Evolutionary Theory Makes Predictions About Cognition Beyond Rational Optimization

Cameron Rouse Turner[1,2], Thomas L. Griffiths[1,2]

[1]Computational Cognitive Science Laboratory, Department of Psychology, Princeton University, Princeton, NJ 08540, USA
[2]Department of Computer Science, Princeton University, Princeton, NJ 08540, USA

Date: June 16, 2026

Correspondence: Cameron Rouse Turner (c.rouse.turner@princeton.edu), Peretsman Scully Hall, Princeton, NJ 08540, USA

Funding: Templeton World Charity Foundation, Grant Number 20648.

Keywords: Cognitive evolution | Inductive bias | Resource rationality | Signal detection theory | Bet-hedging | Environmental uncertainty

Acknowledgments

We thank Thomas J. H. Morgan, Stephen Mann, and Jacob Chisausky for valuable comments on this work.

Evolutionary Theory Makes Predictions About Cognition Beyond Rational Optimization

ABSTRACT

Evolution by natural selection produces the appearance of design in minds, brains, and behavior. While rational models of cognition are usually considered to be consistent with evolution, they are rarely derived from distinctively evolutionary assumptions. Bayesian modeling, reinforcement learning, rational analysis, and resource rational analysis all postulate optimization objectives of coherence in information use or maximization of expected value, potentially under resource constraints. However, cognition is performed by an individual within a lineage where success is determined by the persistence of the lineage across generations. Here, we present a signal detection model that shows how this distinctly evolutionary objective produces predictions that depart from those of rational models. We find that the source of environmental uncertainty determines cognitive design, with weak priors being favored in environments dominated by change. This ensures offspring vary so that some profit against novel challenges, and is favored even at the expense of accurate decision making. Cognitive structure should reflect stable environmental trends, but in fluctuating environments selection strips structure away to release behavioral diversity. We highlight that evolution maximizes the geometric mean number of offspring over generations, an objective that recurs across optimization problems where outcomes accumulate sequentially, including in finance, control engineering, and information theory.

Keywords: Cognitive evolution | Inductive bias | Resource rational analysis | Signal detection theory | Bet-hedging | Environmental uncertainty

1 | Introduction

Viewing a system through the lens of its design illuminates much about how the system operates. A fruitful mode of discovery has been to analyze cognitive systems with respect to the problem they solve, allowing the construction of formal models of their underlying operations (Anderson, 1990; Marr, 1982; Shepard, 1987; Smaldino, Pietraszewski, & Wertz, 2023). In other words, considering the *computational-level* optimization problem facilitates understanding the *algorithmic-level* cognitive architecture (Marr, 1982; Varma, 2014). The earliest attempts at looking through the lens of design assumed quasi-omniscient rational actors with perfect knowledge of their own preferences and the probability of outcomes (von Neumann & Morgenstern, 1944). Accordingly, cognition was analyzed as a system that produces actions that maximize expected utility. Subsequent proposals have suggested a more biologically plausible picture: cognition is designed to perform under the constraints of limited access to information, time, and computational resources (Anderson, 1990; Griffiths, 2020; Lieder & Griffiths, 2020; Lieder, Callaway, & Griffiths, 2026).

While major computational frameworks for understanding cognition are couched in evolutionary terms, there is little consideration of the evolutionary dynamics that generate adaptations (Rich, Blokpoel, de Haan, & van Rooij, 2020; Smaldino et al., 2023). Theories aimed at characterizing design objectives repeatedly reference the notion that cognition belongs to animals attempting to overcome challenges within their environments under resource constraints. For instance, H. A. Simon (1956) presented a mathematical model showing how an animal with limited perception, storage, and planning horizon differs in its decision making from an ideal rational actor. While this aligns with evolutionary theory, it is not uniquely evolutionary. A vacuum-cleaning robot is a cognitive system that reflects both the structure of the owner's house, and the limitations of its onboard hardware. Evolution by natural selection not only suggests the appearance of design in nature, but is a commitment to a particular dynamic process by which that design is produced. Namely, design comes about by the proliferation of individuals within a population with heritable variations that improve their Darwinian fitness (Gardner, 2017).

Despite the appeal to evolutionary adaptation in the framing of rational analysis (Anderson, 1990), many models of cognition based on rational analysis are derived based solely upon principles of coherence in use of information and rational maximization of rewards, subject to limited resources. For instance, models based on Bayesian inference or reinforcement learning are typically written down without any reference to a reproducing population. There are notable exceptions involving explicitly evolutionary models applied to questions in cognitive science (e.g., Morgan, Suchow, & Griffiths, 2020; Nowak, Komarova, & Niyogi, 2001; Thompson, Kirby, & Smith, 2016; Todd & Miller, 1991). However, these models have not been directly compared with non-evolutionary rational models in a way that makes clear how evolutionary dynamics can alter how we understand cognition.

Here, we show that evolutionary theory makes predictions about the design of cognition that cannot be derived from the standard rationality-plus-resource-constraint framing. Cognition is hypothesized to evolve to allow animals to cope with a complex and changing environment by responding to information (Fawcett et al., 2014; Godfrey-Smith, 1996; Sol, 2009; Turner, Morgan, & Griffiths, 2024). This means environmental uncertainty has a crucial role in producing selection on cognition. However, environmental uncertainty arising from fluctuations across generations affects evolution in ways that are difficult to intuit, and depart from standard rational models that assume maximization of expected utility over outcomes (Seger & Brockmann, 1987; Starrfelt & Kokko, 2012). Rather than a single maximizing strategy, temporal fluctuations select for a variety of strategies. By presenting a novel signal detection model, we demonstrate how different forms of environmental stochasticity shape cognition. Our idealized model clarifies the special conditions under which evolutionary and rational models make the same predictions, and when they depart. In particular, in an environment that changes over time, agents should have weaker inductive biases. The optimal design thereby produces fewer correct decisions on aggregate, but causes offspring to vary as a way of coping with the unknown.

2 | Model

2.1 | Rational analysis of simple decisions

We begin by outlining signal detection theory, as it provides a minimal instantiation of an information-using system that implements a policy for responding to partially reliable cues (Green & Swets, 1966; Haselton & Nettle, 2006; Trimmer & Houston, 2014; Turner et al., 2024; Wickens, 2001). It also illustrates how rational analysis is typically employed in cognitive science, allowing comparison to the evolutionary case presented later. Cognition is conceived of as being an optimal solution to a problem posed by the environment, potentially taking into account resource constraints (Anderson, 1990; Lieder et al., 2026). Therefore, in deriving formal models it is standard for researchers to analyze the cognitive design that maximizes expected utility, subject to biological and computational limitations.

The observer is assumed to receive a noisy cue and make a binary decision based on an internal threshold or *criterion* that determines their actions (Figure 1). To illustrate, consider a prey animal that must avoid predators. The important distal state is whether the predator is present or absent, $S \in \{1, 0\}$. The prey animal's success hinges on appropriately performing the actions of fleeing or remaining, $A \in \{1, 0\}$. Therefore, the structure in the environment that cognition should be optimized to respect is the frequency with which predators are encountered, μ, such that $S \sim \text{Bernoulli}(\mu)$.

The animal's policy for fleeing influences the trade-off between missing predator detections versus unnecessary fleeing. Prey must make their decision based on partially reliable information. For instance, a predator may cast a larger shadow than a harmless animal on average, but there is variation due to lighting conditions. If a predator is present, the observed cue is drawn from $O \mid S = 1 \sim \mathcal{N}(d, 1)$. If instead the predator is absent the cue is drawn from $O \mid S = 0 \sim \mathcal{N}(0, 1)$. Therefore, the *discriminability* between states d is the difference in the mean intensity of the cues.

The prey animal's action depends on their policy. Prey flee when they observe a cue with greater intensity than their criterion c, such that $A = \mathbb{1}\{O > c\}$. Here, $\mathbb{1}\{\cdot\}$ is the indicator function, taking the value 1 when its argument is true and 0 otherwise. Prey only gain utility b if they correctly match fleeing to predator presence or remaining to predator absence. Therefore, payoffs are given by $b\,\mathbb{1}\{A = S\}$.

A feature of signal detection theory is that the observer's ability to detect and process information is constrained such that types of errors trade off. The probability of a missed detection is $\text{P}(A = 0 \mid S = 1) = \Phi(c - d)$, with a false alarm being $\text{P}(A = 1 \mid S = 0) = 1 - \Phi(c)$. Here, Φ is the standard cumulative distribution function of the normal distribution, which arises because outcomes are determined by regions of cue intensity either side of the criterion. Changing the criterion decreases one type of error but increases the other; only improving discriminability decreases both errors.

The expected utility of actions is a function of the observer's criterion and discriminability, which it is rational to maximize subject to resource constraints. We have formulated a version of the classic equal-variance normal model (Green

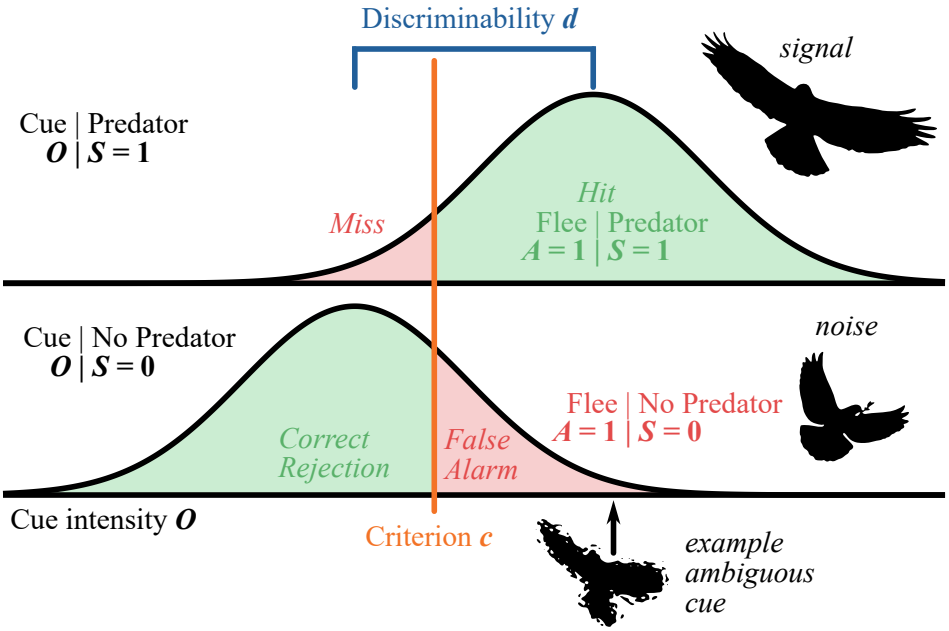

FIGURE 1 | The signal detection model assumes the intensity of cues differ between states. Predator presence results in more intense cues on average than a harmless animal (e.g., greater visual salience of a larger wingspan). The rates of correct detections and errors depend on the *criterion* and *discriminability* of an observing prey animal. Cues are noisy leading to ambiguity about the underlying state.

& Swets, 1966; Wickens, 2001) in which the expected utility is

$$U(c,d) = b\big(1-\Phi(c-d)\big)\mu + b\,\Phi(c)(1-\mu). \tag{1}$$

This encapsulates the expected payoffs gained by fleeing in predator presence and remaining in predator absence. We assume there are resource constraints on detection such that discriminability is limited, $d \in [0, d_{\max}]$. The decision rule is unrestricted, $c \in (-\infty, \infty)$. We can now make predictions about the optimal design of cognition by analyzing

$$(\hat{c}, \hat{d}) = \arg\max_{c,d} U. \tag{2}$$

This can be solved using the second partial derivative test (C. P. Simon & Blume, 1994).

We find it optimal to have as good a discriminability as constraints allow, while the criterion should reflect environmental structure in terms of the frequency of predators. The expected utility increases monotonically with discriminability, meaning it is optimal to take it to its maximum under constraints, $\hat{d} = d_{\max}$. This is because gaining more reliable information always allows more correct decisions by reducing both types of errors. With regard to policy, the optimal criterion is

$$\hat{c} = \frac{\hat{d}}{2} + \frac{1}{\hat{d}} \ln\left(\frac{1-\mu}{\mu}\right). \tag{3}$$

The term $\hat{d}/2$ is the midpoint between observation distributions, a criterion that produces equal amounts of false alarms and missed detections. The second term, involving the log odds of predator frequency, determines the strength of the response bias. If predators are abundant, a gullible criterion ($\hat{c} < 0$) that causes fleeing upon faint evidence is optimal, leading to more false alarms. If predators are rare, a fastidious criterion ($\hat{c} > 0$) is optimal, leading to often remaining and more missed detections.

Notably, we have produced these predictions about the optimal design of cognition without any mention of evolutionary dynamics. While we have discussed a situation that arises in nature, prey avoiding predators, we have used nothing outside standard optimization of expected utility under resource constraints. We now consider how to capture the influence of the evolutionary process.

2.2 | Evolution is shaped by environmental fluctuations

Selection favors producing offspring who vary when the conditions faced by a lineage change over time. Standard rational models consider the expected utility taken over the distribution of outcomes, and this always leads to a single optimal

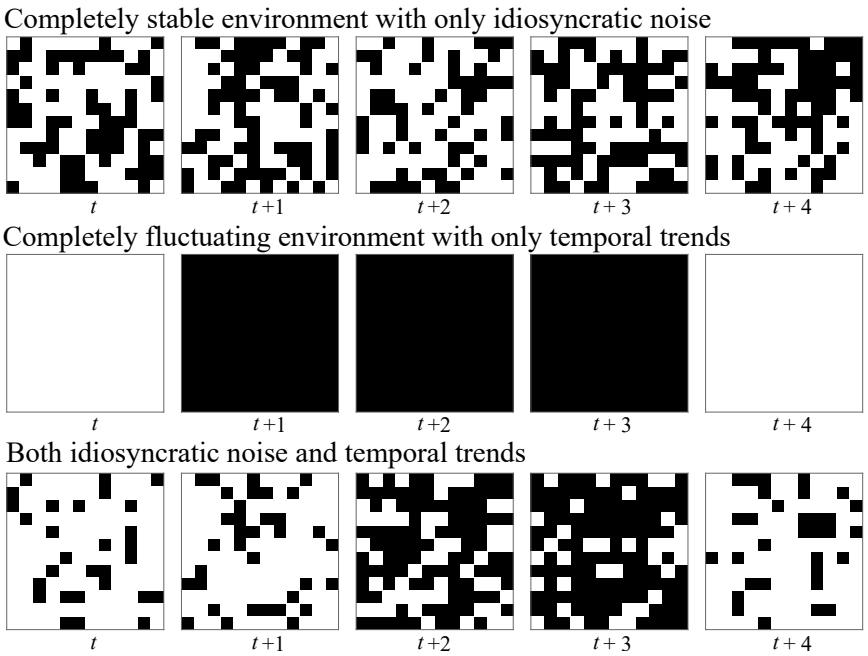

FIGURE 2 | Environmental uncertainty may be generated by stable or temporally fluctuating conditions. Encountered states may vary either within or between generations.

strategy (excluding games). In the evolutionary setting, each generation may face different circumstances produced by temporal fluctuations, such as due to changes in climate. Consequently, fitness is improved by producing a diversity of offspring so a generation facing an unlucky set of circumstances does not send the lineage extinct (Seger & Brockmann, 1987; Starrfelt & Kokko, 2012). This is analogous to diversifying a stock portfolio under compounding financial growth. We can think about environmental stochasticity as existing on a continuum (Figure 2). On one end, we have completely stable environments where each individual faces an unknown state, but there are no temporal fluctuations across generations. On the other end, we have completely fluctuating environments in which all individuals within a generation face the same conditions, but stochasticity occurs only between generations. Realistic environments exist in the middle of the continuum, but focusing on the ends facilitates understanding.

We will now bring in notions from evolutionary modeling to make clear how to think about the change in the composition of a population over time. The replicator dynamics is a canonical model of selection that tracks how proportions of trait carriers shift due to differences in fitness (Otto & Day, 2007). Here, we illustrate using two competing traits, but conclusions apply to an arbitrary number of traits. For instance, the trait could be the degree of neophobia underlying the propensity to explore. In our signal detection model, a trait is the value of the criterion. The proportion of individuals p carrying trait x, compared with trait x', changes according to:

$$p_{t+1} = \frac{U(x)\, p_t}{U(x)\, p_t + U(x')(1 - p_t)}. \tag{4}$$

Here, $U(x)$ is the expected number of offspring of an individual carrying x. To the degree that x leads to more offspring than average, the proportion of carriers p will increase from generation t. There is a general solution as an explicit function of time:

$$p(t) = \frac{U(x)^t\, p_0}{U(x)^t\, p_0 + U(x')^t\, (1 - p_0)}. \tag{5}$$

This holds in the non-game scenario where U depends only on x, and not on p_t. In the long run ($t \to \infty$), the population will be entirely composed of individuals with x when $U(x) > U(x')$, such that the stable equilibrium is $\hat{p} = 1$. This implies the optimal trait $\hat{x}$ that becomes ascendant in any population obeys $\forall x,\ U(\hat{x}) \geq U(x)$. Given this appears to be the usual condition for maximization, it is tempting to now interpret U as an expected utility and assume evolutionary theory gives the same conclusions. However, this only holds under special assumptions.

A more general measure of long-run evolutionary success is the geometric mean of the number of offspring across generations. It was tacitly assumed in the derivation of the previous paragraph that the reproductive success U was exchangeable between each generation. This is not generally true if the environment fluctuates, as the number of offspring

will in fact be a stochastic process that takes on different values depending on the generation, $\{U_t\}$. The solution becomes:

$$p(t) = \frac{p_0 \prod_{i=0}^{t-1} U_i(x)}{p_0 \prod_{i=0}^{t-1} U_i(x) + (1-p_0) \prod_{i=0}^{t-1} U_i(x')}. \tag{6}$$

Therefore, the composition of the population depends on the product of the number of offspring across generations of each type, e.g. $\prod_{i=0}^{t-1} U_i(x)$. This makes sense because lineages grow multiplicatively. If you have two children, both of whom themselves have three, you have six grandchildren. Analyzing this point has led evolutionary theorists to understand that a more accurate measure of fitness is the geometric mean of the number of offspring across generations into the distant future (Otto & Day, 2007; Seger & Brockmann, 1987; Starrfelt & Kokko, 2012):

$$W = \lim_{t\to\infty} \left(\prod_{i=1}^{t} U_i \right)^{\frac{1}{t}}. \tag{7}$$

The geometric mean is the product of a sequence of variables taken to the root of the sequence length. Unlike the arithmetic mean used in expected utility theory that combines outcomes additively, it captures how quantities accumulate via multiplicative growth. Evolution is expected to favor individuals whose growth rate is greater than others such that $W(x) > W(x')$. This condition is only equivalent to $U(x) > U(x')$ if the population is in a stable environment with absolutely no temporal stochasticity.

Evolution typically responds to environmental fluctuations by favoring a diversity of offspring (Seger & Brockmann, 1987; Starrfelt & Kokko, 2012). A lineage that experiences a few poor years will suffer a dramatic long-run decline. A feature of multiplicative growth is that variation leading to a single disastrous year negates many good years. This is because small numbers have an outsized effect. An extreme example is when a generation has no offspring and multiplying by zero ends the lineage. A strategy to persist through hard times is to have a variety of offspring so some are adapted to prevailing circumstances. For instance, let A be the phenotype generated from genotype x, such that $A \sim \text{Bernoulli}(x)$. This could be approaching a novel food or not, depending on neophobia. Whereas expected utility theory implies a single deterministic strategy is optimal, $\hat{x} \in \{0, 1\}$; evolution in a fluctuating environment favors phenotypic variety, $\hat{x} \in (0, 1)$.

The objective W repeatedly appears in optimization problems across finance, control engineering, and information theory (Donaldson-Matasci, Bergstrom, & Lachmann, 2010; Krakauer & Rockmore, 2015; Rivoire & Leibler, 2011). An investor attempting to maximize their long-run wealth under compounding should diversify their portfolio to spread risk. Similarly, decoding a message requires each symbol is transmitted, so a communicator should add redundancy to prevent errors from accumulating. In the evolutionary context, considering W suggests the optimal design of cognition may be different in a stochastically fluctuating environment in a manner that cannot be derived from the standard rational approaches.

2.3 | Optimal cognition in a fluctuating environment

Given that a central role of cognition is coping with environmental uncertainty, it is crucial we understand how different forms of uncertainty shape cognitive evolution (Fawcett et al., 2014; Godfrey-Smith, 1996; Sol, 2009; Turner et al., 2024). Our earlier signal detection model examined environmental uncertainty within a completely temporally-stable environment, with the previous section showing this leads to the same solutions as standard expected utility maximization. For comparison, we now turn to uncertainty arising from environmental fluctuations. That is, the scenario where there are large-scale changes in circumstances affecting all members of a generation, with the idealization of no idiosyncratic randomness in the state encounter within a generation. For example, a habitat may fluctuate due to wildfires causing substantial shifts in predator encounters.

We make the same assumption that the observer gains a noisy cue which they process to decide to flee; the key difference is the treatment of environmental stochasticity. Now the state variable capturing predator presence is a Bernoulli process drawn each generation, $S_t \sim \text{Bernoulli}(\mu)$. The prey animal still flees upon receiving a cue more intense than their criterion c, with the mean difference in the intensity of cues indicating predator presence being the discriminability d. The geometric mean of offspring numbers (Eq. 7) that gives the correct evolutionary objective is:

$$W(c, d) = b\big(1 - \Phi(c - d)\big)^{\mu} \Phi(c)^{(1-\mu)}. \tag{8}$$

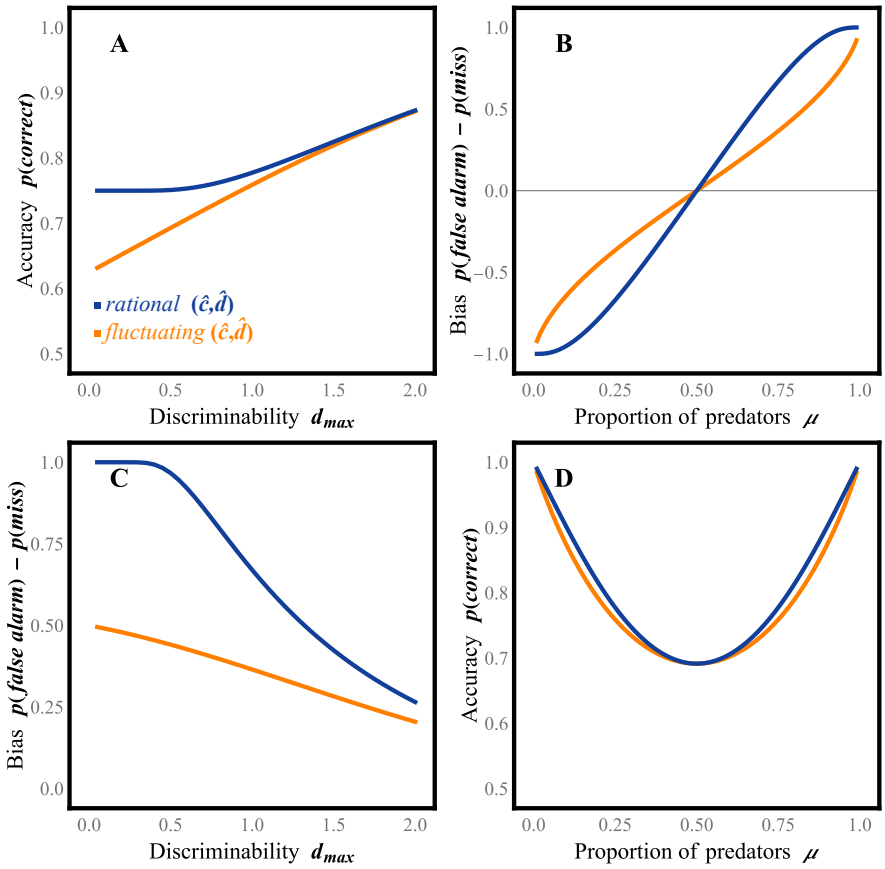

FIGURE 3 | Accuracy and response bias from the optimal strategy under rationality assumptions versus fluctuating evolutionary dynamics. **A.** Accuracy improves as the maximum discriminability increases. Accuracy is always lower under the optimal strategy in a fluctuating environment. **B.** Response bias measured as the difference between error rates tracks the expected frequency of environmental states. **C.** Fluctuating environments favor less biased responses than those predicted for the absolutely stable environment implied by standard rationality assumptions. Bias reduces in both cases when cues are more reliable. **D.** In highly uncertain environments $\mu \approx .5$ error rates are high, so accuracy is low. Representative trends depicted; when not plotted, parameter $\mu = .75$ and $d_{\max} = 1$.

This is because in a sequence of predator states $\{S_t\}$, a proportion μ of them are expected to have the predator present, $S_t = 1$. In these cases, the benefit b is gained upon a correct detection, $1 - \Phi(c - d)$. The same logic applies for the $1 - \mu$ proportion of generations in which $S_t = 0$.

We can now analyze the form of cognition that maximizes the geometric mean objective. As in the previous case of a temporally-stable environment, W monotonically increases with discriminability d. Accordingly, it is optimal for an observer to gain as much information as constraints allow, $\hat{d} = d_{\max}$. However, in this case the optimal criterion $\hat{c}$ cannot be solved for in closed form, and obeys the transcendental equation:

$$0 = \mu \frac{\phi(d - \hat{c})}{\Phi(d - \hat{c})} - (1 - \mu) \frac{\phi(\hat{c})}{\Phi(\hat{c})}. \tag{9}$$

The standard normal probability density function is ϕ. We now turn to numerical solutions to present the distinct predictions about the design of cognition that arise from evolutionary dynamics, along with those that match the standard rational approach.

3 | Results

3.1 | Gaining reliable information is always beneficial

Computational frameworks often assume that cognition is designed to acquire information. We find this holds regardless of whether uncertainty results from environmental fluctuations, or idiosyncratic noise in a stable environment as tacitly assumed by standard rational objectives. The reason is that reducing uncertainty always produces more correct decisions and fewer errors (Figure 3A). Previous work suggests this finding is not an artifact of the current stylized *benefit-if-correct*

model, and also occurs when outcomes have different payoffs (Turner et al., 2024). Reliable information is a broadly valuable resource that cognition is designed to obtain.

3.2 | Biases should reflect environmental structure

Cognitive structure is thought to track environmental structure, but it is unclear how this relationship is affected by environmental change. In our signal detection model the strength of response bias is measured by the propensity to avoid false alarms relative to missed detections. Figure 3B shows that it is always favorable for cognition to track the structure of the environment. In particular, if predators are frequent, the bias should be to flee often, setting a criterion that triggers fleeing on faint evidence of a predator. Conversely, if predators are rare, it is favorable to often remain. We find this pattern holds in both fluctuating and stable environments. This implies that the notion that cognitive structure should reflect environmental structure is robust across rational and evolutionary modeling assumptions.

3.3 | Environmental fluctuations favor weak biases

Standard rational objectives may produce misleading predictions about biases in information processing. Figure 3C shows the strength of response bias as a function of the reliability of information. Weaker less-committal biases are optimal if the environment is marked by temporal changes leading to unforeseen circumstances. This is because of the evolutionary benefits of diversifying offspring to hedge against catastrophic losses in a single generation. The actions of offspring with weak biases are more variable, reflecting weaker prior commitments, and therefore generate greater behavioral diversity. This is contrary to the usual rational analysis that predicts comparatively strong biases. The prioritization of variation in offspring is favored even though it reduces overall decision accuracy (Figure 3A). The degree of divergence in predictions can be substantial. Consider a situation with a moderately likely and discriminable state ($\mu = .75, d_{\max} = 1$). An individual adapted to a fluctuating environment has 12% greater odds of making an error, with a 29% reduction in false alarms and a 200% increase in missed detections. When the environment has few stable trends, selection favors stripping out cognitive structure to diversify responses.

3.4 | Biases are insurance against unreliable information

Our model allowed us to examine how abilities for acquiring information interact with biases for processing information. We found that when cues are reliable, weaker biases are favored, across both fluctuating and stable environments (Figure 3C). This is because when observers gain high certainty from cues, decisions can be driven by acquired information. In particular, as both types of error become rarer, the difference in their relative frequency also diminishes. In the limit of errorless perception, the probabilities of a false alarm and a missed detection are both zero, eliminating the need for any bias. Broadly, biases are useful to the degree the state cannot be determined using cues, by producing a propensity to match the historically likely state.

4 | Discussion

Evolution is often cited as underpinning frameworks for analyzing the design of cognition, but the resulting formalisms rarely involve any distinctly evolutionary assumptions. Here, we presented a mathematical model that showed how adding evolutionary dynamics can alter conclusions about the form of optimal cognition. If the environment does not change with time, so that each individual experiences idiosyncratic noise, decisions are like exchangeable trials. We showed this leads to the recovery of the familiar results from the standard rational approaches. Cognition should produce a response bias that maximizes accuracy in terms of the overall proportion of correct decisions. This is achieved by a Bayes-optimal criterion such that the fraction of fleeing responses matches the marginal frequency of predator encounters (Green & Swets, 1966; Wickens, 2001). By contrast, evolution favors diversity in offspring behavior if the environment fluctuates over time, with whole generations experiencing shifts in conditions. This is achieved by producing offspring with weaker biases so that actions are more varied. Consequently, optimal cognition in a fluctuating environment leads to fewer correct decisions, but diversifies the risk of the lineage, protecting against extinction in any single generation. This reduction in accuracy is a clear example of a prediction that arises distinctly from evolutionary dynamics.

We found several principles about the design of cognition that are robust across both sources of environmental uncertainty. Rational models often assume the objective of acquiring information and tracking environmental structure; we rigorously examined whether these objectives are converged on by evolution. First, we confirmed that reliable cues are a

valuable resource because they allow correct decisions, implying cognition should be designed to acquire as much information as is expedient. That is, information should be acquired until costs in terms of time or metabolic energy dominate (Marzen & DeDeo, 2017; Turner et al., 2024). Second, we found that although environmental fluctuations produce weaker response biases, the structure of cognition should broadly reflect the structure of important environmental contingencies. For instance, if predators are frequent, biases should support fleeing more often on weaker evidence. Third, we found response biases are favorable to the degree information is unreliable. When cues are a poor guide to the environmental state, priors shape behavior toward the likely scenario. However, biases are unnecessary when cues provide certainty, as behavior can be guided by acquired information.

4.1 | Coping with the unknown

Our models allowed us to understand how evolution combats previously unencountered shifts in conditions. Environmental fluctuations mean that while present environments continue past trends, they also fail to resemble the past perfectly. The standard rational objective based on expected utility has the unique properties of capturing a coherent ordering of preferences, and is the value converged on in the long-run across many repeated decisions (Lieder et al., 2026; von Neumann & Morgenstern, 1944). However, expected utility does not capture the fact that cognition of animals is shaped by their position in a lineage. Respecting that evolution works at the level of lineages implies the appropriate objective is the geometric mean of numbers of offspring across generations, rather than the arithmetic mean underlying expected utility. Accordingly, we find that a strategy to cope with the unknown-unknowns generated by a changing environment is to have offspring with weak biases, so some will produce actions that happen to be favorable. Decorrelating the fate of offspring makes the lineage robust. By contrast, environments dominated by stable trends leading to known-unknowns favor strong priors to exploit likely structure. Finally, evolution should invest in information gathering abilities, such as superior sensory systems, as a strategy that combats the unknown, regardless of the underlying source.

4.2 | Evolutionary considerations

Although environmental uncertainty is a primary driver of cognitive evolution, explicitly evolutionary models within cognitive science appear to consider only idiosyncratic uncertainty (e.g., Morgan et al., 2020; Nowak et al., 2001; Thompson et al., 2016; Todd & Miller, 1991). Randomness in these models—from noisy cues, available partners, or reproduction—averages out over many simulations to give the same optimal cognitive designs predicted by rational analysis. Nonetheless, such evolutionary models still provide additional insight into the basins of attraction in design space, revealing which designs evolving populations typically reach. Our stylized model incorporating temporally fluctuating uncertainty shows how evolutionary theory generates predictions that cannot in principle be derived under rational analysis. Notable evolutionary considerations that warrant continued examination include kin selection theory that suggests cognitive design should help genetic relatives (Gardner, West, & Wild, 2011; Lieberman, Tooby, & Cosmides, 2007), and genetic drift through which historical contingency shapes cognitive evolution (Gould & Lewontin, 1979). Novel predictions also arise from considering the developmental timescale of evolutionary life-history theory (Frankenhuis & Walasek, 2020; Nettle, Frankenhuis, & Rickard, 2013), and the intergenerational timescale of non-genetic cultural inheritance (Bender, 2020; Mesoudi, 2011). Moreover, evolutionary foraging theory has been particularly informative for investigations of the cognitive mechanisms underlying exploration and search (Harhen & Bornstein, 2023; Hills et al., 2015).

4.3 | Bias, bias, and more bias

Thinking in evolutionary terms suggests that a clear understanding of cognitive design requires disambiguating three notions of bias. First, there is what we call *rational-relative bias*, for the usage in the bias and heuristic literature (Tversky & Kahneman, 1974). It was a Nobel-prize-winning discovery to show that decision making often departs from rational models, with people relying on simple and error-prone heuristics. The resulting deviations from rational benchmarks are referred to as biases; but whether they are sub-optimal distortions depends on what is included in the objective. One approach to addressing this is to include resource constraints as part of the optimization problem, showing that rational-relative biases are generated by limitations of time, information, or computation (Lieder et al., 2026). We have shown that evolutionary theory gives a very different objective function that can explain departures from even this standard resource rational analysis. This highlights the broad fact that the normative framework chosen by the theorist determines what counts as optimal cognition (Khemlani, Byrne, & Johnson-Laird, 2018; Pothos & Busemeyer, 2013).

Second, *inductive biases* are the pre-existing structure built into an information-using system that determines how incoming data influences behavior (Mitchell, 1997). From an evolutionary perspective, we can see that inductive biases arise partially from the genetic structure inherited by individuals that is shaped by a history of selection.

Third, *response bias* refers to the strength of priors in guiding actions relative to acquired information. In our signal detection model, the extent of discriminability and setting of the criterion determine the observer's inductive bias. By contrast, the degree to which the criterion induced false fleeing rather than missed detection was the observer's response bias. Care must be taken not to confuse these notions, and our work provides clarity by showing how they relate in a mathematical model.

4.4 | Conclusion

A core challenge in understanding the mind is simultaneously construing it as an information-processing system and a product of evolution. Our models point to the work still required to integrate insights across disciplines focused on different sides of this dual status of cognition. Traditional computational frameworks have been successful because they have tracked the qualities required for decision-making under uncertainty. This focus on the first status implies coherence in use of information, maximizing rewards, and expediency given limitations. Yet, animal minds also evolve, and this fact bends optimal design towards that which allows the persistence of the carrier's lineage. The objective of long-run growth we presented arises not only in evolution, but throughout systems where outcomes form a path-dependent sequence (Donaldson-Matasci et al., 2010; Krakauer & Rockmore, 2015; Rivoire & Leibler, 2011). It occurs in the transmission of information, and accumulation of resources. A basic principle of optimization is that changing the objective typically changes the solution. Bridging evolutionary considerations and computational frameworks offers a unifying perspective for reevaluating what constitutes optimal cognitive design.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bender, A. (2020). The role of culture and evolution for human cognition. *Topics in Cognitive Science*, *12*(4), 1403–1420. doi: 10.1111/tops.12449

Donaldson-Matasci, M. C., Bergstrom, C. T., & Lachmann, M. (2010). The fitness value of information. *Oikos*, *119*(2), 219–230. doi: 10.1111/j.1600-0706.2009.17781.x

Fawcett, T. W., Fallenstein, B., Higginson, A. D., Houston, A. I., Mallpress, D. E. W., Trimmer, P. C., & McNamara, J. M. (2014). The evolution of decision rules in complex environments. *Trends in Cognitive Sciences*, *18*(3), 153–161. doi: 10.1016/j.tics.2013.12.012

Frankenhuis, W. E., & Walasek, N. (2020). Modeling the evolution of sensitive periods. *Developmental Cognitive Neuroscience*, *41*, 100715. doi: 10.1016/j.dcn.2019.100715

Gardner, A. (2017). The purpose of adaptation. *Interface Focus*, *7*(5), 20170005. doi: 10.1098/rsfs.2017.0005

Gardner, A., West, S. A., & Wild, G. (2011). The genetical theory of kin selection. *Journal of Evolutionary Biology*, *24*(5), 1020–1043. doi: 10.1111/j.1420-9101.2011.02236.x

Godfrey-Smith, P. (1996). *Complexity and the function of mind in nature*. Cambridge, UK: Cambridge University Press.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *205*(1161), 581–598. doi: 10.1098/rspb.1979.0086

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons.

Griffiths, T. L. (2020). Understanding human intelligence through human limitations. *Trends in Cognitive Sciences*, *24*(11), 873–888. doi: 10.1016/j.tics.2020.09.001

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences*, *120*(13), e2216524120. doi: 10.1073/pnas.2216524120

Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, *10*(1), 47–66. doi: 10.1207/s15327957pspr1001_3

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D., & the Cognitive Search Research Group. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, *19*(1), 46–54. doi: 10.1016/j.tics.2014.10.004

Khemlani, S. S., Byrne, R. M. J., & Johnson-Laird, P. N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science*, *42*(6), 1887–1924. doi: 10.1111/cogs.12634

Krakauer, D. C., & Rockmore, D. N. (2015). The mathematics of adaptation (or the ten avatars of Vishnu). In N. J. Higham (Ed.), *The Princeton companion to applied mathematics* (pp. 591–597). Princeton, NJ: Princeton University Press.

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature*, *445*(7129), 727–731. doi: 10.1038/nature05510

Lieder, F., Callaway, F., & Griffiths, T. L. (2026). *The rational use of cognitive resources*. Princeton, NJ: Princeton University Press.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. doi: 10.1017/S0140525X1900061X

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W.H. Freeman & Co.

Marzen, S. E., & DeDeo, S. (2017). The evolution of lossy compression. *Journal of the Royal Society Interface*, *14*(130), 20170166. doi: 10.1098/rsif.2017.0166

Mesoudi, A. (2011). *Cultural Evolution: How Darwinian theory can explain human culture and synthesize the social sciences*. Chicago, IL: University of Chicago Press.

Mitchell, T. M. (1997). *Machine learning*. New York, NY: McGraw-Hill.

Morgan, T. J. H., Suchow, J. W., & Griffiths, T. L. (2020). What the baldwin effect affects depends on the nature of plasticity. *Cognition*, *197*, 104165. doi: 10.1016/j.cognition.2019.104165

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131343. doi: 10.1098/rspb.2013.1343

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, *291*(5501), 114–118. doi: 10.1126/science.291.5501.114

Otto, S. P., & Day, T. (2007). *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton, NJ: Princeton University Press.

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, *36*(3), 255–274. doi: 10.1017/S0140525X12001525

Rich, P., Blokpoel, M., de Haan, R., & van Rooij, I. (2020). How intractability spans the cognitive and evolutionary levels of explanation. *Topics in Cognitive Science*, *12*(4), 1382–1402. doi: 10.1111/tops.12506

Rivoire, O., & Leibler, S. (2011). The value of information for populations in varying environments. *Journal of Statistical Physics*, *142*(6), 1124–1166. doi: 10.1007/s10955-011-0166-2

Seger, J., & Brockmann, H. J. (1987). What is bet-hedging? In P. H. Harvey & L. Partridge (Eds.), *Oxford surveys in evolutionary biology* (Vol. 4, pp. 182–211). Oxford, UK: Oxford University Press.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. doi: 10.1126/science.3629243

Simon, C. P., & Blume, L. (1994). *Mathematics for economists*. New York: W. W. Norton & Company.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. doi: 10.1037/h0042769

Smaldino, P. E., Pietraszewski, D., & Wertz, A. E. (2023). On the problems solved by cognitive processes. *Cognitive Science*, *47*(6), e13297. doi: 10.1111/cogs.13297

Sol, D. (2009). Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biology Letters*, *5*(1), 130–133. doi: 10.1098/rsbl.2008.0621

Starrfelt, J., & Kokko, H. (2012). Bet-hedging — a triple trade-off between means, variances and correlations. *Biological Reviews*, *87*(3), 742–755. doi: 10.1111/j.1469-185X.2012.00225.x

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535. doi: 10.1073/pnas.1523631113

Todd, P. M., & Miller, G. F. (1991). Exploring adaptive agency II: Simulating the evolution of associative learning. In J.-A. Meyer & S. W. Wilson (Eds.), *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (pp. 306–315). Cambridge, MA: MIT Press/Bradford Books. doi: 10.7551/mitpress/3115.003.0042

Trimmer, P. C., & Houston, A. I. (2014). An evolutionary perspective on information processing. *Topics in Cognitive Science*, *6*(2), 312–330. doi: 10.1111/tops.12085

Turner, C. R., Morgan, T. J. H., & Griffiths, T. L. (2024). Environmental complexity and regularity shape the evolution of cognition. *Proceedings of the Royal Society B: Biological Sciences*, *291*(2033). doi: 10.1098/rspb.2024.1524

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. doi: 10.1126/science.185.4157.1124

Varma, S. (2014). The subjective meaning of cognitive architecture: A Marrian analysis. *Frontiers in Psychology*, *5*, 440. doi: 10.3389/fpsyg.2014.00440

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Wickens, T. D. (2001). *Elementary signal detection theory*. New York: Oxford University Press.