threat must consider their ecological conditions and the natural diversity of dangers they face (Fig. 1A and B; Fanselow and Lester, 1988; Mobbs, 2018). Predators, for example, have different attack strategies, such as stealth and speed, which are often tied to the environment (e.g., open plains or dense forests; Fig. 1C). As we discuss below, defensive behaviors are molded and elicited depending on the specifics of the threat, including whether a predator is spatially close/distant, volatile/stable, of high/low potency, and recurrent/novel dangers (Fig. 1B). When combined with the physical phenotypes of the predator (e.g., venom, speed, camouflage, and power; Fig. 1D), their behavioral strategies (e.g., stealth, group attack; Fig. 1D) and the structure of the environments (Fig. 1C), these dimensions produce almost limitless ways

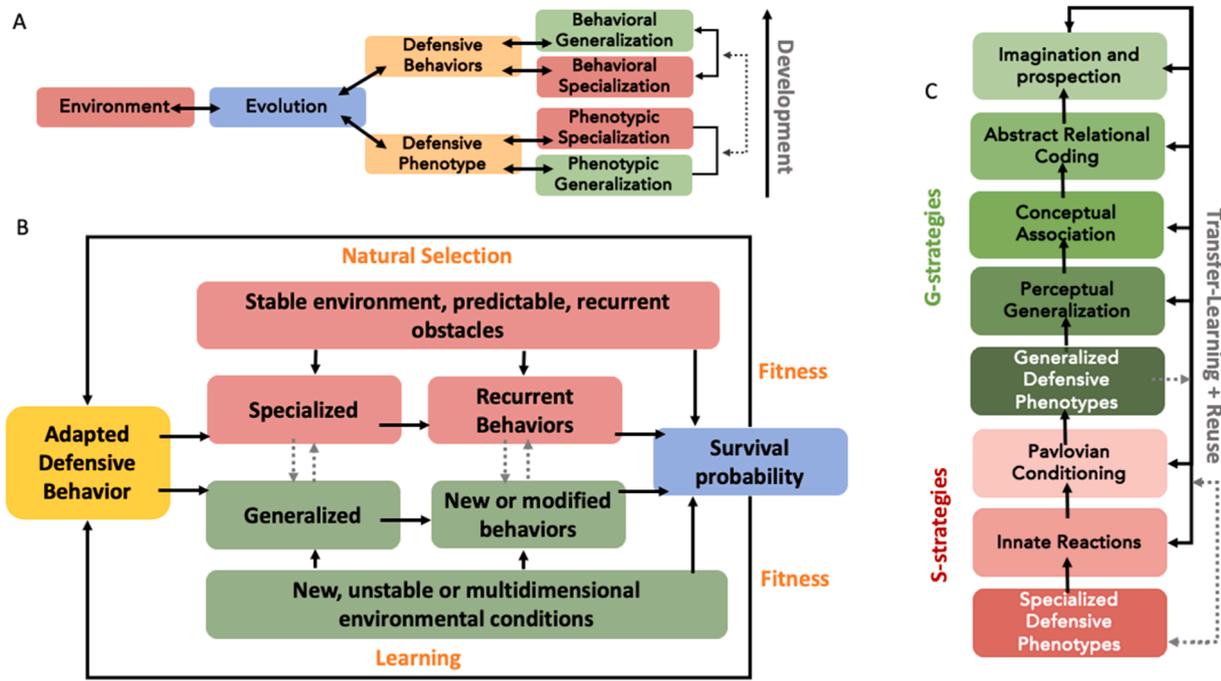
that predators can kill their prey. Further, the brain never experiences the exact same threatening situation twice. To account for these complexities, we hypothesize the need for a system that creates a meta-representation of all the internal and external states and matches them to previous experiences. Here a pattern detection and similarity comparator are critical to devising a successful survival plan. A flexible system is also critical to most species as the natural predators an organism faces may change throughout its life (Sih et al., 2000) and the shift may occur from seeking parental protection to avoiding the threat (Sullivan and Opendak, 2018).

One influential model is the Threat Imminence Continuum proposed by Fanselow and Lester (1988), which classifies threat levels into three core categories based on their imminence (Fig. 2A). However, while instrumental, the Threat Imminence Continuum does not encompass every aspect of natural conditions, notably factors including threat predictability, potency, and organism’s perception of safety (Fig. 2B). Together, these environmental conditions will determine the kind of defensive strategy an organism might employ (Mobbs et al., 2020; Tashjian et al., 2021).

Based on environmental and sensory statistics, specialized organisms will use recurrent hard-wired and learned Pavlovian responses that are highly predictive of threats in their ecology. When a threat is present, sensory cues – like smell, sound, or visual stimuli – will prompt the organism’s reflexive reactions toward danger. In rodents, for example, these sensory cues result in reflexive defensive behaviors such as freezing and, if attacked, escape. On the other hand, generalizers not only tailor their strategies to adapt to new and unpredictable environments but also plan for potential encounters with threats with which they may have no prior direct experience. This proactive approach to danger, or pre-empting danger, implies that they rely on inference, even when sensory cues are absent. Within the realm of reinforcement learning, these responses overlap with both model-free (e.g., habitual) and model-based (e.g., deliberative) learning (Section 3.3.1). Indeed, having a broader repertoire of strategies increases the ability for introspection and prospectively anticipate a larger variety of situations. To



**Fig. 1.** Examples of the multidimensional complexity of danger. (A) three core levels of threat, each providing different survival strategies; (B) Each level of threat has four dimensions each with different threat statistics; (C) Examples of abiotic features landscapes including forests, open savannas, and water across day-night cycles; (D) Examples of the different types of predator strategies used in nature. Note that predators can combine or use more than one strategy. Together A-D provides examples of the complexity of predicting and responding to threats, therefore resulting in a set of survival states that need both Specialization (S)-strategies and Generalization (G)-strategies. Note that we focus on predation and other dangers such as threats to, for example, reproductive and nutritional needs, which are not discussed, but would add to the dimensionality of danger.



**Fig. 2.** Examples of specialization and generalization in defensive phenotypes and learning. (A) Examples of different cognitive, behavioral, and phenotypic strategies associated with generalization and specialization. Defensive behaviors are determined by the predator landscape. The defensive phenotypes (i.e., shell, speed, or spikes) will have constraints that will impact other defensive responses. Although species-specific, these different strategies change as the organisms go from being dependent on the parent to being independent; (B) Adaptive behaviors are segregated into generalized and specialized systems (grey dotted lines) via learning (Modified from [Ricklefs and Wikelski, 2002](#)). These behaviors increase survival probability and fitness. (C) Examples of G-strategies and S-strategies and learning. Specialization begins at the level of the physical phenotype – this, in turn, will influence the type of learning needed in the organism niche. Basic innate and Pavlovian learning are used when encountering recent threats. Generalization learning strategies are used when predations become more complex and high-dimensional. Transfer learning facilitates the dynamic cross-fertilization of knowledge.

streamline this concept, we consolidate the Threat Imminence Continuum model into three key modes each leading to specific defensive sets and actions.

### 1.2. Before danger

The period without danger can be placed into two modes, as identified by [Fanselow and Lester \(1988\)](#). The first is the “safe mode”, during which an animal finds itself in a danger-free zone, such as being in a safe refuge or home dwelling. While there is no danger present in this state animals, particularly humans, could employ episodic and prospective systems to imagine past dangers and anticipate the likelihood of potential future dangers. By doing so, they can formulate preventative behaviors that minimize future encounters with dangers. These prospective systems are present in both specialized and generalized encounter scenarios. For recurrent encounters, prospective systems facilitate fast responses by priming neurons (e.g., meta-plasticity; [Abraham, 2008](#)) to respond in an adaptive manner. Additionally, prospective systems allow us to stimulate and work through possible dangerous scenarios that can occur, creating G-strategies for a variety of threats. To illustrate, one might think, “If the predator attacks this way, I have a pre-planned defensive response ready.

The second mode, known as the “pre-encounter threat mode”, is characterized by an environment with an absence of immediate threats but with a heightened possibility of encountering one. This mode signifies minimal anxiety and avoidance due to threats. During this mode of threat, various cognitive and behavioral systems are activated, including potential attack predictions (i.e., predator location), vigilance, and cautious behaviors (e.g., intermittent locomotion; [Kramer and McLaughlin, 2001](#)). Before danger, organisms can harness remote temporal threat associations to build G-strategies observed in neuronal populations linked to slower and more deliberate cognitive processes.

Prediction systems, like imagination, are advantageous; they enable the organism to preempt attack, therefore minimizing surprises and reducing the immediate need to seek protection ([Mobbs et al., 2015](#)). The prediction system further manifests protective behaviors, ranging from risk dilution behaviors like group living ([Hamilton, 1971; Tedeschi et al., 2021](#)) to niche construction, where organisms create safer environments such as nests, burrows, or walls ([Mobbs et al., 2015](#)).

### 1.3. Present danger

The presence of a predator prompts an assessment spanning multiple dimensions: volatility, potency, spatiotemporal proximity, and novelty ([Fig. 1B](#)). The factorial nature of these four components of threat assessment will determine the configuration of neural systems, in turn determining the defensive strategies and subsequent behaviors elicited. They involve a mixture of generalized and specialized representations, where the distal threat will result in behaviors resembling those seen during pre-encounter phases, yet increasing urgency pushes the organism to more reactive fear-like processes (e.g., active vs. passive avoidance). These responses are modulated by competing variables, including mating and nourishment needs, proximity to safety, and the presence of offspring ([Mobbs, 2018](#)).

Present and non-attacking threats characterize [Fanselow and Lester’s](#) (i) post-encounter threat, where a threat is present in the environment yet there is no direct predator-prey interaction. Within this context, specialized neurons will result in freezing as an innate defensive response. Indeed, [Fanselow and Hoffman \(2024\)](#) suggest that freezing is adaptive as movement triggers attack responses in predators. However, the prospect of flight reaction becomes plausible if a safety refuge is close ([Blanchard and Blanchard, 1989](#)). Such survival decisions might be governed by neuronal populations stimulating potential predator actions and strategizing escape routes, potentially supported by neural

reactivation and replay of prior attack encounters (Wise et al., 2021, Wu et al., 2017). Here, G-strategies will collect information on the danger and use for the current and future threat (See 5.1)

The highest level of threat is the (ii) circa-strike level, marked by the predator's attack on the prey. During a circa strike, S-strategies like protean escape (e.g., zigzagging; Humphries and Driver, 1970), come into play to a greater extent. These unpredictable movements impede the predator's chase. It is worth noting that these are effective for predators that the organism has an evolutionary history (i.e., S-strategies may fail when encountering new dangers). Proximal threats also induce an uncoordinated panic in the prey, further adding unpredictability to the prey's movements and challenging the predator's ability to anticipate their next move (Humphries and Driver, 1967). When all escape routes are blocked, the prey resorts to defensive attacks (e.g., fight responses). The goals of the prey are to increase the distance between itself and the predator and to escape to safety. Finally, as the predator draws closer, the prey will prepare for contact by eliciting analgesic responses that facilitate escape (Mobbs et al., 2015).

#### 1.4. After danger

Following an encounter with the threat, humans (and potentially other mammals), will go through two post-threat learning processes: conscious post-danger evaluation and non-conscious post-threat replay. (i) Post-threat evaluation. After a threat, the organism taps into episodic memory systems to reimagine the encounter. This process will be used for explicit or conscious prospection strategies. Not only does it help in reflecting on the successful escape from the threat, but it also aids in re-imagining alternative future scenarios that could be detrimental or beneficial in the future. An illustrative thought might be: "The next time I go to the forest, I will take a gun or knife." (ii) Post-threat replay. This process is marked by the reactivation of spatially-tuned neurons in sequences that previously resulted in successful threat avoidance (Louie and Wilson, 2001; Pezzulo et al., 2014). Interestingly, the same neuronal sequences can activate even before encountering a similar threat in the future, a phenomenon that could be termed "pre-threat pre-play". In turn, this replay mechanism reinforces threat-related neural activity, thus optimizing the organism's decision for future encounters. One hypothesis is that replay increases survival in response to recurrent threats by sharpening action-outcome predictions.

In sum, these different modes of danger will result in a set of contextually appropriate defensive actions in organisms. Selection pressures have honed the nervous system to respond as quickly and accurately as possible. Nonetheless, learning often unfolds under conditions that are novel or ambiguous, necessitating deeper information-seeking and integration. In the context of natural selection, defensive behaviors will alter in response to the amount of knowledge the organism has about the environment and its dangers. This will result in either S-strategies that have been successful in combating similar predators in the past, or new threats will require new and creative approaches as captured in G-strategies. In turn, the chosen actions will increase or decrease survival probability and fitness (Fig. 2A), which are embedded in adaptive physical phenotypes and behaviors associated with G-strategies and S-strategies (Fig. 2B). Additionally, mechanisms such as transfer learning (Fig. 2C) and cognitive reuse (Anderson, 2010) are crucial for developing G-strategies that can be refined into specialized tactics when necessary.

## 2. Multiple roads to survival: the concepts of G-strategies and S-strategies

The natural world encompasses a spectrum of predictability and unpredictability, featuring environments ranging from low to high and multidimensional ecologies. In low dimensional ecologies, where the predictability is high, organisms have limited cognitive capacity and survive by utilizing a set of specialized behavioral and physiological

phenotypes that genetically endow the organism with a successful set of tools to combat ecological dangers (Fig. 1A, B, and C). These tools represent what we call S-strategies. Conversely, high or multidimensional ecologies drive the organism towards physical phenotypes that combat a diversity of dangers, while the ability to have multiple behavioral strategies allows for flexibility and creativity when facing a new threat. We refer to these as G-strategies.

### 2.1. G-strategies

At their core, G-strategies are dependent upon an organism's ability to grasp the sensory and abstract structure of the world and to extrapolate beyond direct experiences. This yields cognitive heuristics and maps that link previous experiences, enabling the organism to creatively anticipate and counter novel threats. Models of brain representations of abstract relational structure (Kaplan and Friston, 2018; Kaplan et al., 2017; Whittington et al., 2019) suggest that hippocampal-entorhinal systems play a pivotal role. Place and grid cells in the hippocampal-entorhinal systems recast spatial and relational memory into abstraction, facilitating G-strategies (Moser, Kropff, and Moser, 2008). Further, this suggests that the cortex is involved in constructing generalizable (or domain-general) responses to many new challenges.

Contemporary models of the brain's capacity to represent abstract relational structure cast the process as an extension of basic spatial cognitive maps. This is inspired by Edward Tolman's work in the 1930s. Tolman showed that rodents learn about the structure of a maze, and can use that knowledge to take shortcuts – a demonstration of internal world modeling, or a cognitive map. More recent work has demonstrated that humans extend this cognitive mapping process, instantiated in the activity of grid cells, to even non-spatial, abstract knowledge (Constantinescu, O'Reilly, and Behrens, 2016). Such abstract relational mapping of the environment offers two key advantages when avoiding threats. First, it lets one transfer learned abstract relationship structures to new contexts. Second, it facilitates conscious imagination and prospection through episodic reconstruction (Hassabis et al., 2007). Here, humans can creatively use their imagination to anticipate danger and plan the best future strategy (Mobbs et al., 2015). Prospection aids in estimating threats by forming relational structures based on potential future scenarios. Additionally, it diminishes the necessity for multidimensional phenotypic protection. This ability is grounded in episodic memory, where previous experiences are stored and then re-used to visualize future events (Hassabis et al., 2007). The process of episodic reconstruction probably relies on the hippocampal-entorhinal systems, the same systems that enable the brain to navigate and represent the structure of its environment. These systems might employ methods such as replay to achieve this (Wise et al., 2021; Whittington et al., 2019). Additionally, encounters with threats promote the generalization of episodic memory as an adaptive cognitive strategy to avoid various threats (Starita et al., 2019).

As a key part of human G-strategies, imagination, and prospection empower us to shape our future with strategic intent. (Suddendorf and Corballis, 2007). For example, I may never have experienced a tiger attacking me from a tree, but I can imagine that they can climb. Therefore, I can imagine being attacked from above and I can construct counter-strategies to evade attack. Likewise, prospection also allows me to avoid unmatched predators by imagining where they might appear and avoiding that space. We propose that abstract relationships, combined with imagination and prospection, provide a comprehensive threat assessment – even for dangers never faced. Further, we speculate that G-strategies would be important for complex threats particularly those that involved danger from other humans. An abstract system that can model the states, ideologies, and intentions of others would be key to avoiding human danger. G-strategies would involve complex simulation systems that can navigate the complex ways that other humans can attack us. Vicarious social learning such as story-telling can further enrich the imagination (Mobbs et al., 2015). G-strategies also sit closely

with the idea that humans have advanced domain-general abilities suggesting that we have a unique ability to flexibly adapt to many different goals. This is in opposition to more domain-specific abilities observed in other animals where survival is restricted to a specific goal (Premack, 2010).

We speculate that G-strategies would be important for complex threats particularly those that involved danger from other humans. An abstract system that can model the states, ideologies, and intentions of others would be key to avoiding human danger. Further, G-strategies would involve complex simulation systems that can navigate the complex ways that other humans can attack us. We have previously stated (Mobbs et al., 2015) that humans have enhanced two core systems to avoid such dangers – imagination and vicarious learning. These two systems allow us to avoid dangers without personally experiencing them. Imagination provides a playground from which to understand other's minds (mentalizing) which is key to understanding the nefarious intentions of others. Further, vicarious learning allows us to learn through storytelling and gossip.

## 2.2. S-strategies

Ecological selection pressures shape the physical, behavioral, and cognitive phenotypes of organisms both over evolutionary periods and individual lifetimes. These ecological conditions often tune the organism towards S-strategies, optimizing energy and time efficiency. Examples of S-strategies are ubiquitous in the natural world. For example, Koalas and Panda bears have evolved extreme diets limited to eucalyptus leaves and bamboo, respectively. Similarly, specific adaptive responses to threats are observed when it comes to physical defenses against sympatric predators. For example, the honey badger's resistance to puff adder venom (Drabeck, Dean, and Jansa, 2015), the Rove Beetle's appeasement gland that discourages the ant attack. As Futuyma and Moreno state:

“It is not surprising that narrow physiological tolerances are sometimes found in organisms that inhabit relatively constant environments” (Futuyma and Moreno, 1988; p241)

In conjunction with defensive physical phenotypes, S-strategies form behavioral adaptations to recurrent threats. For example, terrestrial animals are often imprinted with innate responses to aerial threats. Monkeys, when faced with looming stimuli, respond with escape and defensive calls (Schiff, Caviness, and Gibson, 1962). Likewise, rats and mice exhibit freezing behaviors when exposed to 2,4,5-trimethylthiazoline (TMT)-an odor reminiscent of fox secretion (Wallace and Rosen, 2000; Brechbühl et al., 2013). Additionally, they exhibit thigmotaxis in open mazes, a tactic to minimize their visibility during pre-encounter states.

Defensive behaviors can become specialized in two ways: through learned and innate mechanisms. Learned specialization occurs when organisms consistently employ a generalization mechanism upon encountering familiar threats. Relevant work in reward learning has shown how the human brain chooses between model-based and model-free computation strategies based on each method's success (Lee, Shimojo, and O'Doherty, 2014). Consequently, as threats repeat, S-strategies refine defensive actions, with efficiency potentially enhanced by better neural coding. For instance, motor learning or procedural learning exemplifies this – complex actions, with repetition, become automatic, implicit, and ingrained in long-term memory. This idea sits close to Johnson's (2011) 'Interactive Specialization' framework where the brain is not just statically maturing based on rigid genetic programs but is constantly in flux with a dynamic interaction between genes and environment that results in neural specialization.

Innate components of S-strategies may evolve from genetic preparedness (Ohman and Mineka, 2001) or potentially from parental experiences that are passed to the offspring (Dias and Ressler., 2014). The idea of an innate specialized defensive circuit is also captured by the

theoretical work of Jaak Panksepp (Panksepp, 1998), a notion also explored from different perspectives by researchers like Joseph LeDoux, (LeDoux, 2012) and others (Anderson, 2016; Mobbs et al., 2015, see Mobbs et al., 2019). It is crucial to recognize that while certain traits might be shared across species, as noted by Barsbai et al., (2021), S-strategies are tailored to specific species and contexts. This is often captured by the idea that most species evolved domain-specific abilities where the animals have a limited range of adaptations (Premack, 2010).

## 2.3. G-strategies and S-strategies are not mutually exclusive

We speculate that G-strategies and S-strategies complement each other through transfer learning. Likewise, repeated G-strategies can evolve into specialized behaviors. Yet, S-strategies may also play an active role in the construction of G-strategies, shaping more immediate, online generalizations. Research has shown that even in the realm of habitual responses, there is an influence on deliberative processes (Moran, Keramati, and Dolan, 2021) and innate anxiety behaviors including thigmotaxis that occurs when there is an increased chance of encountering a threat and may be part of information seeking or exploration strategies (Simon et al., 1994). Indeed, such spatial mapping provides a cognitive map that will facilitate later escape. Indeed, one study has shown that the midbrain periaqueductal gray (PAG) encodes sequential motor programs (Yu et al., 2021) and the hypothalamus is involved in versatile escape behaviors (Wang et al., 2021) and memory coordination (Burdakov and Peleg-Raibstein, 2020). Although structures like the hippocampus, entorhinal, and prefrontal cortices are implicated in more elaborate processing, there is evidence that the hypothalamus and PAG would be informed by teaching signals to update their neuronal structure (Roy et al., 2014). This evidence underlies the potential of subcortical structures, often associated with automaticity, in shaping representations of the environment and supporting the prospective action selection and scaffolding needed for G-strategies.

## 3. Models of survival decision-making

Models of survival often fall under the rubric of defensive decision-making, emotion, and computation. We propose that no single model looks under the hood to characterize: (i) how the organism adapts to complex multidimensional environments; (ii) the processes involved in survival decisions and behaviors (e.g., G-Strategies and S-Strategies); (iii) synthesize current models of emotion and (iv) incorporate computational models of survival. Below we briefly summarize each model and set the stage for the DSC model by integrating these models and theories and planting them in the context of low, high, or multidimensional ecologies.

### 3.1. Ethological models of survival

Behavioral ecologists have long theorized about the role of predation in survival decision-making that encompasses the dynamic interaction and equilibrium between predator and prey (Abrams, 2000). For example, Lima and Dill (1990) proposed a flow model of the different levels of encounters between the predator and prey. In their model, predation risk can be captured by the following equation:

$$P(\text{death}) = 1 - \exp(-\alpha d T)$$

where  $\alpha$  is the rate of predator and prey encounters,  $d$  is the probability of death given an encounter and  $T$  is the time spent in situations of predation risk. Lima and Dill proposed that  $\alpha$ ,  $d$ , and  $T$  are the basic building blocks of predation risk and are accessible to be used by the prey for its benefit. If the predator is predictable or stays in the same patch for long periods  $\alpha$  will be low. Access to  $\alpha$  may also be adjusted by contingencies of the environment including safety distance (Lima and Dill, 1990). Therefore, according to this model, minimizing the number

of threat encounters, and reducing the level of danger and duration of these encounters will maximize survival.

Others have considered the active role of additional survival behaviors when encountering predators. For example, Moody, Houston, and McNamara (1996) included foraging, energy, and lifetime reproduction. They propose that the term  $y$  is the cost of being killed and hence the loss of reproductive success. In the context of increased energy per unit, the rate  $V$  is the value of the food to the animal. As the authors state: "When an animal chooses a foraging option with a net rate of energetic gain and predation rate its net rate of increase in reproductive value is"

$$gV - My.$$

Together,  $gV$  is the rate at which food consumed increases reproduction, and  $My$  is the risk of decreased reproductive value if the predator kills them. This simple equation produces a shorthand of several variables that impact the survival of the organism. Others have extended their models to optimal flight distance from approaching predators (Cooper and Frederick, 2007), protection of offspring (Moller and Nielsen, 2014), and nutritional state (Moran et al., 2021). However, these models do not consider the strategies the organism uses to evade the threat.

To fill this gap, the survival optimization system (SOS) model links animal, ecological, and decision-theoretic models to lay out the strategies that humans use to evade danger (Mobbs et al., 2015). The SOS model proposes that five core strategies are used when facing threats of different spatial and temporal proximity. These include: (i) Prediction strategies including imagination and prospection. If a threat is predicted, the organism can pursue (ii) Prevention strategies to avoid the future threat. If a potential threat enters the environment, (iii) threat detection strategies are engaged including bottom-up and top-down attention. Once the potential threat is detected, the organism goes into (iv) threat assessment and determines the value of the threat, how dangerous it is, tracks the threat's movements, makes predictions about what the threat will do next, searches for safety, and determine the best survival action (e.g., crawl, sprint, climb). Once the threat begins to attack, the organism goes into (v) reactive defensive strategies such as flight or fight.

Feeding into these five strategies is a set of monitoring systems including the appraisal, reappraisal, and cognitive control of emotions and actions, as well as a suite of learning systems (Mobbs et al., 2015). Its strength is that the danger will elicit a decisive response or neural circuit that reflects the context (Fanselow and Lester, 1988) and shift to and from the subcortical to cortical systems depending on the dimensionality and complexity of the threat (Mobbs et al., 2007; 2009; 2010; Qi et al., 2018; Faul et al., 2020). However, the model fails to capture several dimensions of the threat (e.g., high or low predictability), nor does it consider the information and how these defensive states arise (Fig. 1B and C). Further, like most models of survival, it fails to link emotional states and the dimensionality of the threats.

### 3.2. Affective models of survival

Ecological models of survival focus on the variables and strategic defensive decision-making processes, yet the emotional state that the organism is in and why this is useful for survival are rarely discussed. However, implicit in every model of emotion is how the emotion provides some survival advantage. Keltner et al. have proposed that:

"Emotions have the hallmarks of adaptations: They are efficient, coordinated responses that help organisms to reproduce, to protect offspring, to maintain cooperative alliances, and to avoid physical threats" (Keltner, Haidt, & Shiota, 2006, p. 117).

This sentiment is captured in Panksepp's theory (1988) of primary affective systems which posits that survival states like FEAR, RAGE, and

PANIC are genetically determined and laid down in subcortical circuits that are critical to survival. While the configuration and specificity Panksepp ascribe to the mammalian brain are debated (e.g., not natural kinds; Feldman-Barrett, 2006), it is clear that the FEAR system he proposes is supported by a multitude of empirical studies across species (See reviews of empirical literature: Gross and Canteres, 2012; Janak and Tye, 2015; Pellman and Kim, 2016). The observed ubiquity and convergence (Barsbai et al., 2021) of defensive behaviors, such as flight and freezing in terrestrial mammals, supports the theory that these reactions are intrinsic, hard-wired responses specialized for combating carnivore predators. Further, empirical studies show that targeted chemical or electrical stimulation of specific sectors of the midbrain or hypothalamus elicits a consistent set of defensive actions across species (Tovote et al., 2016; Fadok et al., 2017). Panksepp's theory provides support for S-strategies in that it suggests that neural circuits underlying defensive behaviors are innate. However, these reactive motor output systems offer a glimpse into the vast populations of neurons underpinning decision-making, factoring in contextual cues, memory, planning, and execution. A question remains on how effectively these hard-wired FEAR systems can navigate the more complex environments.

In contrast, Feldman-Barrett's Emotional Construction Theory (Feldman-Barrett, 2006) postulates that the emotional circuits, as conceptualized by Panksepp (1988), are purely motor and visceral and thus are not directly representative of emotions. Feldman-Barrett's theory also suggests that the brain is a dynamical system where neural networks interact with one another based on internal models that are pattern generators. This system aims to maintain an equilibrium, known as allostasis. These internal models are influenced by past experiences, shaping how the neural network predicts the sensory landscape of the environment. In turn, these predictions become conscious reflections of feeling states and emotions. Unpredictable events result in prediction errors, prompting adjustments in the central and autonomic nervous systems. These adjustments become perceptions that help us categorize the events, leading to defensive behaviors suitable for the situation at hand and therefore sit well with the concept of G-strategies. Feldman-Barrett further suggests that each emotion does not have a dedicated neural circuit. To support this idea, Feldman-Barrett suggests that fear (like other emotions) has been elusive in pinpointing biologically, and little is known about how circuits communicate, are integrated, and why they so often overlap, particularly in humans.

In the context of defensive strategies, Feldman-Barrett's model has a notable concern. Animals, having evolved under specific ecological conditions, inherit survival instincts that have been shaped over generations by both ecological pressures and natural selection. This is captured by Lorenz's concept of innate releasing mechanisms where the organism performs some survival behavior that has not been learned (e.g., bird of paradise dance). From this perspective, animals are born with certain predispositions towards potential threats in their ecology, contradicting the idea of them being a "blank slate" (McNally, 1987). As evident in various physical and behavioral traits, biological systems, such as the central and autonomic nervous systems, are primed for the specific dangers that are present in an organism's ecology. While such biological readiness aids learning, an animal too specialized in its behavior is likely to be impacted when the ecology changes. Under such circumstances, having the ability to adaptively construct survival strategies would be advantageous. Drawing from diverse information sources, including memory, constructionism allows the formulation of new defensive strategies and would be advantageous in multidimensional ecologies. While these models of emotion were designed for different purposes, they do not consider the G-strategies and S-strategies associated with different levels of threat imminence or the dimensions within (Fig. 1B-D).

### 3.3. Computational models of survival

Computational theorists have become increasingly interested in the

algorithms that underlie survival decisions and learning (Bach and Dayan, 2017). We propose that learning can take many forms that can be characterized under the rubric of G-strategies and S-strategies, although these may often be amalgamated. The most dominant framework for exploring the computational basis of decision-making and survival has been reinforcement learning (RL). RL builds on two biological foundations: the innate preparedness for specific classes of threat (LeDoux and Daw, 2018) and the physical phenotype of the organism (Fig. 2B). These two factors have evolved and represent the clearest form of S-strategies. RL through Pavlovian conditioning builds on S-strategies, gifting the organism the ability to flexibly learn about its environment. When the dimensionality of the environment increases, a shift to G-strategies occurs. In these scenarios, cognitive and higher-order systems will elicit a defensive response (Fig. 2A-D). Below we discuss both model-free and model-based types of RL and how these account for survival across low to high and multidimensional ecologies.

### 3.3.1. Model-free RL and survival

The simplest form of survival decision-making plays out in innate reflexes and Pavlovian reactions that are formed through learning (LeDoux and Daw, 2018). These include fixed action patterns, ranging from simple behaviors such as freezing and thigmotaxis to more complex behaviors like the exotic dance of the bird of paradise. These simple responses to threats are implicit, not consciously goal-directed, and often fast. The neural pathways for innate reactions to danger that come from visual and olfactory sensory systems, merge in specialized amygdala nuclei to create a defensive behavior via the midbrain PAG (i.e., freezing or flight). Instrumental behaviors, on the other hand, are based on action outcomes. Here, the defensive avoidance behavior becomes habitual via sensory systems and the dorsolateral striatum (Daw et al., 2005). These types of computations fall under model-free RL, characterized by stimulus-response association where a particular behavior has been reinforced in a particular context. In the context of reward, model-free RL is insensitive to outcome devaluation (Adams, 1982), thus linking the behavior to habit. Further, evidence shows that dopamine neurons, which are linked to RL, show specialized organization for sensory, motor, and cognitive variables (Engelhard et al., 2019). Model-free RL is one example of how S-strategies are formed.

### 3.3.2. Model-based RL and survival

While simple, model-free avoidance strategies can be effective, particularly when encountering easily predictable threats, they will often fail in situations where more complex avoidance strategies are warranted (e.g., G-Strategies). Furthermore, the requisite experience that informs model-free learning may not always be available. For example, while it is straightforward to learn from the experience of finding edible prey, it is less feasible to learn from the experience of being eaten. Model-based RL algorithms exploit an internal model of the world to enable more flexible avoidance behaviors that do not depend on direct experience, and which can generalize to new scenarios more easily. It is now well established that humans are capable of deploying model-based strategies to flexibly avoid danger and seek out safety (Tashjian et al., 2022). Such processes likely depend on neural reactivation of relevant state representations (Wise et al., 2021), where the agent perspective considers the likely long-run outcome of their actions. Recent findings demonstrate that humans are capable of simulating complex scenarios incorporating predictions about predators' goal-directed behavior (Wise et al., 2023), suggesting that our model-based avoidance abilities are both powerful and flexible in allowing us to simulate, and hence determine, how to avoid a multitude of threatening situations. Model-based strategies also provide for flexible recalculation of the optimal avoidance behavior when aspects of the environment change, without the need for relearning (although there are simpler strategies that can achieve similar performance in certain situations, see Russek et al., 2017). Therefore, Model-based RL algorithms may set the foundation for G-strategies.

In reality, model-based and model-free systems likely operate in concert to provide a balance between optimality and computational cost depending on the current context, potentially informed by uncertainty around the predictions of each system (Lee et al., 2014), the relative advantage of using the more computationally costly model-based system (Kool et al., 2016), or physiological stress (Otto et al., 2013). Model-based approaches also rely on the decision-maker maintaining an accurate model of the environment, which may not always be straightforward in complex, naturalistic environments (Mobbs et al., 2021; Wise, Emery & Radulescu, 2024), and may itself need to be learned via direct experience using model-free learning strategies (Sharp et al., 2023). In the context of G-strategies and S-strategies, we expect similar processes, yet model-free and model-based RL lacks a relationship between the strategies needed across changing environments and the link to internal states. In the subsequent sections, we will delve into the DSC framework and attempt to integrate the aforementioned models.

## 4. The basic architecture of the dynamic survival coding (DSC) framework

The essence of the DSC model is that the survival goal of organisms is to increase their safety by mitigating threats and uncertainty and adapting their behavior to low, high, and multi-dimensional environments. The DSC posits that the breadth of defensive responses depends on the dimensionality of the ecological niche(s) in which has organism has evolved—and survived. Put simply, the organism evolves to reduce uncertainty and danger by, on one hand, utilizing G-strategies notably exploration, avoidance, and curiosity, and on the other hand, by exploiting knowledge and utilizing S-strategies. The larger the dimensionality, the larger the uncertainty and the need for more sensory evidence, resulting in increased inference, cognition, and information-seeking behaviors. We reason that humans emphasize changes in behavior and cognition—rather than relying on specialized physical phenotypes—and will continually update their neural software and move towards more construction and higher-order representations (LeDoux and Brown, 2017). Importantly, these representations not only provide an advantage to escaping current threats but also to avoiding future ones.

### 4.1. Internal and external states

#### 4.1.1. Sensory information

The ecology in which the organism exists will come with specific statistics. Sensory systems serve as an organism's gateway to the external world, allowing the prey to detect and monitor danger. Indeed, Donald Hoffman (2010) has proposed that we do not see the world as it is, but perceive the world in a way that maximizes survival. As the late Richard Gregory (1965) stated:

“Eyes are biological early warning systems. By giving information of events distant in space they serve to probe the immediate future, allowing brains to transcend simple reflexes and control strategic behavior” p16.

Each sense provides unique input across levels of threat imminence. For instance, vision allows for the detection of distant or hidden threats, and sound and smell allow the prey to detect non-visible threats (Mobbs, Garg, and Tashjian, In Press). Depending on the situation and the species, attentional systems direct an organism's senses to focus on specific locations in the environment using either instinctual bottom-up or volitional top-down approaches. Across all species, the senses consistently provide the empirical information needed to refine predictions and guide actions.

In mammals, vision is predominantly relied upon during the post-encounter and circa-strike danger. For pre-encounter, a combination of senses, such as hearing, becomes crucial – particularly in occluded environments such as dense forests (Mugan and MacIver, 2020). For

some mammals, olfaction has a pivotal role during pre-encounter. A predator’s scent can signal that the threat has been in the environment, prompting the prey to engage other senses such as auditory and visual systems for a more accurate detection of the potential presence of the threat. The integration of multisensory information provides a comprehensive perspective, aiding in the construction of an appropriate defensive response. In addition to sensory information, internal states, and inferences about the threat (Fig. 3) will also determine if a defensive action is executed. Importantly, organisms that have a long history in a circumscribed location will have sensory systems that are tuned to their ecology, and this is a case of an S-strategy phenotype (Fig. 1A).

4.1.2. Internal states, and drives

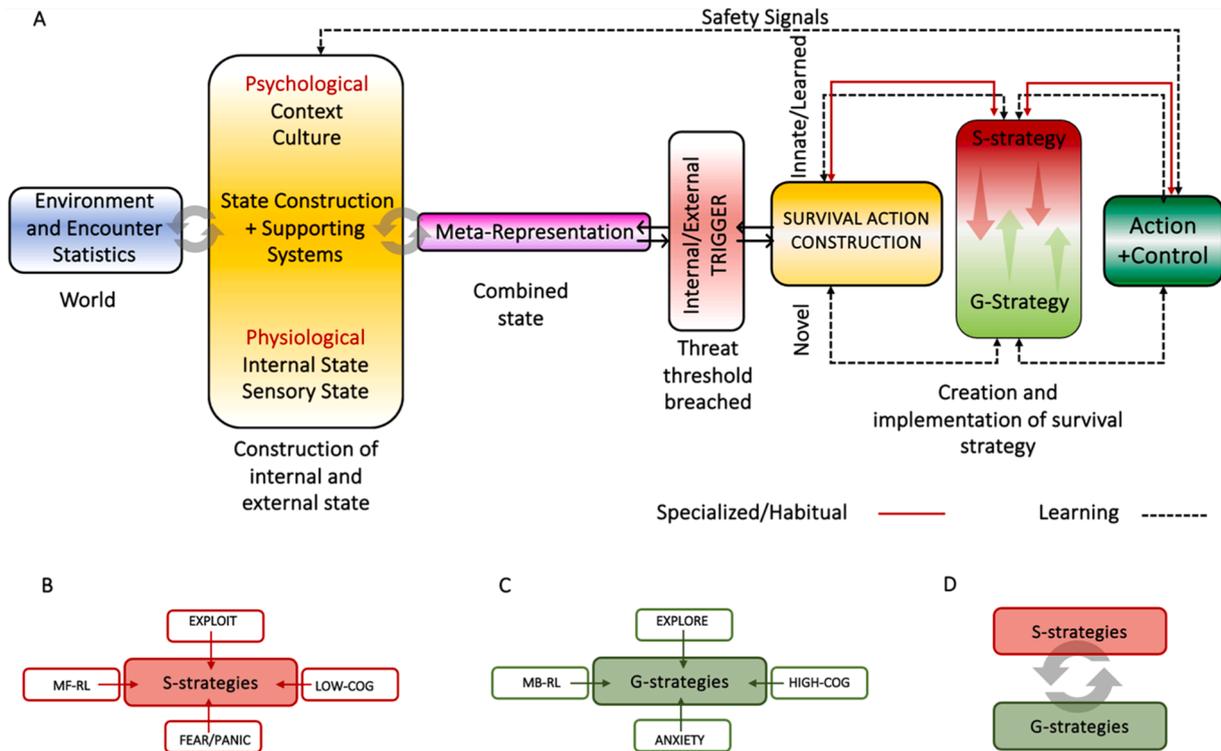
Internal states play a critical role in the instigation, control, and execution of the survival state. These states partly form what LeDoux (2012) has termed “survival circuits.” These include fundamental survival drives like hunger, thirst, and reproductive drives. The internal states of thirst and hunger deprivation can sway animals, humans included, between risk-taking and risk aversion. For example, work on animals has shown that hunger can prompt riskier foraging behaviors, pushing them to areas where predators might be present (Sih, 1980). The potential opportunity costs of losing crucial resources like food or water can result in the riskiest flight in mammals (Lagos et al., 2009).

Illustrating this, one study made rats excessively hungry by stimulating AgRP neurons in their hypothalamus. This hunger drove the rats to ignore danger signals (i.e., Fox urine) in their search for food (Burnett et al., 2016). Furthermore, hunger has also been observed to amplify extinction to threats (Verma et al., 2016). Conversely, if the threat seems particularly potent, or the internal state is only moderate, the organism will suppress its activities to avoid the threat (Gray, 1987).

Survival circuits, as LeDoux (2012) points out, are “closely intertwined”. It is the urgency of each internal state that will partly determine whether a defensive state is initiated. Neural circuits tied to survival states work in tandem with sensory, predictive, and conscious states associated with danger to construct defensive states and behavioral actions. This interplay parallels the concept of allostasis where the organism attempts to maintain a balance between internal states (Sterling and Eyer in, 1988) and minimize allostatic load (McEwen, 1998). Extending on this is the allostatic-interoceptive system theory which suggests that the brain contains a large-scale set of interconnected circuits that support allostasis (Kleckner et al., 2017). Consistent with LeDoux’s survival circuits, the DSC proposes that circuits involved in basic life functions are key in formulating defensive responses.

4.1.3. Meta-affective awareness

In humans, internal states, inferences, and sensory input will produce



**Fig. 3.** The Dynamic Survival Coding (DSC) model. (A) streamlined model of the DSC - A hazardous event will occur when there is an appearance or apprehension of a threat. The level of threat imminence, potency, and particular statistics that include novelty predictability, and volatility, will instigate the synthesis of sensory information, and cultural, internal, and subjective states. Together, these variables construct a defensive state that propagates a specialized defensive response or incites the organism to use knowledge of previous encounters with similar threats to elicit a defensive action. G-strategies will draw on cognitive systems involved in psychological space, memory, and learning. Further, this will cause the organism to explore internal and external environments for information which is driven by states of anxiety. Further, model-based RL will be imposed. The construction of the S-strategies will exploit information, promote intense emotions such as fear and panic, have low cognitive demand, and utilize model-free RL. Both individually, and the synthesis of, G-strategies and S-strategies can determine the defensive action. The defensive action results in a response to the threat which will update further defensive actions. Crucially, one hypothesis of the model is that each part of the DSC synergistically influences each other. They are not mutually exclusive. For instance, while representations combine to help construct the defensive state, S-strategies can scaffold new cognitive strategies or heuristics, and G-strategies can be fine-tuned to create specialized defensive systems. The arrow from action+control to state construction signifies how previous actions can influence how state construction is formed; Once the defensive strategy is successful, a safety signal will be broadcasted throughout the system and stored in memory; (B) the sub-processes linked to S-strategies and include lower cognition, model-free computations, exploitative behaviors and strategies and the emotion of fear and panic (e.g., protean escape, flight, freezing); (C) G-strategies are supported by higher levels of information processing and cognition, model-based computations, prospective emotions such as anxiety that promote exploration (D) S-strategies and G-strategies are a dynamic process across time scales including in the model decisions making and longer-term learning (e.g., transfer learning).

a subjective and conscious representation of the threat. This results in meta-affective awareness or the ability to consciously reflect on one's emotional state. This allows for the conscious control and evaluation of one's emotional state. These emotional states will represent the individual's introspective state of anxiety, horror, and panic. Some researchers have proposed these subjective feelings of 'fear' states emanate from autonomic states (Damasio, 1999; Craig, 2002). However, others, such as LeDoux and Hoffman (2018), see them more as a modulator than the cause. In treatments of interoceptive inference, conscious states are often regarded as high-level hypotheses about 'the emotional state I am in' that best explain both the current interoceptive and exteroceptive sensations (Seth and Friston, 2016; Smith et al., 2019). The conscious state, however, may rely on explicit memory and act as the mnemonic system that draws connections between experiences, allowing individuals to relate and compare similar events (Lau et al., 2022). Conscious prospection would also rely on these memory systems, allowing the organism to foresee danger and create a playground for scenarios to avoid it (Mobbs et al., 2015). Others have suggested that consciousness is advantageous to survival because it allows us to correct behavior (Solms, 2021). Indeed, conscious awareness of one's external environment and internal state provides an advantage. It allows for the distillation of complex internal and external data, guiding behavior for flexible and creative strategizing – a process important to combating threats from other humans. Importantly, meta-affective awareness plays a continuous role from the initial state to the final action constructions.

#### 4.1.4. Culture and social norms

“Basic survival-related behaviors might contribute to, but are not isomorphic with the complex emotional categories experienced by humans” (Lindquist et al., (2022).

Ashley Montague (1956) proposed there are two types of nature - primary nature, which is inborn instincts, and secondary nature which is the cultural environment. Indeed, it is well-accepted that cultural adaptations, which include codes of conduct and social norms, are critical to human survival (Plotkin, 2011). These codes and norms play a role in the construction of defensive states because cultural norms shape how humans perceive and interact with the environment. In all societies, cowardice is shunned, while virtues like courage and self-sacrifice are lauded. This is captured in the maritime codes such as "women and children first" or "the captain goes down with the ship" and is often the storyline behind many of the greatest pieces of fiction (FeldmanHall et al., 2016). This reverence for courage, especially in the form of dignified self-sacrifice, is also evident in wartime battles. In Japanese culture, the samurai ends his life to preserve honor or what is called seppuku, while self-immolation (altruistic suicide), a form of martyrdom, is observed across many cultures and religions. These examples suggest that human survival responses have shifted beyond predation to survival in, and of, social groups (Wilson and Wilson, 2007).

While it is tempting to view cultural adaptations as being independent of biology, Lindquist et al. (2022) proposed the Dual Inheritance Model to bridge this gap, explaining cultural variations, propagations of emotions, and biology. For example, the Dual Inheritance Model proposes that emotions are embedded in neural, autonomic, and motor-action systems, yet the categories of each emotion, such as fear, anxiety, and anger are defined by culture propagated through social transmission. This sits with the Social Heuristics Hypothesis, where cultural norms become internalized, automatic (Rand et al., 2014), and arguably specialized. According to this hypothesis, when decisions must be made quickly, prosocial actions, which are fundamentally adaptive for group living, occur automatically whereby subjects cooperate more (Rand et al., 2016). When deciding whether to behave in prosocial or selfish ways, one must consider not only the distress of the other but the

risk to oneself. This appraisal of danger to self is likely an underlying mechanism for selfish behaviors under a longer decision timeframe (Vieira and Olsson, 2022). Although our discussion here is brief, it is evident that social norms are likely to sway defensive responses and should be integrated into any human model of survival.

#### 4.1.5. State constructions

State constructions form through the perpetual and dynamic interactions between internal and external information states. Cognitive systems that support information states, representations, and S-strategies and G-strategies include working memory, attention, and episodic memory systems. LeDoux and Brown's notion of general networks of cognition and how lower-order 'nonconscious' units fuse with cortical circuits to produce the conscious or subjective experiences of emotion (LeDoux and Brown, 2017). Building on the details laid out in Section 2, Feldman-Barrett, (2006) posits these states are organized as concepts and construct cognition, emotion, and behavior. Each state will contextualize other states. Feldman-Barrett, also suggests that the construction of emotion involves interoception, representations, and culture. From a Jamesian perspective, the autonomic reaction will influence the subjective or conscious response to danger. However, concepts in the form of cultural learning or context will alter the perception of danger (e.g., "I'm in a battle" or "I'm protecting my family"). Like a recipe, the ingredients can be combined in various ways to create a different dish or in our case, a survival strategy. Therefore, in a dynamic system, information is constructed based on the concepts that they represent and form a meta-representation. We call this state construction because it concerns the dynamic construction of informational states that correspond between internal and external milieu. In turn, the meta-representation of the information states dynamically interacts with and triggers S-strategies and G-strategies systems to construct the action plan. These two stages of construction are separated because at the moment strategic planning is based on incoming information states and inference, while Action Construction (as detailed in Section 5.1.12) reflects the formation of S-strategies and G-strategies.

Active inference formulates planning as inference under a generative or world model (Attias, 2003, Botvinick and Toussaint, 2012). This leads to the notion that any agent must infer states of affairs in the world—and then infer the most likely course of action to which it should commit (e.g., state and action constructions). In this view, choices and action selection supervene inference and learning about the sensed world. As Feldman-Barrett (2017) states, internal models are predictive, and not reactive. Feldman-Barrett (2017) suggests that incoming sensory input is processed through Bayesian filters and these in turn construct the perception and drive the actions. This suggests that forward models are at play, where we process the world through top-down mechanisms. Feldman-Barrett further hypothesizes that we use our experiences, and our memories, to simulate and compare similar patterns that act as a template from which to guide actions. Importantly, if the organism is faced with unpredictable information, this will elicit bottom-up processes that produce error signals or prediction errors. In this setting, prediction errors report the mismatch between the outcomes of a behavior and that predicted under the generative model, in this formulation, planning, and action selection rest upon minimizing the free energy expected following a course of action (Friston, 2010).

#### 4.1.6. Meta-Representation, and Dimensionality Reduction

A critique of LeDoux and Pine's and other dual-system models is that they separate survival circuits from the conscious experiences of fear (LeDoux and Pine, 2016). Yet models such as Fanselow and Pennington (2018) Central Fear Generator model, propose that various effectors, whether they be physiological, cognitive, or behavioral responses, converge to produce fear as an integrated response. Likewise, Herry and Jercog (2022) stated that meta-representations are constructed by combining various elements: the representations of the threat, sensory cues, the context, and the behavioral repertoire of the organism. For

S-strategies in our model, it's crucial that meta-representations quickly reduce the high dimensionality of the internal and external milieu. This can occur through many processes, all of which must provide a value signal for each variable. Similar, but slower processes occur for G-strategies. For example, encountering an unfamiliar threat prompts the organism to search its internal and external states and generalize from previous experiences (Fig. 4).

The precise mechanism of how dimensionality reduction forms a meta-representation is unclear. Yet what is clear is that we experience the world through Gestalt lenses. Statistically, dimensionality reduction is performed via several methods including General Linear Models, Boltzmann machines, and Principal Component Analysis. However, how dimensionality reduction is performed biologically is unclear. From a neuroscientific perspective, many adhere to the idea that the brain uses population codes to compute internal and external representations. Here populations of neurons, not a single neuron, are the fundamental level of computation (See Saxena and Cunningham, 2019; Headley et al., 2019; Mobbs et al., 2020; Ebitz and Hayden, 2021). Populations, within populations of neurons each represent some feature, yet for a unified, meta-representation, of the world, these neurons must undergo dimensionality reduction.

In sum, meta-representations can range from simple to complex. In the simple case, meta-representations are the accumulation of sensory, somatic, abstract, and predictive states that create a unified picture of the world (Herry and Jercog, 2022). In other species that do not have conscious experiences, this is their unified perception of the inner and external world. In humans, the conscious experience we have of the world highlights certain states (e.g., my heart is beating fast) and appraisal and reappraisal that can modify the meta-representation through reconstrual and repurposing the goal set (Uusberg et al., 2023). Therefore, by this definition, we can differentiate the meta-representation from the simple accumulation of information to the affective awareness and conscious appraisal.

#### 4.2. Triggers

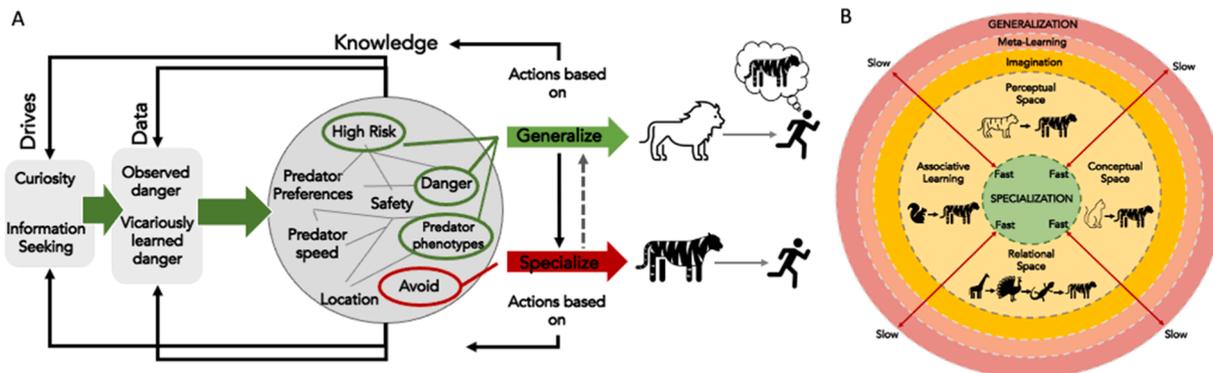
It is the resulting output of the state construction and meta-representation that provides the organism with a unified picture of danger and safety. If a threat threshold is breached, it will trigger the Action Construction to form G-strategies and S-strategies (Section 4.5.2). How are triggers formed? We propose that there are several ways – (i) Sensory triggers or stimuli characteristics or physics that instigate a

defensive response. Robert Hinde (1966) refers to these as ‘key stimuli’ or more recently what LeDoux calls “survival trigger stimuli”. These sit close to Konrad Lorenz’s idea of “innate releasing mechanisms” where the central nervous system either inhibits or activates fixed and instinctive action patterns based on environmental conditions.

In the case of innate programming, several examples of stimuli elicit a defensive response. While the types of trigger stimuli may be context and species-specific, some stimuli seem to produce similar behavioral responses across species. For example, visual systems have evolved to detect motion and seem particularly tuned to looming stimuli (Gibson, 1979; Heinemans and Moita, 2022). Further, increasing sound intensity, a marker for looming stimuli, results in increased arousal, alertness, and elevated activity in both the rodent and human amygdala (Bach et al., 2008). Another example of innate triggers is open spaces, where rodents exhibit innate responses like thigmotaxis (Simon et al., 1994) or open-arm avoidance in the elevated plus maze (Carobrez and Bertoglio, 2005). These are just a few examples of how sensory information can trigger innate defensive responses. This is captured by LeDoux (2012) who states:

“Survival circuits detect key trigger stimuli based on innate programming or experience” p655

Another type of trigger is what we call (ii) Prospective triggers. These are triggers that are instigated by future threats. Defensive responses according to the threat imminence continuum, during pre-encounter danger, where there is an increased likelihood of encountering a threat but not a direct threat now, the organism must be able to use internal models to trigger a defensive response. Indeed, past experiences are critical triggers. This is demonstrated by the Pavlovian condition where a stimulus (e.g., a beep sound) that is paired with a shock can result in a defensive response. In humans, PTSD symptoms can be triggered by stimuli that recapitulate the traumatic experience (Dalgleish, 2004). Importantly, it seems that internal states alone can instigate a trigger via memory, imagination, and prospection without changes to the external world. For example, we may have a threshold concerning the statistical estimation of danger. One example is knowing that there is a 5 in 10 chance of being bitten by a snake if you pick it up. If you pick it up 4 times, the sensory information stays the same, but your estimation of being bitten increases. These triggers, where thought alone can place the organism into a state of anxiety. In the context of defensive space (see Mobbs et al., 2020), a threat can be in the distant future, but elicit a threat response (e.g., given a big public talk). This seems particularly common among humans.



**Fig. 4.** Examples of G-strategies and S-strategies and their convergence. (A) The relationship between learning and drives, knowledge of threats (also see Fig. 3), and how the information is used when facing recurrent threats—where S-strategies are used—and novel threats where generalization is needed. G-strategies will draw on selected knowledge (e.g., see connections in green) to produce a defensive response to a novel encounter with danger. (B) We theorize that there are five determinants of G-strategies and S-strategies. These include perceptual space or similarity (e.g., a tiger looks like a puma), conceptual space (e.g., visually different objects can be of the same category), relational space (e.g., A giraffes, might predict a peacock, which predicts a lizard that predicts a tiger) and associative learning (e.g., a squirrel predicts the presence of a tiger). These associations can be creatively molded by imagination for things that have never been experienced. Transfer learning involves the process by which learning in one domain, will facilitate or teach other learning systems. This can occur independently and through social interaction. G-strategies typically employ cortical systems involved in slow information processes, while recurrent threats will engage S-strategies that are quickly associated and responded to with fast actions. Arrows signify defensive space (i.e., time to respond).

Finally, (iii) autonomic triggers provide an internal alarm signal. This is exemplified by several theories and empirical studies in neuroscientific and psychological studies showing that stimulation of autonomic systems can trigger defensive responses (Signoret-Genest et al., 2023). The idea that autonomic signals play a role in emotion goes back to William James (1884) who stated the idea bodily arousal is first experienced and the emotion follows, for example:

“We feel sorry because we cry, angry because we strike, afraid because we tremble” (James 1884: 189–190).

The role of bodily arousal in the emotional state was elaborated on by Schachter and Singer, who proposed the two-factor theory of emotion. Here the emotion one feels is determined by the cognitive interpretation of the bodily state. They hypothesized that when the subjects feel a bodily state of arousal (e.g., increased heart rate due to injection of epinephrine) for which they have no explanation, they will look for explanations for the arousal and thereby label the emotion with the best explanation. As Schachter and Singer state:

“... Subjects who had no explanation for the bodily state thus produced, gave behavioral and self-report indications that they had been readily manipulable into the disparate feeling states of euphoria and anger” p396.

Later research on misattribution of arousal (e.g., Dutton and Aron, 1974), and contextual effects on the perception of facial emotion (Mobbs et al., 2007) to support the two-factor theory. And lay the foundation for modern theories of emotional construction (Feldman-Barrett, 2017).

In the context of the DSC framework, there are several important things to consider (i) the meta-representation should integrate these triggers with context, safety levels, and history with the threat. Therefore, a threat may not trigger a response if the organism is in a state of safety (e.g., seeing a bear through a car window). (ii) Like James Gibson’s notion of affordances (Gibson, 1966), the DSC proposes that there is a complementary relationship between the organism and the environment. Affordances are what the ecology offers the organism in terms of protection or danger or what possibilities the environment offers or affords. Stimuli in the environment are perceived by their survival value. Finally, (iii) construction of the defensive action state depends on the context, affordances in the world, and autonomic and psychological state. If the internal states result in threat signals, an action will be constructed based on the dimensionality of experience resulting in G-strategies and S-strategies.

#### 4.3. Action Construction: G-strategies and S-strategies and Related Sub-States

The question of how psychological states, including emotion and defensive states, are constructed is a topic of intense debate within psychology and neuroscience (see Mobbs et al., 2019). We hypothesize that the state constructions and meta-representation dynamically interact to construct the survival actions and their related sub-states associated with G-strategies and S-strategies. As discussed in Section 3.1, the most notable, Feldman-Barrett’s theory of constructed emotion has proposed that defensive states are constructed dynamically at the moment, which in turn directs motor and visceral systems (Barrett, 2006; Barrett et al., 2016; Barrett and Simmons, 2015). However, we suggest that this construction depends on several other factors most notably the time pressure (or defensive space Section 5.2.1) and uncertainty. Here, one would expect model-free reactive systems that produce fast, often innate, defensive responses as implemented in S-strategies. Most evident during the interaction with novel or uncertain threats is the construction of G-strategies, which would search for a larger psychological space (Section 5.2) to create a defensive plan (Fig. 4).

## 5. Levels of ecological dimensionality: the roles of G-strategies and S-strategies

In the case of multidimensional ecologies, organisms will evolve various strategies, including changes in their physical phenotypes, underscoring the idea that ‘form follows function’. Consider the difference in S-strategies vs. G-strategies camouflage strategies: the consistent stripe pattern of tigers versus the adaptable camouflages of chameleons or octopuses. Presumably, the latter is only useful when they need to disguise themselves in different environments, and it comes with an energy cost. Therefore, natural selection will in tandem evolve physical and behavioral phenotypes that are defined under the rubric of G-strategies and S-strategies.

Arguably, human evolution has focused on G-strategies. As highlighted by theories like the Cognitive trade-off hypothesis (Matsuzawa, 2009), humans have enhanced the adaptability of their brains to process information, make decisions, and generalize both across and within environments. Indeed, Cantlon and Piantodosi (2024) have recently proposed that the unique intelligence that is observed in humans comes from the expansion in information capacity.

Even though the human brain boasts remarkable flexibility in prospecting and adapting to complex environments, evidence suggests that defensive optimization occurs more gradually. For instance, the optimization of defensive weapons has been a steady process, evolving over generations in response to salient features of the environment rather than undergoing abrupt change (Derex et al., 2019). Having a brain attuned to ecological complexity and frequent change provides a major survival advantage. Such adaptation can be accomplished by internally replaying previous experiences, using them to construct relational structures, and predicting potential threats in unfamiliar environments (Linson et al., 2020). Among other reasons, this cognitive prowess would have been a major advantage for nomadic hunter-gatherers traversing varied landscapes with new dangers.

#### 5.1. G-strategies and S-strategies result in exploration, curiosity, and exploitation of the environment

What are the behavioral manifestations of G-strategies and S-strategies? Futuyama and Moreno, (1988) have suggested that:

“The probability of specialization increases with the decrease in the cost of search” (p211).

Under safe conditions, G-strategies will sometimes result in information-seeking behaviors. This is evident in rodents that, when placed in a location (e.g., a new cage), exhibit exploratory movements. Such exploration in novel environments provides vital information about potential threats, both immediate and prospective. From a motivational perspective, G-strategies essentially act as a long-term strategy that promotes curiosity. According to novelty-based theories, curiosity promotes information-seeking and is therefore intrinsically rewarding (Kang et al., 2009; Barto et al., 2013; Berlyne, 1950; Friston et al., 2017; Oudeyer and Kaplan, 2007; Schmidhuber, 2006, 2010; Schwartenbeck et al., 2019; Still and Precup, 2012; Vigorito and Barto, 2010). Research in humans and other animals has supported the role of the dopaminergic system in underpinning curiosity and information-seeking behaviors (Daw et al., 2006; Schwartenbeck et al., 2013; Schwartenbeck et al., 2015a; Schwartenbeck et al., 2015b).

Explorative behaviors should peak when information is needed for current or future survival (Bromberg-Martin and Hikosaka, 2009). Notably, uncertainty has been linked to anxiety, suggesting that anxious individuals may exhibit higher explorative information-seeking behaviors. Past studies in rodents suggest that both high and low fear states correlate with reduced exploration rates. Conversely, an intermediate level of threat increases exploration, suggesting a U-shaped behavioral relationship between fear and exploration (Lester, 1968). In humans, information seeking is increased when avoidance or escape routes are not available (Restrepo-Castro et al., 2023). Further, the ability to

adaptively seek out information may depend on contextual factors that can increase or decrease survival in individuals with anxiety. Charpentier and colleagues (2022) have shown that information-seeking is valence-dependent, with information associated with reward fostering a heightened information-seeking behavior than information associated with potential loss. Therefore, information-seeking would benefit anxious individuals by increasing knowledge and reducing uncertainty. Further, anxiety reduces model-based control in contexts that involve avoiding negative outcomes or seeking out protection to mitigate harm (Tashjian et al., 2022).

Conversely, S-strategies would rely on the exploitation of knowledge (Fig. 3A), the use of habitual behavior, and model-free computations. Rooted in S-strategies are habitual behaviors that result in energy conservation, rapid and accurate defensive responses, and consistent outcomes. Intriguingly, the automaticity of information-seeking in the face of uncertainty can be viewed as a fairly instinctive and hard-wired response. Yet, it serves a pivotal role within the broader framework of flexible, G-strategies. This underscores the idea that S-strategies and G-strategies are not distinct constructs but rather form a fluid continuum, each dynamically complementing the other.

It is important to state that anxiety is a visceral and cognitive state that is supported by generalization systems including prospection. Speculating, exploration can be behavioral (e.g., information seeking and curiosity), and internal (e.g., searching for connections). Therefore, the exploratory, information-seeking state is part of the G-strategies but is different from how the brain organizes information (e.g., abstract structures).

## 6. Components of G-strategies and S-strategies

We propose that the activation of S-strategies and G-strategies hinges on several key processes, which we categorize into distinct domains (see Fig. 2C). These processes are optimized through their interactions. Further, uncertainty within any of these spaces will drive the system to allocate resources towards cognitive processes (e.g., vigilance, memory search) or inefficient avoidance, where the organism flees upon recognizing a potential threat. These processes increase efficiency by applying previous knowledge to new or recurring threats (Fig. 4).

## 7. Defensive Space

Defensive space is a determinant of S-strategies and G-strategies. Defensive space reflects the amount of time or space available to an organism to determine the best defensive action. The longer the time the organism is given to decide, the larger the defensive space (or knowledge. Fig. 4B) that can be searched (Mobbs et al., 2020; Qi et al., 2018). This can be advantageous because a larger amount of time allows for a more detailed search through memory and imagination to strategize and plan a course of action. With increasing time, the organism can draw on experiences, simulate potential outcomes, and choose the best option. Likewise, defensive space can be reduced when other tasks impinge on the time available to plan defensive actions, a notion supported by the finding that subjects show less reliance on model-based planning when performing a second task simultaneously (Otto et al., 2013). One downside of a larger defensive space is what we call the “time-uncertainty principle”: as time increases, so does the defensive space and the variance in, for example, the predator’s attack possibilities and the prey’s number of behavioral options. If there is no clear solution to the decision problem at hand, this causes conflict and uncertainty in the defensive circuits (Mobbs et al., 2020). Others have shown that rumination can impair faster reinforcement learning (Hitchcock et al., 2022), suggesting that such cognitively demanding processes may impair S-strategies (e.g., reactive escape; Qi et al., 2018).

To combat fast-attacking threats that allow little time for a deliberative response, an organism needs a specialized, reactive system that can coordinate a quick response (Mobbs et al., 2020). Under such conditions,

deliberation can prove fatal. S-strategies therefore provide an advantage when rapid responses are essential – for instance, during unexpected or imminent threats. This will result in the elicitation of learned action plans or fast innate defensive responses like freezing or flight. These reflexive responses are observed across the animal world and by all accounts are successful. Yet, predators will ultimately learn to overcome these defensive behaviors. In addition, a mismatch between the predator’s behavior and the defensive strategy devised for it can lead to the failure of the reflexive response (and prediction error). This underscores the utility of S-strategies, evolving from a G-strategies input, to devise novel and unpredictable defensive strategies against predators. Therefore, we speculate that the dynamic interaction between S-strategies, G-strategies, and other systems in the DSC encapsulates the adoption of adaptive survival behaviors.

### 7.1. Defensive emotions

The genesis of emotions such as anxiety and fear are formed through state constructions, yet triggered through action construction. The resultant emotion depends upon several variables including the context, moods, and the autonomic nervous system. For example, increased heart rate and autonomic reactivity can be elicited by the sensory states in the detection of a threat in the environment. Further, they can be modulated by meta-affective awareness where one can consciously think about past, present, and future threats. This information is accumulated to form a meta-representation, which if a threat threshold is reached, will trigger action construction. According to the DSC framework, emotions like fear and anxiety are actions accumulated based on state representations. The resultant emotion is therefore a key part of the S-strategies and G-strategies. Anxiety drives the organism to search internal environments to estimate the best action and explore the external world to increase knowledge (e.g., thigmotaxis). S-strategies will elicit reactive emotions such as fear and panic, which behavior manifests as rapid protean escape and uncoordinated movement. Importantly, the DSC framework sees emotion as a critical part of the survival strategy, invigorating the organism to quickly react, increase vigilance, or explore internal memory systems.

### 7.2. Psychological space: perceptual and cognitive maps

Shepard’s ‘law of generalizability’ proposes the idea of a ‘psychological space’. According to this view, every stimulus creates a metric space, and the distance between one stimulus and related stimuli determines the probability of generalization (Shepard, 1987). That is, the probability of a learned response generalizing from one stimulus to another depends on the perceptual and cognitive similarity between them. Shepard states:

“We generalize from one situation to another not because we cannot tell the difference between the two situations but because we judge that they are likely to belong to a set of situations having the same consequence”. pp237 Roger Shepard, (1987).

Shepard points out that the probability of generalization decays with distance, but what determines these distances? Examples of determining psychological distance include Euclidean, Minkowski metrics, and the Markov tree, although their application to abstract psychological concepts is challenging. As we discuss below, we use the term psychological space to capture G-strategies across perceptual, conceptual, and relational spaces (Fig. 3B).

### 7.3. Perceptual space

In nature, the roar of a lion is similar to the roar of a tiger, and the sight of a rattlesnake is similar to a copperhead. Such similarities in sensory properties allow for generalizations. For instance, a rabbit might identify a fox as a threat, but it would likely recognize other similar-

sized four-legged creatures, such as a coyote or bobcat, as potential dangers too. Perceptual space can be visualized geometrically, with attributes like size and color as dimensions. Thus, a fox and a coyote have a closer perceptual space than a fox and a sparrow. In this context, S-strategies would result from having a specific behavioral response for each type of predator, while G-strategies would use a similar defensive response to the fox and the coyote. It is worth noting that the ecological evidence for perceptual space comes from Batesian mimicry and Pavlovian generalization. Batesian mimicry is where a non-harmful species mimics a harmful species, as seen with the Californian Mountain King Snake adopting visual patterns similar to the poisonous Eastern Coral Snakes.

Meanwhile, Pavlovian generalization allows for simpler generalization based on similar stimuli, aligning it with perceptual space (Lashley and Wade, 1946). Another example is one study in Fathead Minnows showed that when they are trained to recognize the odor of a natural predator – a lake trout – they can generalize their defensive response to phylogenetically closer species of trout. For example, the Brook Trout, which is phylogenetically closer to the lake trout resulted in higher intensity of antipredator responses as compared to a Rainbow Trout. However, more distant trout in the family tree did not elicit such defensive responses (Ferrari et al., 2007). This study supports the idea that G-strategies extend across the senses and are species-specific.

#### 7.4. Conceptual space

If perceptual space is guided by sensory similarity, then one must also consider how stimuli are associated by their categorical inclusion. For example, both Chihuahuas and English Setters fall under the “dogs” category, which further belongs to broader classifications like sporting dogs, herding dogs, or the genus of *Canis*. Object categories, as defined by both supervised neural networks trained to discriminate between categories and neural networks trained on linguistic data, are represented in the inferior temporal cortex (Khaligh-Razavi and Kriegeskorte, 2014). This is in line with theoretical work suggesting that this region functions as a modality-independent hub for semantic knowledge (Carlson et al., 2014), indicating a dedicated system for conceptual space representation that facilitates category-based judgments. It’s also evident that conceptual and perceptual spaces sometimes blur. Research shows that the brain represents abstract, semantic object properties grounded in perceptual information, such as the shape and size of a bird, in a grid-like neural code, using them for G-strategies similar to spatial information (Lambon-Ralph et al., 2017; Constantinescu, et al., 2016).

While it is true that perceptual space captures how physically similar stimuli are clustered, humans can conceptually categorize stimuli through processes such as induction (Dunsmoor and Murphy, 2014). Conceptual space, as defined here, concerns how stimuli are conceptually categorized. For example, how a leaf is linked to a tree, or similar to our example, how a house cat is in the same category as a lion. One study by Dunsmoor and colleagues conducted a simple memory experiment, where they showed that conditioning on a particular stimulus enhances retroactive memories for related stimuli (Dunsmoor et al., 2015). This shows a link between associative learning where inconsequential information is reclassified as relevant. This interplay between associative learning and conceptual categorization suggests that high-order processes, such as concepts and inductive reasoning, underpin even what are considered low-level processes. This extends our understanding beyond mere perceptual similarities, shedding light on the complex mechanisms humans use to generalize (Dunsmoor and Murphy, 2014).

#### 7.5. Relational space

As discussed in Section 2.1, the abstract relational structures (e.g., Cognitive maps), or for consistency in our terminology, “relational space”, depends upon relations between stimuli. These structures potentially utilize the same neural systems that facilitate flexible

avoidance behavior using abstract cognitive maps of the organism’s environment. A prime example is the hippocampus-entorhinal cortex. While it’s pivotal in representing abstract relational structure, it’s also involved in perceptually-guided fear generalization (Webler et al., 2021). Grid-like neural representations of relational structure are also implicated in the representation of abstract concepts such as the shape of birds (Constantinescu et al., 2016) - concepts that seem analogous to those guiding avoidance based upon perceptual distance. However, when we speak of “psychological space”, we primarily refer to concepts whose similarity can be automatically and immediately judged without abstraction, typically focused on perceptual attributes such as shape, size, or color, rather than more complex abstract relations.

#### 7.6. Imagination and prospection

David Hume stated, “Nothing is more free than the imagination of man” (p47). Indeed, in the context of defensive space, imagination allows one to volitionally and creatively draw on previous experiences (i. e., episodic memory) and prospectively apply them to new or future contexts (Hassabis, et al., 2007; Suddendorf and Corballis, 2007; Mobbs et al., 2015). Rather than a mere recapitulation of past events, imagination facilitates a malleable and inventive reshaping of these memories. This offers the capacity to forecast diverse scenarios, even those never encountered before. This is a critical aspect of G-strategies and survival in novel and high-dimensional contexts but flexibly combines associative learning, perceptual space, conceptual space, and relational space. Doing so, allows one to either preemptively alter behaviors in anticipation of danger (e.g., avoidance) or, upon facing a danger, prepare a planned defensive action. Moreover, imagination and prospection may also use episodic RL, where memories are used to estimate the value of actions (Botvinick et al., 2019).

#### 7.7. Underlying learning and transfer learning

When encountering threats, rapid adjustments are imperative, calling for plasticity in defensive decision-making. Some of these adjustments in S-strategies can become habitual (Model-Free RL), while others are more contemplative (Model-Based RL). While RL provides the building blocks for most forms of generalization, it does not capture every facet, especially inferential associations. Pavlovian fear conditioning is extremely powerful as it allows the organism to learn positive or negative associations about completely separate stimuli. While complex associations can be made through associative learning (e.g., Occasion setting; Zbozinek et al., 2022) in its basic form, associative learning does not provide a good explanation for associations that occur in humans, including prospection and imagination. However, recent work has shown how learning and extinction can occur via imagination (Reddan, Wager, and Schiller, 2018) and the potential for vicarious learning and advanced cognition to form associations between underlying categories (Mobbs et al., 2015). From the perspective of active inference and learning, one important kind of generalization rests upon structural learning (Gershman and Niv, 2010; Ghahramani and Jordan, 1997; Smith et al., 2020; Tenenbaum et al., 2011); namely, getting the right structure in generative or internal models that ‘carve nature at its joints’. This is an important aspect of transfer learning and G-strategies, which can be framed in terms of Bayesian model selection during introspection, sleep, neurodevelopment, and, indeed, natural selection (Campbell, 2016; Frank, 2012; Friston et al., 2017; Vanchurin et al., 2022).

## 8. Conclusions

In this article, we have stated that predictable and recurrent ecological threat fosters the evolution of narrow, S-strategies. In contrast, complex, multidimensional environments will drive the development of G-strategies that not only generalize but also simulate,

construct, and predict danger. Together, this spectrum reflects defensive responses that can combat a large variety of threats across a multitude of environments. Further, the DSC model provides a possible solution to dichotomous theories of emotion by synthesizing hard-wired defensive models (e.g., Panksepp's FEAR model and LeDoux's survival circuits) with higher-order and construction theories (e.g., Barrett, 2006; LeDoux and Pine, 2016). The DSC proposes several testable hypotheses in humans and other animals:

1. G-strategies promote exploratory behaviors driven by curiosity or information-seeking and can reduce entropy and uncertainty. S-strategies will utilize exploitative habitual behaviors. This should be tested using novel methods where movement can be measured (e.g., computational ethology; Mobbs et al., 2021).
2. Defensive space, or the amount of time the organism has to make value and action decisions, will determine the strategy. For example, if the organism has little time, the S-strategies will take control and determine the action to be selected. S-strategies elicit urges and reactive and habitual model-free behaviors such as freezing or flight. Importantly, given sufficient time, reactive urges, or the desire to avoid the situation, are [pre]consciously perceived and controlled (e.g., inhibition of flight). This process may take the form of urgency-gating (Cisek, Puskas, and El-Murr, 2009). Therefore, experiments should manipulate the time that subjects can deliberate/avoid or escape from threat (e.g. Qi et al., 2018; Qi et al., 2024).
3. At the extreme, G-strategies cast a wide net to draw on high-dimensional representations or merge several representations to produce what it believes to be the best defensive response. This process includes the construction of high-dimensional cognitive, affective, and behavioral states that, upon repetition, become low-dimensional and specialized. Thus, researchers should aim to alter the complexity of the threats to examine the switch from hippocampal-prefrontal systems to subcortical and motor systems.
4. Humans face complex social dangers that are based on abstract information such as culture, social, and language embedded in cognition systems (See Ortony, Clore, and Collins, 1988). Based on the state construction, G-strategies would implement the appropriate defensive response that could include verbal strategies to calm the agent (e.g., reasoning) or shouting for help.
5. Transfer learning supports the claim of the efficient 'reuse' and the economical cross-fertilization of abstract, relational, and perceptual space when generalizing across threats. Creating experiments that examine similar attack strategies should be about to see how to learning transfers in the brain. Here one could use modern statistical approaches such as representational similarity analyses (Kriegeskorte et al, 2008).

Although briefly discussed, one question not extensively addressed here is how the DSC is instantiated in the central nervous system (Price, 2005; Price and Drevets, 2010). Given the multidimensional nature of the real world, complex representations may be created and maintained in populations of neurons (Headley et al., 2019; Mobbs et al., 2019; Mobbs et al., 2020; Ebitz and Hayden, 2021). For example, populations of neurons, store similar representations and compare and combine them across modalities (e.g., vision and smell). Further, abstract representations of stimuli can be creatively manipulated to form, for example, dynamic and novel representations. Abstract forms of safety (e.g., my confidence in my abilities to escape a predator) may also be represented in these codes (Tashjian et al., 2021). The conjunction of these representations along with repeated exposure to similar experiences will result in S-strategies. One goal of neuroscience should be to understand how neural populations shift from high to low dimensions to elicit the most adaptive defensive response.

## Acknowledgments

DM is supported by National Institute of Mental Health grants R01MH133730-01, 2P50MH094258, and Templeton Foundation grant TWCF0366. KF is supported by funding for the Wellcome Centre for Human Neuroimaging (Ref: 205103/Z/16/Z), a Canada-UK Artificial Intelligence Initiative (Ref: ES/T01279X/1), and the European Union's Horizon 2020 Framework Program for Research and Innovation under the Specific Grant Agreement No. 945539 (Human Brain Project SGA3). TW is supported by a Wellcome Trust Career Development Award (225945/Z/22/Z). DBH is supported by the grant R01NS123396. We would like to thank Zachary Diamandis, Mac Thurston, and Ketika Garg. Correspondence concerning this article should be addressed to Dean Mobbs, Department of Humanities and Social Sciences California Institute of Technology, 1200 E California Blvd, HSS 228-77, Pasadena, California 91125, USA; dmobbs@caltech.edu.

## References

- Abraham, W., 2008. Metaplasticity: tuning synapses and networks for plasticity. *Nat. Rev. Neurosci.* 9, 387.
- Adams, C.D., 1982. Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Q. J. Exp. Psychol. Sect. B* 34, 77–98.
- Anderson, M.L., 2010. Neural reuse: A fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–313.
- Anderson, D.J., 2016. Circuit modules linking internal states and social behavior in flies and mice. *Nat. Rev. Neurosci.* 17 (11), 692–704.
- Attias, H., 2003. Planning by probabilistic inference. *Proc. 9th Int. Workshop Artif. Intell. Stat.*
- Bach, D.R., Dayan, P., 2017. Algorithms for survival: a comparative perspective on emotions. *Nat Rev Neurosci* 18 (5), 311–319. May.
- Barrett, L.F., 2006. Solving the emotion paradox: categorization and the experience of emotion. *Personal. Soc. Psychol. Rev.: Off. J. Soc. Personal. Soc. Psychol., Inc.* 10, 20–46.
- Barrett, L.F., Quigley, K.S., Hamilton, P., 2016. An active inference theory of allostasis and interoception in depression. *Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* 371, 20160011.
- Barrett, L.F., Simmons, W.K., 2015. Interoceptive predictions in the brain. *Nat. Rev. Neurosci.* 16, 419–429.
- 9 Barsbai, T., Lukas, D., Pondorfer, A. (2021) Local convergence of behavior across species. *Science*, Vol 371, Issue 6526, pp. 292-295.
- Barto, A., Mirolli, M., Baldassarre, G., 2013. Novelty or surprise? *Front Psychol.* 4, 907.
- Berlyne, D.E., 1950. Novelty and curiosity as determinants of explanatory behaviour. *British Journal of Psychology-General Section*, 41, 68–80.
- Blanchard, R.J., Blanchard, D.C., 1989. Attack and defense in rodents as ethoexperimental models for the study of emotion. *Prog Neuropsychopharmacol Biol Psychiatry* 13. Suppl:S3-14.
- Botvinick, M., Ritter, S., Wang, J.X., Kurth-Nelson, Z., Blundell, C., Hassabis, D., 2019. Reinforcement learning, fast and slow. *Trends Cogn. Sci.* 23 (5), 408–422.
- Botvinick, M., Toussaint, M., 2012. Planning as inference. *Trends Cogn. Sci.* 16, 485–488.
- Brechbühl, J., Moine, F., Klaey, M., Nenniger-Tosato, M., Hurni, N., Sporkert, F., Giroud, C., Broillet, M.C., 2013. Mouse alarm pheromone shares structural similarity with predator scents. *Proc. Natl. Acad. Sci. USA* 110 (12), 4762–4767.
- Bromberg-Martin, E.S., Hikosaka, O., 2009. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* 63 (1), 119–126.
- Burdakov, D., Peleg-Raibstein, D., 2020. The hypothalamus as a primary coordinator of memory updating. *Physiol. Behav.* 223, 112988.
- Burnett, C.J., Li, C., Webber, E., Tsaousidou, E., Xue, S.Y., Brüning, J.C., Krashes, M.J., 2016. Hunger-driven motivational state competition. *Neuron* 92 (1), 187–201.
- Campbell, J.O., 2016. Universal darwinism as a process of bayesian inference. *Front. Syst. Neurosci.* 10, 49.
- Cantlon, J.F., Piantadosi, S.T., 2024. Uniquely human intelligence arose from expanded information capacity. *Nat. Rev. Psychol.* 3, 275–293.
- Carlson, T.A., Simmons, R.A., Kriegeskorte, N., Slevc, L.R., 2014. The emergence of semantic meaning in the ventral temporal pathway. *J. Cogn. Neurosci.* 26 (1), 120–131.
- Carobrez, A.P., Bertoglio, L.J., 2005. Ethological and temporal analyses of anxiety-like behavior: the elevated plus-maze model 20 years on. *Neurosci. Biobehav Rev.* 29 (8), 1193–1205.
- Carthey, A.J.R., Bucknall, M.P., Wierucka, K., et al., 2017. Novel predators emit novel cues: a mechanism for prey naivety towards alien predators. *Sci. Rep.* 7, 16377.
- Charpentier, C.J., Cogliati Dezza, I., Vellani, V., et al., 2022. Anxiety increases information-seeking in response to large changes. *Sci. Rep.* 12, 7385. <https://doi.org/10.1038/s41598-022-10813-9>.
- Cisek, P., Puskas, G.A., El-Murr, S., 2009. Decisions in changing conditions: the urgency-gating model. *J. Neurosci.* 29 (37), 11560–11571.
- Constantinescu, A.O., O'Reilly, X., Behrens, T.E.J., 2016. Organizing conceptual knowledge in humans with a gridlike code. *Science* 352, 1464–1468.

- Craig, A.D., 2002. How do you feel? Interoception: the sense of the physiological condition of the body. *Nat. Rev. Neurosci.* 3, 655–666.
- Dalgleish, T., 2004. Cognitive approaches to posttraumatic stress disorder: the evolution of multirepresentational theorizing. *Psychol. Bull.* 130 (2), 228–260.
- Damasio, A.R., 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace, New York.
- Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: Murray.
- Daw, N.D., Niv, Y., Dayan, P., 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8 (12), 1704–1711.
- Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., Dolan, R.J., 2006. Cortical substrates for exploratory decisions in humans. *Nature* 441 (7095), 876–879.
- Derex, M., Bonnefon, J.F., Boyd, R., Mesoudi, A., 2019. Causal understanding is not necessary for the improvement of culturally evolving technology. *Nat. Hum. Behav.* 3, 446–452.
- Dias, B.G., Ressler, K.J., 2014. Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat. Neurosci.* 17 (1), 89–96.
- Drabeck, D.H., Dean, A.M., Jansa, S.A., 2015. Why the honey badger don't care: Convergent evolution of venom-targeted nicotinic acetylcholine receptors in mammals that survive venomous snake bites. *Toxicon* 99, 68–72.
- Dunsmoor, J.E., Murphy, G.L., 2014. Stimulus typicality determines how broadly fear is generalized. *Psychol. Sci.* 25 (9), 1816–1821.
- Dunsmoor, J.E., Murty, V.P., Davachi, L., Phelps, E.A., 2015. Emotional learning selectively and retroactively strengthens memories for related events. *Nature* 520 (7547), 345–348.
- Ebitz, R.B., Hayden, B.Y., 2021. The population doctrine in cognitive neuroscience. *Neuron* 109 (19), 3055–3068.
- Engelhard, B., Finkelstein, J., Cox, J., et al., 2019. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* 570, 509–513.
- Fadok, J., Krabbe, S., Markovic, M., et al., 2017. A competitive inhibitory circuit for selection of active and passive fear responses. *Nature* 542, 96–100.
- Fanselow, M.S., Hoffman, A.N., 2024. Fear, defense, and emotion: A neuroethological understanding of the negative valence research domain criteria. *Am Psychol* 79 (5), 725–734. Jul-Aug.
- Fanselow, M.S., Lester, L.S., 1988. A functional behavioristic approach to aversively motivated behavior: predatory imminence as a determinant of the topography of defensive behavior. In: Beecher, R.C.B.M.D., Hillsdale, N.J. (Eds.), *Evolution and Learning*. Erlbaum, pp. 185–211.
- Fanselow, M.S., Pennington, Z.T., 2018. A return to the psychiatric dark ages with a two-system framework for fear. *Behav. Res Ther.* 100, 24–29.
- Faul, L., Stjepanović, D., Stivers, J.M., Stewart, G.W., Graner, J.L., Morey, R.A., LaBar, K.S., 2020 Jul 14. Proximal threats promote enhanced acquisition and persistence of reactive fear-learning circuits. *Proc Natl Acad Sci U S A* 117 (28), 16678–16689.
- FeldmanHall, O., Dalgleish, T., Evans, D., Navrady, L., Tedeschi, E., Mobbs, D., 2016 Aug. Moral Chivalry: Gender and Harm Sensitivity Predict Costly Altruism. *Soc Psychol Personal Sci* 7 (6), 542–551.
- Ferrari, M.C.O., Gonzalo, A., Messier, F., Chivers, D.P., 2007. Generalization of learned predator recognition: an experimental test and framework for future studies. *Proceedings of the Royal Society B* 274, 1853–1859.
- Finlay, B.L., Darlington, R.B., Nicastro, N., 2001. Developmental structure in brain evolution. *Behav. Brain Sci.* 24 (2), 263–278.
- Frank, S.A., 2012. Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *J. Evol. Biol.* 25, 2377–2396.
- Friston, K.J., Lin, M., Frith, C.D., Pezzulo, G., Hobson, J.A., Ondobaka, S., 2017. Active Inference, Curiosity and Insight. *Neural Comput.* 29, 2633–2683.
- Futuyma, D.J., Moreno, G., 1988. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* 19, 207–233.
- Gershman, S.J., Niv, Y., 2010. Learning latent structure: carving nature at its joints. *Curr. Opin. Neurobiol.* 20, 251–256.
- Ghahramani, Z., Jordan, M.I., 1997. Factorial hidden Markov models. *Mach. Learn.* 29, 245–273.
- Gibson, J.J., 1966. *The Senses Considered as Perceptual Systems*. Allen and Unwin, London.
- Gibson, J.J., 1979. *The ecological approach to visual perception*. Houghton Mifflin, Boston, MA.
- Gray, J.A., 1987. *The psychology of fear and stress*, 2nd ed. Cambridge University Press, Cambridge.
- Gross, C.T., Canteras, N.S., 2012. The many paths to fear. *Nat. Rev. Neurosci.* 13 (9), 651–658.
- Gurevitch, J., Padilla, D.K., 2004. Are invasive species a major cause of extinctions? *Trends Ecol. Evol.* 19 (9), 470–474.
- Hamilton, W.D., 1971. Geometry for the selfish herd. *J. Theor. Biol.* 31 (2), 295–311. [https://doi.org/10.1016/0022-5193\(71\)90189-5](https://doi.org/10.1016/0022-5193(71)90189-5).
- Hassabis, D., Kumaran, D., Vann, S.D., Maguire, E.A., 2007. Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. USA* 104 (5), 1726–1731.
- Headley, D.B., Kanta, V., Kyriazi, P., Paré, D., 2019. Embracing Complexity in Defensive Networks. *Neuron* 103 (2), 189–201.
- Heinemans, M., Moita, M.A., 2022. (preprint) Looming stimuli reliably drive innate, but not learned, defensive responses in rats. [bioRxiv](https://arxiv.org/abs/2207.14943), 479432.
- Herry, C., Jercog, D., 2022. Decoding defensive systems. *Curr. Opin. Neurobiol.* 76, 102600.
- Hinde, R.A., 1966. *Animal behaviour: a synthesis of ethology and comparative psychology*. McGraw-Hill Book Company, New York, NY.
- Hitchcock, P., Forman, E., Rothstein, N., Zhang, F., Kounios, J., Niv, Y., Sims, C., 2022. Rumination details reinforcement learning with possible implications for ineffective behavior. *Clin. Psychol. Sci.* 10, 714–733.
- Hoffman, D.D., 2010. Sensory experiences as cryptic symbols of a multimodal user interface. *Act. Nerv. Super.* 52 (3), 95–104.
- Humphries, D.A., Driver, P.M., 1967. Erratic display as a device against predators. *Science* 156 (3783), 1767–1768.
- Humphries, D.A., Driver, P.M., 1970. Protean defence by prey animals. *Oecologia* 5 (4), 285–302.
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.* 22, 415–427.
- Ingram, T., Costa-Pereira, R., Araújo, M.S., 2018. The dimensionality of individual niche variation. *Ecology* 99 (3), 536–549.
- Janak, P.H., Tye, K.M., 2015. From circuits to behaviour in the amygdala. *Nature* 517 (7534), 284–292.
- Johnson, M.H., 2011. Interactive specialization: a domain-general framework for human functional brain development? *Dev Cogn Neurosci* 1 (1), 7–21. Jan.
- Kang, M.J., Hsu, M., Krajbich, I.M., Loewenstein, G., McClure, S.M., Wang, J.T., Camerer, C.F., 2009. The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychol. Sci.* 20 (8), 963–973.
- Kaplan, R., Friston, K.J., 2018. Planning and navigation as active inference. *Biol. Cybern.* 112, 323–343.
- Kaplan, R., King, J., Koster, R., Penny, W.D., Burgess, N., Friston, K.J., 2017. The neural representation of prospective choice during spatial planning and decisions. *PLoS Biol.* 15, e1002588.
- Khaligh-Razavi, S.-M., Kriegeskorte, N., 2014. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Comput. Biol.* 10 (11), e1003915.
- Kleckner, I.R., et al., 2017. Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nat. Hum. Behav.* 1, 0069.
- Kool, W., Cushman, F.A., Gershman, S.J., 2016. When Does Model-Based Control Pay Off. *PLoS Comput. Biol.* 12 (8), e1005090. <https://doi.org/10.1371/journal.pcbi.1005090>.
- Kramer, D.L., McLaughlin, R.L., 2001. The behavioral ecology of intermittent locomotion. *American Zoologist* 41, 137–153.
- Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst. Neurosci.* 2, 4.
- Lagos, P.A., Lagos, A.M., Ortiz, L., Tolhuysen, L.O., Castro, R.A., Bozinovic, F., Ebensperger, L.A., 2009. Flight initiation distance is differentially sensitive to the costs of staying and leaving food patches in a small-mammal prey. *Can. J. Zool.* 87 (11), 1016–1023.
- Lambon-Ralph, M., Jefferies, E., Patterson, K., et al., 2017. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18, 42–55.
- Lashley, K.S., Wade, M., 1946. The pavlovian theory of generalization. *Psychol. Rev.* 53 (2), 72–87.
- Lau, H., Michel, M., LeDoux, J.E., et al., 2022. The mnemonic basis of subjective experience. *Nat. Rev. Psychol.* 1, 479–488.
- LeDoux, J., 2012. Rethinking the emotional brain. *Neuron* 73 (4), 653–676.
- LeDoux, J.E., Brown, R., 2017. A higher-order theory of emotional consciousness. *Proc. Natl. Acad. Sci. USA* 114 (10), E2016–E2025.
- LeDoux, J.E., Hofmann, S.G., 2018. The subjective experience of emotion: A fearful view. *Curr. Opin. Behav. Sci.* 19, 67–72.
- LeDoux, J.E., Pine, D.S., 2016. Using neuroscience to help understand fear and anxiety: a two-system framework. *Am. J. Psychiatry* 173, 1083–1093.
- Lee, S.W., Shimojo, S., O'Doherty, J.P., 2014. Neural computations underlying arbitration between model-based and model-free learning. *Neuron* 81 (3), 687–699. Feb 5.
- Lester, D., 1968. Two tests of a fear-motivated theory of exploration. *Psychon. Sci.* 10, 385–386.
- Lindquist, K.A., Jackson, J.C., Leshin, J., et al., 2022. The cultural evolution of emotion. *Nat Rev Psychol* 1, 669–681.
- Linson, A., Parr, T., Friston, K.J., 2020. Active inference, stressors, and psychological trauma: A neuroethological model of (mal)adaptive explore-exploit dynamics in ecological context. *Behav. Brain Res.* 380, 112421.
- Louie, K., Wilson, M.A., 2001. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* 29, 145–156.
- Matsuzawa, T., 2009. Symbolic representation of number in chimpanzees. *Curr. Opin. Neurobiol.* 19 (1), 92–98.
- Mayr, E., 2001. *What Evolution Is*. Basic Books, New York.
- McEwen, B.S., 1998. Stress, adaptation, and disease. Allostasis Allostasis Load. *Ann. N. Y. Acad. Sci.* 840, 33–44.
- McNally, R.J., 1987. Preparedness and phobias: A review. *Psychol. Bull.* 101 (2), 283–303.
- Mobbs, D., 2018. The ethological deconstruction of fear(s). *Curr Opin. Behav. Sci.* 24, 32–37.
- Mobbs, D., \*Garg, K., and \*Tashjian, S. (In Press). *Defensive Decisions: From perception to action. Handbook of Affective Neuroscience (2nd edition)*; Cambridge University Press (Eds. Vuilleumier P. and Armony, J.).
- Mobbs, D., Adolphs, R., Feldman-Barrett, L., Fanselow, M., LeDoux, J., Ressler, K., Tye, K., 2019. Viewpoint: approaches to defining and investigating fear. *Nat. Neurosci.* 22, 1205–1216.
- Mobbs, D., Hagan, C.C., Dalgleish, T., Silston, B., Prevost, C., 2015. The ecology of human fear: survival optimization and the nervous system. *Front. Neurosci.* 9 (55), 1–22.

- Mobbs, D., Headley, D., Ding, W., Dayan, P., 2020. Space, Time, and Fear: Survival Computations Along Defensive Circuits. *Trends in Cogn. Sci.* 24 (3), 228–241.
- Mobbs, D., Marchant, J., Hassabis, D., Seymour, B., Gray, M., Tan, G., Petrovic, P., Dolan, R.J., Frith, C.D., 2009. From Threat to Fear: The neural organization of defensive fear systems in humans. *J. Neurosci.* 29, 12236–12243.
- Mobbs, D., Petrovic, P., Marchant, J.L., Hassabis, D., Weiskopf, N., Seymour, B., Dolan, R.J., Frith, C.D., 2007. When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science* 317 (5841), 1079–1083. <https://doi.org/10.1126/science.1144298>.
- Mobbs, D., Wise, T., Suthana, N., Guzman, N., Kriegeskorte, N., Leibo, J., 2021. The Promises and Challenges of Human Computational Ethology. *Neuron* 109 (14), 2224–2238.
- Mobbs, D., Yu, R., Rowe, J., Eich, H., Feldmanhall, O., Dalgleish, T., 2010. Neural activity associated with monitoring the oscillating threat value of a Tarantula. *PNAS* 107, 20582–20586.
- Moran, R., Keramati, M., Dolan, R.J., 2021. Model based planners reflect on their model-free propensities. *PLoS Comput. Biol.* 17 (1), e1008552.
- Moser, E.I., Kropff, E., Moser, M.B., 2008. Place cells, grid cells, and the brain's spatial representation system. *Annu Rev. Neurosci.* 31, 69–89.
- Mugan, U., MacIver, M.A., 2020. Spatial planning with long visual range benefits escape from visual predators in complex naturalistic environments. *Nat Commun* 11 (1), 3057. Jun 16.
- Ohman, A., Mineka, S., Fears, 2001. phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychol. Rev.* 108 (3), 483–522.
- Ortony, A., Clore, G.L., Collins, A., 1988. *The cognitive structure of emotions*. Cambridge University Press, New York.
- Otto, A.R., Gershman, S.J., Markman, A.B., Daw, N.D., 2013. The curse of planning: dissecting multiple reinforcement learning systems by taxing the central executive. *Psychol. Sci.* 24 (5), 751–761.
- Oudeyer, P.-Y., Kaplan, F., 2007. What is intrinsic motivation? a typology of computational approaches. *Front. Neuroinformatics* 1, 6.
- Panksepp, J., 1998. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press, New York, NY.
- Pellmar, B.A., Kim, J.J., 2016. What can ethobehavioral studies tell us about the brain's fear system? *Trends Neurosci.* 39 (6), 420–431.
- Pezzulo, G., van der Meer, M.A., Lansink, C.S., Pennartz, C.M., 2014. Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 18, 647–657.
- Plotkin, H., 2011. Human nature, cultural diversity and evolutionary theory. *Phil. Trans. R. Soc. B* 366, 454–463.
- Poist, T., Bever, J.D., Nemri, A., Thrall, P.H., Hochberg, M.E., 2011. A conceptual framework for the evolution of ecological specialization. *Ecol. Lett.* 14, 841–851.
- Premack, D., 2010. Why Humans Are Unique: Three Theories. *Perspect. Psychol. Sci.* 5 (1), 22–32.
- Price, J.L., 2005. Free will versus survival: brain systems that underlie intrinsic constraints on behavior. *J. Comp. Neurol.* 493, 132–139.
- Price, J.L., Drevets, W.C., 2010. Neurocircuitry of mood disorders. *Neuropsychopharmacology* 35, 192–216.
- Qi, S., Cross, L., Wise, T., Sui, X., O'Doherty, J., Mobbs, D., 2024. The Role of the Medial Prefrontal Cortex in Spatial Margin of Safety Calculations. *J. Neurosci* 44 (34). Aug 21.
- Qi, S., Hassabis, D., Sun, J., Guo, F., Daw, N., Mobbs, D., 2018. How cognitive and reactive fear circuits optimize escape decisions in humans. *PNAS* 115 (12), 3186–3191.
- Rand, D.G., Brescoll, V.L., Everett, J.A., Capraro, V., Barcelo, H., 2016 Apr. Social heuristics and social roles: Intuition favors altruism for women but not for men. *J. Exp Psychol Gen* 145 (4), 389–396.
- Reddan, M.C., Wager, T.D., Schiller, D., 2018. Attenuating Neural Threat Expression with Imagination. *Neuron* 100 (4), 994–1005 e4.
- Restrepo-Castro, J.C., Labrador, F.J., Gantiva, C., Camacho, K., Castro-Camacho, L., 2023. The effect of information seeking behaviors in fear control. *J. Behav. Ther. Exp. Psychiatry* 78, 101797.
- Ricklefs, R.E., Wikelski, M., 2002. *Physiol. /life-Hist. Nexus Trends Ecol. Evol.* 17 (10), 462–468.
- Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G.E., Wager, T.D., 2014. Representation of aversive prediction errors in the human periaqueductal gray. *Nat. Neurosci.* 17 (11), 1607–1612.
- Russek, E.M., Momennejad, I., Botvinick, M.M., Gershman, S.J., Daw, N.D., 2017. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput. Biol.* 13 (9), e1005768. <https://doi.org/10.1371/journal.pcbi.1005768>.
- Saxena, S., Cunningham, J.P., 2019 Apr. Towards the neural population doctrine. *Curr Opin Neurobiol* 55, 103–111.
- Schiff, W., Caviness, J.A., Gibson, J.J., 1962. Persistent fear responses in rhesus monkeys to the optical stimulus of "looming". *Science* 136 (3520), 982–983.
- Schmidhuber, J., 2006. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connect. Sci.* 18, 173–187.
- Schmidhuber, J., 2010. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Trans. Autonomous Mental Dev.* 2, 230–247.
- Schwartenbeck, P., FitzGerald, T., Dolan, R.J., Friston, K., 2013. Exploration, novelty, surprise, and free energy minimization. *Front Psychol.* 4, 710.
- Schwartenbeck, P., FitzGerald, T.H., Mathys, C., Dolan, R., Friston, K., 2015a. The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cereb. Cortex* 25, 3434–3445.
- Schwartenbeck, P., FitzGerald, T.H., Mathys, C., Dolan, R., Kronbichler, M., Friston, K., 2015b. Evidence for surprise minimization over value maximization in choice behavior. *Sci. Rep.* 5, 16575.
- Schwartenbeck, P., Passecker, J., Hauser, T.U., FitzGerald, T.H., Kronbichler, M., Friston, K.J., 2019. Computational mechanisms of curiosity and goal-directed exploration. *eLife* 8, e41703.
- Seth, A.K., Friston, K.J., 2016. Active interoceptive inference and the emotional brain. *Philos. Trans. R. Soc. Lond. Ser. B, Biol. Sci.* 371, 20160007.
- Sharp, P., Dolan, R., Eldar, E., 2023. Disrupted state transition learning as a computational marker of compulsivity. *Psychol. Med.* 53 (5), 2095–2105. <https://doi.org/10.1017/S0033291721003846>.
- Shepard, R., 1987. "Toward a universal law of generalization for psychological science". *Science* 237 (4820), 1317–1323.
- Signoret-Genest, J., Schukraft, N., L. Reis, S., et al., 2023. Integrated cardio-behavioral responses to threat define defensive states. *Nat Neurosci* 26, 447–457.
- Sih, A., 1980. Optimal behavior: can foragers balance two conflicting demands? *Science* 210 (4473), 1041–1043.
- Sih, A., Ziemba, R., Harding, K.C., 2000. New insights on how temporal variation in predation risk shapes prey behavior. *Trends Ecol. Evol.* 15 (1), 3–4.
- Simon, P., Dupuis, R., Costentin, J., 1994. Thigmotaxis as an index of anxiety in mice. *Influ. dopaminergic Transm. Behav. Brain Res* 61 (1), 59–64.
- Smith, R., Lane, R.D., Parr, T., Friston, K.J., 2019. Neurocomputational mechanisms underlying emotional awareness: Insights afforded by deep active inference and their potential clinical relevance. *Neurosci. Biobehav. Rev.* 107, 473–491.
- Smith, R., Schwartenbeck, P., Parr, T., Friston, K.J., 2020. An Active Inference Approach to Modeling Structure Learning: Concept Learning as an Example Case. *Front. Comput. Neurosci.* 14, 41.
- Solms, M. (2021) *The Hidden Spring: A Journey to the Source of Consciousness* W.W. Norton, New York.
- Starita, F., Kroes, M.C.W., Davachi, L., Phelps, E.A., Dunsmoor, J.E., 2019. Threat learning promotes generalization of episodic memory. *J. Exp Psychol Gen* 148 (8), 1426–1434. Aug.
- Sterling, P., Eyer, J., 1988. Allostasis: A new paradigm to explain arousal pathology. In *Handbook of Life Stress*. In: Fisher, S., Reason, J. (Eds.), *Cognition and Health*. John Wiley & Sons, New York, pp. 629–649.
- Still, S., Precup, D., 2012. An information-theoretic approach to curiosity-driven reinforcement learning. *Theory Biosci. = Theor. Den. Biowiss.* 131, 139–148.
- Suddendorf, T., Corballis, M.C., 2007. The evolution of foresight: What is mental time travel, and is it unique to humans? *Behav. Brain Sci.* 30 (3), 299–313.
- Sullivan, R.M., Opendak, M., 2018. Developmental and neurobehavioral transitions in survival circuits. *Current Opinion in Behavioral Sciences* 24, 50–55.
- Tashjian, S.M., Wise, T., Mobbs, D., 2022. Increased model-based control for acquiring protection. *PLoS-Computational. Biology* 18 (12), e1010805.
- Tashjian, S.M., Zbozinek, T.D., Mobbs, D., 2021. A decision architecture for safety computations. *Trends Cogn. Sci.* 25 (5), 342–354. <https://doi.org/10.1016/j.tics.2021.01.013>.
- Tedeschi, E., Armand, S., Buyalskaya, A., Silston, B., Mobbs, D., 2021. Fear in groups: increasing group size reduces perceptions of danger. *Emotion* 21 (7), 1499–1510.
- Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D., 2011. How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285.
- Tovote, P., Esposito, M., Botta, P., et al., 2016. Midbrain circuits for defensive behaviour. *Nature* 534, 206–212.
- Uusberg, A., Ford, B., Uusberg, H., Gross, J.J., 2023. Reappraising reappraisal: an expanded view. *Cogn. Emot.* 37 (3), 357–370.
- Vanchurin, V., Wolf, Y.I., Katsnelson, M.I., Koonin, E.V., 2022. Toward a theory of evolution as multilevel learning. *Proc. Natl. Acad. Sci. USA* 119, e2120037119.
- Verma, D., Wood, J., Lach, G., Herzog, H., Sperk, G., Tazan, R., 2016. Hunger promotes fear extinction by activation of an amygdala microcircuit. *Neuropsychopharmacology* 41 (2), 431–439.
- Vieira, J.B., Olsson, A., 2022. Neural defensive circuits underlie helping under threat in humans. *eLife* 11, e78162.
- Vigorito, C.M., Barto, A.G., 2010. Intrinsically motivated hierarchical skill learning in structured environments. *IEEE transactions on autonomous mental. Development* 2, 132–143.
- Wallace, K.J., Rosen, J.B., 2000. Predator odor as an unconditioned fear stimulus in rats: elicitation of freezing by trimethylthiazoline, a component of fox feces. *Behav. Neurosci.* 114 (5), 912–922.
- Wang, W., Schuette, P.J., Nagai, J., Tobias, B.C., Cuccovia, V., Reis, F.M., Ji, S., de Lima, M.A.X., La-Vu, M.Q., Maesta-Pereira, S., Chakerian, M., Leonard, S.J., Lin, L., Severino, A.L., Cahill, C.M., Canteras, N.S., Khakh, B.S., Kao, J.C., Adhikari, A., 2021. Coordination of escape and spatial navigation circuits orchestrates versatile flight from threats. *Neuron* S0896-6273 (21), 00203–00208.
- Webler, R.D., Berg, H., Phong, K., Tuominen, L., Holt, D.J., Morey, R.A., Lange, I., Burton, P.C., Fullana, M.A., Radua, J., Lissek, S., 2021. The neurobiology of human fear generalization: meta-analysis and working neural model. *Neurosci Biobehav Rev* 128, 421–436. Sep.
- Whittington, J.C.R., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N., Behrens, T.E. J., 2019. The Tolman-Eichenbaum machine: unifying space and relational memory through generalisation in the hippocampal formation. *bioRxiv*, 770495.
- Wilson, D.S., Wilson, E.O., 2007. *Evol.: Surviv. selfless. N. Sci. Volume* 196 (Issue 2628), 42–46.
- Wise, T., Emery, K., Radulescu, A., 2024 Feb. Naturalistic reinforcement learning. *Trends Cogn Sci* 28 (2), 144–158.
- Wise, T., Liu, Y., Chowdhury, F., Dolan, R.J., 2021. Model-based aversive learning in humans is supported by preferential task state reactivation. *Sci. Adv.* eabf9616 1–14.

Wu, C.-T., Haggerty, D., Kemere, C., Ji, D., 2017. Hippocampal awake replay in fear memory retrieval. *Nat. Neurosci.* 20, 571–580.

Yu, H., Xiang, X., Chen, Z., et al., 2021. Periaqueductal gray neurons encode the sequential motor program in hunting behavior of mice. *Nat. Commun.* 12, 6523.

Zbozinek, T.D., Perez, O.D., Wise, T., Fanselow, M., Mobbs, D., 2022. Ambiguity drives higher-order Pavlovian learning. *PLoS Comput. Biol.* 18 (9), e1010410.