Trade-Offs Between Tasks Induced by Capacity Constraints Bound the Scope of Intelligence

Cameron Rouse Turner
Departments of Psychology & Computer Science
Princeton University
c.rouse.turner@princeton.edu

Dilip Arumugam
Department of Computer Science
Princeton University
dilip.a@cs.princeton.edu

Logan Nelson
Department of Psychology
Princeton University

Thomas L. Griffiths
Departments of Psychology & Computer Science
Princeton University

Abstract

A core challenge in cognitive science is understanding the barriers to intelligence and the circumstances that favor cognitive specialization. General intelligence requires a cognitive architecture that is successful across tasks. However, improving an architecture for a given task is often observed to hinder performance on others. Although trade-offs between tasks are a recurring element of explanations in cognitive science, they have received little direct theoretical attention. We present a formal framework that provides a principled understanding of when trade-offs emerge. In particular, we build on recent advances in applying rate-distortion theory to reinforcement learning. This allows us to formalize the representational capacity an agent can call upon in approaching tasks in terms of information. We find trade-offs occur when components of a task conflict in ways that cannot be easily coarse-grained by an agent's encoding scheme. Further, cognition may be general, specialized, or implement a coverall strategy, depending on conditions.

Keywords: Trade-offs; Intelligence; Rate-Distortion Theory; Reinforcement Learning; Resource Rationality

Introduction

While some cognitive systems display general-purpose intelligence, others are specialized and domain-specific. Among the mammals, chimpanzees display at least 20 different foraging techniques customized to local conditions (Whiten et al., 1999), while koalas only eat gumtree leaves. A prominent explanation for why general intelligence is difficult to attain is that improved performance on one task produces worse performance on others. A famous result appearing to support this proposition is the No Free Lunch Theorem, which states that an algorithm cannot perform well across every possible task (Wolpert & Macready, 1997). Unfortunately, the theorem may be limited in explaining constraints on real-world intelligence as its conclusion follows from a key assumption that tasks are encountered within uniform probability. By contrast, structure in nature implies that real-world tasks are not drawn from a uniform distribution, suggesting other factors are responsible for constraining the generality of cognition. Indeed, across many areas of cognitive science it has been found in practice that tasks trade off with one another (Del Giudice & Crespi, 2018). For example, research on cognitive control suggests there are expediency benefits to reusing the same cognitive architecture, but this leads to conflicts that hurt general performance (Musslick & Cohen, 2021). Understanding when tasks trade off appears central to when cognition can be successful across many domains, rather than domain-specific.

General intelligence appears bound by the availability of resources to distribute across tasks. Notably, the Space Hierarchy Theorem suggests that as capacity is reduced, so is the range of tasks a system can solve (Papadimitriou, 1994). To illustrate, if an agent can only hold in memory two symbols they cannot detect the pattern in 'AABAAB', which requires three. Indeed, the expansion of capacity may underpin many sophisticated forms of human intelligence, such as higher-order causal reasoning and language (Cantlon & Piantadosi, 2024). Similarly, 'resource rationality' shows how limited cognitive resources can explain departures from rational action (Callaway et al., 2022; Bhui et al., 2021; Gershman et al., 2015; Lieder et al., 2018; Lieder & Griffiths, 2020).

In this paper, we present a formal model that provides principled understanding of why trade-offs emerge between tasks as a consequence of their structure. In particular, we examine how limiting an agent's representational capacity affects the domain generality of its cognition. We formalize this issue by capitalizing on recent advances in applying rate-distortion theory to reinforcement learning (RL) (Abel et al., 2019; Prystawski et al., 2023; Arumugam et al., 2024). The problem formulation of RL provides a rigorous way to understand tasks: the relationship between states, actions and rewards. Rate-distortion theory builds on information theory to provide an equally rigorous method for examining how changes in representational capacity affect the quality of decision-making, and thereby underlying cognition. An advantage of our approach is that it makes few assumptions about cognitive architecture (for example, connectionist versus symbolic). Consequently, we are able to focus on the minimal problem intelligent agents need to solve. Our model affirms that capacity bounds general intelligence; however, it locates the cause of task conflict in the mismatch between the agents representation and its environment. We go on to identify aspects of task structure that explain when tasks will trade off.

Trade-offs, Capacity, and Task Structure

First, we informally lay out our central ideas through the simplest possible examples, providing formal definitions later. Consider a one-dimensional gridworld in which the agent must navigate to rewards (Figure 1). The interaction between the agent and environment can be seen through the lens of a Markov Decision Process (MDP), as used in RL (Sutton &

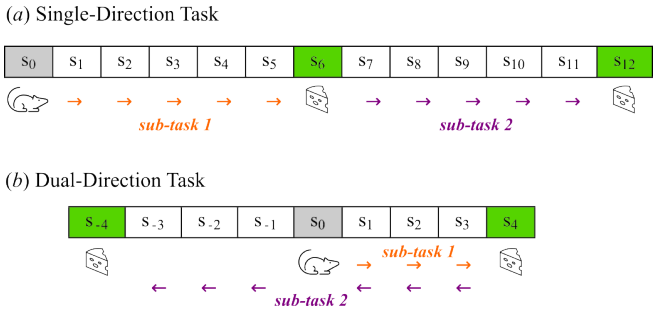

Figure 1: Comparison of tasks that are simple (a) versus complex (b). This complex environment sets conditions for trade-offs when capacity constraints are present. To highlight capacity limitations, tasks are contrived to make movement costs equal.

Barto, 1998). Informally, this means the configuration of the environment at a point in time is represented by its state. The agent produces actions that move it through the state space according to its policy, with the aim of achieving rewards. In Figure 1, the state is the physical location of the agent on a grid, available actions are moving left or right, and rewards occur when the green states are entered. Figure 1 assumes that rewards are consumed when encountered, so technically the state space must include indicators of which rewards remain to be Markovian. A wide variety of quantities can be considered states, making our findings about trade-offs between tasks broadly applicable. Examples of states include pixels in visual input, forces applied to effectors, others' beliefs, and stock prices.

We propose that trade-offs between tasks occur when the agent has a limited capacity and multiple goals that are configured in a conflicting way. Trade-offs cannot occur if there is a single goal. Further, task conflicts arise from the interplay between the state space, possible actions, and how rewards can be attained. Compare the Single-Direction Task (Figure 1a) to the Dual-Direction Task (Figure 1b). In the Single-Direction Task, the location of both rewards with respect to the state space and available actions mean that the optimal policy can take a single action (always move right) to gain both rewards. By contrast, in the Dual-Direction Task, the components of the MDP conspire such that the optimal policy must switch direction to gain both rewards. This means that the Dual-Direction Task has necessary structure to produce a trade-off, if there is also limitations to the capacity underpinning its policy. In particular, to switch between two actions the agent's policy must encode 1 bit of information. With a lower capacity, the agent can only take a single action, so must specialize and choose which way to move. Consequently, a trade-off is produced because moving right comes at the expense of moving left.

Our model focuses on the aspects of general intelligence that lead to success on complex tasks on the basis of a policy generated by a sophisticated representation requiring high capacity. Broadly, domain-general cognition is capable of having success on complex tasks because it can succeed in combining heterogeneous subtasks. For instance, the general intelligence of chimpanzees allows them to engage in nut-cracking, involving the subtasks of (1) selecting appropriate nuts, (2) finding rock 'hammers' and 'anvils', and (3) precisely striking. To tractably examine general versus specialized cognition, we produce a stylized model with a small number of states in a low number of dimensions. Nonetheless, 'complex tasks' can be modeled as circumstances where subtasks are dissimilar, so that their informational encoding is hindered. Illustrating by example, the Dual-Direction Task is complex because its subtasks are dissimilar, as available actions and states necessitate moving in opposite directions. Within the MDP formulation, environments are inherently agential, so that 'environmental complexity' actually depends on the states the agent can detect, performable actions due to effectors, and what constitutes reward. Therefore, when task trade-offs occur in our model, it is by influencing which policy architectures can be successful given how the agent can internally represent the external world. Our model does not allow the agent to flexibly manage its capacity, like the a resource underpinning attention, instead we look at the upper-limit of available information. Finally, we ignore tasks where multiple goals cannot be achieved regardless of capacity, as these equate to impossible challenges.

Theoretical Framework

Here we present a theoretical framework for analyzing how a capacity-limited agent represents the external world. In the next section, we present a concrete instance of this framework that allows us to test its predictions.

Markov Decision Process

We formulate a task as an infinite-horizon, discounted Markov Decision Process (MDP) (Bellman, 1957; Puterman, 1994) defined by $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}, \mu, \gamma \rangle$. $\mathcal{S}$ denotes a set of states, $\mathcal{A}$ is a set of actions, $\mathcal{R} : \mathcal{S} \times \mathcal{A} \to [0,1]$ is a reward function, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is a transition function, $\mu \in \Delta(\mathcal{S})$ is an initial state distribution, and $\gamma \in [0,1)$ is the discount factor. Starting with an initial state $s_0 \sim \mu$, at each timestep t an agent observes the current state $s_t \in \mathcal{S}$, selects an action $a_t \in \mathcal{A}$, enjoys a reward $r_t = \mathcal{R}(s_t, a_t) \in [0,1]$, and transitions to the next state $s_{t+1} \sim \mathcal{T}(\cdot \mid s_t, a_t) \in \mathcal{S}$.

An agent's behavior is defined by a policy $\pi : \mathcal{S} \to \Delta(\mathcal{A})$ such that $a_t \sim \pi(\cdot \mid s_t)$. Agent performance in $\mathcal{M}$ is given by $V^\pi(\mu) \triangleq \mathbb{E}_{s_0 \sim \mu}[V^\pi(s_0)] = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t, a_t)\right]$. The RL objective (Sutton & Barto, 1998) is to learn an optimal policy $\pi^\star$, defined as achieving supremal value $V^\star(\mu) = \sup_\pi V^\pi(\mu)$. To any policy π, the stationary state visitation of π is $d^\pi(s) = (1-\gamma)\sum_{t=0}^{\infty} \gamma^t \mathbb{P}^\pi(s_t = s)$, where each $\mathbb{P}^\pi(s_t = \cdot) \in \Delta(\mathcal{S})$ denotes the per-step distribution over states visited by policy π. d^π represents the γ-weighted distribution over possible future

states encountered by π.

Rate-Distortion Theory

Rate-distortion theory (RDT) (Shannon, 1959) is the area of information theory (Shannon, 1948; Cover & Thomas, 2012) that studies lossy compression. A lossy compression problem requires two inputs: an information source and a distortion function. An information source is a probability distribution over uncompressed data. A distortion function is a loss function quantifying error between compressed and uncompressed data. RDT provides the fundamental limit on the minimum expected distortion (error) that must be incurred given an upper limit on rate (preserved information).

We take the states visited by $\pi^\star$ as an information source $S \sim d^{\pi^\star}$. We model a bounded agent's internal abstract representation of MDP states as a channel $\phi : \mathcal{S} \to \Delta(\widetilde{\mathcal{S}})$ given by a conditional probability distribution over abstract states in $\widetilde{\mathcal{S}}$ given each state $s \in \mathcal{S}$. For simplicity, our experiments will assume $|\mathcal{S}| = |\widetilde{\mathcal{S}}|$ such that an agent with sufficiently high capacity may always recover the identity mapping as ϕ. Given the optimal policy $\pi^\star$, each channel ϕ induces an abstract policy $\pi^\phi : \widetilde{\mathcal{S}} \to \Delta(\mathcal{A})$ (Tishby et al., 1999; Abel et al., 2019). Consequently, we define distortion as the KL-divergence between the policies at each state:

$$d(s,\widetilde{s}) = D_{\mathrm{KL}}(\pi^\star_s \,||\, \pi^\phi_{\widetilde{s}}) \triangleq D_{\mathrm{KL}}(\pi^\star(\cdot \mid s) \,||\, \pi^\phi(\cdot \mid \widetilde{s})).$$

An agent's capacity is modeled as a rate limit $R \in \mathbb{R}_{\geq 0}$ on how many bits of information an agent's internal representation retains about original MDP states. Following Abel et al. (2019), RDT informs us that an agent with a capacity of R bits cannot behave optimally with expected error smaller than distortion-rate function $\mathcal{D}(R)$, defined as:

$$\mathcal{D}(R) = \underbrace{\inf_{\phi(\widetilde{S} \mid S)}}_{\text{Encoding}} \underbrace{\mathbb{E}\left[D_{\mathrm{KL}}(\pi^\star_S \,||\, \pi^\phi_{\widetilde{S}})\right]}_{\text{Exp. Distortion}} \text{ s.t. } \underbrace{\mathbb{I}(S;\widetilde{S})}_{\text{Rate}} \leq R. \qquad (1)$$

The mutual information (Cover & Thomas, 2012) $\mathbb{I}(S;\widetilde{S})$ measures how much information is retained by the abstract states $\widetilde{S}$ about the original states S visited by $\pi^\star$. As KL-divergence measures the closeness between distributions, our choice of $d(s,\widetilde{s})$ models an agent who endeavors to emulate $\pi^\star$ as closely as possible with capacity limited to R bits. We may interpret ϕ as a partition of ground states into abstract states (Li et al., 2006; Abel et al., 2016), but it is a soft partition (Singh et al., 1994). The ϕ solution to $\mathcal{D}(R)$ is the information-theoretically optimal state abstraction for an agent with capacity R to approximately preserve $\pi^\star$, and is computable via the Blahut-Arimoto algorithm (Blahut, 1972; Arimoto, 1972). This can be shown with a value-loss bound that benchmarks the performance of the expected abstract policy $\widetilde{\pi}^\phi(A \mid S) = \mathbb{E}_{\widetilde{S} \sim \phi(\cdot|S)}\left[\pi^\phi(A \mid \widetilde{S})\right]$ against $\pi^\star$: [1]

Theorem 1. *For an agent with rate limit $R \in \mathbb{R}_{\geq 0}$, let ϕ be the state abstraction that achieves the distortion-rate limit (Equation 1). Then, $V^\star(\mu) - V^{\widetilde{\pi}^\phi}(\mu) \leq \frac{\sqrt{\mathcal{D}(R)}}{\sqrt{2}(1-\gamma)^2}$.*

As $\widetilde{\pi}^\phi$ follows directly from Equation 1, $V^{\widetilde{\pi}^\phi}$ is the value of the best policy a capacity-limited agent can execute with only R bits of information at its disposal. Therefore, the LHS of the inequality quantifies the worst-case gap in performance between a capacity-unlimited optimal agent, $\pi^\star$, and an optimal capacity-limited agent, $\widetilde{\pi}^\phi$. Overall, this result implies that the performance shortfall of the latter agent is upper-bounded by an expression that depends on the expected distortion $\mathcal{D}(R)$ incurred at a rate R. Our theorem makes a clear prediction about how limits on capacity translate into limits on the realizability of optimal behavior. Moreover, it affords a principled notion of environment complexity as the minimum rate needed to incur zero expected distortion: $R^\star \triangleq \inf\{\bar{R} \in \mathbb{R}_{\geq 0} \mid \mathcal{D}(\bar{R}) = 0\}$. A severely capacity-limited agent should not have access to policies that yield high performance in a complex environment. There is a triadic relationship between: **(1)** the complexity of the environment $R^\star$ (how much information is needed to encode the world for optimal behavior), **(2)** the capacity of the agent R (how much information is actually available to encode the world), and **(3)** the discrepancies between $\pi^\star$ and $\widetilde{\pi}^\phi$ that may emerge when $R \ll R^\star$ such that complexity exceeds capacity.

Testing the Framework's Predictions

In this section, we describe how our framework applies to a set of concrete problems that differ in complexity, varying agent capacity to illustrate how capacity constraints translate into trade-offs. We consider an agent acting in a family of (contextual) MDPs (Hallak et al., 2015) that capture multi-task RL: 3×3 multi-task gridworlds (Figure 2). The agent always begins in the center with a goal located in one of the surrounding 8 cells, providing a $+1$ reward. The set of actions $\mathcal{A}$ is $\{\uparrow, \downarrow, \leftarrow, \rightarrow\}$ with deterministic transitions. To accommodate multiple tasks, the agent's goal may randomly appear in one of two possible locations, so the state space has a binary indicator denoting which location to pursue $\mathcal{S} = \{0,1\} \times \{1,2,3\} \times \{1,2,3\}$; this yields 6 possible reward configurations. Further, we create the possibility of asymmetries in the expected payoffs of subtasks by having the probability of each goal occurring over trials being either $(.50,.50)$ or $(.70,.30)$.

As mentioned, rate-distortion theory provides a method to introduce capacity limitations into the reinforcement learning set-up. To unpack, the objective of performing well at the task enters via the expected distortion term $\mathbb{E}\left[D_{\mathrm{KL}}(\pi^\star_S \,||\, \pi^\phi_{\widetilde{S}})\right]$, which ensures the abstract-state policy closely matches the optimal policy. The objective of compression enters via the mutual information $\mathbb{I}(S;\widetilde{S})$, which measures how much information in the agent's abstracted representation corresponds to the optimal representation. The free-parameter R is how we manipulate the agent's capacity, bounding the informa-

[1] Please see the Appendix for the full proof.

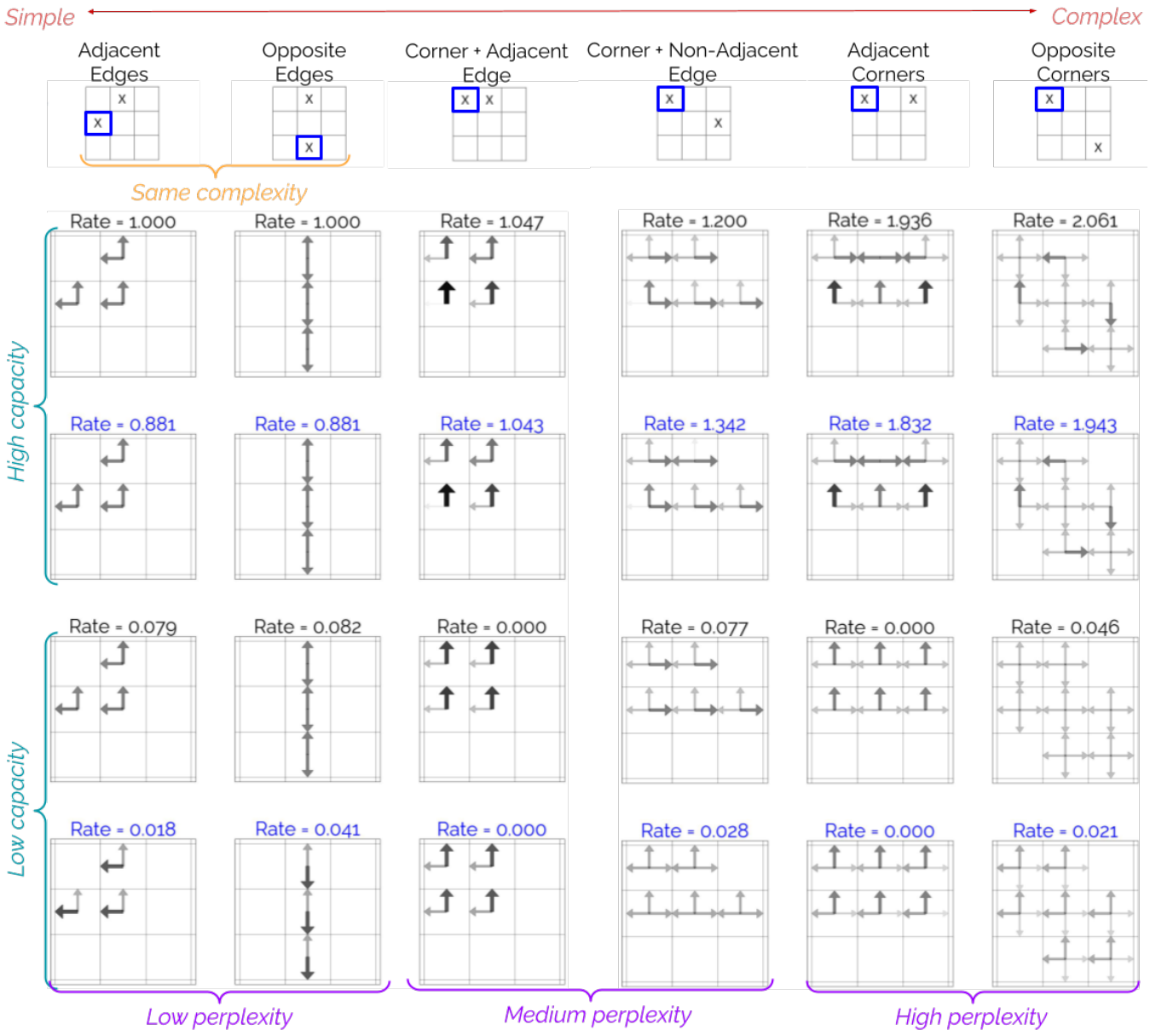

Figure 2: Task complexity versus agent's information capacity. Top row: each task (*e.g.*, Adjacent Edges) ordered by complexity due to their two subtasks. Arrows show the agent's movement policy, with darker arrows indicating higher probability of action. Policies displayed by task complexity (columns) versus capacity (rows). To consider asymmetries between subtasks, we varied the probability of goals. Each capacity level is broken down into symmetric (top row) and asymmetric (bottom row) subtasks. Symmetric condition had goals appear with probability .50. Asymmetric condition had the blue goal appear with probability .70. 'Perplexity' is the average number of abstract states per ground state (labeled as in Figure 3).

tion available. When R is high, a low-compression and high-fidelity encoding occurs allowing a close correspondence to ground states. The distortion-rate function definition enforces that the encoding ϕ is chosen to minimize expected distortion (the infimum) given constraints. In summary, the distortion-rate function $\mathcal{D}(R)$ formalizes the best trade-off between performance and compression an agent with capacity R can achieve.

Complex tasks contain dissimilar subtasks

A conceptual challenge is providing a quantity summarizing the difficulty we expect an agent to have in approaching a task. We solved this problem by formally defining 'complexity' $R^\star$ as the amount of information required to encode a task without loss. Our notion of complexity arises solely from the configuration of state, actions, and rewards via the optimal policy; the informational content is measured, but no lossy

compression has been enforced. We find that more complex tasks have subtasks that are dissimilar in that they required taking conflicting actions (Figure 2, Right). Further, complex tasks contain subtasks that themselves are more difficult, requiring more steps and changing actions. For instance, the Opposite Corners task is the most complex and least compressible because its subtasks require going to opposite edges of the state space. This creates maximum conflict, while also requiring more moves from the center. Broadly, complex tasks require using a variety of actions that depend on state, whereas simple tasks allow using the same actions more frequently. Our formal definition of complexity fits intuition: chess is challenging because its countless positions permit of conflicting strategies, while tic-tac-toe is simple because its solvable with a few action patterns.

Cognition varies in response to limited capacity

As capacity reduces, the agent must compromise on the details it represents, this leads to the intuition that agents may lose ability across domains, causing specialization. By contrast, we find qualitatively different strategies emerge depending on task structure (Figure 2, Left). In particular, specialization only occurs when there are asymmetries in the ability to exploit subtasks, so that lower capacity makes focusing on one subtask favorable. For instance, in the Corner + Adjacent Edge task a capacity-limited agent specializes by prioritizing attaining the closer reward; this specialization is more pronounced when that reward occurs more frequently. By contrast, when subtasks are symmetrical in their exploitability the agent instead produces a bet-hedging strategy. That is, with no reason to prioritize either subtask, a coverall strategy is favored. For instance, in Opposite Edges and Adjacent Edges the optimal strategy is moving to each reward with probability ½. This result suggests that the relative attainability of subtasks determines if capacity constraints lead to genuine specialization or a coverall strategy.

Greater representational chunking occurs in complex environments

As capacity decreases the agent is forced to represent its environment with only a few abstract states. Formally, this means there are more ground states mapped to abstract states on average (Figure 3, Right); this measure is known as 'perplexity' in information theory, as the observer would be more confused about their ground state. We find chunking is the cause of task conflicts, even when little is assumed about cognitive architecture. Complex tasks require fine distinctions, and these must necessarily be more severely chunked when capacity is limited. For example, Opposite Corners is the most complex task requiring an intricate production of actions depending on state. As capacity becomes limited the agent cannot encode this intricate trajectory so favors simply moving randomly in any direction, producing a strategy that approaches a random walk. This result affirms the hypothesis that the cause of task conflict is the attempt to reuse representations that are not rich enough for the environment, suggesting it holds across different cognitive architectures.

Performance drops when effective representations are exhausted

The degree to which task performance is lessened by compression depends on the specifics of task structure and current encoding scheme. Distortion from optimal performance always increases as capacity decreases (Figure 3 Left). However, the severity of this dropoff varies. A decrease in capacity causes a severe drop in performance when there are no longer effective ways to chunk states. This is evident in the gradient of the distortion slope becoming steeper as capacity reduces. In particular, the distortion gradient can be thought of as a measure of the degree to which tasks trade off. The exhaustion of effective ways to chunk also accounts for why simpler tasks may lead to greater distortion than complex tasks. For example, Opposite Edges can nearly be perfectly encoded with 1 bit because the agent simply needs to determine moving up or down. However, performance drastically drops with less capacity. Indeed, an agent with low capacity in Opposite Edges produces worse performance than an equivalent attempting Adjacent Corners. This occurs because although Adjacent Corners is more complex, subtasks are similar with more shared structure that can be shared.

Discussion

Being intelligent across domains appears to hinge on how much performance on one task conflicts with others. We examined trade-offs between tasks using a stylized model based on the application of rate-distortion theory to reinforcement learning. This led to a formal definition of task complexity as the minimum amount of information an agent must compress to encode the optimal solution. Intuitively, complex tasks turn out to be those with subtasks that are dissimilar, requiring conflicting actions. Our model affirmed that trade-offs occur when capacity limitations are present, but found that the form of cognition depends on the details of the subtasks involved. In particular, limited capacity results in specialization if one subtask can more fruitfully be exploited; by contrast, a coverall strategy is favored if subtasks are similarly profitable. When capacity is high, the agent can possess general cognitive that performs effectively across the entire domain.

Our model made few assumptions about cognitive architecture, but nonetheless recovered the notion that task conflict is caused by the coarse-graining of representations in ways that mismatch the environment. Further, we provided a precise way to understand representational chunking as compressing a high-dimensional environment into a few abstract states. In particular, the severity of task conflict can be thought of as the decline in performance with a reduction in capacity, and formalized as the gradient of the distortion-rate function. Surprisingly, complex tasks do not always lead to worse performance. What matters is the degree of exploitable overlapping structure between tasks, compared to the agent's encoding scheme. In our simplest tasks, optimal behavior could be encoded with 1 bit, but performance plummeted when capacity

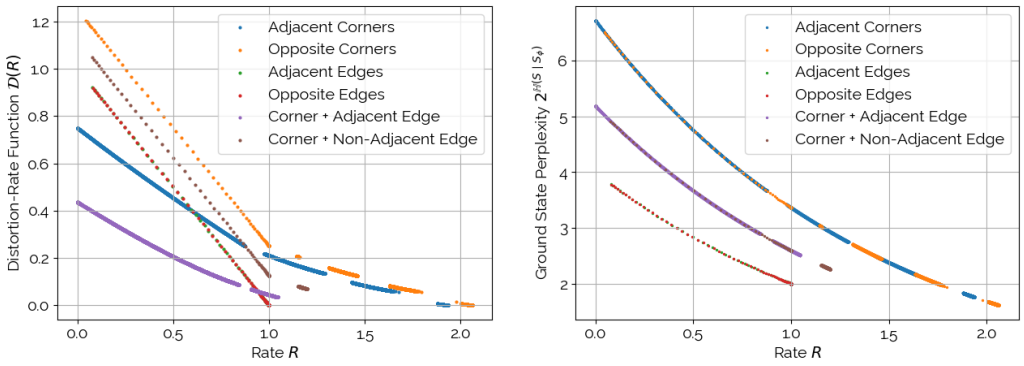

Figure 3: Agent performance and representations as a function of capacity (Rate). Left: As capacity (Rate) decreases, performance becomes worse (Distortion). Right: Reducing capacity increases the average number of ground states per abstract state (*i.e.*, 'Perplexity').

was reduced further because avenues to maintain successful representations were actually more tenuous in such a simple environment. Taken together, our work highlights that trade-offs between tasks arise from a subtle interplay between task structure and the representational resources of the agent.

Our model highlights particular aspects of tasks that influence when trade-offs will occur. Multiple lines of research produce explanations based on conflict between tasks (Musslick & Cohen, 2021; Cohen et al., 1990; Pashler, 1994; Vandierendonck et al., 2010). A prominent example is the study of continual learning in neural networks which grapples with the fact that networks perform poorly on previous tasks when trained for new tasks (McCloskey & Cohen, 1989; Kirkpatrick et al., 2017). We provide principles that describe when task structure leads to conflict. In particular, our model highlighted that two properties of subtasks dictate the form of cognition. In particular, 'similarity' is the property that actions that move towards the goal of one subtask also bring the agent closer to the other. Similarity is important because it allows the possibility of effective chunking and representation sharing. Secondly, subtask 'asymmetry' is the degree to which one goal is more profitable to exploit. Asymmetry produces a representational priority for the agent, allowing them to more readily discard information about the less valuable subtask. Broadly, we found all tasks do not trade off equally, with the severity of task trade-off depending on specific conflicts in structure between subtasks. In other words, whether a 'Jack-of-all-trades is a master of none' depends on the trades.

Although our model is highly stylized, our results affirm and extend prior research on cognitive constraints. Theories of resource and bounded rationality emphasize that limited cognitive resources lead to imperfect heuristics (Gershman et al., 2015; Lieder & Griffiths, 2020). Similarly, our model characterizes the optimal policy an agent with limited capacity should endeavor to learn. While our model does not address learning this resource-rational policy, we anticipate that future work studying capacity-limited RL (Tishby & Polani, 2010; Rubin et al., 2012; Lai & Gershman, 2021) will benefit from these information-theoretic tools and insights. Furthermore, while our model does not assume observational noise, extending it to partially observable MDPs (Poupart & Boutilier, 2002), where noisy observations constrain cognition (Doshi et al., 2012), is another plausible avenue for future work.

Our results aid in understanding the diversity of animal intelligences that have evolved. Some animals appear to have entered a cognitive-niche investing more in their brains, with humans being the most prominent example (Laland, 2017). Recently, it has been proposed that increases in informational capacity explains a wide variety of sophisticated forms of human intelligence, such as our capacity for symbolic language (Cantlon & Piantadosi, 2024). Our model supports this hypothesis, suggesting that general intelligence is bounded by capacity. By contrast, other animals have more specialized cognition. For example, while the mosquito most have encountered is actually a human-specialist, a generalist mosquito exists in Africa that feeds on a wide-range of species including cows, birds, and lizards (McBride et al., 2014). Our model predicts that specialization should be favored when a capacity-limited agent faces an environment where some subtasks produce greater profit for effort; this matches hypotheses about mosquito evolution that suggest specialization occurred because humans are outliers compared to other animals (*e.g.*, humans are hairless). Our findings support the expectation that small brained animals should display either greater specialization or a coverall strategy, depending on the landscape of subtasks. Environmental complexity is already considered an important factor in cognitive evolution (Turner & Walmsley, 2021; Turner et al., 2024). Our work here underscores how the design of cognition can be understood with reference to the tasks it attempts to solve.

Acknowledgments We thank two anonymous reviewers and the metareviewer for their comments. Further, Josh Thomas provided a useful discussion. This research was supported by the Templeton World Charity Foundation (grant number 20648) and supported by ONR MURI N00014-24-1-2748.

References

Abel, D., Arumugam, D., Asadi, K., Jinnai, Y., Littman, M. L., & Wong, L. L. (2019). State abstraction as compression in apprenticeship learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 3134–3142).

Abel, D., Hershkowitz, D., & Littman, M. (2016). Near optimal behavior via approximate state abstraction. In *International Conference on Machine Learning* (Vol. 48, pp. 2915–2923).

Achiam, J., Held, D., Tamar, A., & Abbeel, P. (2017). Constrained Policy Optimization. In *International conference on machine learning* (pp. 22–31).

Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. In *IEEE Transactions on Information Theory* (Vol. 18, pp. 14–20). IEEE.

Arumugam, D., Ho, M. K., Goodman, N. D., & Van Roy, B. (2024). Bayesian reinforcement learning with limited cognitive load. *Open Mind*, *8*, 395–438.

Bellman, R. (1957). A Markovian decision process. *Journal of Mathematics and Mechanics*, *6*(5), 679–684.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15-21.

Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. In *IEEE Transactions on Information Theory* (Vol. 18, pp. 460–473). Los Alamitos, CA: IEEE.

Bretagnolle, J., & Huber, C. (1978). Estimation des Densités: Risque Minimax. *Séminaire de Probabilités de Strasbourg*, *12*, 342–363.

Callaway, F., van Opheusden, B., Gul, S., Wilson, G. M., Schulz, E., Cohen, J. D., ... Griffiths, T. L. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, *6*, 1112–1125.

Canonne, C. L. (2022). A short note on an inequality between KL and TV. *arXiv preprint arXiv:2202.07198*.

Cantlon, J. F., & Piantadosi, S. T. (2024). Uniquely human intelligence arose from expanded information capacity. *Nature Reviews Psychology*, *3*, 275–293.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361.

Cover, T. M., & Thomas, J. A. (2012). *Elements of Information Theory*. Hoboken, NJ: John Wiley & Sons.

Del Giudice, M., & Crespi, B. J. (2018). Basic functional trade-offs in cognition: An integrative framework. *Cognition*, *179*, 56-70.

Doshi, P., Qu, X., Goodie, A. S., & Young, D. L. (2012). Modeling Human Recursive Reasoning Using Empirically Informed Interactive Partially Observable Markov Decision Processes. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, *42*(6), 1431–1442.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*, 273–278.

Hallak, A., Di Castro, D., & Mannor, S. (2015). Contextual Markov Decision Processes. *arXiv preprint*.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, *114*(13), 3521-3526.

Lai, L., & Gershman, S. J. (2021). Chapter five - policy compression: An information bottleneck in action selection. In K. D. Federmeier (Ed.), *The psychology of learning and motivation* (Vol. 74, p. 195-232). Academic Press.

Laland, K. N. (2017). *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton, NJ: Princeton University Press.

Li, L., Walsh, T. J., & Littman, M. L. (2006). Towards a unified theory of state abstraction for MDPs. In *International Symposium on Artificial Intelligence and Mathematics* (Vol. 4, p. 5).

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, 1-60.

Lieder, F., Griffiths, T. L., Huys, M., & Huys, Q. J. (2018). Empirical evidence for resource-rational anchoring and adjustment. *Psychonomic Bulletin & Review*, *25*, 775–784.

McBride, C., Baier, F., Omondi, A., Spitzer, S., Lutomiah, J., Sang, R., ... Vosshall, L. (2014). Evolution of mosquito preference for humans linked to an odorant receptor. *Nature*, *515*, 222–227.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 24). Academic Press.

Müller, A. (1997). Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, *29*(2), 429–443.

Musslick, S., & Cohen, J. D. (2021). Rationalizing constraints on the capacity for cognitive control. *Psychonomic Bulletin & Review*, *25*(9), 757–775.

Papadimitriou, C. H. (1994). *Computational complexity*. Boston, MA: Addison-Wesley.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244.

Pinsker, M. S. (1964). *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day.

Poupart, P., & Boutilier, C. (2002). Value-Directed Compression of POMDPs. *Advances in Neural Information Processing Systems*, *15*.

Prystawski, B., Arumugam, D., & Goodman, N. (2023). Cultural reinforcement learning: A framework for modeling cumulative culture on a limited channel. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Puterman, M. L. (1994). *Markov Decision Processes—Discrete Stochastic Dynamic Programming*. New York, NY: John Wiley & Sons.

Rubin, J., Shamir, O., & Tishby, N. (2012). Trading Value and Information in MDPs. In *Decision making with imperfect decision makers* (pp. 57–74). Springer.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Shannon, C. E. (1959). Coding Theorems for a Discrete Source with a Fidelity Criterion. In *Institute of Radio Engineers, International Convention Record* (Vol. 4, pp. 142–163).

Singh, S., Jaakkola, T., & Jordan, M. (1994). Reinforcement learning with soft state aggregation. In *Advances in Neural Information Processing Systems* (Vol. 7).

Sriperumbudur, B. K., Fukumizu, K., Gretton, A., Schölkopf, B., & Lanckriet, G. R. (2009). On Integral Probability Metrics,ϕ-Divergences and Binary Classification. *arXiv preprint arXiv:0901.2698*.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to Reinforcement Learning*. Cambridge, MA: MIT Press.

Tishby, N., Pereira, F. C., & Bialek, W. (1999). The information bottleneck method. In *The 37th Annual Allerton Conference on Communication, Control, and Computing* (Vol. 37, p. 368-377).

Tishby, N., & Polani, D. (2010). Information Theory of Decisions and Actions. In *Perception-action cycle: Models, architectures, and hardware* (pp. 601–636). Springer.

Turner, C., Morgan, T., & Griffiths, T. (2024). Environmental complexity and regularity shape the evolution of cognition. *Proceedings of the Royal Society B*, *291*, 20241524.

Turner, C., & Walmsley, L. (2021). Preparedness in cultural learning. *Synthese*, *199*, 81–100.

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin*, *136*(4), 601–626.

Whiten, A., Goodall, J., McGrew, W. C., Nishida, T., Reynolds, V., Sugiyama, Y., ... Boesch, C. (1999). Cultures in chimpanzees. *Nature*, *399*, 682–685.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*, 67–82.

Proof of Theorem 1

Theorem 1. *For an agent with rate limit $R \in \mathbb{R}_{\geq 0}$, let ϕ be the state abstraction that achieves the distortion-rate limit (Equation 1). Then,*

$$V^{\star}(\mu) - V^{\widetilde{\pi}^{\phi}}(\mu) \leq \frac{\sqrt{\mathcal{D}(R)}}{\sqrt{2}(1-\gamma)^2}.$$

Proof. For notational convenience, let χ^{π} denote the stationary state-action visitation distribution of any policy π, defined as $\chi^{\pi}(s,a) = d^{\pi}(s)\pi(a \mid s)$. Intuitively, as d^{π} is the state visitation distribution of π that captures the distribution over states encountered by executing π in the environment, χ^{π} simply captures the distribution over state-action pairs obtained from executing π.

Proving our result requires Lemma 3 of Achiam et al. (2017), which establishes that for any two policies π and π'

$$||d^{\pi} - d^{\pi'}||_1 \leq \frac{2\gamma}{(1-\gamma)} \mathbb{E}_{s \sim d^{\pi}} \left[D_{\text{TV}}(\pi(\cdot \mid s) \,||\, \pi'(\cdot \mid s)) \right].$$

Here $D_{\text{TV}}(p \,||\, q) = \frac{1}{2}||p-q||_1$ denotes the total variation distance between probability distributions. We may then establish an analogous result for any two policies π and π' in terms of the stationary state-action visitation distributions:

$$||\chi^{\pi} - \chi^{\pi'}||_1 \leq \frac{2}{(1-\gamma)} \mathbb{E}_{s \sim d^{\pi}} \left[D_{\text{TV}}(\pi(\cdot \mid s) \,||\, \pi'(\cdot \mid s)) \right].$$

This follows as

$$\begin{aligned}
||\chi^{\pi} - \chi^{\pi'}||_1 &= \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} |\chi^{\pi}(s,a) - \chi^{\pi'}(s,a)| \\
&= \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} |d^{\pi}(s)\pi(a \mid s) - d^{\pi'}(s)\pi'(a \mid s)| \\
&= \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} |d^{\pi}(s)\pi(a \mid s) - d^{\pi}(s)\pi'(a \mid s) + d^{\pi}(s)\pi'(a \mid s) - d^{\pi'}(s)\pi'(a \mid s)| \\
&\leq \sum_{s \in \mathcal{S}} d^{\pi}(s) \sum_{a \in \mathcal{A}} |\pi(a \mid s) - \pi'(a \mid s)| + \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} \pi'(a \mid s) |d^{\pi}(s) - d^{\pi'}(s)| \\
&= \sum_{s \in \mathcal{S}} d^{\pi}(s) \cdot 2 \cdot \frac{1}{2} \sum_{a \in \mathcal{A}} |\pi(a \mid s) - \pi'(a \mid s)| + \sum_{s \in \mathcal{S}} |d^{\pi}(s) - d^{\pi'}(s)| \underbrace{\sum_{a \in \mathcal{A}} \pi'(a \mid s)}_{=1} \\
&= 2\mathbb{E}_{s \sim d^{\pi}} \left[D_{\text{TV}} \left(\pi(\cdot \mid s) \,||\, \pi'(\cdot \mid s) \right) \right] + ||d^{\pi} - d^{\pi'}||_1 \\
&\leq 2\mathbb{E}_{s \sim d^{\pi}} \left[D_{\text{TV}} \left(\pi(\cdot \mid s) \,||\, \pi'(\cdot \mid s) \right) \right] + \frac{2\gamma}{(1-\gamma)} \mathbb{E}_{s \sim d^{\pi}} \left[D_{\text{TV}} \left(\pi(\cdot \mid s) \,||\, \pi'(\cdot \mid s) \right) \right] \\
&= \frac{2}{(1-\gamma)} \mathbb{E}_{s \sim d^{\pi}} \left[D_{\text{TV}} \left(\pi(\cdot \mid s) \,||\, \pi'(\cdot \mid s) \right) \right].
\end{aligned}$$

To make use of the above result in a value-loss bound, we will leverage the fact that the total variation distance is an integral probability metric (IPM) (Müller, 1997; Sriperumbudur et al., 2009) and so, for any $p, q \in \Delta(\mathcal{X})$,

$$D_{\text{TV}}(p \,||\, q) = \sup_{f \in \mathcal{F}_{\infty}} |\mathbb{E}_p[f(X)] - \mathbb{E}_q[f(X)]|,$$

where $\mathcal{F}_{\infty} = \{f : \mathcal{X} \to \mathbb{R} \mid ||f||_{\infty} \leq 1\}$. Recall that we assume all rewards are in the unit interval such that $||\mathcal{R}||_{\infty} = 1$. Moreover, it is a well-known fact that the value function of any policy can be written as

$$V^{\pi}(\mu) = \frac{1}{(1-\gamma)} \mathbb{E}_{(s^{\pi}_{\gamma}, a) \sim \chi^{\pi}} \left[\mathcal{R}(s^{\pi}_{\gamma}, a) \right].$$

So, applying the definitions of the total variation distance as an IPM and as the halved L_1-distance between probability distri-

butions alongside the above lemma, we obtain

$$\begin{aligned}
V^\star(\mu) - V^{\widetilde{\pi}^\phi}(\mu) &= \frac{1}{(1-\gamma)} \left(\mathbb{E}_{\chi^{\pi^\star}} \left[\mathcal{R}(s_\gamma^{\pi^\star}, a) \right] - \mathbb{E}_{\chi^{\widetilde{\pi}^\phi}} \left[\mathcal{R}(s_\gamma^{\widetilde{\pi}^\phi}, a) \right] \right) \\
&\leq \frac{1}{(1-\gamma)} \left| \mathbb{E}_{\chi^{\pi^\star}} \left[\mathcal{R}(s_\gamma^{\pi^\star}, a) \right] - \mathbb{E}_{\chi^{\widetilde{\pi}^\phi}} \left[\mathcal{R}(s_\gamma^{\widetilde{\pi}^\phi}, a) \right] \right| \\
&\leq \frac{1}{(1-\gamma)} \sup_{f \in \mathcal{F}_\infty} \left| \mathbb{E}_{\chi^{\pi^\star}} \left[f(s_\gamma^{\pi^\star}, a) \right] - \mathbb{E}_{\chi^{\widetilde{\pi}^\phi}} \left[f(s_\gamma^{\widetilde{\pi}^\phi}, a) \right] \right| \\
&= \frac{1}{(1-\gamma)} D_{\mathrm{TV}}(\chi^{\pi^\star} \,||\, \chi^{\widetilde{\pi}^\phi}) \\
&= \frac{1}{2(1-\gamma)} ||\chi^{\pi^\star} - \chi^{\widetilde{\pi}^\phi}||_1 \\
&\leq \frac{1}{(1-\gamma)^2} \cdot \mathbb{E}_{s_\gamma^{\pi^\star} \sim d^{\pi^\star}} \left[D_{\mathrm{TV}}(\pi^\star(\cdot \mid s_\gamma^{\pi^\star}) \,||\, \widetilde{\pi}^\phi(\cdot \mid s_\gamma^{\pi^\star})) \right]
\end{aligned}$$

All that remains is to relate this upper bound on value loss back to the agent capacity used to obtain the underlying state abstraction ϕ. To do that, we employ Pinsker's inequality (Pinsker, 1964) and two successive applications of Jensen's inequality (where the second use leverages the fact that the KL-divergence is jointly convex in its arguments).

$$\begin{aligned}
V^\star(\mu) - V^{\widetilde{\pi}^\phi}(\mu) &\leq \frac{1}{(1-\gamma)^2} \cdot \mathbb{E}_{s_\gamma^{\pi^\star} \sim d^{\pi^\star}} \left[D_{\mathrm{TV}}(\pi^\star(\cdot \mid s_\gamma^{\pi^\star}) \,||\, \widetilde{\pi}^\phi(\cdot \mid s_\gamma^{\pi^\star})) \right] \\
&\leq \frac{1}{(1-\gamma)^2} \cdot \mathbb{E}_{s_\gamma^{\pi^\star} \sim d^{\pi^\star}} \left[\sqrt{\frac{1}{2} D_{\mathrm{KL}}(\pi^\star(\cdot \mid s_\gamma^{\pi^\star}) \,||\, \widetilde{\pi}^\phi(\cdot \mid s_\gamma^{\pi^\star}))} \right] \\
&\leq \frac{1}{(1-\gamma)^2} \cdot \sqrt{\frac{1}{2} \cdot \mathbb{E}_{s_\gamma^{\pi^\star} \sim d^{\pi^\star}} \left[D_{\mathrm{KL}}(\pi^\star(\cdot \mid s_\gamma^{\pi^\star}) \,||\, \widetilde{\pi}^\phi(\cdot \mid s_\gamma^{\pi^\star})) \right]} \\
&= \frac{1}{(1-\gamma)^2} \cdot \sqrt{\frac{1}{2} \cdot \mathbb{E}_{s_\gamma^{\pi^\star} \sim d^{\pi^\star}} \left[D_{\mathrm{KL}}(\pi^\star(\cdot \mid s_\gamma^{\pi^\star}) \,||\, \mathbb{E}_{s_\phi \sim \phi(\cdot \mid s_\gamma^{\pi^\star})} \left[\pi_\phi(\cdot \mid s_\phi) \right]) \right]} \\
&\leq \frac{1}{(1-\gamma)^2} \cdot \sqrt{\frac{1}{2} \cdot \mathbb{E}_{s_\gamma^{\pi^\star} \sim d^{\pi^\star}} \left[\mathbb{E}_{s_\phi \sim \phi(\cdot \mid s_\gamma^{\pi^\star})} \left[D_{\mathrm{KL}}(\pi^\star(\cdot \mid s_\gamma^{\pi^\star}) \,||\, \pi_\phi(\cdot \mid s_\phi)) \right] \right]} \\
&= \frac{\sqrt{\mathcal{D}(R)}}{\sqrt{2}(1-\gamma)^2}.
\end{aligned}$$

□

One observation is that, for a small rate R, the distortion-rate function $\mathcal{D}(R)$ may grow large enough to make this bound vacuous, largely stemming from the use of Pinsker's inequality and the fact that total variation distance is always upper bounded by 1. In such cases, a more meaningful bound may be obtained by following the same steps above but employing the Bretagnolle-Huber inequality (Bretagnolle & Huber, 1978; Canonne, 2022) in lieu of Pinsker's to yield

$$V^\star(\mu) - V^{\widetilde{\pi}^\phi}(\mu) \leq \frac{1}{(1-\gamma)^2} \cdot \sqrt{1 - \exp\left(-\mathcal{D}(R)\right)}.$$