Attention when you need

Anonymous authors
Paper under double-blind review

Abstract

Paying attention improves performance, but attention is metabolically costly, so how should a resource-efficient agent allocate it? We study optimal allocation strategies using a normative model of a signal detection task in which attention comes at a cost. The model reveals that optimal attention is temporally structured in one of two patterns depending on task conditions: when attention costs penalize intense focus, optimal attention ramps up as evidence for the signal accumulates, but with less prohibitive attention costs, optimal attention fluctuates rhythmically. In cases with rhythmic attention, allocation frequency increases with task parameters such as reward magnitude for successful detection, signal brevity, and signal frequency. We argue that rhythmic attention emerges naturally in an agent whose belief updates are dominated by a stable temporal prior and not by the likelihood of new sensory observations. Our results characterize the conditions under which rhythmic vs. ramping attention is optimal and offer a normative account of attentional fluctuations observed in sustained attention tasks.

1 Introduction

Good investment pays off. Across tasks, performance improves as the resources devoted to them, such as time, effort, money, and compute, increase from zero. These resources are not free, which raises the question of how an agent should allocate them to maximize performance while minimizing expenditure. In machine learning, this question motivates a large literature on conditional computation, which spends compute selectively rather than uniformly (Scardapane et al., 2024): mixture-of-experts layers route each token to a subset of parameters (Shazeer et al., 2017), early-exit and mixture-of-depths mechanisms vary how many layers an input passes through (Teerapittayanon et al., 2016; Raposo et al., 2024), and adaptive computation time lets a network decide how long to process each input (Graves, 2016; Banino et al., 2021). The question becomes harder for agents that process streams, where allocation must be decided moment to moment without knowing what is coming: a video model deciding which frames to encode (Wu et al., 2019), a recurrent network learning when to skip state updates (Campos et al., 2018), or a reinforcement learning agent that must pay to observe its environment (Nam et al., 2021; Krale et al., 2023).

We study this allocation problem for a resource that brains and machines alike must budget: attention. The term has numerous meanings across machine learning and neuroscience, from self-attention in transformers to top-down and bottom-up attention in the brain. We use it in the sense closest to hard attention in machine learning (Mnih et al., 2014) and to attention as studied in neuroscience: a controllable resource that an agent invests to improve the quality of its observations. More attention improves performance when directed at the relevant features of a task (Cohen and Maunsell, 2010; 2011; de Gee et al., 2017), but this benefit comes at a cost. In the brain, tasks requiring sustained attention over long periods cause fatigue, leading to variable performance and an overall decline as time on task increases (Gilden and Wilson, 1995; Warm et al., 2008; Fortenbaugh et al., 2017; Mackworth, 1948; Broadbent, 1971). In artificial systems, the cost is paid in compute, latency, or energy.

We formalize this trade-off with a signal detection task, in which an agent must identify an ambiguous signal amid noise and chooses, at each moment, how much to pay for a more reliable observation. We designed the task to model sustained attention in animals, with timing chosen to match an existing rodent vigilance experiment (de Gee et al., 2022).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

However, the core task demands generalize naturally to any agent that must identify sparse, partially observed signals under resource constraints, from a driver waiting for a red light to turn green, to a trader watching the market for signs to buy or sell, to an artificial agent choosing which frames or tokens warrant more expensive, more capable computation.

To find solutions to our task, we develop a normative model, which reveals that optimal attention is never constant. Instead, it follows one of two consistent within-trial patterns, depending on task parameters. In the "rhythmic" pattern, attention is non-zero only sparsely, at regular intervals. In the "ramping" pattern, attention rises as evidence for the signal accumulates and drops when there is a lack of evidence. Which pattern emerges depends on how steeply attention costs grow with its informational benefit.

This result bears on a common design choice. Fixed-stride schedules, such as encoding every kth frame of a video, are widespread heuristics in systems that process streams. Our model identifies when such a schedule is not merely convenient but optimal, when it should give way to evidence-driven allocation, and how the optimal stride should scale with the task.

Our contributions are as follows:

- We formulate sustained attention as a POMDP in which the agent pays to control the reliability of its observations, and we solve for optimal policies by value iteration over beliefs.
- We show that optimal attention allocation policies exhibit discrete jumps as a function of belief, producing two regimes, rhythmic and ramping, selected by the curvature of the attention cost.
- We show that rhythm emerges when belief updates between checks are dominated by the temporal prior rather than by new observations.
- We characterize how the rhythmic stride scales with reward, signal duration, and noise duration, and we show that optimal policies substantially outperform constant attention at matched mean attention.

2 Related Work

Sustained attention. There are several classes of attention studied in neuroscience (Ocasio, 2011; Lindsay, 2020). We focus on sustained attention, also known as vigilance: a subject's preparedness to detect infrequent and unpredictable signals over extended periods (Sarter et al., 2001). Despite the term "sustained," there is evidence that attention fluctuates between optimal and suboptimal states over time, due to factors such as arousal, mind wandering, and competing behavioral demands (Esterman et al., 2014; Esterman and Rothlein, 2019; Busch and VanRullen, 2010). These fluctuations are often interpreted as lapses. By assigning attention a cost, our work explores a setting in which attention that is suboptimal for *performance* may optimally solve the *trade-off* between the performance gained by paying attention and the price of paying it.

Attention as a costly action. Balkenius and Hulth (1999) argued that attention should be viewed as an action rather than solely a sensory process. Several works have adopted this perspective, building optimal control models that treat attention as an action choice to determine where (Bandera et al., 1996; Paletta et al., 2005; Minut and Mahadevan, 2001) and how often (Di Nocera et al., 2012) to attend. The same perspective underlies hard attention in deep learning, where a recurrent model learns through reinforcement where to direct a limited-resolution glimpse (Mnih et al., 2014). Normative accounts in economics and cognitive science likewise model attention as a resource with explicit costs, deriving behavior from the trade-off between the value of information and the price of acquiring it (Howard, 1966; Sims, 2003; Lieder and Griffiths, 2020). In models of perceptual decision-making, costly evidence accumulation determines when an agent should stop sampling and decide (Wald, 1945; Frazier and Yu, 2007; Drugowitsch et al., 2012); there, the quality of each sample is fixed, whereas our agent also chooses how much to invest in each one. The work most relevant to ours is Chebolu et al. (2022), which trained reinforcement learning agents on an abstract sustained attention task to study the trade-off between the benefits of high-quality

observations and the attentional costs involved. We construct a conceptually similar model, but we extend the analysis to characterize patterns of attention and why they emerge.

Costly observation in reinforcement learning. A line of work in reinforcement learning considers agents that must pay to observe their environment. In action-contingent noiselessly observable MDPs, the agent chooses at each step whether to pay for an exact observation of the state (Nam et al., 2021), and related work develops algorithms for this and closely related settings (Bellinger et al., 2021; Krale et al., 2023). These works aim to learn efficient policies; our aim is to characterize what optimal policies look like and why. Our formulation also differs in that observation quality is graded rather than all-or-nothing: the agent chooses how much to invest in each observation, not only whether to observe. Graded, noisy observations are what make the ramping regime possible. When a single observation resolves all uncertainty, there is no evidence to accumulate.

Allocating computation over streams. Systems that process streams must decide when to spend computation, and two families of strategies recur. The first is fixed-stride sampling; video recognition models, for example, commonly process frames sampled at uniform intervals (Wang et al., 2016; Feichtenhofer et al., 2019). The second adapts allocation to the input. Adaptive frame selection policies decide which frames merit processing (Wu et al., 2019; Korbar et al., 2019), recurrent networks can learn to skip state updates under a computational budget (Campos et al., 2018), and cascades escalate from cheap to expensive models only when earlier stages find the input promising or uncertain, from the classic detection cascade of Viola and Jones (2001) to cascades of large language models (Chen et al., 2023). The same dichotomy appears in control theory as periodic versus event-triggered sampling, where event-triggered schemes can achieve lower estimation error at the same average sampling rate (Åström and Bernhardsson, 2002). Our two regimes map onto these families. Rhythmic policies behave as fixed-stride schedules, and ramping policies escalate investment in stages as evidence accumulates, much like a cascade. In machine learning, the choice between these strategies is usually made empirically. Our model instead derives both as optimal solutions to a single problem and identifies the conditions that determine which is optimal: chiefly the curvature of the attention cost, and whether the agent's beliefs between observations are driven by its prior or by what it observes.

3 Task and Model

3.1 Intuition

We consider an agent in a world whose state is latent: the agent cannot see the state directly and must infer it from observations. It therefore maintains a *belief*, a posterior probability distribution over world states, which it updates as it *observes* the world. In signal detection, the relevant distinction is binary: the world is in either a *signal* or a *noise* (non-signal) state, and the agent must report when a signal is present. To define an optimization objective, we assign values to the possible outcomes. Correctly detected signals ("hits") contribute positively to reward, representing value gained by success at the task. Undetected signals ("misses") forgo this value, and detections reported while the world is in a noise state ("false alarms") incur a penalty, representing value lost by failure at the task.

3.2 Implementation

We model the task as a partially observable Markov decision process (POMDP; Fig 1A) (Kaelbling et al., 1998) in discrete time. At each step, the agent decides how much to attend and whether or not to report, or "*commit*," to a signal's presence.

We construct task trials as a sequence beginning in a start state. The start state transitions with 100% probability to the noise state in the next step, and at each step thereafter, the noise state has a small, fixed probability q of transitioning to the first signal state, so noise durations are geometrically distributed. The number of signal states m can be varied to change the distribution of signal durations. Each signal state transitions to the next (or, for the final signal state, out of the signal period) with probability p, so signal durations follow a negative binomial distribution with mean $\frac{m}{p}$. For $m = 1$, this distribution is geometric, like

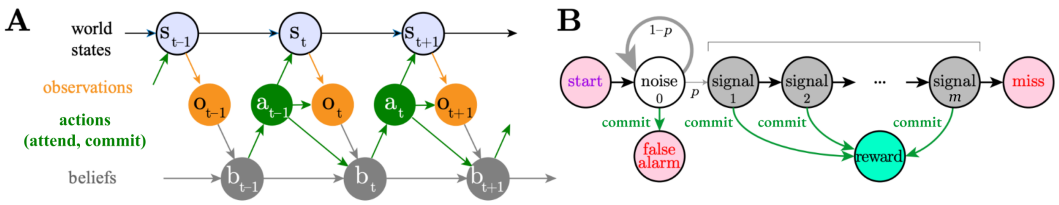


Figure 1: Formal task structure. (**A**) In this POMDP, the agent constructs beliefs b_t that reflect inferences about the latent world state s_t using observations o_t controlled by the "attend" action a_t. The agent uses these beliefs to plan "commit" actions that may obtain rewards. (**B**) Specific world states and transitions in the finite state machine modeling this task. The noise state has a geometric duration. The signal lasts throughout a sequence of m signal states, with a tighter distribution the higher m is. Commit actions produce rewards when taken during the signal period and otherwise lead to a false alarm and penalty.

the distribution of noise durations; as m increases with the mean held fixed, it concentrates around the mean. Task trials end in one of three ways. If the agent does not commit before the world transitions out of the final signal state, the trial ends as a miss. If the agent commits while the world is in the noise state, the trial is a false alarm. If the agent commits during any signal state, the trial is a hit (Fig. 1B).

The agent maintains a belief b_t, a probability vector over world states, which it updates exactly at each step by Bayesian filtering (the forward recursion of a hidden Markov model), combining its previous belief, the known transition dynamics, and the likelihood of its current observation. Attention increases the reliability of the agent's next observation, with higher attention yielding higher reliability but also higher cost. Specifically, observations are binary, either signal or noise, and for attention $a \geq 0$, the observation matches the true state (signal or noise) with probability $\sigma(a) = \frac{1}{1+e^{-a}}$. At $a = 0$, observations are correct with probability $\frac{1}{2}$ and therefore carry no information. The cost of attention is $C(a) = c_1 a^{c_2}$, with scale $c_1 > 0$ and exponent $c_2 > 0$.

The agent maximizes expected discounted reward, combining attention costs, the hit reward, and the false alarm penalty. Because beliefs are sufficient statistics for the agent's history, the POMDP can be recast as a fully observed MDP over beliefs (Kaelbling et al., 1998), which we solve by value iteration on a discretized belief space (Appendix A.2). The resulting policy maps each belief to an attention level and a commit rule.

4 Results

4.1 Rhythmic and ramping attention emerge in optimal policies.

With one signal state ($m = 1$), the belief space is one-dimensional, so we can visualize optimal policies in full (Fig. 2A). These policies exhibit *jumps* in optimal attention across signal belief. At low beliefs, no attention is paid because the signal is unlikely to be present; at intermediate beliefs, attention is allocated; and at high beliefs, attention again drops to zero because the signal is nearly certain, so attending to collect more information is unnecessary. Although policies become harder to visualize as m grows, these jumps persist for higher numbers of signal states; we visualize up to $m = 4$ in Fig. 2B.

Together with the commit rule, these jumps partition belief space into regions of qualitatively different behavior. Where signal belief is low, the chance that the signal is present is too low to justify engaging, so the agent neither attends nor commits: it *waits*. Where signal belief is high, attention is an added cost that would likely only confirm what the agent already believes, so it commits without attending. In the regions where the agent does attend, its commit rule takes one of two forms. Either it does not commit regardless of what it observes, *gathering information* that moves its belief without ending the trial, or it commits only if it observes signal, using the observation to *decide*. Transitions between these behaviors occur at jumps in the policy. If there is a single jump, the policy switches directly from waiting

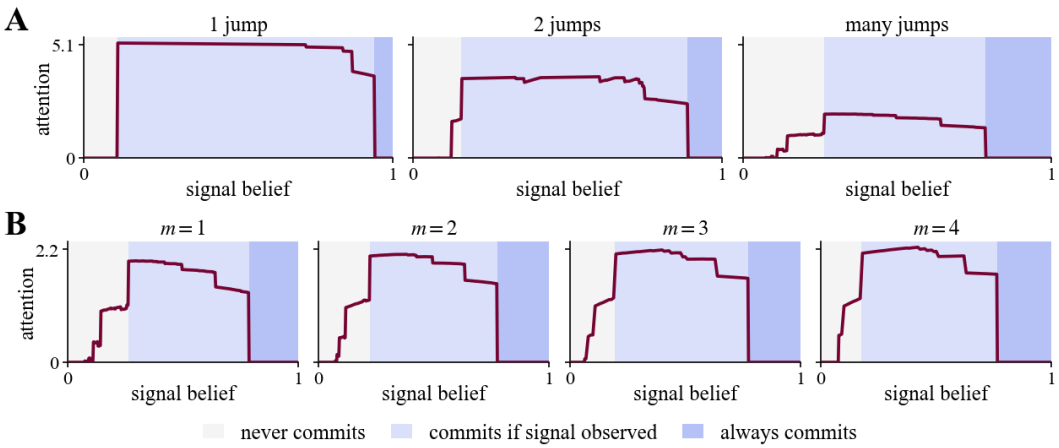


Figure 2: Attention policies over lower-dimension belief spaces. Attention over signal belief is shown in maroon; the background color shows if this signal belief is a level where the agent will not commit (light grey), will commit only if it observes signal (light blue), or will always commit (darker blue). (**A**) When we model a geometric distribution for signal duration using a single signal state ($m = 1$), belief is one-dimensional, so we may visualize full policies as a level of attention over total signal belief. We show a single-jump (left), 2-jump (middle) and many-jump (right) example here. (**B**) We can visualize a *slice* of policies with $m > 1$ by fixing all but one signal state at 0. Here, we show an example of many-jumps policies where signal belief is concentrated in the final signal state for, from left to right, 1, 2, 3, and 4 signal states. We show and discuss more examples in the Appendix A.4.

to deciding. If there are several, the policy switches from waiting to information gathering at the first jump, then from information gathering to deciding at the jump to maximum attention. The number of jumps is closely tied to the two patterns of attention allocation we observe across trials: rhythmic and ramping.

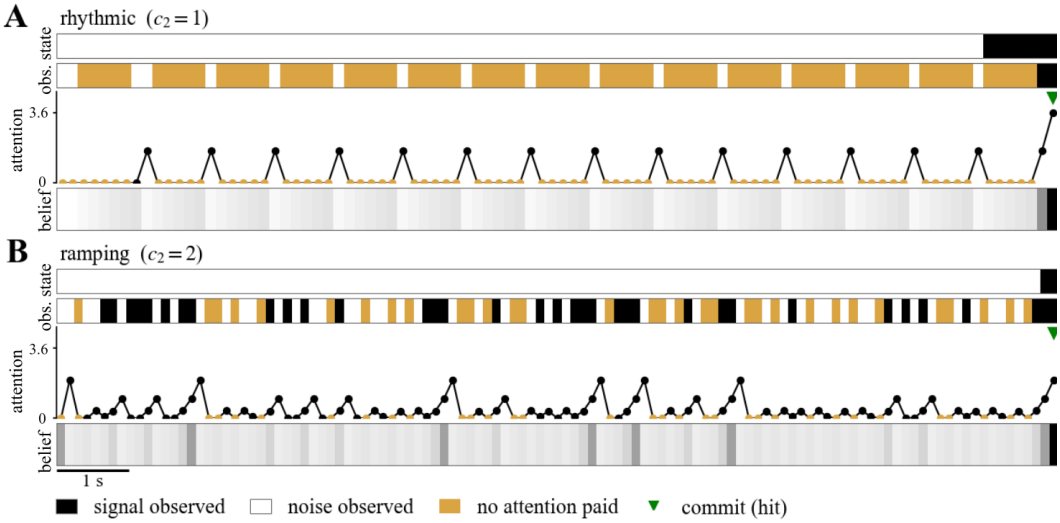


Figure 3: Illustration of agents' interactions with simulated trials. Trial phase (noise: white, signal: black), observations (amber: no attention; white/black: noise/signal observation), belief, and commit action (green triangle). (**A**) agent with a rhythmic policy over a single trial of approximately 100 steps. (**B**) Same as **A** but with a ramping policy.

Rhythmic policies (Fig. 3A)**.** At the start of a trial, when signal belief is 0, the agent pays no attention and waits. Because an observation with 50% accuracy carries no information,

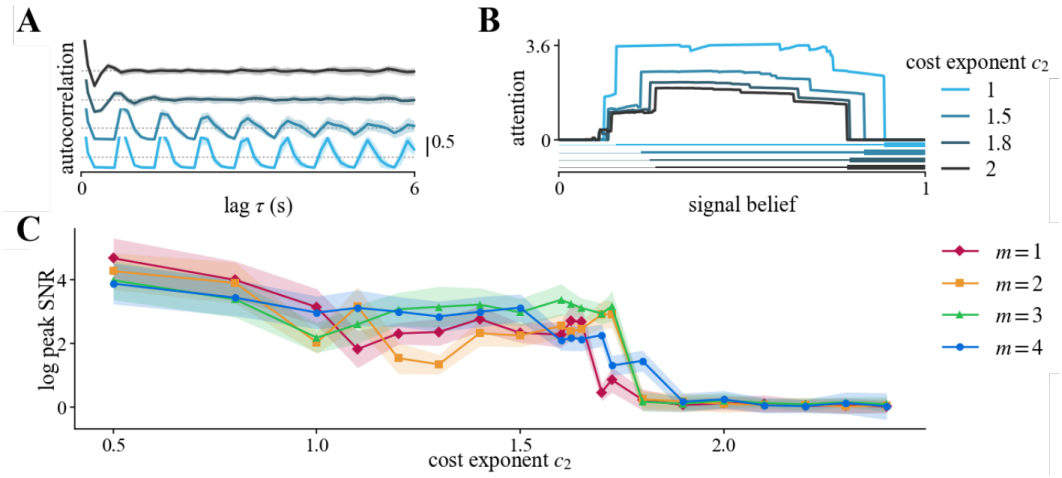


Figure 4: Rhythmicity and policy over attention cost exponent c_2. (**A**) Autocorrelation of attention allocations within a session for optimal policies under the attention cost function $C(a) = 0.2a^{c_2}$, for different c_2. Rhythmic policies show periodic spikes in autocorrelation; ramping policies are flat. (**B**) The corresponding policies producing the autocorrelations in **A**. The lines below show commit behavior (thin: never; medium: when observing signal; thick: always). (**C**) The rhythmicity metric plotted over c_2 for 1, 2, 3, and 4 signal states. Rhythmicity drops sharply at approximately $c_2 = 1.7$: below this value, policies are rhythmic, and above it, policies ramp.

belief during disengagement evolves deterministically under the state transition prior alone. Driven by this prior, the agent's signal belief increases over time until the signal is probable enough to be worth checking. In rhythmic policies, this check is a large installment of attention. The agent is confident in the accuracy of the resulting observation, so if it observes noise, its signal belief drops to near zero, resetting the cycle of wait-then-check. After the first check, the belief just before each check and the belief just after a noise observation are the same every cycle, so the interval between checks is identical every cycle: during the noise period, a rhythmic policy behaves as a fixed-stride schedule, with a stride determined by the task parameters (Section 4.4). The cycle repeats until a check yields a signal observation, which raises signal belief substantially. Rhythmic policies often have only one jump because, with highly accurate observations, a single signal observation suffices to confirm the signal's presence with high confidence.

Ramping policies (Fig. 3B)**.** A ramping policy begins the same way, but its first check is a low-attention one. The resulting observation is less accurate and moves belief less, so a signal observation does not justify committing. It does, however, make the signal more probable, which justifies more attention in the next step. Each further signal observation raises attention again, until the agent reaches its highest attention level and uses the observation to decide, or until it observes noise. Because low-attention observations are unreliable, a noise observation lowers belief without resetting it to 0. Without a common reset point, and with a stochastic observation shaping belief at every stage, the regularity of rhythmic policies gives way to irregular ramps. Like a cascade, the agent escalates only while the evidence continues to favor the signal.

Ramping policies have many jumps, one per increment of attention. Some two-jump policies are nonetheless rhythmic: when the first attention level already yields very accurate observations, the agent rarely reaches the second unless the signal is present, so its behavior during noise remains rhythmic.

4.2 The rhythmicity of a policy depends on its attention cost function.

The exponent c_2 of the attention cost function $C(a) = c_1 a^{c_2}$ determines whether the optimal policy is rhythmic or ramping. When the attention cost is linear ($c_2 = 1$), optimal attention

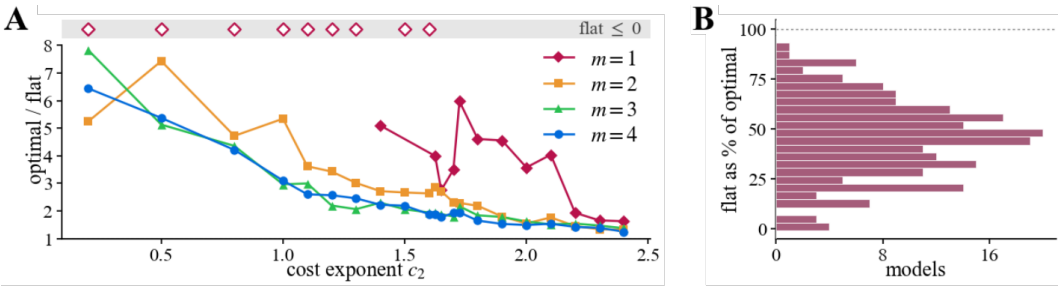


Figure 5: Comparison of optimal vs. constant-attention policies. (**A**) The ratio, over c_2, of mean net reward per step for optimal policies to that of a constant-attention control with the same mean attention. To keep the rest of the plot to scale, we show points where the optimal policy is ten times better, or where the constant policy has a negative net reward, on a grey bar above. (**B**) The inverse comparison: a histogram, over all solved policies, of how much of an optimal policy's net reward a constant policy with the same mean attention captures. For visual clarity, we show only comparisons where the constant policy achieved non-negative mean reward.

is rhythmic; when it is quadratic ($c_2 = 2$), optimal attention ramps. The change is visible both in the autocorrelation of attention within trials (Fig. 4A) and in trial visualizations (Fig. 3), and we quantify it with a rhythmicity metric (Appendix A.3) to characterize the relationship between c_2 and rhythmicity (Fig. 4C). The transition is sharp: rhythmicity drops abruptly at approximately $c_2 = 1.7$.

When costs are linear or sublinear ($c_2 \leq 1$), a large installment of attention costs no more per unit than a small one, so the agent saves by concentrating attention into sparse, decisive checks, which produces rhythmic policies. When costs are convex ($c_2 > 1$), the agent saves by spreading attention into smaller installments, but small installments yield less accurate observations. For mildly convex costs, the accuracy of a single large check remains worth its premium, and policies stay rhythmic. Beyond $c_2 \approx 1.7$, the premium becomes too steep, and ramping policies resolve the cost-accuracy trade-off differently: they keep attention low most of the time, raise it only at critical moments when the signal may be present, and raise it in small increments. By the time an agent with a ramping policy commits, it has observed signal on multiple consecutive occasions, which together constitute stronger evidence than any single low-attention observation.

4.3 Structured attention outperforms constant attention.

Fluctuating attention is not merely a feature of optimal policies but a substantial source of their value. We compared each optimal policy with a control that allocates constant attention equal to the optimal policy's mean attention.[1] The size of the gain varies with c_2 and the other task parameters, but across all parameters tested, 12.45% of constant-attention controls had negative net reward: allocating the same average attention uniformly rather than selectively yielded a net loss. Among controls with non-negative reward, even the lowest-resolution optimal policies, with 3 and 4 signal states at discretizations of 101 belief levels per state, earned a median of 1.95 and 1.99 times the controls' net reward, respectively. The fraction of optimal net reward captured by the controls ranged from 0.7% to 89.6%, with a median of 46.4% and a mean of 45.4%, and 59.3% of controls earned less than half the net reward of the corresponding optimal policy (Fig. 5). In the language of conditional computation, spending the same budget uniformly forfeits, in the median case, more than half of the achievable net reward.

[1]For $c_2 \neq 1$, matching mean attention does not match mean cost. Appendix A.6 compares optimal policies with a control matched on mean attention cost instead, with comparable results.

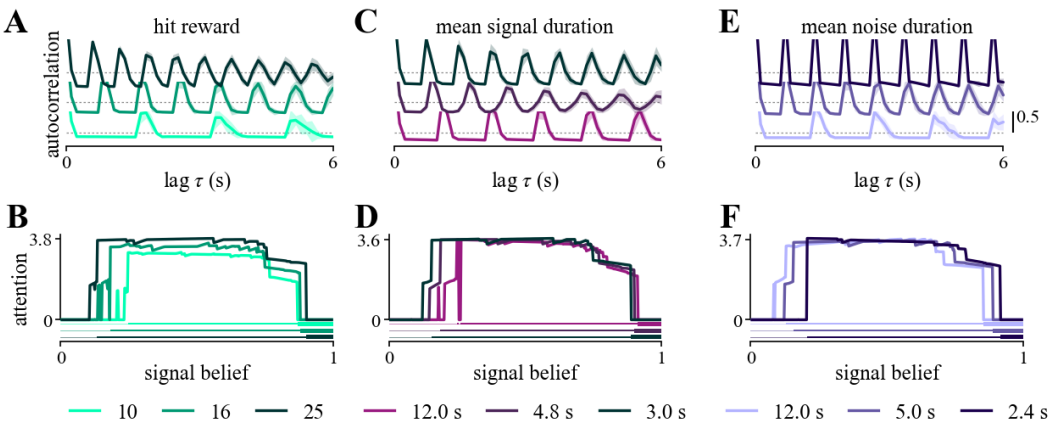


Figure 6: Autocorrelations and policies over model parameters. All models shown here used a single signal state, allowing us to fully visualize their policies. Autocorrelations for models with more signal states are included in Appendix A.5. (**A**) Autocorrelation of attention allocations within a session for optimal policies under different hit reward values. (**B**) The corresponding policies producing the autocorrelations in **A**. (**C**) and (**D**): the same for mean signal duration. (**E**) and (**F**): the same for mean noise duration.

4.4 In rhythmic policies, attention frequency depends on other task parameters.

In a rhythmic policy, the stride between checks is itself a function of the task. To characterize this dependence, we regressed the peak frequency of the attention autocorrelation's power spectrum on all swept model parameters jointly by ordinary least squares ($n = 181$). We restricted the analysis to models with a rhythmicity metric above 2.0, since the peak frequency of an aperiodic policy reflects only noise, and we report standardized coefficients so that parameters with different swept ranges are comparable.

Effect of signal duration. The shorter the mean signal duration, the more frequently the agent attends ($\beta = -.234$, $p < .001$), because it cannot afford to miss a brief signal window (Fig. 6C). The effect is subtle with a single signal state because signal durations are then broadly distributed: a long mean duration does not rule out a brief signal. The effect becomes more pronounced as the number of signal states increases, since this narrows the distribution of signal durations and makes short signals less likely (Appendix A.5).

Effect of noise duration. Similarly, the shorter the mean noise duration, the more frequently the agent attends ($\beta = -.211$, $p < .001$; Fig. 6E). A short mean noise duration means a high probability of transitioning to the signal at each step, so signal belief climbs quickly to the level at which attention pays off. A long mean noise duration means a slow climb and, accordingly, less frequent checks.

Effect of hit reward. Higher hit rewards increase both the frequency of an agent's attention allocations ($\beta = .475$, $p < .001$) and the average attention per allocation (Fig. 6A-B). Greater potential returns justify the cost of more frequent high-quality observations.

5 Discussion

When is a fixed stride optimal? Our results identify two conditions under which fixed-stride schedules, a common heuristic for allocating computation over streams, are optimal: the cost of attention must be close to linear, and between checks the agent's belief must be driven by its prior rather than by informative observations. When both hold, each decisive check that finds no signal leaves the same belief, so the interval never varies. When either fails, a fixed stride is no longer optimal: steep costs favor graded escalation, and informative observations make the timing of attention depend on what the agent sees. Even when a

fixed stride is optimal, the stride should be set by the task rather than treated as a generic hyperparameter: checks should be more frequent when detection is more valuable and when events are brief or frequent (Section 4.4).

Cost curvature and cascades. Ramping policies share their structure with cascades, which escalate from cheap to expensive computation only while evidence remains favorable (Viola and Jones, 2001; Chen et al., 2023). Our model suggests that the choice between a fixed stride and a cascade-like strategy should depend on how steeply the cost of processing rises with the reliability it buys. This is a property of the whole system, not of compute alone. If reliability shows diminishing returns in compute, as it often does, then each additional increment of reliability costs more than the last, and the effective cost function is convex even when compute is billed linearly. In such settings, our results favor graded escalation over periodic, all-or-nothing checks. Testing this prediction in a deployed system would require measuring its cost-reliability curve, which we leave to future work.

Temporal structure in correlates of attention. Our theory predicts how neural and physiological correlates of attention could change with task parameters, including rhythmic allocations or attentional ramps. Several studies have found rhythmic structure in biological attention. In vision, attention is deployed periodically even to a single target (VanRullen et al., 2007), facilitates perception rhythmically (Busch and VanRullen, 2010), is intrinsically rhythmic in its top-down form (Helfrich et al., 2018), and cycles faster under greater attentional demand (Merholz et al., 2022). In audition, performance peaks at a specific beat frequency, suggesting a natural sampling rate for periodic attention (Zalta et al., 2020), and temporal attention can be allocated selectively for short-term consolidation (Shen and Alain, 2011).

Limitations and future work. Observations in our model are binary, and attention can change instantaneously at no cost; richer observations or switching costs (Chebolu et al., 2022) may yield different dynamics. Fatigue could be added as a gradual increase in c_1. Value iteration makes our results exact up to discretization, but the belief space grows with m, limiting us to $m \leq 4$. In preliminary checks, point-based POMDP solvers and an independent value iteration implementation recovered both regimes, whereas deep reinforcement learning agents did not do so reliably. We suspect this reflects the approximation error of learned policies rather than the regimes themselves, but whether learned policies can recover them remains an open question. Finally, comparing optimal policies with other heuristics, such as the best fixed-stride schedule in the ramping regime, would quantify the cost of choosing the wrong strategy.

Broader impacts. This work contributes to the theoretical understanding of how biological and artificial agents should allocate costly attention. In neuroscience, it offers a normative framework for interpreting fluctuations in attention. In machine learning, it may inform the design of systems that must process large or continuous inputs under computational budgets. These systems face an equivalent trade-off: applying maximum processing everywhere is prohibitively expensive, but sparse processing risks missing relevant content. Our framework predicts that the optimal strategy in such settings depends on the cost structure of the processing pipeline.

6 Conclusion

We proposed a normative model of sustained attention with explicit attention costs and solved it by value iteration. Optimal agents either attend rhythmically, in sparse, decisive checks at regular intervals, or ramp attention as evidence accumulates, depending chiefly on how steeply attention costs grow. These results offer a normative account of attention fluctuations in sustained attention tasks and a basis for deciding when a system that allocates computation over a stream should use a fixed stride and when it should escalate adaptively.

AI, Ethics, and Reproducibility Statements

AI use statement. In this work, we used generative AI tools to revise the manuscript, to identify related literature in machine learning, and to improve code. All AI-modified text was reviewed and edited by the authors, and all suggested references were checked by the authors against the original sources. In the code, we used AI tools to optimize the value iteration algorithm for shorter run-times and to augment the code we wrote to generate figures—re-organizing the code for more modularity and modifying the code to reformat figures.

We did not use generative AI tools to formulate our model, to analyze our results, to prove or assist in proving any mathematical claims, or to propose hypotheses/research questions.

Ethics statement. We do not believe the paper violates any part of the ICLR Code of Ethics.

Reproducibility statements; We plan to publicly release code for building, training, simulating, and assessing the models on GitHub.

References

Karl Johan Åström and Bo M. Bernhardsson. Comparison of Riemann and Lebesgue sampling for first order stochastic systems. In *Proceedings of the 41st IEEE Conference on Decision and Control*, 2002.

Christian Balkenius and Nils Hulth. Attention as selection-for-action: a scheme for active perception. In *1999 Third European Workshop on Advanced Mobile Robots (Eurobot'99). Proceedings (Cat. No. 99EX355)*, pages 113–119. IEEE, 1999.

Cesar Bandera, Francisco J Vico, José Manuel Bravo, Mance E Harmon, and Leemon C Baird. Residual q-learning applied to visual attention. In *ICML*, pages 20–27, 1996.

Andrea Banino, Jan Balaguer, and Charles Blundell. PonderNet: Learning to ponder. *arXiv preprint arXiv:2107.05407*, 2021.

Colin Bellinger, Rory Coles, Mark Crowley, and Isaac Tamblyn. Active measure reinforcement learning for observation cost minimization. In *Proceedings of the Canadian Conference on Artificial Intelligence*, 2021.

Donald Eric Broadbent. Decision and stress. 1971.

Niko A Busch and Rufin VanRullen. Spontaneous eeg oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences*, 107(37):16048–16053, 2010.

Víctor Campos, Brendan Jou, Xavier Giró-i Nieto, Jordi Torres, and Shih-Fu Chang. Skip RNN: Learning to skip state updates in recurrent neural networks. In *International Conference on Learning Representations*, 2018.

Sahiti Chebolu, Peter Dayan, and Kevin Lloyd. Vigilance, arousal, and acetylcholine: Optimal control of attention in a simple detection task. *PLOS Computational Biology*, 18 (10):e1010642, 2022.

Lingjiao Chen, Matei Zaharia, and James Zou. FrugalGPT: How to use large language models while reducing cost and improving performance. *arXiv preprint arXiv:2305.05176*, 2023.

Marlene R Cohen and John HR Maunsell. A neuronal population measure of attention predicts behavioral performance on individual trials. *Journal of Neuroscience*, 30(45): 15241–15253, 2010.

Marlene R Cohen and John HR Maunsell. When attention wanders: how uncontrolled fluctuations in attention affect performance. *Journal of Neuroscience*, 31(44):15802–15806, 2011.

Jan Willem de Gee, Olympia Colizoli, Niels A Kloosterman, Tomas Knapen, Sander Nieuwenhuis, and Tobias H Donner. Dynamic modulation of decision biases by brainstem arousal systems. *elife*, 6:e23232, 2017.

JW de Gee, Z Mridha, M Hudson, Y Shi, H Ramsaywak, S Smith, N Karediya, M Thompson, K Jaspe, H Jiang, et al. Strategic stabilization of arousal boosts sustained attention. *bioRxiv*, pages 2022–03, 2022.

Dario Di Nocera, Alberto Finzi, Silvia Rossi, and Mariacarla Staffa. Attentional action selection using reinforcement learning. In *From Animals to Animats 12: 12th International Conference on Simulation of Adaptive Behavior, SAB 2012, Odense, Denmark, August 27-30, 2012. Proceedings 12*, pages 371–380. Springer, 2012.

Jan Drugowitsch, Rubén Moreno-Bote, Anne K. Churchland, Michael N. Shadlen, and Alexandre Pouget. The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11):3612–3628, 2012.

Michael Esterman and David Rothlein. Models of sustained attention. *Current opinion in psychology*, 29:174–180, 2019.

Michael Esterman, Monica D Rosenberg, and Sarah K Noonan. Intrinsic fluctuations in sustained attention and distractor processing. *Journal of Neuroscience*, 34(5):1724–1730, 2014.

Christoph Feichtenhofer, Haoqi Fan, Jitendra Malik, and Kaiming He. SlowFast networks for video recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6202–6211, 2019.

Francesca C Fortenbaugh, Joseph DeGutis, and Michael Esterman. Recent theoretical, neural, and clinical advances in sustained attention research. *Annals of the New York Academy of Sciences*, 1396(1):70–91, 2017.

Peter Frazier and Angela J. Yu. Sequential hypothesis testing under stochastic deadlines. In *Advances in Neural Information Processing Systems*, volume 20, 2007.

David L Gilden and Stephanie Gray Wilson. On the nature of streaks in signal detection. *Cognitive psychology*, 28(1):17–64, 1995.

Alex Graves. Adaptive computation time for recurrent neural networks. *arXiv preprint arXiv:1603.08983*, 2016.

Randolph F Helfrich, Ian C Fiebelkorn, Sara M Szczepanski, Jack J Lin, Josef Parvizi, Robert T Knight, and Sabine Kastner. Neural mechanisms of sustained attention are rhythmic. *Neuron*, 99(4):854–865, 2018.

Ronald A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966.

Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

Bruno Korbar, Du Tran, and Lorenzo Torresani. SCSampler: Sampling salient clips from video for efficient action recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

Merlijn Krale, Thiago D. Simão, and Nils Jansen. Act-then-measure: Reinforcement learning for partially observable environments with active measuring. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS)*, volume 33, 2023.

Falk Lieder and Thomas L. Griffiths. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1, 2020.

Grace W Lindsay. Attention in psychology, neuroscience, and machine learning. *Frontiers in computational neuroscience*, 14:516985, 2020.

Norman H Mackworth. The breakdown of vigilance during prolonged visual search. *Quarterly journal of experimental psychology*, 1(1):6–21, 1948.

Garance Merholz, Laetitia Grabot, Rufin VanRullen, and Laura Dugué. Periodic attention operates faster during more complex visual search. *Scientific Reports*, 12(1):6688, 2022.

Silviu Minut and Sridhar Mahadevan. A reinforcement learning model of selective visual attention. In *Proceedings of the fifth international conference on Autonomous agents*, pages 457–464, 2001.

Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. Recurrent models of visual attention. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

Hyunji Alex Nam, Scott Fleming, and Emma Brunskill. Reinforcement learning with state observation costs in action-contingent noiselessly observable Markov decision processes. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

William Ocasio. Attention to attention. *Organization science*, 22(5):1286–1296, 2011.

Lucas Paletta, Gerald Fritz, and Christin Seifert. Q-learning of sequential attention for visual object recognition from informative local descriptors. In *Proceedings of the 22nd international conference on Machine learning*, pages 649–656, 2005.

David Raposo, Sam Ritter, Blake Richards, Timothy Lillicrap, Peter Conway Humphreys, and Adam Santoro. Mixture-of-depths: Dynamically allocating compute in transformer-based language models. *arXiv preprint arXiv:2404.02258*, 2024.

Martin Sarter, Ben Givens, and John P Bruno. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain research reviews*, 35(2):146–160, 2001.

Simone Scardapane, Alessandro Baiocchi, Alessio Devoto, Valerio Marsocci, Pasquale Minervini, and Jary Pomponi. Conditional computation in neural networks: Principles and research trends. *Intelligenza Artificiale*, 18(1):175–190, 2024. ISSN 2211-0097. doi: 10.3233/ia-240035. URL `http://dx.doi.org/10.3233/IA-240035`.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *International Conference on Learning Representations*, 2017.

Dawei Shen and Claude Alain. Temporal attention facilitates short-term consolidation during a rapid serial auditory presentation task. *Experimental brain research*, 215:285–292, 2011.

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003.

Surat Teerapittayanon, Bradley McDanel, and H. T. Kung. BranchyNet: Fast inference via early exiting from deep neural networks. In *International Conference on Pattern Recognition (ICPR)*, pages 2464–2469, 2016.

Rufin VanRullen, Thomas Carlson, and Patrick Cavanagh. The blinking spotlight of attention. *Proceedings of the National Academy of Sciences*, 104(49):19204–19209, 2007.

Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–511–I–518, 2001.

Abraham Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

Limin Wang, Yuanjun Xiong, Zhe Wang, Yu Qiao, Dahua Lin, Xiaoou Tang, and Luc Van Gool. Temporal segment networks: Towards good practices for deep action recognition. In *European Conference on Computer Vision (ECCV)*, pages 20–36, 2016.

Joel S Warm, Raja Parasuraman, and Gerald Matthews. Vigilance requires hard mental work and is stressful. *Human factors*, 50(3):433–441, 2008.

Zuxuan Wu, Caiming Xiong, Chih-Yao Ma, Richard Socher, and Larry S. Davis. AdaFrame: Adaptive frame selection for fast video recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1278–1287, 2019.

Arnaud Zalta, Spase Petkoski, and Benjamin Morillon. Natural rhythms of periodic temporal attention. *Nature communications*, 11(1):1051, 2020.

A Appendix

A.1 Signal Detection Model Defaults

The default probability q of transitioning from the noise state to the first signal state was 2.4%, for a mean noise duration of approximately 42 time steps. The default mean signal duration was 25 time steps, so each signal state transitions out with probability $p = \frac{m}{25}$. For the default attention cost function, we used $c_1 = 0.2$ and $c_2 = 1.0$. For the default outcome rewards, we used a hit reward of 20, a commit cost of 5, and no explicit miss or false alarm penalty. As a result, the effective hit reward is 15, the difference between the reward collected for a correct commit and the commit cost, while the false alarm penalty is 5 and a miss leads to neither a reward nor a penalty. Our default temporal discount factor was 0.995, and we used an inter-trial interval of 16.67 to 25 time steps, which encouraged agents to perform well within the current trial without neglecting future trials. Rhythmic and ramping policies are optimal both with and without this inter-trial interval.

We took inspiration for these task parameters from de Gee et al. (2022), in which mice performed a sustained attention task for many consecutive trials in each session. With time steps of 120 milliseconds, many of our default parameters match the timing of their experimental task, in which noise durations were exponentially distributed with a mean of 5 seconds, signals lasted 3 seconds, and inter-trial intervals ranged from 2 to 3 seconds. We used different discretizations for different values of m. For $m = 1$, we ran value iteration on 10001 evenly spaced signal beliefs. For $m = 2$, we used every combination of 1001 probability levels in the first signal state and 1001 in the second, excluding pairs whose sum exceeded 1. For $m = 3$ and $m = 4$, we used 101 probability levels per signal state.

A.2 Signal Detection Model Solution Through Value Iteration

Because the belief b is a sufficient statistic for the agent's history, we solve the POMDP as a fully observed MDP over beliefs. Within a step, the agent chooses an attention level $a \geq 0$ and a commit rule $\kappa : \{\text{noise}, \text{signal}\} \to \{0, 1\}$ specifying whether to commit after each possible observation. The optimal value function satisfies

$$V(b) = \max_{a \geq 0,\, \kappa} \Big\{ -C(a) + \sum_{o} P(o \mid b, a) \Big[\kappa(o)\, R(b^o) + \big(1 - \kappa(o)\big)\, \gamma\, V\big(\tau(b^o)\big) \Big] \Big\}, \qquad (1)$$

where b^o is the posterior belief after observing o, τ propagates a belief through the transition dynamics, and $R(b^o)$ is the expected reward for committing under belief b^o: the hit reward weighted by the posterior probability of the signal, minus the commit cost. Committing additionally yields the discounted value of subsequent trials after the inter-trial interval, which we omit from Eq. 1 for clarity.

We solve Eq. 1 by value iteration on the belief grid described above. For any belief at the start of a step, the expected reward within that step is an analytic function of attention, which we maximize to obtain the optimal attention and its value. In each subsequent iteration, the value function from the previous iteration serves as the discounted continuation value, so that after n iterations the solution looks n steps into the future. We iterate either until convergence, when optimal attention is identical between steps, or until $n = 919$, when looking further into the future confers almost no additional benefit due to temporal discounting ($0.995^{919} < 0.01$). For off-grid beliefs, optimal attention and expected value are computed from the grid via linear interpolation.

A.3 Rhythmicity Metric

To analyze the relationship between task parameters and the rhythmicity of the resulting policies, we use a numerical metric for rhythmicity. We compute the within-trial autocorrelation of attention from trials simulated under the optimal policy and take the log of the signal-to-noise ratio at the peak of the autocorrelation's power spectral density. A sharp spectral peak yields a higher log ratio and indicates consistent rhythmic structure; a broad, featureless spectrum indicates aperiodic or ramping allocation.

A.4 Other policy visualizations

In addition to one-dimensional policy visualizations, which plot optimal attention on the y-axis against signal belief on the x-axis, we can plot two-dimensional visualizations that show optimal attention as color over belief in two signal states, one on each axis (Fig. 7). Attention jumps are evident in these visualizations as well.

In the main text, we focus on policy slices—in which we vary only total signal belief to show a slice of a policy over a higher-dimensional belief space—where total signal belief is concentrated in the last signal state. We made this choice because it yields the cleanest visualizations and thus conveys the consistency of attention jumps most efficiently. However, we also visualized policy slices where signal belief is concentrated in the first signal state (Fig. 8A) and where signal belief is spread uniformly across all signal states (Fig. 8B). These slices all exhibit jump-like behavior, but the overall structure of the higher-dimensional policy is less cleanly discernible. For instance, one can easily see bands of different attention levels in the $m = 2$ two-dimensional visualization (Fig. 7, left), while the policy slice concentrated on the first signal state, for this exact same policy, appears erratic (Fig. 8A; $m = 2$).

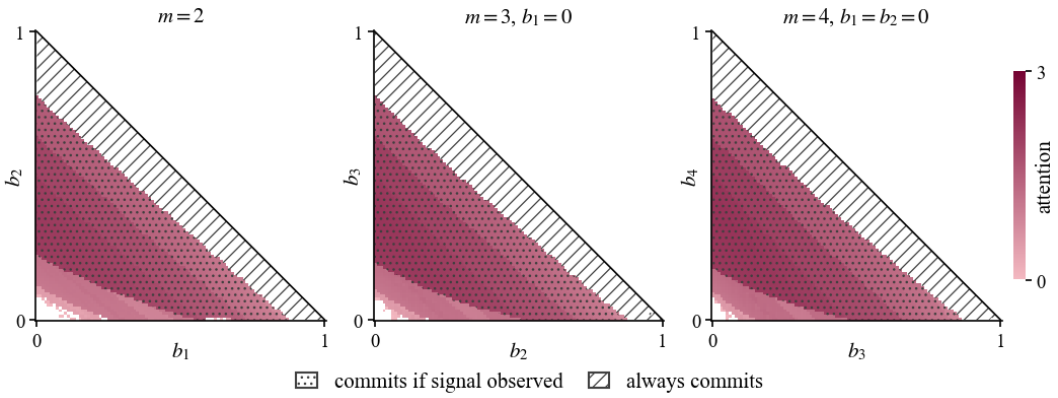


Figure 7: We may also visualize policies over two dimensions, allowing a look at how attention should be allocated over belief distributed in two signal states. For $m = 2$, this covers the full policy space. For $m > 2$, this enables visualizations of the policy over belief in two of the signal states while the others are fixed at 0. The same "jumps" in optimal attention can be observed here in the diagonal bands.

We postulate that policy slices appear more erratic for signal belief concentrated in earlier signal states for the following reason. When signal belief is concentrated in earlier states, there are more options for optimal strategies. The signal does not end until the world transitions out of the last signal state, so if the agent thinks the last signal state is unlikely or impossible, it does not need to act urgently to collect reward. It could wait, attend only slightly to refine its estimate, or attend a lot to decide whether it can collect an early reward. Many of these options may be equally good. When signal belief is concentrated in the last signal state, by contrast, waiting and slight refinements lose value compared to commitment decisions, and the balance of good vs. bad performance becomes more immediate, with fewer opportunities to act anticipated in the near future. This may restrict the possible optimal levels of attention. Optima over beliefs concentrated in early signal states may therefore have "noise" introduced, where one of two equally good options is selected for one exact belief and

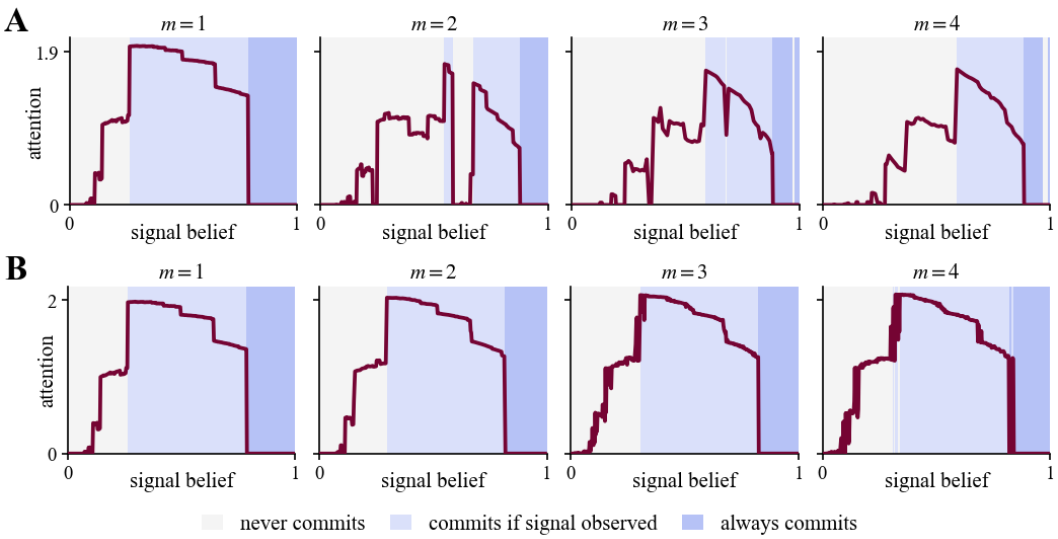


Figure 8: Visualization of policies and policy slices, as in Fig. 2, except that signal belief in the slices is (**A**) concentrated in the first signal state or (**B**) distributed uniformly across all modeled signal states.

the other is selected for a similar, adjacent belief. Viewing the policy over these adjacent beliefs results in a pattern that looks random. By contrast, there may be few optima, or only one, over beliefs concentrated in the late signal states, resulting in more consistency between the optimal attention at adjacent beliefs. The central point made in the paper, that policies exhibit jumps, holds whether or not this explanation is correct.

A.5 Autocorrelations for $m \geq 1$

Fig. 9 shows autocorrelations for policies representing the signal with different numbers of signal states. Results hold as m increases, and the effects of varying temporal parameters generally become easier to read for higher m. As m increases, allocation frequency increases for short signal durations and decreases for long signal durations, consistent with the expected effect of sharpening the duration distribution. Across all m, increases in hit reward similarly increase allocation frequency, and linear attention costs yield rhythmicity, while quadratic costs do not.

A.6 Flat controls with equal mean attention cost instead of equal mean attention

Because attention cost depends nonlinearly on attention whenever $c_2 \neq 1$, matching a control's mean attention does not match its mean cost. We therefore also compared optimal policies with a second constant-attention control, whose attention level was chosen so that its mean attention cost matched that of the optimal policy. Results were comparable to those for the mean-attention control (Fig. 10). Across all parameters tested, 11.62% of cost-matched controls had negative net reward. Among controls with non-negative reward, the lowest-resolution optimal policies ($m = 3$ and $m = 4$) earned a median of 1.54 and 1.61 times the controls' net reward, respectively. The fraction of optimal net reward captured by the controls ranged from 0.6% to 89.6%, with a median of 54.7% and a mean of 50.2%, and 46.4% of controls earned less than half the net reward of the corresponding optimal policy.

The largest advantages of optimal policies were greater against the cost-matched control than against the mean-attention control. For the base parameter set, the maximum ratio of optimal to control net reward against the mean-attention control was 5.97, 7.43, 7.83, and 7.80 for $m = 1$, 2, 3, and 4, respectively, compared with 19.77, 9.64, 171.69, and 8.47 against the cost-matched control.

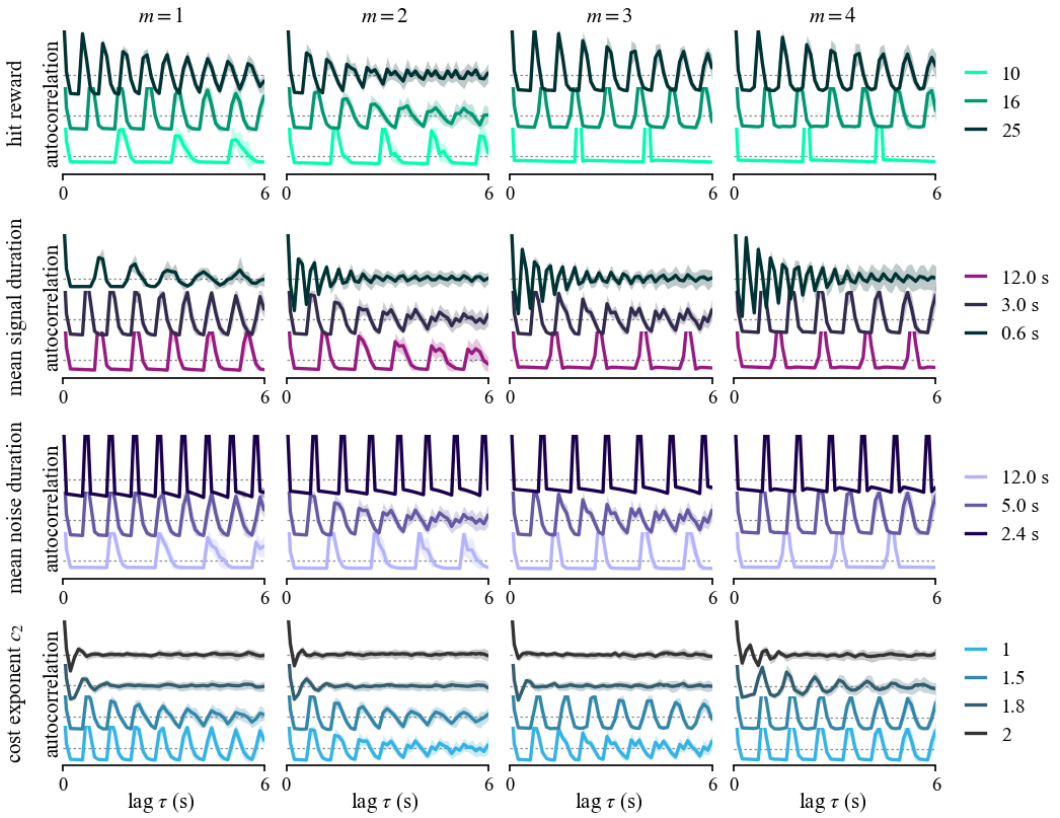


Figure 9: Autocorrelations over model parameters. The columns show models with $m = 1$, 2, 3, and 4, respectively. The rows show autocorrelations over different values of the following parameters, from top to bottom: hit reward, mean signal duration, mean noise duration, temporal discount factor, and attention cost exponent.

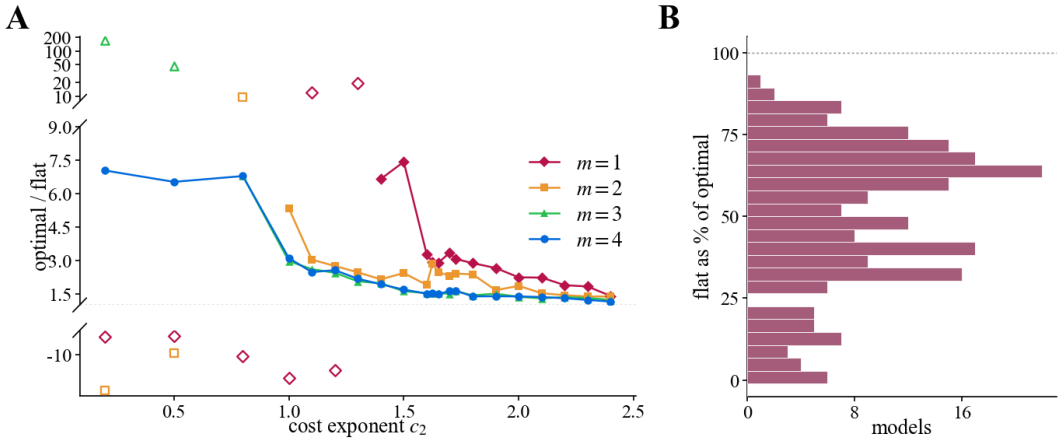


Figure 10: Comparison of optimal vs. cost-matched constant-attention policies. (**A**) The ratio, over c_2, of mean net reward per step for optimal policies to that of a constant-attention control with the same mean attention cost. A broken axis shows negative ratios and ratios above 10. (**B**) The inverse comparison: a histogram, over all solved policies, of the percentage of an optimal policy's net reward captured by a constant-attention control with the same mean attention cost. For visual clarity, we show only comparisons where the control achieved non-negative mean reward.