Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Neurobiology

Optimal models of decision-making in dynamic environments

Zachary P Kilpatrick[1], William R Holmes[2,3,7], Tahra L Eissa[1] and Krešimir Josić[4,5,6]

Nature is in constant flux, so animals must account for changes in their environment when making decisions. How animals learn the timescale of such changes and adapt their decision strategies accordingly is not well understood. Recent psychophysical experiments have shown humans and other animals can achieve near-optimal performance at two alternative forced choice (2AFC) tasks in dynamically changing environments. Characterization of performance requires the derivation and analysis of computational models of optimal decision-making policies on such tasks. We review recent theoretical work in this area, and discuss how models compare with subjects' behavior in tasks where the correct choice or evidence quality changes in dynamic, but predictable, ways.

Addresses
[1] Department of Applied Mathematics, University of Colorado, Boulder, CO, USA
[2] Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA
[3] Department of Mathematics, Vanderbilt University, Nashville, TN, USA
[4] Department of Mathematics, University of Houston, Houston, TX, USA
[5] Department of Biology and Biochemistry, University of Houston, Houston, TX, USA
[6] Department of BioSciences, Rice University, Houston, TX, USA
[7] Quantitative Systems Biology Center, Vanderbilt University, Nashville, TN, USA

Corresponding authors: Kilpatrick, Zachary P (zpkilpat@colorado.edu), Josić, Krešimir (josic@math.uh.edu)

Current Opinion in Neurobiology 2019, **58**:xx-yy

This review comes from a themed issue on **Computational neuroscience**

Edited by **Máté Lengyel** and **Brent Doiron**

https://doi.org/10.1016/j.conb.2019.06.006

Introduction

To translate stimuli into decisions, animals interpret sequences of observations based on their prior experiences [1]. However, the world is fluid: The context in which a decision is made, the quality of the evidence, and even the best choice can change before a judgment is formed, or an action taken. A source of water can dry up, or a nesting site can become compromised. But even when not fully predictable, changes often have statistical structure: Some changes are rare, others are frequent, and some are more likely to occur at specific times. How have animals adapted their decision strategies to a world that is structured, but in flux?

Classic computational, behavioral, and neurophysiological studies of decision-making mostly involved tasks with fixed or statistically stable evidence [1,2,3]. To characterize the neural computations underlying decision strategies in changing environments, we must understand the dynamics of evidence accumulation [4]. This requires novel theoretical approaches. While normative models are a touchstone for theoretical studies [5,6••], even for simple dynamic tasks the computations required to optimally translate evidence into decisions can become prohibitive [7]. Nonetheless, quantifying how behavior differs from normative predictions helps elucidate the assumptions animals use to make decisions [8,9••].

We review normative models and compare them with experimental data from two alternative forced choice (2AFC) tasks in dynamic environments. Our focus is on tasks where subjects passively observe streams of evidence, and the evidence quality or correct choice can vary within or across trials. Humans and animals adapt their decision strategies to account for such volatile environments, often resulting in performance that is nearly optimal on average. However, neither the computations they use to do so, nor their neural implementations are well understood.

Optimal evidence accumulation in changing environments

Normative models of decision-making typically assume subjects are Bayesian agents [14,15] that probabilistically compute their belief of the state of the world by combining fresh evidence with previous knowledge. Beyond normative models, notions of optimality require a defined objective. For instance, an observer may need to report the location of a sound [16], or the direction of a moving cloud of dots [5], and is rewarded if the report is correct. Combined with a framework to translate probabilities or beliefs into actions, normative models provide a rational way to maximize the net rewards dictated by the environment and task. Thus an optimal model combines normative computations with a policy that translates a belief into the optimal action.

Box 1 Normative evidence accumulation in dynamic environments.

Discrete time. At times $t_{1:n}$ an observer receives a sequence of noisy observations, $\xi_{1:n}$, of the state $S_{1:n}$, governed by a two-state Markov process (Figure 1b). Observation likelihoods, $f_{\pm}(\xi) = \mathrm{P}(\xi|S_{\pm})$, determine the belief (log-likelihood ratio: LLR), $y_n = \log\frac{\mathrm{P}(S_n=S_+|\xi_{1:n})}{\mathrm{P}(S_n=S_-|\xi_{1:n})}$, after observation n. If the observations are conditionally independent, the LLR can be updated recursively [5,17•]:

$$y_n = \underbrace{\log\frac{f_+(\xi_n)}{f_-(\xi_n)}}_{\text{current evidence}} + \underbrace{\log\frac{(1-h)\exp(y_{n-1})+h}{h\exp(y_{n-1})+(1-h)}}_{\text{discounted prior belief}}, \quad (1)$$

where h is the hazard rate (probability the state switches between times t_{n-1} and t_n). The belief prior to the observation at time t_n, y_{n-1}, is discounted according to the environment's volatility h. When $h = 0$, Eqn (1) reduces to the classic drift-diffusion model (DDM), and evidence is accumulated perfectly over time. When $h = 1/2$, only the latest observation, ξ_n, is informative. For $0 < h < 1/2$, prior beliefs are discounted, so past evidence contributes less to the current belief, y_n, corresponding to leaky integration. When $1/2 < h < 1$, the environment alternates.

Continuous time. When $t_n - t_{n-1} = \Delta t \ll 1$, and the hazard rate is defined $\Delta t{\cdot}h$, LLR evolution can be approximated by the stochastic differential equation [5,17•]:

$$\mathrm{d}y = \underbrace{g(t)\mathrm{d}t}_{\text{drift}} + \underbrace{\mathrm{d}W_t}_{\text{noise}} \underbrace{-2h\sinh(y)\mathrm{d}t}_{\text{nonlinear filter}}, \quad (2)$$

where $g(t)$ jumps between $+g$ and $-g$ at a rate h, W_t is a zero mean Wiener process with variance ρ^2, and the nonlinear filter $-2h\sinh(y)$ optimally discounts prior evidence. In contrast to the classic continuum DDM, the belief, $y(t)$, does not increase indefinitely, but saturates due to evidence-discounting.

How are normative models and optimal policies in dynamic environments characterized? Older observations have less relevance in rapidly changing environments than in slowly changing ones. Ideal observers account for environmental changes by adjusting the rate at which they discount prior information when making inferences and decisions [17•]. In Box 1 we show how, in a normative model, past evidence is nonlinearly discounted at a rate dependent on environmental volatility [5,17•]. When this volatility [8] or the underlying evidence quality [13••,18] are unknown, they must also be inferred.

In 2AFC tasks, subjects accumulate evidence until they decide on one of two choices either freely or when interrogated. In these tasks, fluctuations can act on different timescales (Figure 1a): on each trial (Figure 1b,c) [5,6••], unpredictably within only some trials [19•,20], between trials in a sequence [11,16], or gradually across long blocks of trials [21]. We review findings in the first three cases and compare them to predictions of normative models.

Within trial changes promote leaky evidence accumulation

Normative models of dynamic 2AFC tasks (Figures 1b,c and 2a, Box 1) exhibit adaptive, nonlinear discounting of prior beliefs at a rate modified by expectations of the environment's volatility (Figure 1c) and saturation of certainty about each hypothesis, regardless of how much evidence is accumulated (Figure 2a). Likewise, the performance of ideal observers at change points — times when the correct choice switches — depends sensitively on environmental volatility (Figure 2aiii). In slowly changing environments, optimal observers assume that changes are rare, and thus adapt slowly after one has occurred. Whereas, in rapidly changing environments, observers quickly update their belief after a change point. In contrast, ideal observers in static environments weigh all past observations equally, and their certainty grows without bound until a decision [3,1].

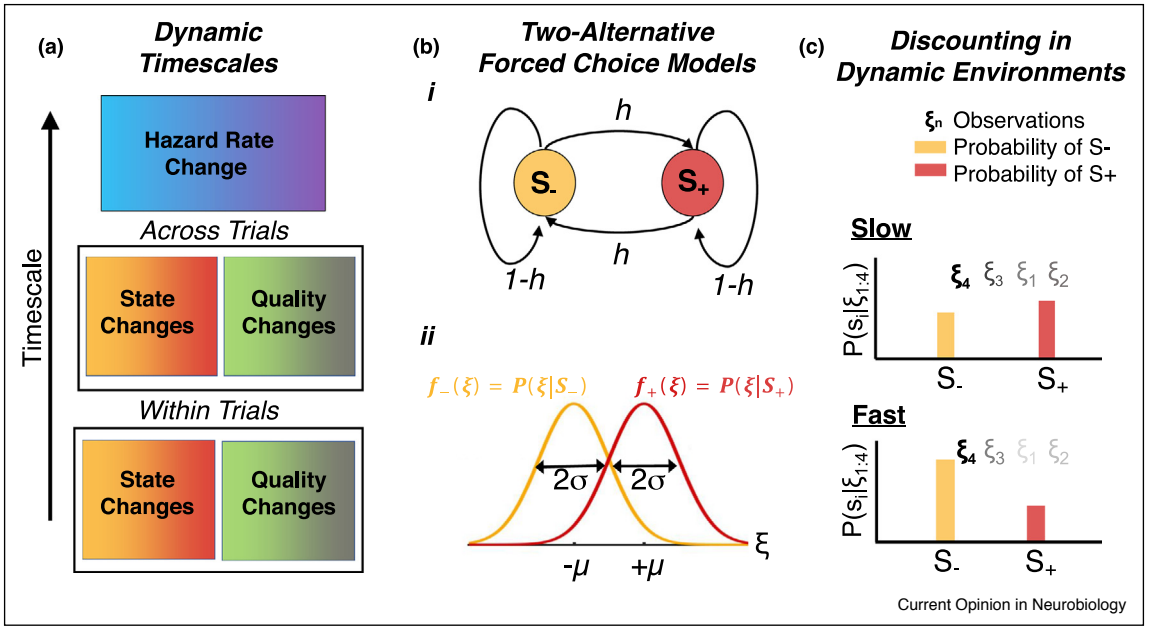
The responses of humans and other animals on tasks in which the correct choice changes stochastically during a trial share features with normative models: In a random dot-motion discrimination (RDMD) task, where the motion direction switches at unsignaled changepoints, humans adapt their decision-making process to the switching (hazard) rate (Figure 2ai) [5]. Yet, on average, they overestimate the change rates of rapidly switching environments and underestimate the change rates of slowly switching environments, possibly due to ecologically adaptive biases that are hard to train away. In a related experiment (Figure 2aii), rats were trained to identify which of two Poisson auditory click streams arrived at a higher rate [22]. When the identity of the higher-frequency stream switched unpredictably during a trial, trained rats discounted past clicks near-optimally on average, suggesting they learned to account for latent environmental dynamics [6••].

However, behavioral data are not uniquely explained by normative models. Linear approximations of normative models perform nearly identically [17•], and, under certain conditions, fit behavioral data well [5,6••,23]. Do subjects implement normative decision policies or simpler strategies that approximate them? Subjects' decision strategies can depend strongly on task design and vary across individuals [5,9••], suggesting a need for sophisticated model selection techniques. Recent research suggests normative models can be robustly distinguished from coarser approximations when task difficulty and volatility are carefully tuned [24].

Subjects account for correlations between trials by biasing initial beliefs

Natural environments can change over timescales that encompass multiple decisions. However, in many experimental studies, task parameters are fixed or generated independently across trials, so evidence from previous trials is irrelevant. Even so, subjects often use decisions and information from earlier trials to (serially) bias future choices [25,26,27•], reflecting ingrained assumptions about cross-trial dependencies [21,28].

Figure 1



Two alternative forced choice (2AFC) tasks in dynamic environments. **(a)** Possible timescales of environmental dynamics: The state (S_+ or S_-), or the quality of the evidence (e.g. coherence of random dot motion stimulus) may switch within a trial [5,6••,10], or across trials [11,12,13••]; the hazard rate (h), can change across blocks of trials [6••,9••]. **(b)** In a dynamic 2AFC task, a two-state Markov chain with hazard rate h determines the state. **(bi)** The current state (correct hypothesis) is either S_+ (red) or S_- (yellow). **(bii)** Conditional densities of the observations, $f_\pm(\xi) = P(\xi|S_\pm)$, shown as Gaussians with means $\pm\mu$ and standard deviation σ. **(c)** Evidence discounting is shaped by the environmental timescale: (Top) In slow environments, posterior probabilities over the states, $P(S_\pm|\xi_{1:4})$, are more strongly influenced by the cumulative effect of past observations, $\xi_{1:3}$, (darker shades of the observations, ξ_i, indicate higher weight) and thus points to S_+. (Bottom) If changes are fast, beliefs depend more strongly on the current observation, ξ_4, which outweighs older evidence and points to S_-.

To understand how subjects adapt to constancy and flux across trials, classic 2AFC experiments have been extended to include correlated cross-trial choices (Figure 2b) where both the evidence accumulated during a trial and probabilistic reward provide information that can be used to guide subsequent decisions [16,29]. When a Markov process [30] (Figure 1b) is used to generate correct choices, human observers adapt to these trial-to-trial correlations, and their response times are accurately modeled by drift diffusion [11] or ballistic models [16] with biased initial conditions.

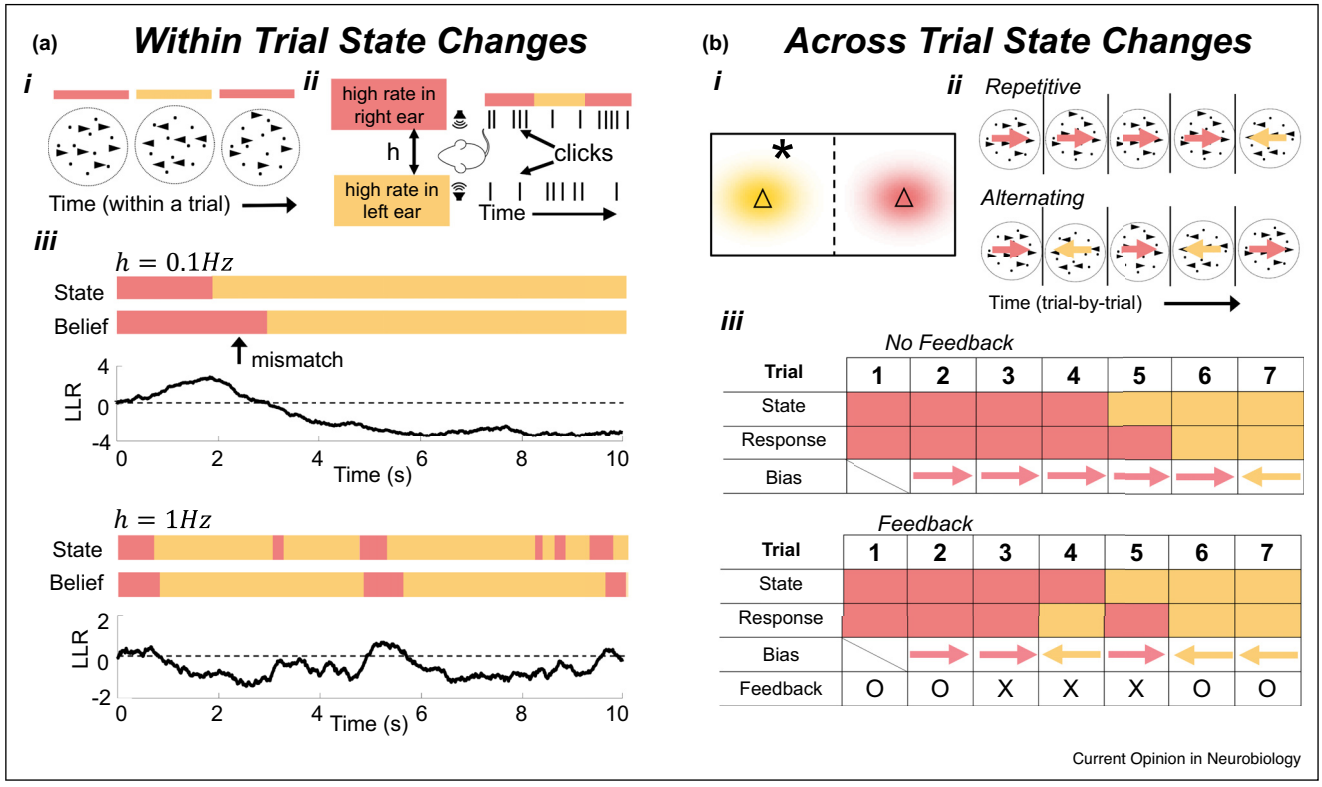
Feedback or decisions across correlated trials impact different aspects of normative models [31] including accumulation speed (drift) [32–34], decision bounds [11], or the initial belief on subsequent trials [12,35,36]. Given a sequence of dependent but statistically identical trials, optimal observers should adjust their initial belief and decision threshold [16,28], but not their accumulation speed in cases where difficulty is fixed across trials [18]. Thus, optimal models predict that observers should, on average, respond more quickly, but not more accurately [28]. Empirically, humans [12,35,36] and other animals [29] do indeed often respond faster on repeat trials, which can be modeled by per trial adjustments in initial belief. Furthermore, this bias can result from explicit feedback or subjective estimates, as demonstrated in studies where no feedback is provided (Figure 2biii) [16,36].

The mechanism by which human subjects carry information across trials remains unclear. Different models fit to human subject data have represented inter-trial dependencies using initial bias, changes in drift rate, and updated decision thresholds [11,16,34]. Humans also tend to have strong preexisting repetition biases, even when such biases are suboptimal [25,26,27•]. Can this inherent bias be overcome through training? The answer may be attainable by extending the training periods of humans or nonhuman primates [5,9••], or using novel auditory decision tasks developed for rodents [6••,29]. Ultimately, high throughput experiments may be needed to probe how ecologically adaptive evidence accumulation strategies change with training.

Time-varying thresholds account for heterogeneities in task difficulty

Optimal decision policies can also be shaped by unpredictable changes in decision difficulty. For instance, task difficulty can be titrated by varying the signal-to-noise ratio of the stimulus, so more observations are required to

Figure 2



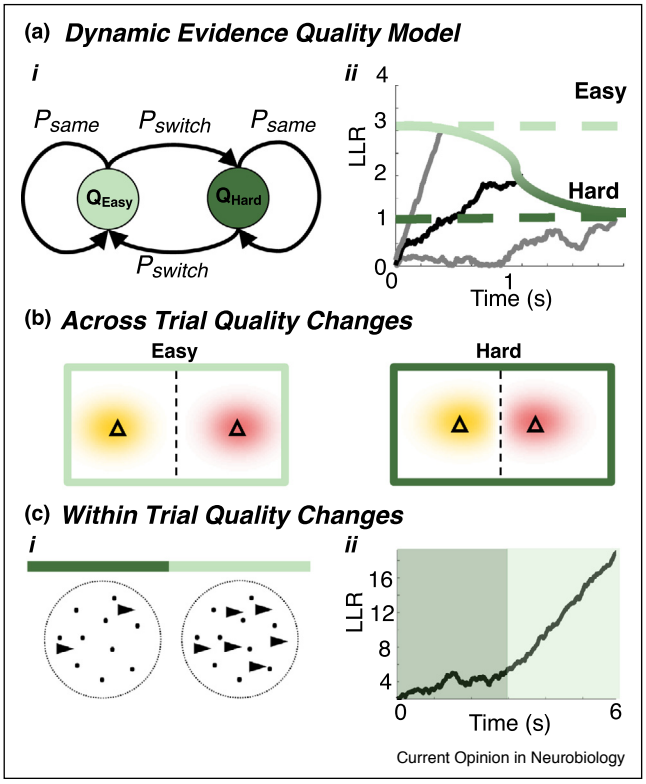
Dynamic state changes. **(a)** State changes within trials in a **(ai)** random dot motion discrimination (RDMD) task, in which drift direction switches throughout the trial [5], and **(aii)** dynamic auditory clicks task, in which the side of the higher rate stream alternates during the trial [6••]. **(aiii)** An ideal observer's LLR (see Eqn 2 in Box 1) when the hazard rate is low (top panels: $h = 0.1$ Hz) and high (bottom panels: $h = 1$ Hz). Immediately after state changes, the belief typically does not match the state. **(b)** State changes across trials. **(bi)** In the triangles task [5], samples (star) are drawn from one of two Gaussian distributions (yellow and red clouds) whose centers are represented by triangles. The observers must choose the current center (triangle). **(bii)** In an RDMD task, dots on each trial move in one of two directions (colored arrows) chosen according to a two-state Markov process. Depending on the switching rate, trial sequences may include excessive repetitions (Top), or alternations (Bottom). **(biii)** (Top) Responses can be biased by decisions from previous trials. (Bottom) Probabilistic feedback ('O': correct; 'X': incorrect) affects initial bias (e.g. trials 3, 4, and 5), even when not completely reliable.

obtain the same level of certainty. Theoretical studies have shown that it is optimal to change one's decision criterion *within* a trial when the difficulty of a decision varies *across* trials [13••,18,37]. The threshold that determines how much evidence is needed to make a decision should vary during the trial (Figure 3a) to incorporate up-to-date estimates of trial difficulty [18]. There is evidence that subjects use time-varying decision boundaries to balance speed and accuracy on such tasks [38,39].

Dynamic programming can be used to derive optimal decision policies when trial-to-trial difficulties or reward sizes change. This method provides an optimal solution to a complex decision-making process by recursively breaking it into a sequence of simpler steps. For instance, when task difficulty changes across trials in a RDMD task, optimal decisions are modeled by a DDM with a time-varying boundary, in agreement with reaction time distributions of humans and monkeys [18,38]. Both dynamic programming [18] and parameterized function [38,40] based models suggest decreasing bounds maximize reward rates (Figure 3a,b). This dynamic criterion helps participants avoid noise-triggered early decisions or extended deliberations [18]. An exception to this trend was identified in trial sequences without trials of extreme difficulty [13••], in which case the optimal strategy used a threshold that increased over time.

Time-varying decision criteria also arise when subjects perform tasks where information quality changes within trials (Figure 3c) [40], especially when initially weak evidence is followed by stronger evidence later in the trial. However, most studies use heuristic models to explain psychophysical data [19•,20], suggesting a need for normative model development in these contexts. Decision threshold switches have also been observed in humans performing changepoint detection tasks, whose difficulty changes from trial-to-trial [41], and in

Figure 3



Dynamic evidence quality. **(a)** Trial-to-trial two-state Markovian evidence quality switching: **(ai)** Evidence quality switches between easy (Q_{easy}) and hard (Q_{hard}) with probability P_{switch}. **(aii)** Optimal decision policies require time-varying decision thresholds. An observer who knows the evidence quality (easy or hard) uses a fixed threshold (gray traces, dashed lines) to maximize reward rate, but thresholds must vary when evidence quality is initially unknown (black trace, green gradient). **(b)** Different triangle task difficulties (from Figure 2ai): Triangles are spaced further apart in easy trials compared to hard trials. **(c)** Changes in quality within trials: **(ci)** An RDMD task in which the drift coherence increases mid-trial, providing stronger evidence later in the trial. **(cii)** The corresponding LLR increases slowly early in the trial, and more rapidly once evidence becomes stronger.

a model of value-based decisions, where the reward amounts change between trials [42]. Overall, optimal performance on tasks in which reward structure or decision difficulty changes across trials require time-varying decision criteria, and subject behavior approximates these normative assumptions.

One caveat is that extensive training or obvious across-trial changes are needed for subjects to learn optimal solutions. A meta-analysis of multiple studies showed that fixed threshold DDMs fit human behavior well when difficulty changes between trials were hard to perceive [43]. A similar conclusion holds when changes occur within trials [44]. However, when nonhuman primates are trained extensively on tasks where difficulty variations were likely difficult to perceive, they appear to learn a time-varying criterion strategy [45]. Humans also exhibit time-varying criteria in reward-free trial sequences where interrogations are interspersed with free responses [46]. Thus, when task design makes it difficult to perceive task heterogeneity or learn the optimal strategy, subjects seem to use fixed threshold criteria [43,44]. In contrast, with sufficient training [45], or when changes are easy to perceive [46], subjects can learn adaptive threshold strategies.

Questions remain about how well normative models describe subject performance when difficulty changes across or within trials. How distinct do task difficulty extremes need to be for subjects to use optimal models? No systematic study has quantified performance advantages of time-varying decision thresholds. If they do not confer a significant advantage, the added complexity of dynamic thresholds may discourage their use.

When and how are normative computations learned and achieved?

Except in simple situations, or with overtrained animals, subjects can at best approximate computations of an ideal observer [14]. Yet, the studies we reviewed suggest that subjects often learn to do so effectively. Humans appear to use a process resembling reinforcement learning to learn the structure and parameters of decision task environments [47]. Such learning tracks a gradient in reward space, and subjects adapt rapidly when the task structure changes [48]. Subjects also switch between different near-optimal models when making inferences, which may reflect continuous task structure learning [9••]. However, these learning strategies appear to rely on reward and could be noisier when feedback is probabilistic or absent. Alternatively, subjects may ignore feedback and learn from evidence accumulated within or across trials [28,46].

Strategy learning can be facilitated by using simplified models. For example, humans appear to use sampling strategies that approximate, but are simpler than, optimal inference [9••,49]. Humans also behave in ways that limit performance by, for instance, not changing their mind when faced with new evidence [50]. This confirmation bias may reflect interactions between decision and attention related systems that are difficult to train away [51]. Cognitive biases may also arise due to suboptimal applications of normative models [52]. For instance, recency bias can reflect an incorrect assumption of trial dependencies [53]. Subjects seem to continuously update latent parameters (e.g. hazard rate), perhaps assuming that these parameters are always changing [21,29].

The adaptive processes we have discussed occur on disparate timescales, and thus likely involve neural mechanisms that interact across scales. Task structure learning occurs over many sessions (days), while the volatility of the environment and other latent parameters can be learned over many trials (hours) [6••,49]. Trial-to-trial

dependencies likely require memory processes that span minutes, while within trial changes require much faster adaptation (milliseconds to seconds).

This leaves us with a number of questions: How does the brain bridge timescales to learn and implement adaptive evidence integration? This likely requires coordinating fast neural activity changes with slower changes in network architecture [8]. Studies of decision tasks in static environments suggest that a subject's belief and ultimate choice is reflected in evolving neural activity [2,3,1,54]. It is unclear whether similar processes represent adaptive evidence accumulation, and, if so, how they are modulated.

Conclusions

As the range of possible descriptive models grows with task complexity [9,49], optimal observer models provide a framework for interpreting behavioral data [5,6••,34]. However, understanding the computations subjects use on dynamic tasks, and when they depart from optimality, requires both careful comparison of models to data and comparisons between model classes [55].

While we mainly considered optimality defined by performance, model complexity may be just as important in determining the computations used by experimental subjects [56]. Complex models, while more accurate, may be difficult to learn, hard to implement, and offer little advantage over simpler ones [8,9••] Moreover, predictions of more complex models typically have higher variance, compared to the higher bias of more parsimonious models, resulting in a trade-off between the two [9••].

Invasive approaches for probing adaptive evidence accumulation are a work in progress [57,58]. However, pupillometry has been shown to reflect arousal changes linked to a mismatch between expectations and observations in dynamic environments [10,27•,59]. Large pupil sizes reflect high arousal after a perceived change, resulting in adaptive changes in evidence weighting. Thus, pupillometry may provide additional information for identifying computations underlying adaptive evidence accumulation.

Understanding how animals make decisions in volatile environments requires careful task design. Learning and implementing an adaptive evidence accumulation strategy needs to be both rewarding and sufficiently simple so subjects do not resign themselves to simpler computations [43,44]. A range of studies have now shown that mammals can learn to use adaptive decision-making strategies in dynamic 2AFC tasks [5,6••]. Building on these approaches, and using them to guide invasive studies with mammals offers promising new ways of understanding the neural computations that underlie our everyday decisions.

Conflict of interest statement

Nothing declared.

Acknowledgements

We are grateful to Joshua Gold, Alex Piet, and Nicholas Barendregt for helpful feedback. This work was supported by an NSF/NIH CRCNS grant (R01MH115557) and an NSF grant (DMS-1517629). ZPK was also supported by an NSF grant (DMS-1615737). KJ was also supported by NSF grant DBI-1707400. WRH was supported by NSF grant SES-1556325.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Gold JI, Shadlen MN: **The neural basis of decision making**. *Annu Rev Neurosci* 2007, **30**.

2. Britten KH, Shadlen MN, Newsome WT, Movshon JA: **The analysis of visual motion: a comparison of neuronal and psychophysical performance**. *J Neurosci* 1992, **12**:4745-4765.

3. Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD: **The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks**. *Psychol Rev* 2006, **113**:700.

4. Gao P *et al.*: **A theory of multineuronal dimensionality, dynamics and measurement**. *bioRxiv* 2017:214262.

5. Glaze CM, Kable JW, Gold JI: **Normative evidence accumulation in unpredictable environments**. *Elife* 2015, **4**:e08825.

6. •• Piet AT, El Hady A, Brody CD: **Rats adopt the optimal timescale for evidence integration in a dynamic environment**. *Nat Commun* 2018, **9**:4265.
Rats can learn to optimally discount evidence when deciding between two dynamically switching auditory click streams, and they adapted to changes in environmental volatility.

7. Adams RP, MacKay DJ: *Bayesian Online Changepoint Detection*. 2007 https://arxiv.org/abs/0710.3742.

8. Radillo AE, Veliz-Cuba A, Josić K, Kilpatrick ZP: **Evidence accumulation and change rate inference in dynamic environments**. *Neural Comput* 2017, **29**:1561-1610.

9. •• Glaze CM, Filipowicz AL, Kable JW, Balasubramanian V, Gold JI: **A bias-variance trade-off governs individual differences in on-line learning in an unpredictable environment**. *Nat Hum Behav* 2018, **2**:213.
Humans performing a dynamic triangles task use decision strategies that suggest a trade-off in which history-dependent adaptive strategies lead to higher choice variability. A sampling strategy best accounted for subject data.

10. Krishnamurthy K, Nassar MR, Sarode S, Gold JI: **Arousal-related adjustments of perceptual biases optimize perception in dynamic environments**. *Nat Hum Behav* 2017, **1**:0107.

11. Goldfarb S, Wong-Lin K, Schwemmer M, Leonard NE, Holmes P: **Can post-error dynamics explain sequential reaction time patterns?** *Front Psychol* 2012, **3**:213.

12. Purcell BA, Kiani R: **Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy**. *Proc Natl Acad Sci U S A* 2016, **113**:E4531-E4540.

13. •• Malhotra G, Leslie DS, Ludwig CJ, Bogacz R: **Overcoming indecision by changing the decision boundary**. *J Exp Psychol: Gen* 2017, **146**:776.
Humans' decision strategies in tasks where difficulty varies trial-to-trial are well approximated by a drift-diffusion model with time-varying decision boundaries. Subjects' deviations from this normative model did little to impact the reward rate.

14. Geisler WS: **Ideal observer analysis**. *The Visual Neurosciences* 2003, **vol 10**12.

15. Knill DC, Pouget A: **The Bayesian brain: the role of uncertainty in neural coding and computation**. *Trends Neurosci* 2004, **27**:712-719.

16. Kim TD, Kabir M, Gold JI: **Coupled decision processes update and maintain saccadic priors in a dynamic environment**. *J Neurosci* 2017:3078-3116.

17. • Veliz-Cuba A, Kilpatrick ZP, Josic K: **Stochastic models of evidence accumulation in changing environments**. *SIAM Rev* 2016, **58**:264-289.
This paper presents derivations and analyses of nonlinear stochastic models of evidence accumulation in dynamic environments for decisions between two and more alternatives. It shows how optimal evidence discounting can be implemented in a multi-population model with mutual excitation.

18. Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A: **The cost of accumulating evidence in perceptual decision making**. *J Neurosci* 2012, **32**:3612-3628.

19. • Holmes WR, Trueblood JS, Heathcote A: **A new framework for modeling decisions about changing information: the piecewise linear ballistic accumulator model**. *Cogn Psychol* 2016, **85**:1-29.
In this study, humans performed a RDMD task in which the direction of dots sometimes switched midtrial. A piecewise linear accumulator model was fit to data and demonstrated that subjects react slowly to new evidence, and that the perceived strength of post-switch evidence is influenced by pre-switch evidence strength.

20. Holmes WR, Trueblood JS: **Bayesian analysis of the piecewise diffusion decision model**. *Behav Res Methods* 2018, **50**:730-743.

21. Yu AJ, Cohen JD: **Sequential effects: superstition or rational behavior?** *Adv Neural Inf Process Syst* 2008, **21**:1873-1880.

22. Brunton BW, Botvinick MM, Brody CD: **Rats and humans can optimally accumulate evidence for decision-making**. *Science* 2013, **340**:95-98.

23. Ossmy O *et al.*: **The timescale of perceptual evidence integration can be adapted to the environment**. *Curr Biol* 2013, **23**:981-986.

24. Tavoni G, Balasubramanian V, Gold JI: **On the complexity of predictive strategies in noisy and changing environments**. *Computational and Systems Neuroscience (CoSyNe)*; *Denver, CO, March 1–4: 2018*.

25. Fernberger SW: **Interdependence of judgments within the series for the method of constant stimuli**. *J Exp Psychol* 1920, **3**:126.

26. Fründ I, Wichmann FA, Macke JH: **Quantifying the effect of intertrial dependence on perceptual decisions**. *J Vis* 2014, **14**:9.

27. • Urai AE, Braun A, Donner TH: **Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias**. *Nat Commun* 2017, **8**:14637.
Increases in pupil diameter can be used to predict choice alternations in serial decisions, providing a promising, non-invasive approach for validating theories of adaptive decision making strategies.

28. Nguyen KP, Josić K, Kilpatrick ZP: **Optimizing sequential decisions in the drift-diffusion model**. *J Math Psychol* 2019, **88**:32-47.

29. Hermoso-Mendizabal A *et al.*: **Response outcomes gate the impact of expectations on perceptual decisions**. *bioRxiv* 2018:433409.

30. Anderson N: **Effect of first-order conditional probability in two-choice learning situation**. *J Exp Psychol* 1960, **59**:73-93.

31. White CN, Poldrack RA: **Decomposing bias in different types of simple decisions**. *J Exp Psychol: Learn Mem Cogn* 2014, **40**:385.

32. Ratcliff R: **Theoretical interpretations of the speed and accuracy of positive and negative responses**. *Psychol Rev* 1985, **92**:212.

33. Diederich A, Busemeyer JR: **Modeling the effects of payoff on response bias in a perceptual discrimination task: bound-change, drift-rate-change, or two-stage-processing hypothesis**. *Percept Psychophys* 2006, **68**:194-207.

34. Urai AE, de Gee JW, Donner TH: **Choice history biases subsequent evidence accumulation**. *bioRxiv* 2018:251595.

35. Olianezhad F, Tohidi-Moghaddam M, Zabbah S, Ebrahimpour R: *Residual Information of Previous Decision Affects Evidence Accumulation in Current Decision*. 2016 https://arxiv.org/abs/1611.03965v2.

36. Braun A, Urai AE, Donner TH: **Adaptive history biases result from confidence-weighted accumulation of past choices**. *J Neurosci* 2018:2189-2217.

37. Deneve S: **Making decisions with unknown sensory reliability**. *Front Neurosci* 2012, **6**:75.

38. Zhang S, Lee MD, Vandekerckhove J, Maris G, Wagenmakers E-J: **Time-varying boundaries for diffusion models of decision making and response time**. *Front Psychol* 2014, **5**:1364.

39. Purcell BA, Kiani R: **Neural mechanisms of post-error adjustments of decision policy in parietal cortex**. *Neuron* 2016, **89**:658-671.

40. Thura D, Beauregard-Racine J, Fradet C-W, Cisek P: **Decision making by urgency gating: theory and experimental support**. *J Neurophysiol* 2012, **108**:2912-2930.

41. Johnson B, Verma R, Sun M, Hanks TD: **Characterization of decision commitment rule alterations during an auditory change detection task**. *J Neurophysiol* 2017, **118**:2526-2536.

42. Tajima S, Drugowitsch J, Pouget A: **Optimal policy for value-based decision-making**. *Nat Commun* 2016, **7**:12400.

43. Hawkins GE, Forstmann BU, Wagenmakers E-J, Ratcliff R, Brown SD: **Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making**. *J Neurosci* 2015, **35**:2476-2484.

44. Evans NJ, Hawkins GE, Boehm U, Wagenmakers E-J, Brown SD: **The computations that support simple decision-making: a comparison between the diffusion and urgency-gating models**. *Sci Rep* 2017, **7**:16433.

45. Hawkins G, Wagenmakers E, Ratcliff R, Brown S: **Discriminating evidence accumulation from urgency signals in speeded decision making**. *J Neurophysiol* 2015, **114**:40-47.

46. Palestro JJ, Weichart E, Sederberg PB, Turner BM: **Some task demands induce collapsing bounds: evidence from a behavioral analysis**. *Psychon Bull Rev* 2018:1-24.

47. Khodadadi A, Fakhari P, Busemeyer JR: **Learning to allocate limited time to decisions with different expected outcomes**. *Cogn Psychol* 2017, **95**:17-49.

48. Drugowitsch J, DeAngelis GC, Angelaki DE, Pouget A: **Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making**. *Elife* 2015, **4**:e06678.

49. Wilson RC, Nassar MR, Gold JI: **Bayesian online learning of the hazard rate in change-point problems**. *Neural Comput* 2010, **22**:2452-2476.

50. Bronfman ZZ *et al.*: **Decisions reduce sensitivity to subsequent information**. *Proc R Soc B: Biol Sci* 2015, **282**.

51. Talluri BC, Urai AE, Tsetsos K, Usher M, Donner TH: **Confirmation bias through selective overweighting of choice-consistent evidence**. *Curr Biol* 2018, **28**:3128-3135.

52. Beck JM, Ma WJ, Pitkow X, Latham PE, Pouget A: **Not noisy, just wrong: the role of suboptimal inference in behavioral variability**. *Neuron* 2012, **74**:30-39.

53. Feldman J, Hanna JF: **The structure of responses to a sequence of binary events**. *J Math Psychol* 1966, **3**:371-387.

54. Hanks T, Kiani R, Shadlen MN: **A neural mechanism of speed-accuracy tradeoff in macaque area lip**. *Elife* 2014, **3**:e02260.

55. Wu Z, Schrater P, Pitkow X: *Inverse POMDP: Inferring What You Think from What You Do*. 2018 https://arxiv.org/abs/1805.09864.

56. Bialek W, Nemenman I, Tishby N: **Predictability, complexity, and learning**. *Neural Comput* 2001, **13**:2409-2463.

57. Thura D, Cisek P: **The basal ganglia do not select reach targets but control the urgency of commitment**. *Neuron* 2017, **95**:1160-1170.

58. Akrami A, Kopec CD, Diamond ME, Brody CD: **Posterior parietal cortex represents sensory history and mediates its effects on behaviour**. *Nature* 2018, **554**:368.

59. Nassar MR *et al.*: **Rational regulation of learning dynamics by pupil-linked arousal systems**. *Nat Neurosci* 2012, **15**:1040.