Complexity and Misspecification*

Drew Fudenberg† Florian Mudekereza‡

March 12, 2026

Abstract

We propose a tractable unified framework to study the evolution and interaction of model-*misspecification* concerns and *complexity* aversion in repeated decision problems. This aims to capture environments where decision makers worry that their models are misspecified while also disliking overly complex models. We find that pathological *cycles* caused by endogenous concerns for misspecification can be eliminated by penalizing complex models and show that such preferences for simplicity tend to favor safety, which can enhance welfare in the long run. We use our framework to provide new microfoundations for pervasive empirical phenomena such as "scale heterogeneity" in discrete-choice analysis, "probability neglect" in behavioral economics, and "home bias" in international finance.

Keywords: learning, complexity, misspecification, robustness.

Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena. – Box (1976, p. 792)

*We thank Giacomo Lanzani, Lasse Mononen, Arthur Robson, Larry Samuelson, Jakub Steiner, and Tomasz Strzalecki for helpful comments, and NSF grant SES-2417162 for financial support.

†Department of Economics, MIT, drew.fudenberg@gmail.com

‡Department of Economics, MIT, florianm@mit.edu

1 Introduction

Economic agents routinely rely on simplified models to make decisions in complex environments such as financial markets, insurance, and macroeconomic forecasting. These models distill reality into a small number of parameters and probabilistic relationships, but agents may be concerned that such structured representations are only approximations of the true data-generating process. A large literature on robust control, beginning with Hansen and Sargent (2001, 2008), captures this concern by modeling decision makers who choose actions to perform well under worst-case distortions of a reference model. In this framework, concerns about misspecification are captured by allowing Nature to distort the reference model, with larger departures penalized according to relative entropy. These penalties depend only on statistical distance and do not consider the complexity of the alternative scenarios the agent fears.

In contrast, a long tradition in statistics and learning theory dating back to Occam's razor emphasizes a preference for simpler explanations over complex ones (Box, 1976; Rissanen, 1978). This raises a natural question: how should an agent trade off concern about misspecification against a preference for simple, plausible worst-case narratives? We model simplicity using Shannon entropy because it is the unique additive regularizer that preserves the convex duality structure underlying robust control and yields a closed-form Gibbs distortion. Importantly, we penalize the entropy of worst-case distortions—not the richness of the model class itself—so simplicity disciplines the construction of adversarial explanations rather than restricting which structured models the agent considers.

Our analysis builds on Cerreia-Vioglio et al.'s (2025) framework for decision making under model misspecification, which distinguishes uncertainty about models from concerns that all of the models being considered may be wrong. Lanzani (2025a) studies the dynamic implications of this approach when misspecification concerns adjust endogenously as model fit changes. We extend this dynamic framework with a preference for simplicity by adding a term that penalizes more disperse beliefs. This penalty forces worst-case beliefs to balance pessimism against simplicity: among all distortions that make payoffs low, Nature prefers those that are more concentrated and easier to describe. In the main text we suppose that the complexity penalty is proportional to the entropy of the distribution, so that the worst-case beliefs admit a closed-form characterization with a Gibbs structure, combining an exponential tilt in utilities with a power transformation of the structured model. However, this functional form is not essential; as we explain in Online Appendix D.1 it suffices to use the more general additive perturbations of Fudenberg et al. (2015).

We first study this criterion in a static environment. We show that, except in knife-edge cases, increasing complexity aversion strictly reduces the entropy of the worst-case distortion. In a simple safe-versus-risky example, this mechanism alters the ranking of actions even when the degree of robustness is held fixed: pessimism about risky actions can be generated using concentrated, low-entropy distortions, whereas pessimism about safe actions necessarily remains diffuse. As a result, complexity aversion tilts choice toward safety.

We then embed the static criterion into the dynamic learning environment of Lanzani (2025a). The agent updates beliefs about structured models using Bayes rule, and their robustness parameter evolves as a function of relative model fit. We show that Lanzani's dynamic selection result carries over to our setting: long-run empirical frequencies of actions correspond to mixed robust equilibria defined by the modified criterion. Complexity aversion reshapes this equilibrium set by penalizing actions whose worst-case explanations are simple and sharply concentrated. These modifications generate sharp dynamic predictions. In a two-action environment with a safe and a risky arm, we show that sufficiently strong complexity aversion eliminates all mixed robust equilibria. Thus, the endogenous *cycles* between safe and risky actions that arise under robustness alone disappear, and long-run behavior converges to a pure action. More generally, when multiple risky actions are available, complexity aversion acts as a selection device: actions whose worst-case distortions have lower entropy are eliminated first as the preference for simplicity increases.

To illustrate the mechanics of robustness-driven cycles, consider a central bank in a small open economy choosing between a hard currency peg (safe) and a managed float (risky). The float offers better stabilization when the bank's structured model is reliable, but leaves the economy exposed to unexpected shocks when the model is misspecified. When the bank's model has a forecast failure, its misspecification concern rises, forcing a retreat to the peg. This leads to a calm period in which realized outcomes are more in line with the bank's model, reducing misspecification concerns and tempting a return to the float, and generates a robustness-driven cycle between a risky and a safe regime. Such policy cycling is costly, as it undermines credibility and creates instability. We show how a preference for simple worst-case narratives eliminates these cycles, ensuring long-run policy stability, and that this stabilization can improve long-run realized payoffs. In our open-economy application, this implies the central bank can enhance welfare by prioritizing simpler economic narratives.

In our framework, the simplicity motive does not play a uniform role across parameter regimes. When the simplicity motive is moderate relative to misspecification concerns, the preference remains within the *average robust control* (ARC) family axiomatized by Lanzani

(2025b): it is observationally equivalent to ARC, but with an endogenously transformed within-model benchmark. However, when the simplicity motive is sufficiently strong, the criterion no longer has an ARC structure: evaluation collapses to the most adverse outcomes within each structured model, and further increases in the simplicity motive do not affect choices. When misspecification concerns vanish, the preference reduces to the subjective version of the *entropy-modified expected utility* criterion axiomatized by Mononen (2025).

In addition to stabilizing learning dynamics, our framework generates new explanations for several well-documented empirical phenomena. First, we apply our framework to discrete choice problems by embedding complexity-augmented robustness into the rational inattention framework of Matějka and McKay (2015). When the agent's worst-case narrative concentrates on specific adverse states, the effective logit scale parameter shrinks in those states, making choices highly deterministic precisely where the pessimistic narrative is most salient. This provides a structural microfoundation for scale heterogeneity in discrete choice (Swait and Louviere, 1993; Fiebig et al., 2010) and probability neglect in behavioral economics (Sunstein, 2002, 2003). Second, we apply our framework to the stochastic-growth setting of Robson et al. (2023). Standard explanations of equity home bias are invariant to relabeling of loss states and therefore cannot distinguish assets that share the same reference payoff distribution. In contrast, a preference for simple worst-case narratives breaks this invariance: a foreign asset whose losses are concentrated in a small number of focal scenarios is perceived as more fragile, generating home bias even without expected-return differences, hedging motives, or market segmentation. We also find that the same degree of misspecification can generate very different growth losses depending on how returns load on states, which matches the empirical observation that large uncertainty shocks can persist with much smaller growth losses than standard models would imply (e.g., Caggiano et al., 2014, 2017).

1.1 Related Work

Our work sits at the intersection of robust control, information theory, and learning theory. We highlight some key connections:

Robust control and misspecification. Our framework is grounded in the literature on variational preferences for decision-making under ambiguity (Maccheroni et al., 2006). Hansen and Sargent (2001, 2008) introduced and developed the multiplier criterion, where agents form worst-case beliefs subject to a relative entropic penalty to account for potential misspecification of a reference model. Strzalecki (2011) provided the axiomatic foundations for this criterion within the class of variational preferences. Subsequently, Cerreia-Vioglio et al. (2025) provided a formal separation between ambiguity aversion and fear of misspecification in a maxmin setting, and also outlined a Bayesian version of their model. Lanzani (2025a) axiomatized a special case of this Bayesian robust approach for the multiplier criterion and developed its learning dynamics, where the concern for misspecification evolves endogenously with model fit. We extend Lanzani's dynamic framework by incorporating a preference for simple worst-case beliefs—while his agent fears any statistically plausible distortion, our agent specifically fears distortions that are both plausible and parsimonious. Specifically, simplicity is penalized using Shannon entropy. This is consistent with Strzalecki's (2024) use of entropic penalty in variational Bayes settings to microfound non-Bayesian updating rules such as the "exponential" updating rule documented in Benjamin (2019) and Augenblick and Rabin (2021). More broadly, the use of worst-case beliefs as a model of ambiguity aversion originates with Gilboa and Schmeidler (1989). Our approach departs from multiple-prior models by endogenizing the selection of pessimistic beliefs through relative entropy and complexity penalties, which yields sharp comparative statics in both static and dynamic environments.

Expected utility and Complexity Aversion An adjacent literature that studies complexity aversion without model uncertainty. On the experimental side, Bernheim and Sprenger (2020) shows that valuations of lotteries are sensitive to event-splitting manipulations, and Enke and Graeber (2023) shows that a measure of cognitive uncertainty predicts systematic biases in risky choice; Enke et al. (2025) extends this perspective to intertemporal choice.

The closest theoretical paper is Gabaix (2025), which introduces "first-order" complexity aversion through imperfect signals about the optimal action; we show that our framework nests this approach. Mononen (2025) provides an axiomatic foundation for the entropy-modified expected utility using event-splitting considerations; our framework nests a subjective version of this model. Puri (2025) studies simplicity in risky choice through the support size of lotteries, a different measure of complexity than entropy. A bit further afield, Oprea (2020) studies the complexity of decision rules using automata theory, and Lipman (1995) models bounded rationality using coarse representations.

Narratives and coarse reasoning. Shiller (2017) defends the inclusion of "narratives" in formal economic analysis, defining them as simple, contagious explanations of events that drive decision-making. This focus on simplicity aligns with theories of coarse reasoning, such as Mullainathan et al. (2008), where agents act on limited representations by grouping distinct states into broad categories. Our framework operationalizes this preference for simple narratives through an entropic penalty on worst-case beliefs. By favoring low-entropy distortions, our model captures an agent who is most concerned by simple, concentrated explanations of failure—effectively "coarsening" the worst-case scenario—rather than by complex, diffused ones.

Information-theoretic preferences. The entropic penalty on worst-case beliefs relates to rational inattention (Sims, 2003) and information-theoretic models of choice (Matějka and McKay, 2015). In those frameworks, agents incur information processing costs proportional to entropy reduction. Here, we interpret the complexity parameter as a shadow price on the entropy of worst-case scenarios, linking preferences for simple explanations to information capacity constraints.[1]

Model selection and learning. While minimum description length (MDL; Rissanen, 1978) is a popular complexity measure for model selection, it is designed for settings with fixed exogenous data. Our framework differs in two ways: (i) data is endogenously generated by the agent's repeated action choices, and (ii) complexity operates through the agent's construction of worst-case beliefs within a robust criterion, not through model selection itself. Grünwald and Roos (2019, Section 6.3) surveys some challenges of using MDL under misspecification. More generally, the principle that simpler models are more robust motivates our approach. Blumer et al. (1987) show that hypotheses with lower VC dimension incur smaller generalization error in learning problems. We operationalize this principle through entropy penalties on worst-case beliefs: agents prefer concentrated explanations for observed data, which embodies the Occam's razor intuition that simpler models are more defensible under misspecification.

2 Static Environment

This section introduces the static complexity-augmented criterion, characterizes the worst-case beliefs in closed form, and shows how complexity aversion shapes pessimistic distortions.

[1] Formally, since our criterion adds entropy linearly to Nature's minimization problem, it is the Lagrangian form of an equivalent problem in which worst-case beliefs are chosen subject to an upper bound on their entropy. Hence, our complexity parameter is naturally interpreted as the shadow price of one extra unit of information-processing capacity, available to construct the adversarial belief.

2.1 Primitives and running example

Let A be a finite set of actions and Y a finite set of outcomes. For each $a \in A$ and $y \in Y$, the agent's per–period utility is $u(a, y) \in \mathbb{R}$.

Following Hansen and Sargent (2001) and Lanzani (2025a), a *structured model* specifies action-dependent outcome distributions $q = (q_a)_{a\in A} \in \Delta(Y)^A$. The agent entertains a finite set of such models $Q \subset \Delta(Y)^A$ and holds a posterior belief $\pi \in \Delta(Q)$.

For distributions $p, q \in \Delta(Y)$ with $p \ll q$, let

$$R(p\|q) := \sum_{y\in Y} p(y) \log \frac{p(y)}{q(y)}, \qquad H(p) := -\sum_{y\in Y} p(y) \log p(y)$$

denote relative entropy and Shannon entropy, respectively.

Assumption 1 (Full support)**.** *For each $q \in Q$, each $a \in A$, and each $y \in Y$, $q_a(y) > 0$.*

Throughout, we will use the following running example of an agent who must choose between a constant "safe" payoff and a state–dependent "risky" payoff under misspecified models. The example is rich enough to illustrate how worst-case beliefs are distorted and how the new parameter μ can flip the ranking between safe and risky, yet simple enough that all expressions can be written in closed form and reused in the dynamic analysis.

Example 1 (Safe vs risky arm)**.** *Let $A = \{r, s\}$, where r is a* risky *action and s is a* safe *action. Let $Y = \{g, b\}$ with g a "good" outcome and b a "bad" outcome. Utilities are*

$$u(r, g) = 1, \quad u(r, b) = 0, \qquad u(s, g) = u(s, b) = \bar{u} \in (0, 1).$$

Thus s delivers a constant payoff $\bar{u}$ and r delivers a high payoff in g and a low payoff in b.

The set of structured models is $Q = \{q^H, q^L\}$, where for each $q \in Q$,

$$q^H_r(g) = p_H, \quad q^H_r(b) = 1 - p_H, \qquad q^L_r(g) = p_L, \quad q^L_r(b) = 1 - p_L,$$

with $p_H > p_L$ so that q^H is "optimistic" and q^L is "pessimistic" about the risky arm. Under the safe arm both models agree: $q^H_s = q^L_s = \frac{1}{2}\delta_g + \frac{1}{2}\delta_b$. We take an arbitrary full–support prior $\pi_0 \in \Delta(Q)$, for instance $\pi_0(q^H) = \pi_0(q^L) = 1/2$.

2.2 Robust criterion with complexity aversion

We now introduce the complexity-augmented criterion. For each action a and model q, Nature chooses a distorted distribution p to minimize payoffs, subject to a KL penalty (robustness) and an entropic penalty (complexity). Fix parameters $\lambda > 0$ and $\mu \in \mathbb{R}$. For each action $a \in A$ and model $q \in Q$, the agent evaluates a by solving

$$v_{\lambda,\mu}(a;q) := \min_{p\in\Delta(Y)} \left\{ \sum_{y\in Y} u(a,y)p(y) + \frac{1}{\lambda} R(p\|q_a) + \mu H(p) \right\}. \tag{2.1}$$

The term $R(p_a\|q_a)/\lambda$ in (2.1) is the standard entropic penalty that discourages Nature from choosing models p_a that are too far from the structured model q_a, while the term $\mu H(p_a)$ penalizes or rewards the entropy of the distortion itself depending on the sign of μ. When $\mu = 0$, (2.1) becomes Hansen and Sargent's (2001) multiplier criterion. For $\mu > 0$, the adversary favors low–entropy (more concentrated) distortions; for $\mu < 0$, the adversary prefers high-entropy (more diffused) distortions. Notice that the penalty on the entropy is consistent with the penalty in Strzalecki (2024, eq. (6)).[2] When $\lambda \to 0$ and $\mu \neq 0$, (2.1) becomes $\mathbb{E}_{q_a}[u(a,y)] + \mu H(q_a)$, which is the subjective version of Mononen's (2025) entropy-modified expected utility preference that captures complexity aversion when the agent has no misspecification concerns. When $\lambda \to \infty$ and $\mu = 0$, (2.1) becomes Gilboa and Schmeidler's (1989) maxmin criterion where the set of subjective models is the entire simplex $\Delta(Y)$, which reflects an agent who is excessively concerned about misspecification. Section 5.3 discusses the interpretation of μ and Online Appendix D.1 shows that all our main results extend to a more general class of complexity functionals that nests Shannon entropy $H(.)$.

Given a posterior $\pi \in \Delta(Q)$, the overall worst-case value of action a is

$$V_{\lambda,\mu}(a;\pi) := \sum_{q\in Q} v_{\lambda,\mu}(a;q)\pi(q). \tag{2.2}$$

The associated static choice rule chooses any action in

$$\mathrm{BR}^{\mu}_{\lambda}(\pi) := \arg\max_{a\in A} V_{\lambda,\mu}(a;\pi).$$

The inner minimization (2.1) represents an adversarial process: given a model q and action a, Nature chooses a distorted distribution p_a that makes payoffs low, but pays a penalty for moving away from q_a and for using a complicated (high–entropy) explanation. The outer maximization over a in (2.2) captures the agent's robust choice given their posterior over Q. When $\mu = 0$ (no complexity penalty), (2.2) coincides with the *average robust control* (ARC) criterion in Lanzani (2025a), which is a special case of the "smooth Bayesian" criterion in Cerreia-Vioglio et al. (2025).

Thus, the new parameter μ trades off misspecification concerns against complexity concerns: a higher μ means that among all ways of being pessimistic, the agent prefers those

[2]Strzalecki (2024, p. 6) also interprets $\mu > 0$ as capturing an agent who prefers to have "simple theories of the world."

with simpler probability distributions. Our complexity-augmented criterion in (2.2) therefore provides a tractable and parsimonious unification of many existing decision criteria in the literature including Gilboa and Schmeidler (1989), Hansen and Sargent (2001), Lanzani (2025a), and Mononen (2025). Section 5.2 shows how our framework nests Gabaix (2025).

The next lemma characterizes the solution to (2.1) in closed form.

Lemma 1. *Suppose Assumption 1 holds and define*

$$\kappa := \frac{1}{\lambda} - \mu, \qquad \beta := \frac{1}{1 - \lambda\mu}.$$

If $\kappa > 0$ (i.e., $\mu < 1/\lambda$), then for every $(a, q) \in A \times Q$, (2.1) has a unique minimizer

$$\hat{p}_{\lambda,\mu}(a; q)(y) = \frac{\exp\{-u(a, y)/\kappa\} q_a(y)^\beta}{\sum_{z \in Y} \exp\{-u(a, z)/\kappa\} q_a(z)^\beta} \qquad (y \in Y). \tag{2.3}$$

Moreover, $v_{\lambda,\mu}(a; q)$ and $\hat{p}_{\lambda,\mu}(a; q)$ are continuous in (λ, μ, q) on $\{\kappa > 0\}$, and the following envelope identities hold for all (λ, μ, q) with $\kappa > 0$:

$$\frac{\partial}{\partial\mu} v_{\lambda,\mu}(a; q) = H\big(\hat{p}_{\lambda,\mu}(a; q)\big), \qquad \frac{\partial}{\partial\lambda} v_{\lambda,\mu}(a; q) = -\frac{1}{\lambda^2} R\big(\hat{p}_{\lambda,\mu}(a; q) \| q_a\big). \tag{2.4}$$

The worst–case distortion $\hat{p}_{\lambda,\mu}(a; q)$ has a Gibbs form: it reweights the model q_a by a payoff tilt $e^{-u(a,y)/\kappa}$ and a power–prior term $q_a(y)^\beta$. When $\mu = 0$, $\beta = 1$, so we recover the standard entropic distortion $\hat{p}_{\lambda,0} \propto q_a e^{-\lambda u}$. When $\mu > 0$, the coefficient κ shrinks and $\beta > 1$, so the distortion places more weight on low–utility outcomes and leans more heavily on the prior model q_a, resulting in a more concentrated and more pessimistic worst–case belief.

The envelope identities show that μ and λ operate through different channels. The marginal effect of μ on the worst–case value equals the entropy of the distortion, while the marginal effect of λ equals minus the KL divergence from q_a. Thus, misspecification concerns operate by allowing the worst–case belief to move farther from the reference model, while complexity concerns operate by restricting how diffuse that belief can be.

The next result shows that increasing μ strictly reduces the entropy of the worst-case distortion except in the knife–edge case where $u(a, \cdot)$ is affine in $\log q_a(\cdot)$.

Proposition 1 (Entropy strictly decreasing in complexity aversion)**.** *Fix $\lambda > 0$. For each $\mu < 1/\lambda$ and each (a, q), let $\hat{p}_{\lambda,\mu}(a; q)$ be the worst-case distortion from Lemma 1. Assume that for all (a, q), payoffs satisfy: $u(a, y) \neq b_1 + b_2 \log q_a(y)$ for any constants b_1, b_2. Then, $\mu \mapsto H(\hat{p}_{\lambda,\mu}(a; q))$ is strictly decreasing.*

Proposition 1 shows that μ really is a complexity parameter: as the agent becomes more complexity averse (higher μ), the “worst-case story” about outcomes becomes simpler and more concentrated. The condition that $u(a, \cdot)$ and q_a are nondegenerate rules out knife–edge cases where the distortion cannot move probability mass in a meaningful way. In those regular cases, the adversary optimally sacrifices entropy to make low–payoff states more likely when the agent puts more weight on simplicity.

2.3 Static illustration in the running example

This section illustrates the static effect of μ using the safe–versus–risky environment from Example 1. For concreteness, we fix some specific values for our primitives: $\lambda = 1$, $p_H = \frac{7}{10}$, $p_L = \frac{3}{10}$, $\bar{u} = \frac{7}{12}$, and consider a single model q for the risky arm with $q_r(g) = \frac{7}{10}$, $q_r(b) = \frac{3}{10}$, so π is degenerate at q.

Example 2. — Risky arm. *For the risky action r, the worst-case distortion $\hat{p}_{\lambda,\mu}(r; q)$ is given by (2.3). A simple numerical calculation yields:*

- *$\mu = 0$ (benchmark). Here $\kappa = 1, \beta = 1$ and*
$$\hat{p}_{\lambda,\mu}(r; q)(g) \approx 0.462, \qquad \hat{p}_{\lambda,\mu}(r; q)(b) \approx 0.538.$$
- *$\mu = \frac{4}{10}$ (complexity aversion). Here $\kappa = 0.6, \beta \approx 1.667$ and*
$$\hat{p}_{\lambda,\mu}(r; q)(g) \approx 0.437, \qquad \hat{p}_{\lambda,\mu}(r; q)(b) \approx 0.563.$$
- *$\mu = -\frac{4}{10}$ (complexity love). Here $\kappa = 1.4, \beta \approx 0.714$ and*
$$\hat{p}_{\lambda,\mu}(r; q)(g) \approx 0.473, \qquad \hat{p}_{\lambda,\mu}(r; q)(b) \approx 0.527.$$

As μ increases, the distortion becomes more concentrated and more pessimistic.

— Safe arm. *For the safe action s, the payoff $u(s, y) = \bar{u}$ is constant across states. The Gibbs form (2.3) then implies that the payoff tilt cancels and $\hat{p}_{\lambda,\mu}(s; q)$ is uniform on $\{g, b\}$ for all μ. Its entropy is therefore maximal and independent of μ, whereas for the risky arm the worst-case distribution becomes less entropic as μ rises.*

— Ranking of actions. *$V_{\lambda,\mu}(r; \pi)$ and $V_{\lambda,\mu}(s; \pi)$ are computed by plugging $\hat{p}_{\lambda,\mu}$ into (2.1) and averaging with respect to π. For the parameter values above, we can verify numerically that:*

- *for $\mu = 0$ the risky arm is strictly preferred to the safe arm;*
- *for $\mu > 0$ sufficiently large, the safe arm becomes strictly preferred because the worst-case distortion against r becomes very concentrated on the bad outcome;*

- *for all $\mu < 0$, the risky arm is more attractive because worst-case beliefs are forced to be diffuse, so there is a unique cutoff $\mu^* \in (0, \frac{4}{10})$ at which the ranking flips.*

This example illustrates that μ reweights the risky–safe tradeoff: complexity aversion allows simple (concentrated) worst-case narratives for risky outcomes, pushing toward safety.

3 Dynamic Environment and Long-Run Behavior

We embed our static criterion into the dynamic learning environment of Lanzani (2025a): the agent chooses actions over time, updates beliefs over a finite set of structured models by Bayes rule, and adjusts her misspecification concern based on relative model fit.

3.1 Dynamic environment

As in the static environment, there is a finite set of actions A and a finite outcome space Y. The agent's per–period payoff is given by a utility function $u : A \times Y \to \mathbb{R}$. The true data-generating process (DGP) is $p^\star = (p^\star_a)_{a\in A} \in \Delta(Y)^A$. The agent entertains a finite set of structured models $Q \subset \Delta(Y)^A$ and begins with a prior $\pi_0 \in \Delta(Q)$. At each time t, the agent observes outcomes, updates their posterior π_t via Bayes, and updates their misspecification concern λ_t based on how much the best model in Q underperforms an unrestricted model.

For $t \geq 0$, let $h_t = (a_1, y_1, \ldots, a_t, y_t) \in (A \times Y)^t$ denote the history up to time t, with $h_0 = \varnothing$. Let $H_t = (A \times Y)^t$ be the set of histories of length t, $\mathcal{F}_t$ the sigma–field generated by H_t, and $\mathcal{H} := \bigcup_{t\geq 0} H_t$ the set of all finite histories.

Following Lanzani (2025a), the agent evaluates the adequacy of the structured model class by comparing its likelihood to that of the unrestricted benchmark $\Delta(Y)^A$.[3] The agent's misspecification concern is allowed to depend on how well their models fit the data. Specifically, at each history h_t, the agent's misspecification concern depends on the log-likelihood ratio (LLR) between the structured model class Q and the unrestricted model class $\Delta(Y)^A$:

$$\mathrm{LLR}(h_t; Q) := -\log \frac{\max_{q\in Q} \prod_{\tau=1}^{t} q_{a_\tau}(y_\tau)}{\max_{p\in\Delta(Y)^A} \prod_{\tau=1}^{t} p_{a_\tau}(y_\tau)},$$

for all t and all $h_t \in H_t$. The numerator is the maximum likelihood of the data under the structured models in Q, whereas the denominator is the maximum likelihood under unrestricted models. A large LLR indicates a poor fit of Q to data, which suggests a high

[3]Since Y is finite and all structured models have full support, we follow Lanzani's (2025a, Section 2.2) suggestion by choosing the entire simplex $\Delta(Y)^A$ as the set of "alternative unstructured" models.

risk of specification risk. The agent normalizes this by time and a constant $c > 0$ to obtain:

$$\lambda_t := \frac{1}{ct}\,\mathrm{LLR}(h_t; Q),$$

which is interpreted as the *average* LLR. At each time t, given the pair (π_t, λ_t), the agent evaluates actions using the criterion of Section 2, with the complexity parameter set to

$$\mu_t := \bar{\mu}\,\lambda_t,$$

where $\bar{\mu} \geq 0$ is a fixed preference parameter. The period–t evaluation for action a becomes

$$v_t(a; q) = \min_{p \in \Delta(Y)} \left\{ \mathbb{E}_p[u(a, y)] + \frac{1}{\lambda_t} R(p \| q_a) + \bar{\mu}\lambda_t H(p) \right\}. \tag{3.1}$$

Scaling the complexity parameter proportionally to λ_t ensures that complexity aversion is active only when misspecification concerns are relevant. In contrast, a fixed $\mu_t \equiv \mu > 0$ would prevent convergence to standard Bayesian choice when the agent is correctly specified, which can reduce welfare in the long run (Online Appendix D.3). Thus, setting $\mu_t = \bar{\mu}\lambda_t$ should be viewed as a reduced-form representation of the idea that complexity aversion is countercyclical: it tightens when model fit deteriorates. When λ_t is small, the entropic penalty is negligible and the agent behaves approximately as a Bayesian; when λ_t is large, worst-case beliefs are both more pessimistic and more concentrated. As $\lambda_t \to 0$, both penalties vanish and the evaluation converges to Bayesian expected utility under π_t (Proposition 3).

More generally, the entropic penalty affects behavior only through the shape of the worst-case beliefs. When λ_t is large, worst-case distortions are both more pessimistic and more concentrated, reflecting a preference for simple adversarial explanations. When λ_t is small, the entropy term becomes negligible and the worst-case belief approaches the structured model. Since the static minimization problem preserves its convex structure, the worst-case belief retains the same closed–form characterization as in Lemma 1, with μ replaced by μ_t. Thus, all envelope and comparative–statics arguments carry over to the dynamic setting.

Assumption 2 (Uniform positivity of κ_t)**.** *There is $\bar{\lambda} < \infty$ such that $0 < \lambda_t \leq \bar{\lambda}$ for all t a.s. Fix $\bar{\mu} \geq 0$ such that $\bar{\mu}\bar{\lambda}^2 < 1$, so that $\kappa_t := \frac{1}{\lambda_t} - \bar{\mu}\lambda_t > 0$ for all t.*

At each date t, given the pair (λ_t, π_t), the agent evaluates action $a \in A$ using the static criterion with complexity aversion introduced in Section 2, where the complexity parameter is now $\mu_t = \bar{\mu}\lambda_t$. The value of action a at $(\lambda_t, \bar{\mu}, \pi_t)$ is

$$V_{\lambda_t,\mu_t}(a; \pi_t) := \sum_{q \in Q} v_{\lambda_t,\mu_t}(a; q)\pi_t(q),$$

where $v_{\lambda,\mu}$ is defined by (2.1) and $\mu_t = \bar{\mu}\lambda_t$. The associated best–reply correspondence is

$$\mathrm{BR}^{\bar{\mu}}_{\lambda_t}(\pi_t) := \arg\max_{a\in A} V_{\lambda_t,\mu_t}(a;\pi_t).$$

A (pure) *policy* is a measurable $\sigma : \mathcal{H} \to A$ that specifies an action for each history. Given a policy σ and the true DGP $p^\star$, the induced probability measure on histories is denoted $\mathbb{P}_\sigma$.

Definition 1 ($\bar{\mu}$–optimal policies)**.** *Fix $\bar{\mu} \geq 0$ satisfying $\bar{\mu}\bar{\lambda}^2 < 1$ and the updating rule (3.1). A policy σ is* $\bar{\mu}$–optimal *if, for every t and every history h_t,*

$$\sigma_t(h_t) \in \mathrm{BR}^{\bar{\mu}}_{\lambda_t}(\pi_t),$$

where $\lambda_t = \Lambda(h_t)$, $\mu_t = \bar{\mu}\lambda_t$, and π_t is the posterior induced by history h_t.

A $\bar{\mu}$–optimal policy is simply one that behaves myopically optimally in every period according to the (λ_t, μ_t)–criterion with $\mu_t = \bar{\mu}\lambda_t$. The agent updates beliefs and λ_t as in Lanzani (2025a), but when choosing actions they now care about the complexity of worst-case explanations, not only about their fit. This makes the dynamic impact of $\bar{\mu}$ transparent: any differences in long–run behavior compared to the baseline model are due to this changed static objective, not to any new learning assumption.

Definition 2 (Λ-limit frequencies)**.** *Let σ be a $\bar{\mu}$–optimal policy and let $\alpha_t(h_t) \in \Delta(A)$ denote the empirical frequency of actions up to time t on history h_t:*

$$\alpha_t(h_t)(a) := \frac{1}{t}\sum_{\tau=1}^{t} \mathbb{1}\{a_\tau = a\}, \qquad a \in A.$$

A random mixed action $\alpha^\Lambda \in \Delta(A)$ is a Λ–limit frequency *if there exists a $\bar{\mu}$–optimal policy σ such that, for every $\varepsilon > 0$,*

$$\mathbb{P}_\sigma\Big(\limsup_{t\to\infty} \big\|\alpha_t(h_t) - \alpha^\Lambda\big\|_1 \leq \varepsilon\Big) > 0.$$

Equivalently, with positive probability the empirical frequencies get and remain arbitrarily close to α^Λ along the realized history.

3.2 Mixed c–robust equilibria for the $\bar{\mu}$–criterion

We adapt Lanzani's static concept of mixed c–robust equilibrium to the $\bar{\mu}$–criterion. Fix the true DGP $p^\star$ and the set of structured models Q. For any mixed action $\alpha \in \Delta(A)$, let

$$D(q;\alpha) := \sum_{a\in A} \alpha(a) R\big(p^\star_a \| q_a\big)$$

denote the average model misfit of q under α, and define the set of best–fit models

$$Q(\alpha) := \arg\min_{q\in Q} D(q;\alpha).$$

Definition 3 (Mixed c–robust equilibrium)**.** *A triple* $(\alpha,\eta,\tau) \in \Delta(A)\times\Delta(Q)\times\mathbb{R}_+$ *is a* mixed c–robust equilibrium for the $\bar{\mu}$–criterion *if*

1. $\operatorname{supp}(\eta) \subset Q(\alpha)$*;*

2. *for every* $a \in A$ *with* $\alpha(a) > 0$*,* $a \in \mathrm{BR}^{\bar{\mu}}_{\tau}(\eta)$*, where the best–reply is evaluated at* $(\lambda,\mu) = (\tau,\bar{\mu}\tau)$*;*

3. *the misspecification concern* τ *equals the average misfit per unit* $c > 0$*:*
$$\tau = \frac{1}{c}\min_{q\in Q} D(q;\alpha) = \frac{1}{c}D(q;\alpha) \quad \textit{for all } q \in \operatorname{supp}(\eta).$$

A mixed c–robust equilibrium is a self–consistent long–run configuration. The mixed action α describes the frequencies with which the agent plays each action; $Q(\alpha)$ is the set of models that best fit the data generated by α; the belief η puts weight only on those well–fitting models; and τ is the strength of misspecification concern suggested by the data. Relative to Lanzani's framework, the only change is that best replies are evaluated using the complexity-augmented criterion.

The next result extends Lanzani (2025a, Theorem 3) to the $\bar{\mu}$–criterion.

Proposition 2. *Fix* $\bar{\mu} \geq 0$ *satisfying* $\bar{\mu}\bar{\lambda}^2 < 1$ *and a* $\bar{\mu}$*–optimal policy* σ*. Let* α^{Λ} *be any* Λ*–limit frequency of* σ*. Then, there exist* $\eta^{\bar{\mu}} \in \Delta(Q)$ *and* $\tau^{\bar{\mu}} \geq 0$ *such that* $(\alpha^{\Lambda},\eta^{\bar{\mu}},\tau^{\bar{\mu}})$ *is a mixed* c*–robust equilibrium for the* $\bar{\mu}$*–criterion.*

Proposition 2 says that the empirical behavior of a $\bar{\mu}$–optimal agent over the long run is described by a mixed c–robust equilibrium, so only the static notion of "best reply" has changed. Thus, any new dynamic phenomena we derive will be due to how $\bar{\mu}$ reshapes the set of mixed c–robust equilibria.

3.3 Correct specification and the Bayesian limit

We now describe what the dynamic framework implies when the agent is *correctly specified*. In this case, the structured family Q contains the true model $p^\star$, so the LLR statistic that determines λ_t eventually stops detecting misspecification.

Assumption 3 (Correct specification)**.** *There exist* $q^\star \in Q$ *such that* $q^\star_a = p^\star_a$*, for all* $a \in A$*.*

Assumption 3 says that the agent's structured models are rich enough to include a model $q^\star$ that coincides with the true DGP. Recall that statistic $\lambda_t = \Lambda(h_t)$ measures how much better the best model in Q fits the data than the best model in $\Delta(Y)^A$. Under correct specification, Q does *not* systematically underperform relative to $\Delta(Y)^A$, so the LLR statistic has no reason to grow, and therefore misspecification concerns will vanish in the long run.

Proposition 3. *Suppose Assumptions 1, 2, and 3 hold. Fix $\bar{\mu} \geq 0$ satisfying $\bar{\mu}\bar{\lambda}^2 < 1$. Let σ be any $\bar{\mu}$–optimal policy and let α^Λ be any Λ–limit frequency of σ. Then, for every action $a \in A$ with $\alpha^\Lambda(a) > 0$, $V_{\lambda_t, \bar{\mu}\lambda_t}(a; \pi_t) \longrightarrow \mathbb{E}_{p_a^\star}[u(a, y)]$ a.s.*

The mechanism has two parts. First, by Proposition 2, the Λ–limit frequency α^Λ corresponds to a mixed c–robust equilibrium $(\alpha^\Lambda, \eta^{\bar{\mu}}, \tau^{\bar{\mu}})$, and the long–run posterior π_t concentrates on the set of best–fitting models $Q(\alpha^\Lambda)$. For actions played infinitely often, the true model $q^\star$ must belong to $Q(\alpha^\Lambda)$ under correct specification, so the posterior eventually assigns all weight to models that coincide with $p^\star$ on the played actions. Second, because $\mu_t = \bar{\mu}\lambda_t \to 0$ as $\lambda_t \to 0$, both the robustness penalty and the complexity penalty vanish. Consequently, the worst-case distortion $\hat{p}_{\lambda_t, \mu_t}(a; q)$ converges to $q_a = p_a^\star$ for any $q \in Q(\alpha^\Lambda)$, and the agent's evaluation coincides with Bayesian expected utility. The specification $\mu_t = \bar{\mu}\lambda_t$ thus serves as "cognitive scaffolding": complexity constraints are instrumental tools for learning that are removed once the true model is identified on path.

3.4 Complexity aversion and elimination of λ–cycles

We now specialize to the two–action case $A = \{r, s\}$ of the running example and show that complexity aversion can eliminate λ–driven cycles. The key object is the value difference between r and s. Fix (λ, μ) with $\mu = \bar{\mu}\lambda$ and $\kappa = 1/\lambda - \mu > 0$, a posterior $\pi \in \Delta(Q)$, and distinguished actions $r, s \in A$. Define the value difference

$$\Delta(\lambda, \bar{\mu}, \pi) := V_{\lambda, \bar{\mu}\lambda}(r; \pi) - V_{\lambda, \bar{\mu}\lambda}(s; \pi). \tag{3.2}$$

For $a \in \{r, s\}$, define the average entropy of the worst-case distortions as

$$H_a(\lambda, \bar{\mu}, \pi) := \sum_{q \in Q} \pi(q) H\big(\hat{p}_{\lambda, \bar{\mu}\lambda}(a; q)\big). \tag{3.3}$$

Define the *switching surface* as

$$S := \big\{(\lambda, \bar{\mu}, \pi) : \Delta(\lambda, \bar{\mu}, \pi) = 0\big\} \tag{3.4}$$

i.e., the set of parameters where the agent is indifferent between r and s.

Lemma 2. *Suppose Assumption 1 holds. Then:*

1. *For each $(\lambda, \bar{\mu}, \pi)$ with $\kappa = 1/\lambda - \bar{\mu}\lambda > 0$, the partial derivative $\partial\Delta/\partial\bar{\mu}$ exists and is given by $\frac{\partial}{\partial\bar{\mu}}\Delta(\lambda, \bar{\mu}, \pi) = \lambda\big(H_r(\lambda, \bar{\mu}, \pi) - H_s(\lambda, \bar{\mu}, \pi)\big)$.*

2. *In particular, on the switching surface S in eq. (3.4), the sign of $\partial\Delta/\partial\bar{\mu}$ equals the sign of the entropy difference $H_r(\lambda, \bar{\mu}, \pi) - H_s(\lambda, \bar{\mu}, \pi)$.*

Lemma 2 shows that the marginal effect of complexity aversion on the risky–safe tradeoff is governed entirely by the difference in worst-case entropies. If the worst-case explanation of the risky arm is more concentrated (lower entropy) than that of the safe arm, then increasing $\bar{\mu}$ tilts the value difference in favor of the safe arm. The factor λ reflects the fact that the effective complexity penalty is $\mu = \bar{\mu}\lambda$: at higher levels of misspecification concern, the same increase in $\bar{\mu}$ has a larger effect on the value difference. More formally, the dynamic evolution $\mu_t = \bar{\mu}\lambda_t$ and the λ-factor in Lemma 2 formalizes the idea that complexity aversion is endogenously *countercyclical*: when data fit is poor and λ_t is high, the simplicity constraint is tighter. To obtain a clean dynamic selection result, we impose two assumptions.

Assumption 4 (Equilibrium range of λ)**.** *There exists a compact interval $[\underline{\lambda}, \bar{\lambda}] \subset (0, \infty)$ with the following property: for every parameter $\bar{\mu} \geq 0$ satisfying $\bar{\mu}\bar{\lambda}^2 < 1$ and every mixed c–robust equilibrium (α, η, τ) for the $\bar{\mu}$–criterion in the sense of Definition 3 such that $\alpha(r) > 0$, the associated misspecification concern τ belongs to $[\underline{\lambda}, \bar{\lambda}]$.*

Assumption 4 restricts attention to environments in which, whenever both r and s are in the support of the equilibrium randomization, the associated misspecification concern τ is neither arbitrarily small nor arbitrarily large. Instead, τ always lies in a fixed interval $[\underline{\lambda}, \bar{\lambda}]$. It holds whenever the true DGP lies in the interior of the convex hull of Q, so that no structured model can fit the long-run data arbitrarily well or arbitrarily poorly.

Assumption 5 (Uniform entropy gap)**.** *Let $[\underline{\lambda}, \bar{\lambda}]$ be as in Assumption 4. There exist constants $\bar{\mu}_0 \geq 0$ and $H^* > 0$ such that for all $(\lambda, \pi) \in [\underline{\lambda}, \bar{\lambda}] \times \Delta(Q)$ and all $\bar{\mu} \geq \bar{\mu}_0$ satisfying $\bar{\mu}\lambda^2 < 1$, the following implication holds: $\Delta(\lambda, \bar{\mu}, \pi) \geq 0 \Longrightarrow H_s(\lambda, \bar{\mu}, \pi) - H_r(\lambda, \bar{\mu}, \pi) \geq H^*$.*

By Lemma 2, Assumption 5 implies $\frac{\partial}{\partial\bar{\mu}}\Delta(\lambda, \bar{\mu}, \pi) \leq -\lambda H^*$ on the same domain. It imposes a uniform lower bound on the entropy gap $H_s(\lambda, \bar{\mu}, \pi) - H_r(\lambda, \bar{\mu}, \pi)$ over the equilibrium range $\lambda \in [\underline{\lambda}, \bar{\lambda}]$ and all admissible $(\bar{\mu}, \pi)$. In particular, whenever the agent is indifferent between risky and safe, the worst-case explanation of the safe arm is strictly more complex than that of the risky arm, and the difference is uniformly bounded. In the running example, this is natural: the safe arm has state–independent payoffs, so pessimism about s cannot be generated by concentrating on any particular state; pessimistic narratives about s must remain relatively diffuse. By contrast, pessimism about r can be implemented by piling probability on the bad state, yielding a low–entropy worst-case distribution. Thus, at any point of indifference, the safe arm is supported by a "richer" worst-case story.

Theorem 1 (Complexity aversion eliminates λ–cycles)**.** *Let the misspecification concern evolve according to the normalization rule $\lambda_t = \Lambda(h_t)$ in (3.1). Suppose $A = \{r, s\}$ and Assumptions 4 and 5 hold. Then, there exists $\bar{\mu}^* > 0$ such that the following holds for every $\bar{\mu} \geq \bar{\mu}^*$ with $\bar{\mu}\bar{\lambda}^2 < 1$:*

(i) *Every mixed c–robust equilibrium $(\alpha^{\bar{\mu}}, \eta^{\bar{\mu}}, \tau^{\bar{\mu}})$ for the $\bar{\mu}$–criterion satisfies $\alpha^{\bar{\mu}}(s) = 1$.*

(ii) *Consequently, for all $\bar{\mu} \geq \bar{\mu}^*$ every Λ–limit frequency α^{Λ} of any $\bar{\mu}$–optimal policy is pure and equals the safe action. In particular, the λ–driven cycles between r and s that arise in Lanzani's (2025a) $\bar{\mu} = 0$ benchmark cannot occur when $\bar{\mu}$ is sufficiently large.*

When $\bar{\mu} = 0$, Lanzani (2025a, Section 4) shows that the dynamic adjustment of λ_t can generate mixed c–robust equilibria in which the agent endogenously cycles between risky and safe: bad outcomes after r raise λ_t and push toward s, while good outcomes after s lower λ_t and favor r. With $\bar{\mu} > 0$, the static comparison at any point of indifference is tilted toward the safe arm because its worst-case explanation is more complex. Theorem 1 shows that, once $\bar{\mu}$ passes a threshold, these small tilts accumulate: mixed equilibria with both actions in support disappear, and only pure equilibria survive. The long–run dynamics therefore settle on a single action. This shows that a static preference for simple worst-case explanations can eliminate dynamic instability that is driven by misspecification concerns alone.

If Assumption 5 is dropped, the complexity penalty need not tilt incentives toward safety: the safe arm may admit an equally simple (low-entropy) worst-case narrative, or the entropy ranking may reverse, so increasing $\bar{\mu}$ need not shrink (and may even expand) the region of indifference and mixing. In that case, the endogenous adjustment of λ_t can continue to sustain mixed play and λ–cycles even for large $\bar{\mu}$, which is precisely what Theorem 1 rules out under Assumption 5. Online Appendix D.2 uses Chamberlain's (2020) portfolio-choice environment to illustrate that Assumption 5 holds naturally in familiar settings.

3.5 Complexity aversion and switching arms

3.5.1 Safe arm and multiple risky arms

We extend the running example (Example 1) to an environment with one safe arm s and $r_1, \ldots, r_K$ risky arms, for $K \geq 2$. For each risky arm r_i, define its risky–safe value difference

$$\Delta_i(\lambda, \bar{\mu}, \pi) := V_{\lambda,\bar{\mu}\lambda}(r_i; \pi) - V_{\lambda,\bar{\mu}\lambda}(s; \pi). \tag{3.5}$$

As in the two–arm case, the effect of complexity aversion on each risky–safe comparison is governed by the entropy gap between their worst-case distortions. When beliefs are parameterized by the posterior weight on an optimistic model, each risky arm admits a one-dimensional comparison against the safe arm. The slope of this comparison with respect to optimistic beliefs is constant, so each risky arm is characterized by a single belief threshold.

Assumption 6 (Uniform entropy gap for each risky arm)**.** *Let $[\underline{\lambda}, \bar{\lambda}]$ be as in Assumption 4. There exist constants $H_1^*, \ldots, H_K^* > 0$ such that for all $(\lambda, \pi) \in [\underline{\lambda}, \bar{\lambda}] \times \Delta(Q)$, all $\bar{\mu} \geq 0$ with $\bar{\mu}\lambda^2 < 1$, and for each risky arm r_i, $H_s(\lambda, \bar{\mu}, \pi) - H_{r_i}(\lambda, \bar{\mu}, \pi) \geq H_i^*$.*

Assumption 6 requires that the safe arm's worst-case distortion be uniformly more entropic than that of each risky arm over the relevant parameter range. This ensures that increasing $\bar{\mu}$ tilts every risky–safe comparison in favor of the safe arm.

Assumption 7 (Monotone effect of the optimistic model)**.** *For every $\lambda \in [\underline{\lambda}, \bar{\lambda}]$, every $\bar{\mu} \geq 0$ with $\bar{\mu}\lambda^2 < 1$, and every risky arm r_i, shifting posterior weight toward the optimistic model strictly raises the risky–safe value difference.*

Assumption 7 guarantees that learning in favor of the optimistic model raises the relative value of each risky arm, so each arm admits a well-defined belief threshold.

Proposition 4. *Suppose Assumptions 1, 4, 6, and 7 hold. Fix any $\lambda \in [\underline{\lambda}, \bar{\lambda}]$ and any $\bar{\mu}$ satisfying $\bar{\mu}\lambda^2 < 1$. Assume further that for each risky arm r_i, $\Delta_i(\lambda, \bar{\mu}, 0) < 0 < \Delta_i(\lambda, \bar{\mu}, 1)$. Then, for each risky arm r_i:*

(i) There exists a unique belief threshold $\theta_i^*(\bar{\mu}) \in (0, 1)$ such that r_i is (weakly) optimal relative to the safe arm if and only if the posterior weight on the optimistic model exceeds $\theta_i^*(\bar{\mu})$.

(ii) The threshold $\theta_i^*(\bar{\mu})$ is strictly increasing in $\bar{\mu}$.

Proposition 4 shows that each risky arm admits a belief threshold $\theta_i^*(\bar{\mu})$, and stronger complexity aversion raises these thresholds. Thus, as $\bar{\mu}$ increases, experimentation with risky arms requires increasingly optimistic beliefs. Moreover, the magnitude of the entropy gap governs how rapidly thresholds shift. Risky arms whose worst-case distortions are especially concentrated (low entropy) experience larger downward shifts in relative value as $\bar{\mu}$ rises, and are therefore eliminated first. Complexity aversion acts as a cross-sectional selection device among risky alternatives. In the dynamic model, these shifting thresholds imply that risky arms disappear sequentially as $\bar{\mu}$ increases.

3.5.2 Only risky arms

We now consider the case in which all arms are risky, so there is no safe benchmark. In this environment, we show that complexity aversion acts as a cross-sectional selection device among risky alternatives without necessarily inducing conservatism.

For any two actions $i, j \in A$, define the pairwise value difference

$$\Delta_{ij}(\lambda, \bar{\mu}, \pi) := V_{\lambda,\bar{\mu}\lambda}(i;\pi) - V_{\lambda,\bar{\mu}\lambda}(j;\pi).$$

As in the safe–risky case, the effect of complexity aversion on this comparison is governed by the entropy gap between the corresponding worst-case distortions.

Lemma 3. *Suppose Assumption 1 holds. Then, for any $i, j \in A$ and any $(\lambda, \bar{\mu}, \pi)$ with $\kappa = 1/\lambda - \bar{\mu}\lambda > 0$, $\frac{\partial}{\partial \bar{\mu}} \Delta_{ij}(\lambda, \bar{\mu}, \pi) = \lambda\big(H_i(\lambda, \bar{\mu}, \pi) - H_j(\lambda, \bar{\mu}, \pi)\big)$.*

Lemma 3 implies that increasing $\bar{\mu}$ tilts comparisons in favor of actions whose worst-case distortions are more entropic, and against those whose pessimistic narratives are simpler (more concentrated).

Pairwise entropy dominance. To obtain a clear elimination result, we require that one risky arm has a uniform entropy advantage over another across the relevant parameter range.

Assumption 8 (Pairwise entropy dominance)**.** *There exist distinct actions $i, j \in A$, a constant $H_{ij}^* > 0$, and a baseline $\bar{\mu}_0 \geq 0$ such that:*

(i) For all $(\lambda, \bar{\mu}, \pi) \in [\underline{\lambda}, \bar{\lambda}] \times [\bar{\mu}_0, 1/\bar{\lambda}^2) \times \Delta(Q)$, let $H_j(\lambda, \bar{\mu}, \pi) - H_i(\lambda, \bar{\mu}, \pi) \geq H_{ij}^*$.

(ii) The initial value advantage of i at $\bar{\mu}_0$ is not too large:

$$K_{ij} := \sup_{(\lambda,\pi)\in[\underline{\lambda},\bar{\lambda}]\times\Delta(Q)} \Delta_{ij}(\lambda, \bar{\mu}_0, \pi) < \underline{\lambda} H_{ij}^* \big(1/\bar{\lambda}^2 - \bar{\mu}_0\big). \tag{3.6}$$

Assumption 8 requires that arm j's worst-case distortions be uniformly more entropic than arm i's over the relevant parameter range, and that this entropy advantage be quantitatively strong enough to overturn any baseline value advantage of i as $\bar{\mu}$ increases.

Proposition 5 (Elimination without a safe arm)**.** *Suppose Assumptions 1 and 8 hold. Then, there exists* $\bar{\mu}^* \in (\bar{\mu}_0, 1/\bar{\lambda}^2)$ *such that for all* $\bar{\mu} \in [\bar{\mu}^*, 1/\bar{\lambda}^2)$, $\lambda \in [\underline{\lambda}, \bar{\lambda}]$, *and* $\pi \in \Delta(Q)$, $V_{\lambda,\bar{\mu}\lambda}(j;\pi) > V_{\lambda,\bar{\mu}\lambda}(i;\pi)$. *In particular, for all such* $\bar{\mu}$, *action* i *is never a best reply.*

Proposition 5 shows that sufficiently strong complexity aversion can eliminate a risky arm even when no safe alternative exists. In this case, complexity aversion does not simply induce conservatism, but instead selects among risky alternatives by favoring those whose worst-case narratives are more diffuse. It is useful to contrast this result with the safe–risky environment. With a safe benchmark, Assumption 6 ensures that the safe arm enjoys a uniform entropy advantage over all risky arms, so increasing $\bar{\mu}$ systematically pushes the agent toward safety. In the present setting, where all arms are risky, there is no canonical benchmark. Complexity aversion does not necessarily make the agent more conservative in any absolute sense; rather, it reshapes the relative ranking of risky actions.

Absent a global entropy ordering as in Assumption 8, Lemma 3 still implies that increasing $\bar{\mu}$ locally tilts pairwise comparisons toward arms with more entropic worst-case distortions. However, no arm need become uniformly dominated across the relevant parameter range. In such environments, the dynamic evolution of λ_t can continue to sustain switching among risky arms even at high levels of complexity aversion.

3.6 Complexity aversion and long–run payoffs

We now isolate a class of environments in which increasing complexity aversion is welfare-improving: it raises the agent's long-run realized payoff under the true DGP, even though it makes the agent more conservative over time. The key is that the true DGP actually favors the safe arm, so eliminating safe–risky cycles prevents the agent from over-experimenting with the risky action. When the safe action is ex-post optimal but the $\bar{\mu} = 0$ benchmark exhibits endogenous safe–risky cycles, sufficiently strong complexity aversion eliminates experimentation with the risky arm and raises long-run realized payoffs. In such environments, simplicity stabilizes behavior and prevents overreaction to temporarily favorable but weakly supported outcomes. Online Appendix D.3 provides the details and a formal welfare comparison. It also shows that complexity aversion alone can create welfare losses.

4 Applications

Sections 4.1 and 4.2 explore applications in discrete-choice analysis and stochastic growth, respectively. Online Appendix D.2 applies our framework to portfolio-choice environments.

4.1 Rational Inattention

This application shows how our complexity-augmented criterion in Section 2 delivers new, empirically relevant predictions when embedded in Matějka and McKay's (2015) rational inattention (RI) framework for discrete-choice analysis. The key is that complexity aversion induces *state-dependent choice sensitivity* even when the underlying information-cost technology is the standard Shannon RI cost.

4.1.1 Setup and definitions

We follow the framework in Matějka and McKay (2015). Let $A = \{1, \dots, n\}$ be a finite set of actions and Ω be a finite set of payoff states. In state $\omega \in \Omega$, action $a \in A$ yields payoff $v(a, \omega) \in \mathbb{R}$. The agent's *structured* or *reference* prior over Ω is $g \in \Delta(\Omega)$.

An RI strategy is a stochastic choice rule $\psi \in \Delta(A)^\Omega$ satisfying

$$\psi(a \mid \omega) \in [0, 1], \quad \text{and} \quad \sum_{a \in A} \psi(a \mid \omega) = 1,$$

for every state $\omega \in \Omega$. Let the (reference) unconditional choice probabilities be

$$\bar{\psi}(a) := \sum_{\omega \in \Omega} g(\omega) \psi(a \mid \omega) \in \Delta(A),$$

and for a cost parameter $\xi > 0$ define the standard Shannon entropy cost as

$$\mathcal{C}(\psi; g) := \xi \sum_{\omega \in \Omega} g(\omega) \sum_{a \in A} \psi(a \mid \omega) \log \frac{\psi(a \mid \omega)}{\bar{\psi}(a)}. \tag{4.1}$$

This quantity is ξ times the mutual information between ω and a under the reference prior g, which coincides with Matějka and McKay (2015, eq. (5)). We now introduce robustness to misspecification of the prior over payoff states, together with complexity aversion over worst-case scenarios. Given a strategy ψ, let the induced state-contingent payoff be

$$U_\psi(\omega) := \sum_{a \in A} v(a, \omega) \psi(a \mid \omega).$$

Then, Nature chooses a distorted distribution $m \in \Delta(\Omega)$ to make payoffs low, paying a KL penalty relative to g and an entropic penalty as in (2.1). Our agent anticipates this distortion and therefore chooses ψ according to our complexity-augmented criterion:

$$\max_{\psi\in\Delta(A)^{\Omega}} \min_{m\in\Delta(\Omega)} \left\{ \sum_{\omega\in\Omega} m(\omega)\, U_{\psi}(\omega) + \frac{1}{\lambda} R(m\|g) + \mu H(m) \; - \; \mathcal{C}(\psi; g) \right\}, \tag{4.2}$$

where $\lambda > 0$ is the misspecification concern and $\mu \geq 0$ is the complexity parameter. When the agent has no concern for misspecification and no complexity aversion, (4.2) coincides with Matějka and McKay (2015, eq. (10)). In the standard RI framework, the prior over payoff states is fixed at g and the optimal ψ has a generalized logit form with a *single* global scale parameter (Matějka and McKay, 2015, eq. (15)). In (4.2), the state distribution entering payoffs is endogenously distorted by robustness and disciplined by complexity. This interaction then implies that the *effective* choice sensitivity becomes state-dependent even though the information-cost technology (4.1) is unchanged.

4.1.2 Worst-case payoff states

Fix a stochastic rule ψ. The inner problem in (4.2) is the static criterion (2.1) applied to the outcome space Ω with reference g and utility U_{ψ}. Thus, the closed-form characterization in Lemma 1 applies directly.

Observation 1. *Suppose the prior g has full support on Ω and let $\kappa := \frac{1}{\lambda} - \mu$ and $\beta := \frac{1}{1-\lambda\mu}$ as in Lemma 1. If $\kappa > 0$, then for every ψ the inner minimization over m in (4.2) has a unique minimizer $m^{\star}_{\lambda,\mu}(\cdot;\psi) \in \Delta(\Omega)$ given by*

$$m^{\star}_{\lambda,\mu}(\omega;\psi) = \frac{\exp\{-U_{\psi}(\omega)/\kappa\}\, g(\omega)^{\beta}}{\sum_{\omega'\in\Omega} \exp\{-U_{\psi}(\omega')/\kappa\}\, g(\omega')^{\beta}} \qquad (\omega \in \Omega). \tag{4.3}$$

For fixed λ, increasing μ reduces the entropy of the worst-case distribution (Proposition 1).

Observation 1 formalizes the central "narrative" effect: robustness shifts weight toward low $U_{\psi}(\omega)$ states, while complexity aversion amplifies concentration by penalizing diffused distortions. Note that $m^{\star}_{\lambda,\mu}(\cdot;\psi)$ depends on ψ through U_{ψ}, so the worst-case distribution and the induced effective scale are determined jointly with the agent's choice rule.

4.1.3 Endogenous state-dependent logit scale

We now state the main new implication for observables in discrete-choice applications: the optimal stochastic choice rule inherits *state-dependent* sensitivity.

Define the state-specific *effective scale* induced by $(g, m^{\star})$:

$$\xi_{\omega}(\psi) := \xi\, \frac{g(\omega)}{m^{\star}_{\lambda,\mu}(\omega;\psi)} \in (0,\infty), \qquad \omega \in \Omega. \tag{4.4}$$

The effective scale $\xi_\omega(\psi)$ captures how pessimistic reweighting of state ω alters choice sensitivity. When the worst-case distribution overweights a state relative to the reference prior ($m^\star_{\lambda,\mu}(\omega;\psi) > g(\omega)$), the effective scale satisfies $\xi_\omega(\psi) < \xi$, so the agent behaves as if information were cheaper and choices were more deterministic. Conversely, states that are underweighted by the worst-case narrative exhibit higher effective noise.

Proposition 6. *Assume $\mu < 1/\lambda$ so that Observation 1 applies, and consider any solution $\psi^\star$ of (4.2) with $\bar{\psi}^\star(a) > 0$ for all $a \in A$. Then for each $\omega \in \Omega$ and $a \in A$,*

$$\psi^\star(a \mid \omega) = \frac{\bar{\psi}^\star(a) \exp\left\{\frac{v(a,\omega)}{\xi_\omega(\psi^\star)}\right\}}{\sum_{b\in A} \bar{\psi}^\star(b) \exp\left\{\frac{v(b,\omega)}{\xi_\omega(\psi^\star)}\right\}}, \tag{4.5}$$

where $\xi_\omega(\psi^\star)$ is defined by (4.4).

The proof of this and the other propositions in this section are in Online Appendix D.4.

Corollary 1. *Under the conditions of Proposition 6, for any $a, b \in \operatorname{supp} \bar{\psi}^\star$ and $\omega \in \Omega$,*

$$\log \frac{\psi^\star(a \mid \omega)}{\psi^\star(b \mid \omega)} = \log \frac{\bar{\psi}^\star(a)}{\bar{\psi}^\star(b)} + \xi_\omega(\psi^\star)^{-1}\big(v(a,\omega) - v(b,\omega)\big).$$

Hence, the slope of log choice odds with respect to payoff differences equals $\xi_\omega(\psi^\star)^{-1} = \xi^{-1}\, m^\star_{\lambda,\mu}(\omega;\psi^\star)/g(\omega)$.

Corollary 1 provides a key empirical prediction of our framework: choice sensitivity endogenously rises in states that the worst-case narrative overweights. This result therefore provides a new foundation for the pervasive practice of allowing scale heterogeneity in logit and mixed-logit estimation (e.g., Swait and Louviere, 1993; Fiebig et al., 2010).

4.1.4 Probability neglect

We next relate our framework to a phenomenon called "probability neglect": when emotions or salience push agents toward worst-case thinking, responses often appear to depend on the severity of a bad outcome more than on its probability (see, Sunstein, 2002, 2003). In our framework, probability neglect arises when complexity aversion drives the worst-case belief $m^\star$ toward a concentrated single scenario, which makes the scale ξ_ω small precisely there.

Corollary 2. *Fix $\lambda > 0$ and $\xi > 0$. Consider any ψ and the associated $m^\star_{\lambda,\mu}(\cdot;\psi)$ from (4.3). As $\mu \uparrow 1/\lambda$, $m^\star_{\lambda,\mu}(\cdot;\psi)$ concentrates on the set $\arg\min_{\omega\in\Omega}\left\{U_\psi(\omega) - \frac{1}{\lambda}\log g(\omega)\right\}$. If the minimizer is unique, say $\hat{\omega}$, then $m^\star_{\lambda,\mu}(\hat{\omega};\psi) \to 1$ and $m^\star_{\lambda,\mu}(\omega;\psi) \to 0$ for all $\omega \neq \hat{\omega}$, then*

$$\xi_{\hat{\omega}}(\psi) = \xi\,\frac{g(\hat{\omega})}{m^\star_{\lambda,\mu}(\hat{\omega};\psi)} \to \xi\, g(\hat{\omega}), \qquad \xi_\omega(\psi) \to \infty \text{ for } \omega \neq \hat{\omega}.$$

Thus, the induced logit rule (4.5) becomes comparatively "sharp" (high payoff sensitivity) in the selected worst-case state $\hat{\omega}$ and comparatively "flat" elsewhere.

Notice, for example, that when the prior g is uniform, the minimizer in Corollary 2 satisfies $\hat{\omega} = \arg\min_{\omega\in\Omega}\{U_\psi(\omega)\}$, i.e., $\hat{\omega}$ is the worst payoff state. Corollary 2 therefore provides a new mechanism for probability neglect without appealing to bounded rationality: a sophisticated agent overly concerned about complexity behaves as if a single adverse scenario dominates evaluation, and allocates decision precision accordingly. Notably, the standard Shannon RI framework does not generate such endogenous state-dependent sharpness because the scale parameter is global.

4.1.5 Discussion

Empirical discrete-choice analysis typically allows the logit scale to vary across contexts or individuals (i.e., scale heterogeneity) and develops flexible specifications to accommodate it (e.g., Swait and Louviere, 1993; Fiebig et al., 2010). Corollary 1 shows that scale heterogeneity can arise *endogenously* even with a fixed information-cost technology: it is driven by the likelihood ratio $m^\star/g$ generated by the robust-complexity criterion. Observable factors that increase pessimistic overweighting of particular states—such as adverse news, forecast failures, regime changes, or time pressure—raise the likelihood ratio $m^\star_{\lambda,\mu}(\omega)/g(\omega)$ and therefore reduce the effective scale ξ_ω. The model predicts that such factors generate state-specific increases in choice sensitivity, rather than a uniform reduction in noise. This provides a structural interpretation of scale heterogeneity in discrete-choice data based on robustness and complexity aversion, rather than preference shocks.

In summary, this application highlights what robustness and complexity aversion adds to the RI literature: it transforms a single global "noise" parameter into a structured, state-dependent object pinned down by primitives (prior g, payoffs, and the (λ, μ) criterion). This delivers a tractable bridge between robustness, Occam-style simplicity of narratives, and empirical evidence of context-dependent choice sensitivity.

4.2 Stochastic growth

We introduce our criterion in the stochastic-growth analysis of Robson et al. (2023).

4.2.1 Home bias

A key empirical regularity in international finance is *equity home bias*: domestic assets are overweighted relative to diversification benchmarks. In our framework, home bias can arise even when domestic and foreign assets have the *same* unconditional payoff distribution under the reference model. The mechanism is that, when $\mu > 0$, the criterion (2.1) penalizes diffuse worst-case distortions; this makes assets whose downside risk is concentrated in a small set of focal crisis scenarios (low-entropy narratives) endogenously more fragile. This channel is absent in entropy-free benchmarks (i.e., $\mu = 0$), including the stochastic-growth formulation in Robson et al. (2023), which depends only on a KL penalty and therefore cannot distinguish assets that share the same payoff distribution under the reference model.

There are two assets $A = \{d, f\}$ and outcomes $Y = \{y^*, y_1, \ldots, y_N\}$ with $N \geq 2$. Utilities are binary and identical across assets: $u(d, y^*) = u(f, y^*) = 1$, and $u(d, y) = u(f, y) = 0$ for all $y \in \{y_1, \ldots, y_N\}$. Hence, outcome y^* is a favorable scenario, whereas the remaining N outcomes are unfavorable scenarios that capture various economic downturns. Let $Q = \{q\}$ be a structured model, with $q_d, q_f \in \Delta(Y)$. Fix $\delta \in (0, 1)$ and $\varepsilon \in (0, \delta/N)$ and define

$$q_d(y^*) = q_f(y^*) = 1 - \delta, \qquad q_d(y) = \frac{\delta}{N} \quad \text{for } y = y_1, \ldots, y_N,$$
$$q_f(y_1) = \delta - (N-1)\varepsilon, \qquad q_f(y) = \varepsilon \quad \text{for } y = y_2, \ldots, y_N.$$

Thus, under q both assets deliver payoff 1 with probability $1 - \delta$ and payoff 0 with probability δ, but the foreign downside mass is concentrated on a single crisis state. Lemma 1 implies the closed-form value for each asset $a \in \{d, f\}$ is $v_{\lambda,\mu}(a; q) = -\kappa \log Z_a$, where $Z_a := e^{-1/\kappa} q_a(y^*)^\beta + \sum_{i=1}^N q_a(y_i)^\beta$, so the *home-bias premium* is $v_{\lambda,\mu}(d; q) - v_{\lambda,\mu}(f; q) = \kappa \log \frac{Z_f}{Z_d}$.

Proposition 7. *For every $\lambda > 0$,*

$$v_{\lambda,0}(d; q) = v_{\lambda,0}(f; q) \qquad \textit{and} \qquad v_{\lambda,\mu}(d; q) > v_{\lambda,\mu}(f; q) \quad \textit{for all } \mu \in (0, 1/\lambda).$$

In Robson et al. (2023), only the likelihood cost of a distorted model and the associated payoff tilt matter. As a result, relabeling or reshuffling which particular states deliver a loss has no effect, because the benchmark criterion depends on *how much* probability mass is shifted, not on *how concentrated* the implied downside scenarios are. In contrast, with our complexity-augmented criterion a foreign asset whose losses are concentrated in a small number of focal states is strictly less attractive when $\mu > 0$ holding fixed the unconditional payoff distribution, because the worst-case distortion can load on a simpler narrative. This delivers a parsimonious preference-based rationale for home bias that does not rely on expected-return differences, hedging motives, or segmentation: investors avoid foreign exposure precisely when the perceived foreign downside is concentrated around a small number of prominent crisis scenarios. Thus, Proposition 7 links home bias to the *structure* of tail risk, which generates sharp comparative statics that the stochastic-growth benchmark cannot deliver. Note also that since $v_{\lambda,\mu}(d;q) \geq v_{\lambda,\mu}(f;q) \iff Z_d \leq Z_f$ in our framework, home bias disappears (or reverses) whenever domestic tail risk is at least as concentrated as foreign tail risk in the power-sum sense $\sum_{i\geq 1} q_a(y_i)^\beta$ induced by $\beta > 1$.

The magnitude of home bias is potentially large even holding the probability of downside scenarios δ fixed: along the extreme-concentration limit $\varepsilon \downarrow 0$,

$$v_{\lambda,\mu}(d;q) - v_{\lambda,\mu}(f;q) \to \kappa(\beta - 1)\log N = \mu \log N.$$

Thus, the mechanism can produce an arbitrarily strong tilt toward home bias as either complexity aversion μ rises or the number of distinct downside scenarios N grows. What matters in our analysis is not "rarity" per se but the low-entropy downside support: for $\mu > 0$, any relabelling that concentrates probability within the loss set $\{y_1, \ldots, y_N\}$ increases the foreign term $\sum_{i\geq 1} q_f(y_i)^\beta$ and therefore decreases $v_{\lambda,\mu}(f;q)$. Rare crises amplify the mechanism when robustness is tight (large μ) or payoffs are very low in those crisis states, because the Gibbs tilt in Lemma 1 loads the worst-case distortion on low-utility states even if they are ex ante unlikely. Also note that whether induced home bias raises or lowers realized returns depends on the true joint distribution of domestic/foreign payoffs and the associated risk premia: home bias improves outcomes precisely when the concentrated foreign tail corresponds to genuine downside risk under the true DGP, and is costly when it induces under-diversification without compensating tail-risk reduction.

4.2.2 Endogenous growth rates

In Robson et al. (2023, Proposition 4), the growth-rate loss from optimizing under misspecification equals the degree of misspecification, measured by the KL divergence between the structured model and true DGP, so long-run growth does not depend on the payoff environment for a given degree of misspecification. We show that with complexity aversion, misspecification losses become *payoff-dependent* because the relevant choice rule is now disciplined by simplicity and therefore interacts with how returns load on states.

Complexity-augmented criterion. Fix $\mu \in [0, 1)$. Following Robson et al. (2023), let A be a finite set of actions and let $Q = \{p'\}$ be a singleton structured model $p' \in \Delta(Y)$. For each mixed strategy (or *portfolio*) $\alpha \in \Delta(A)$ and state $y \in Y$,

$$G_\mu(\alpha, y) = \min_{q \in \Delta(A)} \Big\{ \mathbb{E}_q[-u(a, y)] + R(q \| \alpha) + \mu H(q) \Big\}, \tag{4.6}$$

is our complexity-augmented criterion, where $\lambda = 1$ and $-u$ are imposed to match Robson et al. (2023, eq. (7)) when $\mu = 0$. Given a model $p \in \Delta(Y)$, the portfolio problem is

$$V_\mu(p) := \max_{\alpha \in \Delta(A)} \mathbb{E}_p\big[G_\mu(\alpha, y)\big], \qquad \alpha^*_\mu(p) \in \arg\max_{\alpha \in \Delta(A)} \mathbb{E}_p\big[G_\mu(\alpha, y)\big].$$

Given $p, p' \in \Delta(Y)$, the function

$$L_\mu(p, p') := \mathbb{E}_p\left[G_\mu(\alpha^*_\mu(p), y)\right] - \mathbb{E}_p\left[G_\mu(\alpha^*_\mu(p'), y)\right] \tag{4.7}$$

is the reduction of the growth rate caused by the optimization for the misspecified model p' when the true model is p. When $\mu = 0$, $L_0(p, p')$ is the loss in Robson et al. (2023, p. 371). In Robson et al. (2023, Proposition 4), two environments with the same $R(p\|p')$ exhibit the same growth loss function under their Regularity Condition 2.[4]

Proposition 8. *Under* (4.6)*, model-misspecification costs are not universal: even when the misspecified model satisfies Robson et al. (2023, Regularity Condition 2) and yields* $L_0(p, p') = R(p\|p')$*, the loss* $L_\mu(p, p')$ *in* (4.7) *need not equal* $R(p\|p')$ *for* $\mu > 0$.

In our framework, the same $R(p\|p')$ can generate very different growth losses even when Robson et al. (2023, Regularity Condition 2) is satisfied, because complexity aversion penalizes diffuse sampled-choice rules, so the cost of a given belief error depends on how much randomization is required to implement growth-optimal behavior in that payoff environment. This captures a key empirical pattern in macro-finance and growth: belief disagreements and "narrative shocks" tend to have large effects when portfolios load on a small set of salient states (crisis scenarios), while similarly sized misspecifications in calm environments have negligible effects on growth. Robson et al.'s (2023, Proposition 4) payoff-invariant loss rules out this heterogeneity, whereas our framework is able to capture it in a tractable way.[5]

5 Representation

This section provides a representation result for our criterion by restating it in an Anscombe—Aumann framework as in Lanzani (2025b). The outcome space Y introduced earlier now

[4] Regularity Condition 2 requires that the degree of misspecification is not too large (Assumption 10).

[5] Eden et al. (2026) shows that in long-horizon stochastic-growth problems, a standard expected-utility maximizer is generically driven by atypical, rare-shock frequencies. We do not study how expected utility amplifies rare events, but how adding complexity aversion changes which downside scenarios matter.

plays the role of a state space, and an act is a mapping $f : Y \to X$ from states to consequences in a convex consequence space X. This reformulation does not change the economics: in the static model, each action a induces a payoff profile $y \mapsto u(a, y)$, which can be identified with an act over Y.

Fix an affine, nonconstant utility index $u : X \to \mathbb{R}$. For each $(f, q) \in X^Y \times Q$, define the analogues of (2.1) and (2.2) as

$$v_{\lambda,\mu}(f; q) := \min_{p\in\Delta(Y)} \left\{ \mathbb{E}_p[u(f)] + \frac{1}{\lambda} R(p\|q) + \mu H(p) \right\}, \tag{5.1}$$

and

$$V_{\lambda,\mu}(f; \pi) := \sum_{q\in Q} \pi(q) v_{\lambda,\mu}(f; q). \tag{5.2}$$

The induced preference, denoted $\succeq$, satisfies $f \succeq g$ if and only if $V_{\lambda,\mu}(f; \pi) \geq V_{\lambda,\mu}(g; \pi)$. When $\mu = 0$ (i.e., no complexity penalty), (5.1) becomes Hansen and Sargent's (2001) standard multiplier criterion and (5.2) becomes the ARC criterion in Lanzani (2025b, eq. (6)). This yields the following hierarchy of representations: the multiplier criterion is a special case of the standard ARC criterion when Q is a singleton, and the ARC criterion is a special case of our complexity-augmented criterion when $\mu = 0$.

5.1 ARC representation

Given a constant $\beta \in \mathbb{R}$ and $q \in \Delta(Y)$, let $Z(q) := \sum_{y\in Y} q(y)^\beta$ be a normalizing constant and define the *power-transform* of q as

$$\tilde{q}(.) := \frac{q(.)^\beta}{Z(q)} \in \Delta(Y). \tag{5.3}$$

Let $\widetilde{Q} := \{\tilde{q} : q \in Q\} \subset \Delta(Y)$ and let $\widetilde{\pi} \in \Delta(\widetilde{Q})$ be the pushforward of π under $q \mapsto \tilde{q}$.[6]

Theorem 2. *Let $\mu \in \mathbb{R}$, finite $Q \subset \Delta(Y)$, and $\pi \in \Delta(Q)$. The following hold:*

1. *Suppose $\lambda > 0$, $\mu < 1/\lambda$, and define $\tilde{q}$ by (5.3). Then, $\kappa > 0$ and there exists a constant $\widetilde{C}(\lambda, \mu, \pi) \in \mathbb{R}$, independent of acts, such that for all acts f,*

$$V_{\lambda,\mu}(f; \pi) = \widetilde{C}(\lambda, \mu, \pi) + \sum_{q\in Q} \pi(q) \min_{p\in\Delta(Y)} \left\{ \mathbb{E}_p[u(f)] + \kappa R(p\|\tilde{q}) \right\}. \tag{5.4}$$

Equivalently, $\succeq$ is ordinally equivalent to an ARC preference with parameter $\lambda_{\text{ARC}} := 1/\kappa$ and structured models $\widetilde{Q}$ endowed with belief $\widetilde{\pi}$.

[6] The power-transform distribution $\tilde{q} \propto q^\beta$ in (5.3) is called an "escort distribution" in statistical physics and information geometry (e.g., Beck and Schögl, 1993).

2. *Suppose $\lambda > 0$ and $\mu \geq 1/\lambda$. Then, the preference induced by $V_{\lambda,\mu}(\cdot;\pi)$ is a variational preference but not an ARC preference. In particular, for all acts f,*

$$V_{\lambda,\mu}(f;\pi) = \sum_{q\in Q} \pi(q) \min_{y\in \mathrm{supp}(q)} \Big\{ u(f(y)) + \frac{1}{\lambda} \log \frac{1}{q(y)} \Big\}.$$

which is independent of μ on the entire region $\mu > 1/\lambda$.

3. *Suppose $\lambda \to 0$ and $\mu \in \mathbb{R}$ is fixed. Then, $\lim_{\lambda\to 0} V_{\lambda,\mu}(\cdot;\pi) = \sum_{q\in Q} \pi(q)\big(\mathbb{E}_q[u(\cdot)] + \mu H(q)\big)$ induces a subjective expected utility preference. When $Q = \{q\}$, $\lim_{\lambda\to 0} v_{\lambda,\mu}(\cdot;q)$ induces the subjective version of Mononen's (2025) entropy-modified expected utility preference.*

Theorem 2 classifies the induced preference into three regimes. In the low-simplicity region ($\mu < 1/\lambda$), the preference is equivalent to an ARC criterion computed against a transformed within-model benchmark (5.3), so simplicity affects behavior only through a change in the effective reference distributions. In the high-simplicity region ($\mu \geq 1/\lambda$), the criterion becomes extreme-state within each structured model and no longer induces an ARC preference; moreover, beyond the threshold, choices are unaffected by higher complexity aversion. Finally, in the vanishing-misspecification limit, the induced preference reduces to subjective expected utility preference with an act-independent additive entropy term.

5.2 Special case: Gabaix (2025)

Theorem 2.2 shows that in the extreme-simplicity regime ($\mu \geq 1/\lambda$), worst-case beliefs collapse to Dirac measures—Nature is forced to construct pessimism using a single salient adverse scenario per model. We show that this extreme-state limit coincides with Gabaix's (2025) "first-order" complexity aversion, where dread (a linear utility penalty proportional to choice complexity) makes choice completely insensitive to some characteristics, as in lasso regression and Gabaix (2014). In our model, this extreme underreaction occurs when the simplicity motive is so strong that pessimism is driven entirely by one focal disaster scenario.

In the example in Section 3.1, the agent's utility absent complexity considerations is $\widetilde{V}(b) := \bar{v} + (2\bar{b} - b^2) - (1-m)b^2$, where $\bar{v}$ is an exogenous constant, $\bar{b}$ is the unknown optimal value of b for a rational agent, and $m \in [0,1]$ is an exogenous parameter related to the exogenous precision of the agent's signal. The agent also has first-order "complexity aversion" $\mathrm{C}^{CA}(b) := 2\xi\sigma_a(1-m)|b|$, so their overall objective is

$$V^G(b) = \widetilde{V}(b) - 2\xi\sigma_a(1-m)|b|, \tag{5.5}$$

which is maximized by the threshold rule $b^G = \frac{\max\{b^r - \xi\sigma_a(1-m), 0\}}{2-m}$.[7]

To embed this problem into our static framework, we introduce an auxiliary "mistake narrative" state that Nature can distort. Fix a binary narrative space $Y = \{0, 1\}$. Interpret $y = 1$ as a salient "mistake" narrative (the agent focuses on an error scenario) and $y = 0$ as "no salient mistake." For each $b \in \mathbb{R}$, define the act $f_b : Y \to \mathbb{R}$ by $f_b(0) = 0$, and $f_b(1) = -\gamma|b|$, where $\gamma > 0$ is the marginal dread from exposure $|b|$. We will match γ to $2\xi\sigma_a(1-m)$, and suppose there is a single structured narrative model q and it is symmetric, i.e., $q(0) = q(1) = \frac{1}{2}$.

Proposition 9. *Under high-simplicity regime $\mu \geq 1/\lambda$, our criterion reduces via Theorem 2.2 to the extreme-state form. Applied to the indexation act f_b: $v_{\lambda,\mu}(f_b; q) = \frac{1}{\lambda}\log 2 - \gamma|b|$. When $\gamma = 2\xi\sigma_a(1-m)$, our criterion $V^*(b) := \widetilde{V}(b) + v_{\lambda,\mu}(f_b; q)$ coincides with (5.5) up to an additive constant, and optimal indexation follows Gabaix's threshold rule b^G.*

When $\mu \geq 1/\lambda$, Nature's worst-case distortion becomes a Dirac mass on a single adverse state. Economically, the agent evaluates actions *as if* pessimism is driven by one salient scenario, not by a carefully balanced portfolio of bad outcomes. This generates literal probability neglect: the criterion depends only on the worst payoff. For $\mu < 1/\lambda$, the worst-case distortion $\hat{p}_{\lambda,\mu}(1; b)$ in (2.3) remains interior: $\frac{d}{d|b|} v_{\lambda,\mu}(f_b; q) = -\gamma \hat{p}_{\lambda,\mu}(1; b)$, where $\hat{p}_{\lambda,\mu}(1; b)$ increases continuously in $|b|$. This expression predicts smooth *partial* simplicity (many small but nonzero choices) rather than sharp bunching at exactly zero. This behavior is consistent with empirical evidence, which tends to favor partial dampening as the baseline case, with full dampening arising in more extreme environments.[8]

5.3 Discussion: interpretation of μ

Theorem 2 shows that the role of μ depends critically on its magnitude relative to λ. When $\mu < 1/\lambda$, the preference is ordinally equivalent to an ARC preference computed under the transformed reference models $\widetilde{Q}$ in (5.3). Since ARC does not identify Q uniquely, μ should be understood as selecting a particular effective representation rather than as transforming a uniquely defined set of models.

[7] Note that unlike rational losses this is not second order in the neighborhood of the optimal choice.

[8] In insurance choice, Barseghyan et al. (2013) find probability distortions but only mild insensitivity to changes in probabilities, indicating attenuated rather than shut-down responses. In tax salience, Chetty et al. (2009) show that consumers underreact to taxes that are not salient, again implying partial but nonzero adjustment. More broadly, Handel and Schwartzstein (2018) emphasizes that interventions often move choices only part-way toward frictionless benchmarks.

When $\mu \geq 1/\lambda$, the entropic penalty forces the criterion to collapse to the extreme-state form in Theorem 2.2. In this region, the induced preference is variational but not ARC, and for $\mu > 1/\lambda$ the criterion is independent of μ. Thus, only the region $\mu < 1/\lambda$ delivers a smooth comparative statics of complexity aversion within the ARC family, whereas beyond the threshold, the ARC structure collapses and additional simplicity has no behavioral bite. Notice also that this is the representation that captures probability neglect as illustrated in Sections 4.1.4 and 5.2.

This classification also sharpens our dynamic interpretation. In the ARC case, $\mu > 0$ selects representations that can eliminate robustness-driven cycles. In the extreme-state case, stability is driven by the collapse of worst-case beliefs to the most adverse outcomes.

6 Conclusion

We developed a unified framework that combines a concern for model misspecification with a preference for simplicity. By penalizing the entropy of worst-case robust distortions, our agent balances pessimism against the plausibility of adversarial narratives. This mechanism yields sharp dynamic predictions: it breaks the indifference in standard robust control by penalizing risky actions that admit simple disaster narratives, thereby eliminating cycles, favoring safety and long-run stability. Our framework also offers microfoundations for empirical phenomena such as scale heterogeneity, probability neglect, and home bias. By disciplining robustness with Occam's razor, our framework establishes a formal link between statistical learning and behavioral regularities in decision-making under uncertainty. We focused on myopic decision making to avoid the complications arising from combining misspecification concerns and dynamic consistency flagged by Lanzani (2025a); we hope that future research will extend our framework to forward-looking agents.

A Appendix: Proofs from Section 2

A.1 Proof of Lemma 1

Fix (λ, μ) with $\kappa := 1/\lambda - \mu > 0$, an action $a \in A$, and a model $q \in Q$. Define

$$\Phi(\lambda, \mu, p_a; q) := \sum_{y \in Y} u(a, y) p_a(y) + \frac{1}{\lambda} R(p_a \| q_a) + \mu H(p_a).$$

We solve $\min_{p_a \in \Delta(Y)} \Phi(\lambda, \mu, p_a; q)$.

Step 1: Existence and Uniqueness. Rewrite the objective as

$$\Phi(\lambda, \mu, p_a; q) = \sum_{y \in Y} [u(a,y)p_a(y) + \kappa\, p_a(y) \log p_a(y)] - \frac{1}{\lambda} \sum_{y \in Y} p_a(y) \log q_a(y),$$

where we use $0 \log 0 := 0$. Since $\kappa > 0$, the function $x \mapsto x \log x$ (with $0 \log 0 = 0$) is strictly convex on $[0,1]$, making the objective strictly convex on $\Delta(Y)$. The simplex $\Delta(Y)$ is compact, so a unique global minimizer $p^* \in \Delta(Y)$ exists.

Step 2: Interiority of the Minimizer. Suppose $p^*(y_0) = 0$ for some $y_0 \in Y$. Then there exists $y_1 \in Y$ with $p^*(y_1) > 0$. For $t \in (0, p^*(y_1))$, define the perturbation p_t by

$$p_t(y) = \begin{cases} t & \text{if } y = y_0 \\ p^*(y_1) - t & \text{if } y = y_1 \\ p^*(y) & \text{otherwise.} \end{cases}$$

Note that $p_t \in \Delta(Y)$ by construction. Define the univariate function $\psi(x) := u(a, y_0)x + \kappa x \log x$. Then $\Phi(p_t) - \Phi(p^*) = \psi(t) - \psi(0) + [\psi_{y_1}(p^*(y_1) - t) - \psi_{y_1}(p^*(y_1))]$, where $\psi_{y_1}(x) := u(a, y_1)x + \kappa x \log x$. Dividing by t and taking $t \downarrow 0^+$:

For the first term:

$$\lim_{t \downarrow 0^+} \frac{\psi(t) - \psi(0)}{t} = \lim_{t \downarrow 0^+} [u(a, y_0) + \kappa(\log t + 1)] = u(a, y_0) + \kappa \cdot (-\infty) = -\infty,$$

since $\kappa > 0$ and $\lim_{t \downarrow 0^+} \log t = -\infty$. For the second term: Since $p^*(y_1) > 0$, the function ψ_{y_1} is differentiable in a neighborhood of $p^*(y_1)$ (away from 0), so

$$\lim_{t \downarrow 0^+} \frac{\psi_{y_1}(p^*(y_1) - t) - \psi_{y_1}(p^*(y_1))}{-t} = \psi'_{y_1}(p^*(y_1)) \in \mathbb{R}.$$

Thus the directional derivative in direction e_{y_0} is $-\infty$, contradicting optimality of p^*, so $p^*(y) > 0$ for all $y \in Y$.

Step 3: First-Order Conditions and Closed Form. Since p^* is interior, standard Lagrangian analysis applies. Let γ be the multiplier for $\sum_y p_a(y) = 1$. The Lagrangian is

$$\mathcal{L}(p_a, \gamma) = \sum_{y \in Y} u(a,y)p_a(y) + \frac{1}{\lambda} \sum_{y \in Y} p_a(y) \log \frac{p_a(y)}{q_a(y)} + \mu \sum_{y \in Y} (-p_a(y) \log p_a(y)) + \gamma \Big(\sum_{y \in Y} p_a(y) - 1 \Big).$$

Taking $\partial \mathcal{L} / \partial p_a(y) = 0$ for each y:

$$u(a,y) + \frac{1}{\lambda}(\log p_a(y) - \log q_a(y) + 1) + \mu(-\log p_a(y) - 1) + \gamma = 0.$$

Rearranging: $\left(\frac{1}{\lambda}-\mu\right)\log p_a(y)=\frac{1}{\lambda}\log q_a(y)-u(a,y)+C$, where C is independent of y. Using $\kappa=1/\lambda-\mu>0$ and $\beta=1/(\lambda\kappa)$: $\log p_a(y)=-\frac{u(a,y)}{\kappa}+\beta\log q_a(y)+\text{constant}$. Exponentiating and normalizing yields: $\hat{p}_{\lambda,\mu}(a;q)(y)=\frac{\exp\{-u(a,y)/\kappa\}q_a(y)^\beta}{\sum_{z\in Y}\exp\{-u(a,z)/\kappa\}q_a(z)^\beta}$.

Step 4: Continuity and Envelope Identities. By the implicit function theorem applied to the FOCs at an interior minimizer, $\hat{p}_{\lambda,\mu}(a;q)$ depends continuously on (λ,μ,q) for $\kappa>0$. Since $v_{\lambda,\mu}(a;q)=\Phi(\lambda,\mu,\hat{p}_{\lambda,\mu}(a;q);q)$ is a composition of continuous functions, it is continuous in (λ,μ,q) on $\{\kappa>0\}$. For the envelope identities: since $\hat{p}_{\lambda,\mu}(a;q)$ is the unique interior minimizer, by the envelope theorem (valid for unconstrained optimization when the minimizer is interior and the objective is differentiable in the parameters), we have

$$\frac{\partial v_{\lambda,\mu}}{\partial\mu}(a;q)=\frac{\partial\Phi}{\partial\mu}\Big|_{p_a=\hat{p}_{\lambda,\mu}(a;q)}=H(\hat{p}_{\lambda,\mu}(a;q)),$$

$$\frac{\partial v_{\lambda,\mu}}{\partial\lambda}(a;q)=\frac{\partial\Phi}{\partial\lambda}\Big|_{p_a=\hat{p}_{\lambda,\mu}(a;q)}=-\frac{1}{\lambda^2}R(\hat{p}_{\lambda,\mu}(a;q)\|q_a).$$

which coincide with (2.4). □

A.2 Proof of Proposition 1

Lemma 4 (Entropy in a finite exponential family)**.** *Let Y be a finite set and $\varphi:Y\to\mathbb{R}$ be nonconstant. For each $t>0$ define*

$$Z(t):=\sum_{y\in Y}e^{t\varphi(y)},\qquad A(t):=\log Z(t),\qquad p_t(y):=\frac{e^{t\varphi(y)}}{Z(t)}\quad(y\in Y),$$

and let $H(t):=-\sum_{y\in Y}p_t(y)\log p_t(y)$ be the Shannon entropy of p_t. Then:

1. *The function A is twice continuously differentiable on $(0,\infty)$, and for all $t>0$,*
$$A'(t)=\sum_{y\in Y}\varphi(y)p_t(y),\qquad A''(t)=\sum_{y\in Y}\big(\varphi(y)-A'(t)\big)^2p_t(y).$$
2. *For all $t>0$, $H(t)=A(t)-tA'(t)$.*
3. *For all $t>0$, $H'(t)=-tA''(t)=-t\sum_{y\in Y}\big(\varphi(y)-A'(t)\big)^2p_t(y)$. In particular, since φ is nonconstant and $p_t(y)>0$ for all y, we have $A''(t)>0$ and hence $H'(t)<0$ for all $t>0$. Thus H is strictly decreasing on $(0,\infty)$.*

Proof. Step 1: Formulas for $A'(t)$ and $A''(t)$. Since Y is finite and each map $t\mapsto e^{t\varphi(y)}$ is smooth, the function $Z(t)=\sum_y e^{t\varphi(y)}$ is smooth on $\mathbb{R}$ and in particular C^2 on $(0,\infty)$. Thus $A(t)=\log Z(t)$ is also C^2 on $(0,\infty)$.

Differentiating $A(t) = \log Z(t)$ gives $A'(t) = \frac{Z'(t)}{Z(t)}$. We have $Z'(t) = \sum_{y \in Y} \varphi(y) e^{t\varphi(y)}$, so

$$A'(t) = \frac{\sum_y \varphi(y) e^{t\varphi(y)}}{\sum_z e^{t\varphi(z)}} = \sum_{y \in Y} \varphi(y) \frac{e^{t\varphi(y)}}{Z(t)} = \sum_{y \in Y} \varphi(y) p_t(y).$$

For $A''(t)$, differentiate $A'(t) = Z'(t)/Z(t)$: $A''(t) = \frac{Z''(t)Z(t) - [Z'(t)]^2}{[Z(t)]^2}$. Here, $Z''(t) = \sum_{y \in Y} \varphi(y)^2 e^{t\varphi(y)}$. Therefore,

$$\begin{aligned} A''(t) = \frac{\big(\sum_y \varphi(y)^2 e^{t\varphi(y)}\big)\big(\sum_z e^{t\varphi(z)}\big) - \big(\sum_y \varphi(y) e^{t\varphi(y)}\big)^2}{\big(\sum_z e^{t\varphi(z)}\big)^2} &= \sum_{y \in Y} \varphi(y)^2 \frac{e^{t\varphi(y)}}{Z(t)} - \Big(\sum_{y \in Y} \varphi(y) \frac{e^{t\varphi(y)}}{Z(t)}\Big)^2 \\ &= \sum_{y \in Y} \varphi(y)^2 p_t(y) - \Big(\sum_{y \in Y} \varphi(y) p_t(y)\Big)^2. \end{aligned}$$

Using the expression for $A'(t)$, we can rewrite this as $A''(t) = \sum_{y \in Y} \big(\varphi(y) - A'(t)\big)^2 p_t(y)$. This proves part (1).

Step 2: Representation of $H(t)$. By definition, $\log p_t(y) = t\varphi(y) - \log Z(t) = t\varphi(y) - A(t)$, so

$$\begin{aligned} H(t) = -\sum_{y \in Y} p_t(y) \log p_t(y) = -\sum_y p_t(y)\big(t\varphi(y) - A(t)\big) &= -t \sum_y \varphi(y) p_t(y) + A(t) \sum_y p_t(y) \\ &= -tA'(t) + A(t), \end{aligned}$$

where we used the expression for $A'(t)$ from Step 1 and the fact that $\sum_y p_t(y) = 1$. This proves part (2).

Step 3: Derivative of $H(t)$ and its sign. Differentiating $H(t) = A(t) - tA'(t)$, $H'(t) = A'(t) - A'(t) - tA''(t) = -tA''(t)$. Substituting the expression for $A''(t)$ from part (1) gives $H'(t) = -t \sum_{y \in Y} \big(\varphi(y) - A'(t)\big)^2 p_t(y)$. For each $t > 0$, $p_t(y) > 0$ for all $y \in Y$. Since φ is nonconstant, $\varphi(Y)$ is not a.s. constant under p_t, and thus $\sum_y (\varphi(y) - A'(t))^2 p_t(y) > 0$. Thus, $A''(t) > 0$ and $H'(t) < 0$ for all $t > 0$, showing that H is strictly decreasing on $(0, \infty)$. □

Proof of Proposition 1. By Lemma 1, for each $\mu < 1/\lambda$ we have $\kappa := 1/\lambda - \mu > 0$ and $\hat{p}_{\lambda,\mu}(a; q)(y) = \frac{\exp\{\beta(\mu)\varphi(y)\}}{\sum_{z \in Y} \exp\{\beta(\mu)\varphi(z)\}}$, where $\varphi(y) := \log q_a(y) - \lambda u(a, y)$, $\beta(\mu) := 1/(1 - \lambda\mu)$, and φ is nonconstant by the payoff-likelihood condition. Identify $\hat{p}_{\lambda,\mu}(a; q)$ with the exponential family $p_t(y) \propto \exp\{t\varphi(y)\}$ at $t = \beta(\mu)$. By Lemma 4, the entropy $H(p_t)$ is strictly decreasing in $t > 0$. Since $\beta(\mu)$ is strictly increasing in μ (derivative: $\beta'(\mu) = \lambda/(1 - \lambda\mu)^2 > 0$), the composition $\mu \mapsto H(\hat{p}_{\lambda,\mu}(a; q))$ is strictly decreasing on $(-\infty, 1/\lambda)$. □

B Appendix: Proofs from Section 3

B.1 Proof of Proposition 2

Proof. Fix a $\bar{\mu}$-optimal policy σ and let α^{Λ} be a Λ-limit frequency. We show that $(\alpha^{\Lambda}, \eta^{\bar{\mu}}, \tau^{\bar{\mu}})$ satisfies Definition 3 for some $\eta^{\bar{\mu}} \in \Delta(Q)$ and $\tau^{\bar{\mu}} \geq 0$.

Step 1: Asymptotic identification of λ_t and posterior concentration. By the log-likelihood ratio (LLR) identity, for any belief path π_t and action frequency α_t:

$$\frac{1}{t}\,\mathrm{LLR}(h_t; Q) = \min_{q \in Q} \sum_{a \in A} \alpha_t(a) \cdot R(\hat{p}_{t,a} \| q_a),$$

where $\hat{p}_{t,a}$ is the empirical outcome distribution given action a. By the law of large numbers, for each action a played infinitely often, $\hat{p}_{t,a} \to p^{\star}_a$ a.s. Since Q is finite and the empirical average misfit converges uniformly (by compactness), we have $\lambda_t = \Lambda(h_t) \to \tau(\alpha_t)$ a.s., where $\tau(\alpha) := c^{-1} \min_{q \in Q} D(q; \alpha)$. And because Q is finite, π_t concentrates on $Q(\alpha_t)$ as $t \to \infty$.

Step 2: Set-valued drift and differential inclusion. By Lemma 1, the value function $V_{\lambda,\mu}(a; \pi)$ is continuous in all arguments on $\{\kappa > 0\}$. Consequently, the best-reply correspondence $\mathrm{BR}^{\bar{\mu}}_{\lambda}(\eta) := \arg\max_a V_{\lambda,\bar{\mu}\lambda}(a; \eta)$ has closed graph.

Define the set-valued correspondence

$$\chi^{\bar{\mu}}(\alpha) := \left\{ \beta \in \Delta(A) : \exists\, (\hat{\alpha}, \lambda, \eta) \text{ s.t. } \|\hat{\alpha} - \alpha\| < \varepsilon,\ |\lambda - \tau(\alpha)| < \varepsilon,\ \eta \in M_{\varepsilon}(\hat{\alpha}),\ \beta \in \widehat{\mathrm{BR}}^{\bar{\mu}}_{\lambda}(\eta) \right\},$$

where $M_{\varepsilon}(\alpha)$ is the ε-neighborhood of best-fit beliefs and $\widehat{\mathrm{BR}}^{\bar{\mu}}_{\lambda}$ is the convexified best-reply set. Because A and Q are finite and the best-reply correspondence has closed graph, $\chi^{\bar{\mu}}$ has compact convex values and closed graph.

By standard stochastic approximation theory (Benaïm et al., 2005),[9] the empirical frequency process $\{\alpha_t\}$ forms an asymptotic pseudotrajectory of the differential inclusion

$$\dot{\alpha}(s) \in \chi^{\bar{\mu}}(\alpha(s)) - \alpha(s).$$

Step 3: Non-equilibria cannot be limit frequencies. Suppose $\alpha^* \in \Delta(A)$ is not the α-component of any mixed c-robust equilibrium for the $\bar{\mu}$-criterion. Then there exists an

[9] Here the step size $\gamma_t = 1/(t+1)$ satisfies $\sum_t \gamma_t = \infty$ and $\sum_t \gamma_t^2 < \infty$, the noise terms are bounded martingale differences, and the drift correspondence has compact convex values and closed graph; thus the conditions of Benaïm et al. (2005, Theorem 4.1) are satisfied.

action $a^* \in A$ with $\alpha^*(a^*) > 0$ such that

$$a^* \notin \mathrm{BR}^{\bar{\mu}}_{\tau(\alpha^*)}(\eta) \quad \text{for every } \eta \text{ with } \mathrm{supp}(\eta) \subseteq Q(\alpha^*).$$

This implies that for all α near α^*, all λ near $\tau(\alpha^*)$, and all η in the best-fit neighborhood of α, the action a^* is never optimal. Therefore, $\max_{\beta \in \chi^{\bar{\mu}}(\alpha^*)} \beta(a^*) = 0$.

By the theory of differential inclusions, a solution $\alpha(s)$ starting within ε of α^* satisfies $d\alpha_{a^*}(s)/ds < 0$ on average, so it drifts away from $\alpha^*(a^*)$ at a uniform rate. Since $\{\alpha_t\}$ is an asymptotic pseudotrajectory, it cannot remain in a small neighborhood of α^* forever. Thus α^* cannot be a Λ-limit frequency. □

B.2 Proof of Proposition 3

Proof. Fix a $\bar{\mu}$–optimal policy σ and work under $\mathbb{P}_\sigma$. Let α^Λ be any Λ–limit frequency of σ, and fix an action $a \in A$ with $\alpha^\Lambda(a) > 0$.

Step 1: Correctly specified models are KL–minimizers on-path. Let $q^\star \in Q$ be as in Assumption 3. Then for every $\alpha \in \Delta(A)$, $D(q^\star;\alpha) = \sum_{a' \in A} \alpha(a')\, R(p^\star_{a'} \| q^\star_{a'}) = 0$, so $\min_{q \in Q} D(q;\alpha) = 0$. Moreover, if $q \in Q(\alpha)$, then $D(q;\alpha) = 0$, hence $\alpha(a') > 0 \Rightarrow R(p^\star_{a'} \| q_{a'}) = 0 \Rightarrow q_{a'} = p^\star_{a'}$ for every $a' \in A$.

Step 2: $\lambda_t \to 0$ and $\mu_t \to 0$ a.s. Because A is finite, $\Delta(A)$ is compact. Let $(t_n)_n$ be any subsequence of times. By compactness, there is a further subsequence (not relabeled) and some $\alpha^* \in \Delta(A)$ such that $\alpha_{t_n} \to \alpha^*$. Since (Q, A, Y) are finite and Assumption 1 holds, there exists $K < \infty$ with $\sup_{q \in Q, a \in A, y \in Y}(-\log q_a(y)) \le K$. Thus, the hypotheses needed to apply Lanzani's (2025a) Lemma 7 and the empirical-convergence step in the proof of Theorem 1 are satisfied along (t_n), which yield $\lim_{n\to\infty} \lambda_{t_n} = \lim_{n\to\infty} \frac{1}{ct_n}\mathrm{LLR}(h_{t_n}; Q) = \frac{1}{c}\min_{q \in Q} D(q;\alpha^*)$. By Step 1, the correctly specified model $q^\star$ satisfies

$$D(q^\star;\alpha^*) = \sum_{a' \in A} \alpha^*(a')\, R(p^\star_{a'} \| q^\star_{a'}) = 0,$$

since $q^\star_a = p^\star_a$ for all a under Assumption 3. Hence $\min_{q \in Q} D(q;\alpha^*) = 0$, and therefore every subsequential limit of $(\lambda_t)_t$ equals 0. Consequently, $\mu_t = \bar{\mu}\lambda_t \to 0$ a.s.

Step 3: Posterior mass concentration. Let $Q^\star_a := \{q \in Q : \ q_a = p^\star_a\}$. Since $\alpha^\Lambda(a) > 0$, action a is played infinitely often along any history whose empirical frequencies stabilize at α^Λ. Along such histories, the standard Bayesian likelihood-ratio argument used in the proof of Lanzani (2025a, Theorem 1) implies that models with $q_a \neq p^\star_a$ have strictly worse asymptotic log-likelihood along the subsequence of dates at which a is played, hence

$$\pi_t(Q_a^\star) \longrightarrow 1 \quad \text{a.s.}$$

Equivalently, along any stabilizing subsequence with limit action frequencies α^* such that $\alpha^*(a) > 0$, the posterior concentrates on $Q(\alpha^*) \subseteq Q_a^\star$ by Step 1.

Step 4: convergence to SEU. Fix $q \in Q_a^\star$, so $q_a = p_a^\star$. For each t, since $\mu_t \geq 0$,

$$v_{\lambda_t,0}(a;q) \;\leq\; v_{\lambda_t,\mu_t}(a;q) \;\leq\; \mathbb{E}_{q_a}[u(a,y)] + \mu_t H(q_a), \tag{B.1}$$

where the upper bound uses feasibility of $p = q_a$ in (2.1). Then, $\mu = 0$ criterion in (2.1) becomes $v_{\lambda,0}(a;q) = -\frac{1}{\lambda}\log\big(\sum_{y\in Y} q_a(y)e^{-\lambda u(a,y)}\big)$, so $v_{\lambda,0}(a;q) \to \mathbb{E}_{q_a}[u(a,y)]$ as $\lambda \downarrow 0$. Using (B.1) and $\mu_t \to 0$ (Step 2) yields $v_{\lambda_t,\mu_t}(a;q) \to \mathbb{E}_{q_a}[u(a,y)] = \mathbb{E}_{p_a^\star}[u(a,y)]$ a.s.

Step 5: Average over the posterior. Since Q is finite and u is bounded, the values $v_{\lambda_t,\mu_t}(a;q)$ are uniformly bounded in t and q (e.g., by evaluating at $p = q_a$ in (2.1) and using $\mu_t \leq \bar{\mu}\bar{\lambda}$ and $H(\cdot) \leq \log|Y|$). Hence, $V_{\lambda_t,\mu_t}(a;\pi_t) = \sum_{q\in Q_a^\star}\pi_t(q)v_{\lambda_t,\mu_t}(a;q) + \sum_{q\notin Q_a^\star}\pi_t(q)v_{\lambda_t,\mu_t}(a;q)$. By Step 3, $\pi_t(Q_a^\star) \to 1$ a.s., so the second term converges to 0 a.s. (by boundedness). By Step 4, for each $q \in Q_a^\star$, $v_{\lambda_t,\mu_t}(a;q) \to \mathbb{E}_{p_a^\star}[u(a,y)]$ a.s., so the first term converges to $\mathbb{E}_{p_a^\star}[u(a,y)]$ a.s. Combining the two gives $V_{\lambda_t,\bar{\mu}\lambda_t}(a;\pi_t) \to \mathbb{E}_{p_a^\star}[u(a,y)]$ a.s., for every a with $\alpha^\Lambda(a) > 0$. □

B.3 Proof of Lemma 2

Fix $(\lambda,\bar{\mu},\pi)$ such that $\kappa := 1/\lambda - \bar{\mu}\lambda > 0$. Set $\mu = \bar{\mu}\lambda$. For each action $a \in \{r,s\}$ and model $q \in Q$, Lemma 1 establishes that $v_{\lambda,\mu}(a;q)$ is differentiable in (λ,μ). The aggregate worst-case value is $V_{\lambda,\mu}(a;\pi) = \sum_{q\in Q} v_{\lambda,\mu}(a;q)\pi(q)$. Since Q is finite, we may differentiate term–by–term. By the envelope theorem (Lemma 1, eq. (2.4)):

$$\frac{\partial}{\partial\mu}v_{\lambda,\mu}(a;q) = H(\hat{p}_{\lambda,\mu}(a;q)), \qquad \frac{\partial}{\partial\lambda}v_{\lambda,\mu}(a;q) = -\frac{1}{\lambda^2}R(\hat{p}_{\lambda,\mu}(a;q)\|q_a).$$

Part 1: Derivative with respect to $\bar{\mu}$. Using the chain rule with $\mu = \bar{\mu}\lambda$:

$$\frac{\partial}{\partial\bar{\mu}}v_{\lambda,\bar{\mu}\lambda}(a;q) = \frac{\partial v}{\partial\mu}\cdot\frac{\partial\mu}{\partial\bar{\mu}} = H(\hat{p}_{\lambda,\bar{\mu}\lambda}(a;q))\cdot\lambda.$$

Summing over models: $\frac{\partial}{\partial\bar{\mu}}V_{\lambda,\bar{\mu}\lambda}(a;\pi) = \lambda\sum_{q\in Q}\pi(q)H(\hat{p}_{\lambda,\bar{\mu}\lambda}(a;q)) = \lambda H_a(\lambda,\bar{\mu},\pi)$. Therefore, $\frac{\partial}{\partial\bar{\mu}}\Delta(\lambda,\bar{\mu},\pi) = \frac{\partial}{\partial\bar{\mu}}V_{\lambda,\bar{\mu}\lambda}(r;\pi) - \frac{\partial}{\partial\bar{\mu}}V_{\lambda,\bar{\mu}\lambda}(s;\pi) = \lambda\big(H_r(\lambda,\bar{\mu},\pi) - H_s(\lambda,\bar{\mu},\pi)\big)$.

Part 2: Sign on switching surface. On the switching surface $S = \{(\lambda,\bar{\mu},\pi) : \Delta(\lambda,\bar{\mu},\pi) = 0\}$, the sign of $\partial\Delta/\partial\bar{\mu}$ equals the sign of $H_r(\lambda,\bar{\mu},\pi) - H_s(\lambda,\bar{\mu},\pi)$ (since $\lambda > 0$). □

B.4 Proof of Theorem 1

Define the domain $\mathcal{D} := [\underline{\lambda}, \bar{\lambda}] \times \Delta(Q)$. Let $(\bar{\mu}_0, H^*)$ be as in Assumption 5 and set

$$\hat{m} := \underline{\lambda} H^* > 0, \qquad M_0 := \sup_{(\lambda,\pi)\in\mathcal{D}} \max\{\Delta(\lambda, \bar{\mu}_0, \pi), 0\}.$$

Since Δ is continuous and $\mathcal{D}$ is compact, $M_0 < \infty$.

Define

$$\varepsilon \ := \ \frac{1}{2}\Big(\frac{1}{\bar{\lambda}^2} - \bar{\mu}_0 - \frac{M_0}{\hat{m}}\Big)_+, \qquad \bar{\mu}^* \ := \ \bar{\mu}_0 + \frac{M_0}{\hat{m}} + \varepsilon,$$

where $(x)_+ := \max\{x, 0\}$. Note that if $\varepsilon > 0$, then $\bar{\mu}^* < 1/\bar{\lambda}^2$ and $\bar{\mu}^* > \bar{\mu}_0 + M_0/\hat{m}$; if $\varepsilon = 0$, then $\bar{\mu}^* \geq 1/\bar{\lambda}^2$ and the set of admissible $\bar{\mu}$ satisfying $\bar{\mu} \geq \bar{\mu}^*$ and $\bar{\mu}\bar{\lambda}^2 < 1$ is empty, so the theorem holds vacuously.

Step 1: Equilibrium concerns lie in $[\underline{\lambda}, \bar{\lambda}]$. Let $(\alpha^{\bar{\mu}}, \eta^{\bar{\mu}}, \tau^{\bar{\mu}})$ be any mixed c–robust equilibrium for the $\bar{\mu}$–criterion with $\alpha^{\bar{\mu}}(r) > 0$. By Assumption 4, $\tau^{\bar{\mu}} \in [\underline{\lambda}, \bar{\lambda}]$.

Step 2: Uniform strict negativity of Δ for large $\bar{\mu}$. Fix any $(\lambda, \pi) \in \mathcal{D}$ and define $ff(\bar{\mu}) := \Delta(\lambda, \bar{\mu}, \pi)$ on the admissible range $\{\bar{\mu} : \bar{\mu}\lambda^2 < 1\}$. By Lemma 1 and Lemma 2, $\Delta(\lambda, \bar{\mu}, \pi)$ is continuously differentiable in $\bar{\mu}$ whenever $\bar{\mu}\lambda^2 < 1$, so f is C^1 on its domain. We claim that for every $(\lambda, \pi) \in \mathcal{D}$ and every admissible $\bar{\mu} > \bar{\mu}_0 + M_0/\hat{m}$, we have $f(\bar{\mu}) < 0$.

Case 1: $f(\bar{\mu}_0) \leq 0$. Suppose, toward a contradiction, that there exists $\bar{\mu}_1 > \bar{\mu}_0$ (admissible) with $f(\bar{\mu}_1) \geq 0$. By continuity of f, there exists $\bar{\mu}' \in [\bar{\mu}_0, \bar{\mu}_1]$ such that $f(\bar{\mu}') = 0$. Since $\bar{\mu}' \geq \bar{\mu}_0$ and $f(\bar{\mu}') = 0 \geq 0$, Assumption 5 implies $f'(\bar{\mu}') \leq -\hat{m} < 0$ (using Lemma 2 and $\lambda \geq \underline{\lambda}$). Hence, there exists $\delta > 0$ such that $f(\bar{\mu}' + \delta) < 0$, contradicting the existence of $\bar{\mu}_1 > \bar{\mu}'$ with $f(\bar{\mu}_1) \geq 0$ without another zero-crossing. Therefore, if $f(\bar{\mu}_0) \leq 0$ then $f(\bar{\mu}) < 0$ for all admissible $\bar{\mu} > \bar{\mu}_0$.

Case 2: $f(\bar{\mu}_0) > 0$. Define $\bar{\mu}_1 := \bar{\mu}_0 + f(\bar{\mu}_0)/\hat{m}$ (note $\bar{\mu}_1 \leq \bar{\mu}_0 + M_0/\hat{m}$). We show that $f(\bar{\mu}_1) \leq 0$. If not, suppose $f(\bar{\mu}_1) > 0$. We first claim that then $f(\bar{\mu}) > 0$ for all $\bar{\mu} \in [\bar{\mu}_0, \bar{\mu}_1]$. To see this, if there were $\tilde{\mu} \in [\bar{\mu}_0, \bar{\mu}_1]$ with $f(\tilde{\mu}) = 0$, then by Assumption 5 we would have $f'(\tilde{\mu}) \leq -\hat{m} < 0$, so f would be strictly negative immediately to the right of $\tilde{\mu}$, making it impossible to have $f(\bar{\mu}_1) > 0$ without a subsequent crossing from negative to nonnegative, which would require another zero at which the derivative is strictly negative. This is impossible. Hence, $f > 0$ on the entire interval $[\bar{\mu}_0, \bar{\mu}_1]$.

Since $f(\bar{\mu}) > 0$ on $[\bar{\mu}_0, \bar{\mu}_1]$, Assumption 5 applies throughout this interval, giving $f'(\bar{\mu}) \leq$

$-\hat{m}$ for all $\bar{\mu} \in [\bar{\mu}_0, \bar{\mu}_1]$. Integrating yields

$$f(\bar{\mu}_1) = f(\bar{\mu}_0) + \int_{\bar{\mu}_0}^{\bar{\mu}_1} f'(t)\,dt \le f(\bar{\mu}_0) - \hat{m}(\bar{\mu}_1 - \bar{\mu}_0) = f(\bar{\mu}_0) - \hat{m} \cdot \frac{f(\bar{\mu}_0)}{\hat{m}} = 0,$$

contradicting $f(\bar{\mu}_1) > 0$. Therefore $f(\bar{\mu}_1) \le 0$, and by Case 1 we conclude that $f(\bar{\mu}) < 0$ for all admissible $\bar{\mu} > \bar{\mu}_1$. Combining both cases, for every $(\lambda, \pi) \in \mathcal{D}$ we have $\Delta(\lambda, \bar{\mu}, \pi) < 0$ for all admissible $\bar{\mu} > \bar{\mu}_0 + \frac{M_0}{\hat{m}}$. In particular, for every admissible $\bar{\mu} \ge \bar{\mu}^*$ with $\bar{\mu}\bar{\lambda}^2 < 1$, we have $\bar{\mu} > \bar{\mu}_0 + M_0/\hat{m}$ whenever the admissible set is nonempty, so $\Delta(\lambda, \bar{\mu}, \pi) < 0$ for all $(\lambda, \pi) \in \mathcal{D}$.

Step 3: Elimination of mixed equilibria. Take any admissible $\bar{\mu} \ge \bar{\mu}^*$ with $\bar{\mu}\bar{\lambda}^2 < 1$ and suppose $(\alpha^{\bar{\mu}}, \eta^{\bar{\mu}}, \tau^{\bar{\mu}})$ is a mixed c–robust equilibrium. If $\alpha^{\bar{\mu}}(r) > 0$, then by Step 1 we have $\tau^{\bar{\mu}} \in [\underline{\lambda}, \bar{\lambda}]$, hence $(\tau^{\bar{\mu}}, \eta^{\bar{\mu}}) \in \mathcal{D}$. By Step 2, $\Delta(\tau^{\bar{\mu}}, \bar{\mu}, \eta^{\bar{\mu}}) < 0$, i.e., $V_{\tau^{\bar{\mu}}, \bar{\mu}\tau^{\bar{\mu}}}(r; \eta^{\bar{\mu}}) < V_{\tau^{\bar{\mu}}, \bar{\mu}\tau^{\bar{\mu}}}(s; \eta^{\bar{\mu}})$. By Definition 3(ii), any a with $\alpha^{\bar{\mu}}(a) > 0$ must be a best reply. Since r yields strictly lower value than s, we have $r \notin \mathrm{BR}^{\bar{\mu}}_{\tau^{\bar{\mu}}}(\eta^{\bar{\mu}})$, so $\alpha^{\bar{\mu}}(r) = 0$ in every equilibrium. Thus, $\alpha^{\bar{\mu}}(s) = 1$, proving part (i).

Step 4: Elimination of cycles. By Proposition 2, every Λ–limit frequency α^{Λ} of a $\bar{\mu}$–optimal policy corresponds to a mixed c–robust equilibrium. By Step 3, every such equilibrium satisfies $\alpha^{\bar{\mu}}(s) = 1$. Thus, the unique Λ–limit frequency is $\alpha^{\Lambda}(s) = 1$, and cycles cannot occur, which proves part (ii). □

B.5 Proof of Proposition 4

Fix $\lambda \in [\underline{\lambda}, \bar{\lambda}]$ and $\bar{\mu} \in I$. Using the single–model values $v_{\lambda, \bar{\mu}\lambda}(a; q)$ from (2.1), define the arm–specific slope $\Gamma_i(\lambda, \bar{\mu}) := \big[v_{\lambda,\bar{\mu}\lambda}(r_i; q^H) - v_{\lambda,\bar{\mu}\lambda}(s; q^H)\big] - \big[v_{\lambda,\bar{\mu}\lambda}(r_i; q^L) - v_{\lambda,\bar{\mu}\lambda}(s; q^L)\big]$. Since $V_{\lambda,\bar{\mu}\lambda}(a; \pi)$ is affine in π by (2.2), we then have

$$\frac{\partial}{\partial \theta} \Delta_i(\lambda, \bar{\mu}, \theta) = \Gamma_i(\lambda, \bar{\mu}) \quad \text{for all } \theta \in [0, 1]. \tag{B.2}$$

Step 1: $\Delta_i(\lambda, \bar{\mu}, \theta)$ is strictly increasing in θ. By eq. (B.2): $\frac{\partial}{\partial\theta}\Delta_i(\lambda, \bar{\mu}, \theta) = \Gamma_i(\lambda, \bar{\mu})$. By Assumption 7, $\Gamma_i(\lambda, \bar{\mu}) > 0$, so $\theta \mapsto \Delta_i(\lambda, \bar{\mu}, \theta)$ is strictly increasing.

Step 2: Existence and uniqueness of the threshold. By the assumption $\Delta_i(\lambda, \bar{\mu}, 0) < 0 < \Delta_i(\lambda, \bar{\mu}, 1)$, we have $\Delta_i(\lambda, \bar{\mu}, 0) < 0$ and $\Delta_i(\lambda, \bar{\mu}, 1) > 0$. By Step 1 and the intermediate value theorem, there exists a unique $\theta_i^*(\bar{\mu}) \in (0, 1)$ with $\Delta_i(\lambda, \bar{\mu}, \theta_i^*(\bar{\mu})) = 0$. Monotonicity gives: $\Delta_i(\lambda, \bar{\mu}, \theta) \ge 0 \iff \theta \ge \theta_i^*(\bar{\mu})$.

Step 3: The threshold is strictly increasing in $\bar{\mu}$. By Assumption 6 and Lemma 2: $\frac{\partial}{\partial \bar{\mu}}\Delta_i(\lambda, \bar{\mu}, \pi) = \lambda(H_{r_i} - H_s) \le -\lambda H_i^* < 0$. Thus, $\bar{\mu} \mapsto \Delta_i(\lambda, \bar{\mu}, \theta)$ is strictly decreasing for each fixed θ. For $\bar{\mu}_2 > \bar{\mu}_1$ in I, if θ satisfies $\Delta_i(\lambda, \bar{\mu}_2, \theta) \ge 0$, then since Δ_i is decreasing in $\bar{\mu}$: $\Delta_i(\lambda, \bar{\mu}_1, \theta) > \Delta_i(\lambda, \bar{\mu}_2, \theta) \ge 0$. In particular, $\theta \ge \theta_i^*(\bar{\mu}_2)$ implies $\theta > \theta_i^*(\bar{\mu}_1)$, so $\theta_i^*(\bar{\mu}_2) > \theta_i^*(\bar{\mu}_1)$. □

B.6 Proof of Lemma 3

The proof is identical to Lemma 2, applied to arbitrary actions $i, j \in A$ instead of r, s. Setting $\mu = \bar{\mu}\lambda$ and using the envelope theorem:

$$\frac{\partial}{\partial \bar{\mu}}\Delta_{ij}(\lambda, \bar{\mu}, \pi) = \frac{\partial}{\partial \bar{\mu}}\big[V_{\lambda,\bar{\mu}\lambda}(i;\pi) - V_{\lambda,\bar{\mu}\lambda}(j;\pi)\big] = \lambda\big(H_i(\lambda, \bar{\mu}, \pi) - H_j(\lambda, \bar{\mu}, \pi)\big). \quad \square$$

B.7 Proof of Proposition 5

Step 1: Bound on the derivative. By Assumption 8(i) and Lemma 3, for all $(\lambda, \bar{\mu}, \pi) \in [\underline{\lambda}, \bar{\lambda}] \times [\bar{\mu}_0, 1/\bar{\lambda}^2) \times \Delta(Q)$: $\frac{\partial}{\partial \bar{\mu}}\Delta_{ij}(\lambda, \bar{\mu}, \pi) = \lambda(H_i - H_j) \le -\lambda H_{ij}^* \le -\underline{\lambda} H_{ij}^*$.

Step 2: Integration. For $\bar{\mu} \in [\bar{\mu}_0, 1/\bar{\lambda}^2)$ and $(\lambda, \pi) \in [\underline{\lambda}, \bar{\lambda}] \times \Delta(Q)$:

$$\Delta_{ij}(\lambda, \bar{\mu}, \pi) = \Delta_{ij}(\lambda, \bar{\mu}_0, \pi) + \int_{\bar{\mu}_0}^{\bar{\mu}} \frac{\partial \Delta_{ij}}{\partial t}\, dt \le K_{ij} - \underline{\lambda} H_{ij}^*(\bar{\mu} - \bar{\mu}_0),$$

where K_{ij} is defined in (3.6).

Step 3: Finding the threshold. Set $\Delta_{ij}(\lambda, \bar{\mu}, \pi) < 0$, which requires: $\bar{\mu} > \bar{\mu}_0 + \frac{K_{ij}}{\underline{\lambda} H_{ij}^*}$. Define $\bar{\mu}^* := \bar{\mu}_0 + K_{ij}/(\underline{\lambda} H_{ij}^*)$. By Assumption 8(ii): $K_{ij} < \underline{\lambda} H_{ij}^*(1/\bar{\lambda}^2 - \bar{\mu}_0)$, so $\bar{\mu}^* = \bar{\mu}_0 + \frac{K_{ij}}{\underline{\lambda} H_{ij}^*} < \bar{\mu}_0 + (1/\bar{\lambda}^2 - \bar{\mu}_0) = 1/\bar{\lambda}^2$. Thus, $\bar{\mu}^* \in (\bar{\mu}_0, 1/\bar{\lambda}^2)$.

Step 4: Conclusion. For all $\bar{\mu} \in [\bar{\mu}^*, 1/\bar{\lambda}^2)$, $\lambda \in [\underline{\lambda}, \bar{\lambda}]$, and $\pi \in \Delta(Q)$: $\Delta_{ij}(\lambda, \bar{\mu}, \pi) < 0$, i.e., $V_{\lambda,\bar{\mu}\lambda}(j;\pi) > V_{\lambda,\bar{\mu}\lambda}(i;\pi)$. Hence, $i \notin \mathrm{BR}_{\lambda}^{\bar{\mu}}(\pi)$ for all such parameters. □

C Appendix: proofs from Section 5

The proof of Proposition 9 is in Online Appendix D.4; the remaining results of that section are proved here.

Lemma 5. *If $\lambda > 0$ and $\mu \geq 1/\lambda$, then for every $q \in \Delta(Y)$ and every act f,*

$$v_{\lambda,\mu}(f;q) = \min_{y \in \text{supp}(q)} \left\{ u(f(y)) + \frac{1}{\lambda} \log \frac{1}{q(y)} \right\}. \tag{C.1}$$

The proof of Lemma 5 is standard, so it is in Online Appendix D.4.

Proof of Theorem 2.1. Fix $q \in Q$ and an act $f \in X^Y$. If $p \not\ll q$, then $R(p\|q) = +\infty$, so such p cannot attain the minimum in (5.1). Hence

$$v_{\lambda,\mu}(f;q) = \min_{\substack{p \in \Delta(Y) \\ p \ll q}} \left\{ \mathbb{E}_p[u(f)] + \frac{1}{\lambda} R(p\|q) + \mu H(p) \right\}. \tag{C.2}$$

Fix $p \in \Delta(Y)$ with $p \ll q$. Then,

$$\begin{aligned}
\frac{1}{\lambda} R(p\|q) + \mu H(p) &= \frac{1}{\lambda} \sum_{y \in Y} p(y) \big(\log p(y) - \log q(y) \big) - \mu \sum_{y \in Y} p(y) \log p(y) \\
&= \left(\tfrac{1}{\lambda} - \mu \right) \sum_{y \in Y} p(y) \log p(y) - \frac{1}{\lambda} \sum_{y \in Y} p(y) \log q(y) \\
&= \kappa \sum_{y \in Y} p(y) \log p(y) - \frac{1}{\lambda} \sum_{y \in Y} p(y) \log q(y),
\end{aligned}$$

where $\kappa = \frac{1}{\lambda} - \mu > 0$. Now define $Z(q)$ and $\tilde{q}$ by (5.3). For every y with $q(y) > 0$, $\log \tilde{q}(y) = \beta \log q(y) - \log Z(q)$. Since $p \ll q$, taking expectations under p yields $\sum_{y \in Y} p(y) \log \tilde{q}(y) = \beta \sum_{y \in Y} p(y) \log q(y) - \log Z(q)$, and therefore

$$\begin{aligned}
R(p\|\tilde{q}) = \sum_{y \in Y} p(y) \log \frac{p(y)}{\tilde{q}(y)} &= \sum_{y \in Y} p(y) \log p(y) - \sum_{y \in Y} p(y) \log \tilde{q}(y) \\
&= \sum_{y \in Y} p(y) \log p(y) - \beta \sum_{y \in Y} p(y) \log q(y) + \log Z(q).
\end{aligned}$$

Multiplying by κ gives $\kappa R(p\|\tilde{q}) = \kappa \sum_{y \in Y} p(y) \log p(y) - \kappa\beta \sum_{y \in Y} p(y) \log q(y) + \kappa \log Z(q)$. By $\beta = \frac{1}{1-\lambda\mu}$ and $\kappa = \frac{1}{\lambda} - \mu = \frac{1-\lambda\mu}{\lambda}$ we have $\kappa\beta = \frac{1}{\lambda}$. Hence, $\frac{1}{\lambda} R(p\|q) + \mu H(p) = \kappa R(p\|\tilde{q}) - \kappa \log Z(q)$. Substituting into (C.2) yields

$$v_{\lambda,\mu}(f;q) = -\kappa \log Z(q) + \min_{\substack{p \in \Delta(Y) \\ p \ll q}} \left\{ \mathbb{E}_p[u(f)] + \kappa R(p\|\tilde{q}) \right\}.$$

Since $\tilde{q}$ is supported on $\{y : q(y) > 0\}$, any $p \not\ll q$ also satisfies $p \not\ll \tilde{q}$ and so $R(p\|\tilde{q}) = +\infty$; therefore the restriction $p \ll q$ is redundant:

$$\min_{\substack{p \in \Delta(Y) \\ p \ll q}} \left\{ \mathbb{E}_p[u(f)] + \kappa R(p\|\tilde{q}) \right\} = \min_{p \in \Delta(Y)} \left\{ \mathbb{E}_p[u(f)] + \kappa R(p\|\tilde{q}) \right\}.$$

Multiply by $\pi(q)$ and sum over $q \in Q$ to obtain

$$V_{\lambda,\mu}(f;\pi) = -\kappa \sum_{q \in Q} \pi(q) \log Z(q) + \sum_{q \in Q} \pi(q) \min_{p \in \Delta(Y)} \left\{ \mathbb{E}_p[u(f)] + \kappa R(p \| \tilde{q}) \right\}.$$

Define the act-independent constant $\widetilde{C}(\lambda, \mu, \pi) := -\kappa \sum_{q \in Q} \pi(q) \log Z(q) \in \mathbb{R}$, which gives (5.4). The ARC formulation follows by setting $\lambda_{\text{ARC}} := 1/\kappa$. □

The proof of Theorem 2.2 is standard, so it is relegated to Online Appendix D.4.

Proof of Theorem 2.3. The limit of $V_{\lambda,\mu}(\cdot;\pi)$ as $\lambda \to 0$ is $\lim_{\lambda \to 0} V_{\lambda,\mu}(\cdot;\pi) = \sum_{q \in Q} \pi(q) \big(\mathbb{E}_q[u(\cdot)] + \mu H(q) \big)$, which is a subjective expected utility preference with belief $\sum_{q \in Q} q\pi(q)$ because the additive term $\mu \sum_{q \in Q} H(q)\pi(q)$ is a constant that is independent of acts. When $Q = \{q\}$, π is a Dirac measure on q, so $\lim_{\lambda \to 0} V_{\lambda,\mu}(\cdot;\pi) = \mathbb{E}_q[u(\cdot)] + \mu H(q) = \lim_{\lambda \to 0} v_{\lambda,\mu}(\cdot; q)$, which is the subjective version of Mononen's (2025) entropy-modified expected utility preference. □

References

Augenblick, N. and Rabin, M. (2021). Belief movement, uncertainty reduction, and rational updating. *The Quarterly Journal of Economics*, 136(2):933–985.

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American economic review*, 103(6):2499–2529.

Beck, C. and Schögl, F. (1993). Thermodynamics of chaotic systems. *Thermodynamics of Chaotic Systems: an Introduction*.

Benaïm, M., Hofbauer, J., and Sorin, S. (2005). Stochastic approximations and differential inclusions. *SIAM Journal on Control and Optimization*, 44(1):328–348.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:69–186.

Bernheim, B. D. and Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica*, 88(4):1363–1409.

Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. (1987). Occam's razor. *Information processing letters*, 24(6):377–380.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.

Caggiano, G., Castelnuovo, E., and Figueres, J. M. (2017). Economic policy uncertainty and unemployment in the united states: A nonlinear approach. *Economics Letters*, 151:31–34.

Caggiano, G., Castelnuovo, E., and Groshenny, N. (2014). Uncertainty shocks and unemployment dynamics in us recessions. *Journal of Monetary Economics*, 67:78–92.

Cerreia-Vioglio, S., Hansen, L. P., Maccheroni, F., and Marinacci, M. (2025). Making decisions under model misspecification. *Review of Economic Studies*, page rdaf046.

Chamberlain, G. (2020). Robust decision theory and econometrics. *Annual Review of Economics*, 12(1):239–271.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American economic review*, 99(4):1145–1177.

Eden, M., Samuelson, L., and Steiner, J. (2026). The importance of unlikely events.

Enke, B. and Graeber, T. (2023). Cognitive uncertainty. *The Quarterly Journal of Economics*, 138(4):2021–2067.

Enke, B., Graeber, T., and Oprea, R. (2025). Complexity and time. *Journal of the European Economic Association*, 23(5):1838–1867.

Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science*, 29(3):393–421.

Fudenberg, D., Iijima, R., and Strzalecki, T. (2015). Stochastic choice and revealed perturbed utility. *Econometrica*, 83(6):2371–2409.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gabaix, X. (2025). A theory of complexity aversion. *Available at SSRN*.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of mathematical economics*, 18(2):141–153.

Grünwald, P. and Roos, T. (2019). Minimum description length revisited. *International journal of mathematics for industry*, 11(01):1930001.

Handel, B. and Schwartzstein, J. (2018). Frictions or mental gaps: what's behind the information we (don't) use and when do we care? *Journal of Economic Perspectives*, 32(1):155–178.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*, 91(2):60–66.

Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton University Press.

Lanzani, G. (2025a). Dynamic concern for misspecification. *Econometrica*, 93(4):1333–1370.

Lanzani, G. (2025b). Supplement to “dynamic concern for misspecification”. *Econometrica*.

Lipman, B. L. (1995). Information processing and bounded rationality: a survey. *Canadian Journal of Economics*, pages 42–67.

Maccheroni, F., Marinacci, M., and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498.

Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–298.

Mononen, L. (2025). On preference for simplicity and probability weighting. *working paper*.

Mullainathan, S., Schwartzstein, J., and Shleifer, A. (2008). Coarse thinking and persuasion. *The Quarterly journal of economics*, 123(2):577–619.

Oprea, R. (2020). What makes a rule complex? *American economic review*, 110(12):3913–3951.

Puri, I. (2025). Simplicity and risk. *The Journal of Finance*, 80(2):1029–1080.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465–471.

Robson, A., Samuelson, L., and Steiner, J. (2023). Decision theory and stochastic growth. *American Economic Review: Insights*, 5(3):357–376.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, 107(4):967–1004.

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

Strzalecki, T. (2011). Axiomatic foundations of multiplier preferences. *Econometrica*, 79(1):47–73.

Strzalecki, T. (2024). Variational bayes and non-bayesian updating. *arXiv preprint arXiv:2405.08796*.

Sunstein, C. R. (2002). Probability neglect: Emotions, worst cases, and law. *The Yale Law Journal*, 112:61.

Sunstein, C. R. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty*, 26(2-3):121–136.

Swait, J. and Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30(3):305–314.

D For Online Publication

D.1 General complexity functionals

This appendix demonstrates that most of our qualitative results extend to when the entropy function is generalized to the additive perturbations axiomatized by Fudenberg et al. (2015). Suppose now that the agent's static objective function is

$$v^W_{\lambda,\mu}(a;q) := \min_{p\in\Delta(Y)} \left\{ \sum_{y\in Y} u(a,y)p(y) + \frac{1}{\lambda} R(p\|q_a) + \mu \sum_{y\in Y} W(p(y)), \right\} \tag{D.1}$$

where W is continuous on $[0,1]$, twice continuously differentiable on $(0,1]$, strictly concave, $W(0)=0$, and $C_W := \sup_{x\in(0,1]} -xW''(x) < \infty$.[10]

The conditions on W imply that more diffuse beliefs are more complex, and that the curvature of the complexity term is not strong enough to overturn the strict convexity generated by the KL term. The Shannon case corresponds to $W(x) = -x\log x$, for which $C_W = 1$.

Proposition 10 (Static properties under a general complexity index)**.** *Suppose Assumption 1 holds. Fix $(a,q)\in A\times Q$, $\lambda>0$, and $\mu\geq 0$ with $\mu<1/(\lambda C_W)$. Then:*

(i) *The minimization problem in (D.1) has a unique solution, denoted $\hat{p}^W_{\lambda,\mu}(a;q)$.*

(ii) *The maps $(\lambda,\mu,q)\mapsto v^W_{\lambda,\mu}(a;q)$ and $(\lambda,\mu,q)\mapsto \hat{p}^W_{\lambda,\mu}(a;q)$ are continuous on*

$$\{(\lambda,\mu,q) : \lambda>0,\ \mu<1/(\lambda C_W)\}.$$

(iii) For every $\mu\in(0,1/(\lambda C_W))$, $\frac{\partial}{\partial\mu} v^W_{\lambda,\mu}(a;q) = M\left(\hat{p}^W_{\lambda,\mu}(a;q)\right)$. At $\mu=0$, the right derivative exists and satisfies $\frac{\partial^+}{\partial\mu} v^W_{\lambda,\mu}(a;q)\Big|_{\mu=0} = M\left(\hat{p}^W_{\lambda,0}(a;q)\right)$.

(iv) *If $0\leq\mu_1<\mu_2<1/(\lambda C_W)$, then $M\big(\hat{p}^W_{\lambda,\mu_2}(a;q)\big)\leq M\big(\hat{p}^W_{\lambda,\mu_1}(a;q)\big)$, with strict inequality whenever $\hat{p}^W_{\lambda,\mu_2}(a;q)\neq\hat{p}^W_{\lambda,\mu_1}(a;q)$.*

Proposition 10 identifies the main static driver of our qualitative results. The exact Shannon formula is not essential; what matters is that the complexity term enters linearly in the objective and ranks diffuse distortions above concentrated ones. What is lost relative

[10] The additive structure can further be relaxed, but the analysis would become less transparent because the Hessian of the objective would no longer be diagonal, so off-diagonal terms—which do not have clear economic interpretations—would need to be accounted for.

to entropy is tractability, not the economic mechanism. Notice also that whenever the minimizer is interior, its first-order conditions are

$$u(a,y)+\frac{1}{\lambda}\Big(1+\log\frac{\hat{p}^W_{\lambda,\mu}(a;q)(y)}{q_a(y)}\Big)+\mu W'\big(\hat{p}^W_{\lambda,\mu}(a;q)(y)\big)+\xi=0 \qquad (y\in Y) \tag{D.2}$$

for some multiplier $\xi\in\mathbb{R}$. For $W(x)=-x\log x$, (D.2) reduces to the Gibbs closed-form formula in Lemma 1. For a generic W, (D.2) is implicit, so the closed form and the representation results are lost, but our qualitative results are preserved.

The remaining theoretical results in the paper use only this more general structure. In particular: (i) Proposition 2 only needs continuity of the one-period value and best-reply correspondence, which Proposition 10 provides; (ii) Proposition 3 only needs boundedness of M on the simplex and the fact that $\mu_t=\bar{\mu}\lambda_t\to 0$ under correct specification; and (iii) Theorem 1, Proposition 4, Lemma 3, Proposition 5, and Proposition 12 use entropy only through the envelope identities and the corresponding entropy-gap assumptions. For concreteness, we now state and prove the analog of Theorem 1.

Let $M(p):=\sum_{y\in Y}W(p(y))$, for every $p\in\Delta(Y)$. For any action $a\in A$, $\pi\in\Delta(Q)$, and admissible pair $(\lambda,\bar{\mu})$ with $\bar{\mu}\lambda^2<1/C_W$, define $M_a(\lambda,\bar{\mu},\pi):=\sum_{q\in Q}\pi(q)M\big(\hat{p}^W_{\lambda,\bar{\mu}\lambda}(a;q)\big)$. If $A=\{r,s\}$, define the risky-safe value gap by $\Delta^W(\lambda,\bar{\mu},\pi):=V^W_{\lambda,\bar{\mu}\lambda}(r;\pi)-V^W_{\lambda,\bar{\mu}\lambda}(s;\pi)$.

Assumption 9 (Uniform complexity gap)**.** *Let $[\underline{\lambda},\bar{\lambda}]$ be as in Assumption 4. There exist constants $\bar{\mu}_0\geq 0$ and $M^*>0$ such that for all $(\lambda,\pi)\in[\underline{\lambda},\bar{\lambda}]\times\Delta(Q)$ and all admissible $\bar{\mu}\geq\bar{\mu}_0$ satisfying $\bar{\mu}\lambda^2<1/C_W$, $\Delta^W(\lambda,\bar{\mu},\pi)\geq 0 \Longrightarrow M_s(\lambda,\bar{\mu},\pi)-M_r(\lambda,\bar{\mu},\pi)\geq M^*$.*

Assumption 9 is the analog of Assumption 5 with M in place of H. It says that whenever the risky arm is still competitive, the safe arm can only be supported by a uniformly more diffuse and therefore more complex worst-case narrative.

Lemma 6. *Suppose Assumptions 1 and 9 hold. Then, for any two actions $i,j\in A$, any posterior $\pi\in\Delta(Q)$, and any admissible $(\lambda,\bar{\mu})$ with $\bar{\mu}\lambda^2<1/C_W$,*

$$\frac{\partial}{\partial\bar{\mu}}\Big(V^W_{\lambda,\bar{\mu}\lambda}(i;\pi)-V^W_{\lambda,\bar{\mu}\lambda}(j;\pi)\Big)=\lambda\Big(M_i(\lambda,\bar{\mu},\pi)-M_j(\lambda,\bar{\mu},\pi)\Big).$$

Theorem 3 (General complexity indices eliminate λ-cycles)**.** *Consider the dynamic model with the one-period criterion (D.1) and the same normalization rule $\lambda_t=\Lambda(h_t)$ as in (3.1). Suppose $A=\{r,s\}$ and Assumptions 4 and 9 hold. Then, there exists $\bar{\mu}^*>0$ such that for every admissible $\bar{\mu}\geq\bar{\mu}^*$ with $\bar{\mu}\bar{\lambda}^2<1/C_W$ the following holds:*

(i) *Every mixed c-robust equilibrium based on M satisfies $\alpha(s) = 1$.*

(ii) *Consequently, every Λ-limit frequency of any optimal policy based on M is pure and equals the safe action.*

The same logic yields the other qualitative results in the main text. Proposition 4 and Proposition 5 use entropy only through the pairwise derivative identity in Lemma 3; Lemma 6 gives the corresponding identity for M, so those proofs go through after replacing the relevant entropy-gap assumptions by their M-analogues. Likewise, Proposition 12 uses only the eventual selection of the safe arm, so it continues to hold once Theorem 3 replaces Theorem 1.

The economic content of this extension is therefore simple. The paper does not need Shannon entropy in order to generate conservatism, selection, or the elimination of λ-cycles. It only needs a complexity index that assigns a larger value to diffuse worst-case narratives than to concentrated ones, enters linearly in the objective, and is regular enough that the KL term still pins down a unique minimizer. Thus, Shannon entropy is the tractable benchmark inside this broader class because it turns (D.2) into a closed-form Gibbs distortion and therefore delivers the cleaner representation and application formulas in the main text.

Proof of Proposition 10. Write $F_{\lambda,\mu}(p;a,q) := \sum_{y\in Y} u(a,y)p(y) + \frac{1}{\lambda}R(p\|q_a) + \mu\sum_{y\in Y} W(p(y))$. For $x \in [0,1]$, define $g_{\lambda,\mu}(x) := (1/\lambda)x\log x + \mu W(x)$, with the convention $0\log 0 := 0$. For every $x \in (0,1]$, $g''_{\lambda,\mu}(x) = \frac{1}{\lambda x} + \mu W''(x) \geq \frac{1-\lambda\mu C_W}{\lambda x} > 0$. Hence, $g_{\lambda,\mu}$ is strictly convex on $[0,1]$. Since $F_{\lambda,\mu}(p;a,q) = \sum_{y\in Y} g_{\lambda,\mu}(p(y)) + \sum_{y\in Y}\big(u(a,y) - \frac{1}{\lambda}\log q_a(y)\big)p(y)$, $F_{\lambda,\mu}(\cdot;a,q)$ is strictly convex on the compact convex set $\Delta(Y)$. Therefore, it has a unique minimizer $\hat{p}^W_{\lambda,\mu}(a;q)$, which proves (i).

For (ii), the objective $F_{\lambda,\mu}(p;a,q)$ is jointly continuous in (p,λ,μ,q) on the admissible domain, and the feasible set $\Delta(Y)$ is compact and independent of parameters. Berge's maximum theorem yields continuity of the value function, and uniqueness of the minimizer upgrades upper hemicontinuity of the argmin correspondence to continuity of $(\lambda,\mu,q) \mapsto \hat{p}^W_{\lambda,\mu}(a;q)$.

For (iii), fix admissible (λ,μ,q) and write $p_\mu := \hat{p}^W_{\lambda,\mu}(a;q)$. For $h > 0$ small enough that $\mu + h < 1/(\lambda C_W)$, optimality of $p_{\mu+h}$ for $F_{\lambda,\mu+h}$ and of p_μ for $F_{\lambda,\mu}$ gives

$$F_{\lambda,\mu+h}(p_{\mu+h};a,q) \leq F_{\lambda,\mu+h}(p_\mu;a,q) = F_{\lambda,\mu}(p_\mu;a,q) + hM(p_\mu),$$

and $F_{\lambda,\mu}(p_\mu;a,q) \leq F_{\lambda,\mu}(p_{\mu+h};a,q) = F_{\lambda,\mu+h}(p_{\mu+h};a,q) - hM(p_{\mu+h})$. Subtracting shows $M(p_{\mu+h}) \leq \frac{v^W_{\lambda,\mu+h}(a;q) - v^W_{\lambda,\mu}(a;q)}{h} \leq M(p_\mu)$, where we recall that $M(p) = \sum_{y\in Y} W(p(y))$ for all $p \in \Delta(Y)$. Letting $h \downarrow 0$ and using continuity of $\mu \mapsto p_\mu$ and of M yields $\frac{\partial^+}{\partial \mu} v^W_{\lambda,\mu}(a;q) = M\big(\hat{p}^W_{\lambda,\mu}(a;q)\big)$. If $\mu \in (0, 1/(\lambda C_W))$, the same argument with $h < 0$ and $\mu + h \geq 0$ gives the left derivative, so the two-sided derivative exists and equals $M\big(\hat{p}^W_{\lambda,\mu}(a;q)\big)$.

For (iv), let $p_i := \hat{p}^W_{\lambda,\mu_i}(a;q)$ for $i = 1, 2$. Optimality implies $F_{\lambda,\mu_1}(p_1; a, q) \leq F_{\lambda,\mu_1}(p_2; a, q)$ and $F_{\lambda,\mu_2}(p_2; a, q) \leq F_{\lambda,\mu_2}(p_1; a, q)$. Adding the two inequalities and cancelling the common terms yields $(\mu_2 - \mu_1)\big(M(p_2) - M(p_1)\big) \leq 0$, so $M(p_2) \leq M(p_1)$. If equality held with $p_1 \neq p_2$, then both optimality inequalities above would have to bind, so both p_1 and p_2 would minimize the strictly convex function $F_{\lambda,\mu_1}(\cdot; a, q)$, a contradiction. Hence equality is possible only when $p_1 = p_2$. □

Proof of Lemma 6. Fix $a \in A$. By Proposition 10, $\frac{\partial}{\partial \mu} v^W_{\lambda,\mu}(a;q) = M\big(\hat{p}^W_{\lambda,\mu}(a;q)\big)$. Setting $\mu = \bar{\mu}\lambda$ and using the chain rule gives $\frac{\partial}{\partial \bar{\mu}} v^W_{\lambda,\bar{\mu}\lambda}(a;q) = \lambda M\big(\hat{p}^W_{\lambda,\bar{\mu}\lambda}(a;q)\big)$. Let $V^W_{\lambda,\mu}(a;\pi) := \sum_{q\in Q} \pi(q) v^W_{\lambda,\mu}(a;q)$, so summing the derivative over $q \in Q$ with weights $\pi(q)$ yields $\frac{\partial}{\partial \bar{\mu}} V^W_{\lambda,\bar{\mu}\lambda}(a;\pi) = \lambda M_a(\lambda, \bar{\mu}, \pi)$, and subtracting the identities for i and j proves the claim. □

Proof of Theorem 3. Let $\mathcal{D} := [\underline{\lambda}, \bar{\lambda}] \times \Delta(Q)$, $\hat{m} := \underline{\lambda} M^* > 0$, and define the constant $M_0 := \sup_{(\lambda,\pi)\in\mathcal{D}} \max\{\Delta^M(\lambda, \bar{\mu}_0, \pi), 0\}$. Because Δ^M is continuous by Proposition 10 and $\mathcal{D}$ is compact, $M_0 < \infty$. Set $\varepsilon := \frac{1}{2}\big(\frac{1}{C_W \bar{\lambda}^2} - \bar{\mu}_0 - \frac{M_0}{\hat{m}}\big)_+$ and $\bar{\mu}^* := \bar{\mu}_0 + \frac{M_0}{\hat{m}} + \varepsilon$. If $\varepsilon = 0$, then there is no admissible $\bar{\mu}$ satisfying both $\bar{\mu} \geq \bar{\mu}^*$ and $\bar{\mu}\bar{\lambda}^2 < 1/C_W$, so the theorem is vacuous. Hence, it suffices to consider the case $\varepsilon > 0$, in which $\bar{\mu}^* < 1/(C_W \bar{\lambda}^2)$.

Fix $(\lambda, \pi) \in \mathcal{D}$ and define $f(\bar{\mu}) := \Delta^M(\lambda, \bar{\mu}, \pi)$ on the admissible interval $\{\bar{\mu} : \bar{\mu}\lambda^2 < 1/C_W\}$. By Proposition 10 and Lemma 6, f is continuously differentiable on that interval. Assumption 9 and Lemma 6 imply that whenever $\bar{\mu} \geq \bar{\mu}_0$ and $f(\bar{\mu}) \geq 0$,

$$f'(\bar{\mu}) = \lambda\big(M_r(\lambda, \bar{\mu}, \pi) - M_s(\lambda, \bar{\mu}, \pi)\big) \leq -\underline{\lambda} M^* = -\hat{m}. \tag{D.3}$$

We claim that for every admissible $\bar{\mu} > \bar{\mu}_0 + M_0/\hat{m}$, one has $f(\bar{\mu}) < 0$.

First suppose $f(\bar{\mu}_0) \leq 0$. If there were an admissible $\bar{\mu}_1 > \bar{\mu}_0$ with $f(\bar{\mu}_1) \geq 0$, let

$$\bar{\mu}' := \inf\big\{t \in [\bar{\mu}_0, \bar{\mu}_1] : f(t) \geq 0\big\}.$$

By continuity, $f(\bar{\mu}') = 0$. By definition of the infimum, there exists a sequence $t_n \downarrow \bar{\mu}'$ with $f(t_n) \geq 0$. But (D.3) gives $f'(\bar{\mu}') \leq -\hat{m} < 0$, so for some $\delta > 0$ we have $f(t) < 0$ for all $t \in (\bar{\mu}', \bar{\mu}' + \delta)$, contradicting the existence of such a sequence. Therefore $f(\bar{\mu}) < 0$ for every admissible $\bar{\mu} > \bar{\mu}_0$.

Next suppose $f(\bar{\mu}_0) > 0$. As long as f remains nonnegative, (D.3) implies $f(\bar{\mu}) \leq$

$f(\bar{\mu}_0) - \hat{m}(\bar{\mu} - \bar{\mu}_0)$. Hence, at $\bar{\mu}_1 := \bar{\mu}_0 + \frac{f(\bar{\mu}_0)}{\hat{m}}$, we have $f(\bar{\mu}_1) \leq 0$ whenever $\bar{\mu}_1$ is admissible. Since $f(\bar{\mu}_0) \leq M_0$, every admissible $\bar{\mu} > \bar{\mu}_0 + M_0/\hat{m}$ satisfies $\bar{\mu} > \bar{\mu}_1$. Applying the previous paragraph from the starting point $\bar{\mu}_1$ shows that $f(\bar{\mu}) < 0$ for all such admissible $\bar{\mu}$.

We have thus proved that

$$\Delta^M(\lambda, \bar{\mu}, \pi) < 0 \qquad \text{for all } (\lambda, \pi) \in \mathcal{D} \text{ and all admissible } \bar{\mu} \geq \bar{\mu}^*. \tag{D.4}$$

To prove part (i), fix admissible $\bar{\mu} \geq \bar{\mu}^*$ and let (α, η, τ) be a mixed c-robust equilibrium for the dynamic criterion based on M. If $\alpha(r) > 0$, Assumption 4 implies $\tau \in [\underline{\lambda}, \bar{\lambda}]$. Since (α, η, τ) is a mixed c-robust equilibrium, any action in the support of α must be a best reply at (τ, η), so in particular r must be a best reply. But (D.4) gives $\Delta^M(\tau, \bar{\mu}, \eta) < 0$, which means $V^W_{\tau,\bar{\mu}\tau}(r; \eta) < V^W_{\tau,\bar{\mu}\tau}(s; \eta)$. Thus, r is not a best reply, a contradiction. Hence, $\alpha(r) = 0$, so $\alpha(s) = 1$.

For part (ii), the proof of Proposition 2 uses only continuity of the one-period value and of the best-reply correspondence. Proposition 10 provides exactly those properties for the dynamic criterion based on M, so the same argument applies verbatim: every Λ-limit frequency is the action component of a mixed c-robust equilibrium for the criterion based on M. Part (i) then implies that every such limit frequency is pure safe. □

D.2 Additional Application: Portfolio choice

We apply our complexity-augmented criterion in the two-period portfolio-choice environment of Chamberlain (2020, Sections 2.1 and 2.3) to illustrate a familiar class of existing frameworks in the literature where the uniform entropy gap in Assumption 5 holds naturally.

D.2.1 Setup and primitives

Chamberlain's primitives. There are two dates $t = 0, 1$ and a gross risk-free return $r_f > 0$. The risky gross return is $R_t \in \{R^H, R^L\}$, with $0 < R^L < r_f < R^H$. The realized return history is $y = (R_0, R_1) \in Y$, where $Y := \{(R^H, R^H), (R^H, R^L), (R^L, R^H), (R^L, R^L)\}$. Initial wealth is $w_0 > 0$ and preference over terminal wealth $w > 0$ satisfies constant relative risk aversion (CRRA):

$$U(w) := \begin{cases} \dfrac{w^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1, \\ \log w & \text{if } \gamma = 1, \end{cases} \qquad \gamma > 0.$$

In Chamberlain (2020, Section 2.3), the agent has a multiplier preference, so the set of structured models is a singleton $Q = \{q\}$. The reference distribution $q \in \Delta(Y)$ for (R_0, R_1) is exchangeable: there exist numbers $q_0, q_h, q_l \in (0, 1)$ such that

$$\begin{aligned} q(R^H, R^H) &= q_0 q_h, & q(R^H, R^L) &= q_0(1 - q_h), \\ q(R^L, R^H) &= (1 - q_0) q_l, & q(R^L, R^L) &= (1 - q_0)(1 - q_l). \end{aligned} \tag{D.5}$$

Our primitives. We now add a natural symmetry condition to the setup above.

- *Symmetric exchangeable reference model.* Let $q_0 = \frac{1}{2}$, $q_h = 1 - q_l$, and under the symmetric Beta prior in Chamberlain (2020, Figure 1, eq. (14))), we have $q_h > 1/2$. Combined with (D.5), this implies $q(R^H, R^H) = q(R^L, R^L) > q(R^H, R^L) = q(R^L, R^H)$. This condition requires "no ex-ante directional bias" about up vs. down moves, which is a standard benchmark in portfolio calibration and in binomial-tree approximations.

Embedding into our framework. Let $A = \{r, s\}$ and consider two (pure) portfolio plans:

$$\begin{aligned} &s : \text{invest only in the risk-free asset at both dates,} \\ &r : \text{invest only in the risky asset at both dates.} \end{aligned}$$

Thus, terminal wealth is

$$w_s(y) = w_0 r_f^2 \quad (\text{constant in } y), \qquad w_r(R_0, R_1) = w_0 R_0 R_1,$$

and we identify our one-period payoff index with $u(a, y) := U(w_a(y))$ for $a \in \{r, s\}$.

For $(\lambda, \bar{\mu})$ satisfying $\lambda > 0$ and $\bar{\mu}\lambda^2 < 1$, we evaluate actions with our complexity-augmented criterion, i.e., with $\mu = \bar{\mu}\lambda$ while maintaining the restriction $\kappa := \frac{1}{\lambda} - \mu = \frac{1 - \bar{\mu}\lambda^2}{\lambda} > 0$, so that the worst-case distortions $\hat{p}_{\lambda, \bar{\mu}\lambda}(a; q)$ are given by Lemma 1. Since $Q = \{q\}$ is a singleton, the belief π is a Dirac measure on q, so we write

$$H_a(\lambda, \bar{\mu}) := H\big(\hat{p}_{\lambda, \bar{\mu}\lambda}(a; q)\big), \qquad a \in \{r, s\}.$$

Define also the minimal utility gap above the worst return history:

$$\delta := \min_{y \in Y \setminus \{(R^L, R^L)\}} \big(U(w_r(y)) - U(w_r(R^L, R^L))\big) = U(w_0 R^H R^L) - U(w_0 (R^L)^2) > 0, \tag{D.6}$$

where the last equality uses $R^H > R^L$ and the monotonicity of U.

Entropy gap. Fix any target concentration level $\bar{p} \in (0, 1)$ such that

$$\log 2 \ > \ \bar{H}(\bar{p}) \quad \text{where} \quad \bar{H}(p) := -p \log p - (1-p) \log\left(\frac{1-p}{3}\right). \tag{D.7}$$

Let $\bar{\kappa} := \frac{\delta}{\log(\frac{3\bar{p}}{1-\bar{p}})}$ and define

$$\bar{\mu}_0 := \frac{1/\underline{\lambda} - \bar{\kappa}}{\underline{\lambda}}, \qquad H^* := \log 2 - \bar{H}(\bar{p}) \ > \ 0, \tag{D.8}$$

and we also maintain the condition in Assumption 2:

$$\bar{\mu}_0 \, \bar{\lambda}^2 < 1. \tag{D.9}$$

Condition (D.9) is now a transparent "riskiness vs. robustness" restriction: it requires that the risky payoff spread δ in (D.6) be large enough (relative to the λ-range) so that, for $\bar{\mu} \geq \bar{\mu}_0$, the worst-case narrative for the risky plan concentrates strongly on the worst return history.

D.2.2 Result

Combining all the above, this portfolio-choice environment captures that following: (i) downside risk is sizable (small R^L relative to R^H), (ii) preference curvature γ is nontrivial (so that utility gaps δ are large), and (iii) the horizon compounds risk (here, two-period compounding already creates a gap between history (R^L, R^L) and the next-worst history). The next result shows that the uniform entropy gap in Assumption 5 holds in this environment.

Proposition 11. *In the portfolio-choice environment above, Assumption 5 holds unconditionally with the constants* $(\bar{\mu}_0, H^*)$ *in* (D.8).

Proposition 11 demonstrates that Assumption 5 is not an abstract regularity condition: it is a formal expression of the familiar portfolio-choice logic that robust (multiplier) investors evaluate risky positions under concentrated downside scenarios, whereas safe positions do not trigger such concentration.

Proof of Proposition 11. Since $Q = \{q\}$ is a singleton and we have $q_r = q_s = q$ in the environment, the belief π is degenerate, so we can write

$$H_a(\lambda, \bar{\mu}, \pi) = H_a(\lambda, \bar{\mu}) = H\big(\hat{p}_{\lambda, \bar{\mu}\lambda}(a; q)\big), \qquad a \in \{r, s\},$$

where $\hat{p}_{\lambda,\bar{\mu}\lambda}(a;q)$ is the worst-case belief in (2.3). Fix any $\lambda \in [\underline{\lambda}, \bar{\lambda}]$ and any admissible $\bar{\mu} \geq \bar{\mu}_0$ with $\bar{\mu}\lambda^2 < 1$. Set $\mu := \bar{\mu}\lambda$ and define $\kappa := \frac{1}{\lambda} - \mu = \frac{1-\bar{\mu}\lambda^2}{\lambda} > 0$, and $\beta := \frac{1}{1-\lambda\mu} = \frac{1}{1-\bar{\mu}\lambda^2} > 0$. Throughout, identify the four return histories as $HH := (R^H, R^H)$, $HL := (R^H, R^L)$, $LH := (R^L, R^H)$, and $LL := (R^L, R^L)$, so we can write $Y = \{HH, LL, HL, LH\}$.

Step 1: a uniform lower bound on $H_s(\lambda, \bar{\mu})$. Since $w_s(y) = w_0 r_f^2$ is constant in y, the utility index $u(s,y) = U(w_s(y))$ is constant on Y. Hence, Lemma 1 and (2.3) imply $\hat{p}_{\lambda,\mu}(s;q)(y) = \frac{q(y)^\beta}{\sum_{z\in Y} q(z)^\beta}$ for $y \in Y$. By the symmetry restrictions imposed in the setup (i.e., $q(HH) = q(LL)$ and $q(HL) = q(LH)$), raising to the power $\beta > 0$ preserves equalities, so

$$\hat{p}_{\lambda,\mu}(s;q)(HH) = \hat{p}_{\lambda,\mu}(s;q)(LL) =: a, \qquad \hat{p}_{\lambda,\mu}(s;q)(HL) = \hat{p}_{\lambda,\mu}(s;q)(LH) =: b,$$

with $a, b \geq 0$ and $2a + 2b = 1$. Thus $\hat{p}_{\lambda,\mu}(s;q)$ lies on the line segment

$$\Big\{(a,b,b,a) : a,b \geq 0,\ 2a+2b = 1\Big\} = \mathrm{co}\Big\{\Big(\tfrac{1}{2}, 0, 0, \tfrac{1}{2}\Big),\ \Big(0, \tfrac{1}{2}, \tfrac{1}{2}, 0\Big)\Big\}.$$

Shannon entropy H is concave on $\Delta(Y)$, hence its restriction to this line segment is concave and therefore attains its minimum at an endpoint. At each endpoint the distribution puts probability $\frac{1}{2}$ on exactly two states, so the entropy equals $\log 2$. Consequently,

$$H_s(\lambda, \bar{\mu}) = H\big(\hat{p}_{\lambda,\mu}(s;q)\big) \ \geq\ \log 2 \qquad \text{for all admissible } (\lambda, \bar{\mu}). \tag{D.10}$$

Step 2: the risky distortion concentrates on the worst history. Let $y^{\min} := LL$ denote the disaster history under the risky plan r. By (D.6), the utility gap

$$\delta := \min_{y\in Y\setminus\{y^{\min}\}} \big(u(r,y) - u(r,y^{\min})\big) = u(r,HL) - u(r,LL) > 0$$

is strictly positive since $R^H > R^L$ and U is strictly increasing. By Lemma 1,

$$\hat{p}_{\lambda,\mu}(r;q)(y) = \frac{\exp\{-u(r,y)/\kappa\}\, q(y)^\beta}{\sum_{z\in Y} \exp\{-u(r,z)/\kappa\}\, q(z)^\beta} \qquad (y \in Y).$$

For any $y \neq y^{\min}$, $u(r,y) \geq u(r,y^{\min}) + \delta$, hence $\exp\{-u(r,y)/\kappa\} \leq e^{-\delta/\kappa} \exp\{-u(r,y^{\min})/\kappa\}$. Moreover, since $q(HH) = q(LL) > q(HL) = q(LH)$, so $q(y) \leq q(y^{\min})$ for every $y \neq y^{\min}$ and therefore $q(y)^\beta \leq q(y^{\min})^\beta$ because $\beta > 0$. Combining these two facts yields, for all $y \neq y^{\min}$, $\exp\{-u(r,y)/\kappa\}\, q(y)^\beta \leq e^{-\delta/\kappa} \exp\{-u(r,y^{\min})/\kappa\}\, q(y^{\min})^\beta$. Summing over the three states $y \in Y \setminus \{y^{\min}\}$ gives $\sum_{y\neq y^{\min}} \exp\{-u(r,y)/\kappa\}\, q(y)^\beta \leq 3e^{-\delta/\kappa} \exp\{-u(r,y^{\min})/\kappa\}\, q(y^{\min})^\beta$, and let $w(y) := \exp\{-u(r,y)/\kappa\}q(y)^{\beta}$, then

$$\hat{p}_{\lambda,\mu}(r;q)(y^{\min}) = \frac{w(y^{\min})}{w(y^{\min}) + \sum_{y\neq y^{\min}} w(y)} \;\geq\; \frac{1}{1+3e^{-\delta/\kappa}}. \tag{D.11}$$

Step 3: uniform concentration bound over $(\lambda,\bar{\mu})$. The map $(\lambda,\bar{\mu}) \mapsto \kappa(\lambda,\bar{\mu}) = \frac{1}{\lambda} - \bar{\mu}\lambda$ is strictly decreasing in both arguments on $(0,\infty)\times[0,\infty)$. Therefore, for $\lambda \in [\underline{\lambda},\bar{\lambda}]$ and $\bar{\mu} \geq \bar{\mu}_0$, $\kappa(\lambda,\bar{\mu}) \leq \kappa(\underline{\lambda},\bar{\mu}_0) = \frac{1}{\underline{\lambda}} - \bar{\mu}_0\underline{\lambda} = \bar{\kappa}$, where $\bar{\kappa}$ is defined in (D.8). Hence $e^{-\delta/\kappa} \leq e^{-\delta/\bar{\kappa}}$, and (D.11) implies $\hat{p}_{\lambda,\mu}(r;q)(y^{\min}) \;\geq\; \frac{1}{1+3e^{-\delta/\bar{\kappa}}}$. By definition of $\bar{\kappa}$ in (D.8), we have $3e^{-\delta/\bar{\kappa}} = (1-\bar{p})/\bar{p}$, so the previous display yields the uniform bound

$$\hat{p}_{\lambda,\mu}(r;q)(LL) \;\geq\; \bar{p} \qquad \text{for all admissible } (\lambda,\bar{\mu}). \tag{D.12}$$

Step 4: a uniform upper bound on $H_r(\lambda,\bar{\mu})$. Let $p \in \Delta(Y)$ satisfy $p(LL) \geq \bar{p}$. Since $x \mapsto -x\log x$ is concave on $[0,1]$, Jensen's inequality implies that, holding $\sum_{y\neq LL} p(y) = 1 - p(LL)$ fixed, the quantity $-\sum_{y\neq LL} p(y)\log p(y)$ is maximized when the remaining mass is split equally across the three states $Y \setminus \{LL\}$. Therefore,

$$H(p) \leq -p(LL)\log p(LL) - (1-p(LL))\log\Big(\frac{1-p(LL)}{3}\Big) =: \bar{H}(p(LL)).$$

Using (D.12) and the definition of $\bar{H}$ gives

$$H_r(\lambda,\bar{\mu}) = H\big(\hat{p}_{\lambda,\mu}(r;q)\big) \;\leq\; \bar{H}(\bar{p}). \tag{D.13}$$

Step 5: conclude Assumption 5. Combining (D.10) and (D.13) yields $H_s(\lambda,\bar{\mu}) - H_r(\lambda,\bar{\mu}) \;\geq\; \log 2 - \bar{H}(\bar{p}) = H^*$, where $H^* > 0$ by (D.7)–(D.8). Since this lower bound holds unconditionally for all admissible $(\lambda,\bar{\mu})$ and is independent of π (because Q is a singleton), it holds in particular whenever $\Delta(\lambda,\bar{\mu},\pi) \geq 0$. □

D.3 Welfare Comparison

This appendix complements Section 3.6 by isolating a class of environments in which increasing complexity aversion is *welfare-improving*: it raises the agent's long-run realized payoff under the true DGP, even though it makes the agent more conservative over time. The key mechanism is that when the true DGP actually favors the safe arm, eliminating safe–risky cycles prevents the agent from over–experimenting with the risky action.

Formally, we distinguish the agent's robust objective (which depends on $\bar{\mu}$) from an ex-post payoff criterion evaluated under the true DGP. For any mixed action $\alpha \in \Delta(A)$ define its objective expected payoff by $U^\star(\alpha) := \sum_{a\in A} \alpha(a)\bar{u}^\star(a)$ and $\bar{u}^\star(a) := \sum_{y\in Y} p^\star_a(y)u(a, y)$, where $p^\star = (p^\star_a)_{a\in A} \in \Delta(Y)^A$ is the true DGP. Notice that the criterion above depends only on outcomes in the sense that it does *not* attach any direct welfare value to model simplicity. Proposition 12 is a welfare comparison, whose assumptions are discussed below:

(i) pins down the ex-post ranking of actions under $p^\star$: the safe arm s is strictly better than the risky arm r in objective expected utility. Thus, any positive long–run weight on r is an ex-post payoff loss.

(ii) formalizes the baseline dynamic pathology in the $\bar{\mu} = 0$ model: even though s is ex-post optimal by (i), there exists a 0–optimal policy whose Λ–limit frequency α^0 assigns positive long–run probability to both r and s. This captures robustness–driven endogenous safe–risky cycles, which imply $\alpha^0(r) > 0$.

(iii) is the high–$\bar{\mu}$ stabilization condition: the uniform entropy gap (Assumption 5) together with the additional condition in Theorem 1(ii) ensures that for all $\bar{\mu} \geq \bar{\mu}^*$ the mixed c–robust equilibrium is unique and selects the safe arm in the long run, i.e., $\alpha^{\bar{\mu}}(s) = 1$. In other words, sufficiently strong complexity aversion eliminates the safe–risky cycles of (ii).

Proposition 12. *Suppose $A = \{r, s\}$ and assume:*

(i) $\bar{u}^\star(s) > \bar{u}^\star(r)$*;*

(ii) *For $\bar{\mu} = 0$, there exists a 0–optimal policy whose Λ–limit frequency α^0 satisfies $\alpha^0(r) \in (0, 1)$ and $\alpha^0(s) > 0$;*

(iii) *Assumption 5 holds and the additional condition in Theorem 1(ii) is satisfied.*

Then, for any $\bar{\mu} \geq \bar{\mu}^*$ there exists a $\bar{\mu}$–optimal policy with Λ–limit frequency $\alpha^{\bar{\mu}}$ such that

$$U^\star(\alpha^{\bar{\mu}}) > U^\star(\alpha^0).$$

Under the true DGP, the safe arm dominates the risky one, but the agent does not know this ex ante and entertains a misspecified model. When $\bar{\mu} = 0$, (ii) implies that their misspecification concerns alone can drive them to cycle between safe and risky in the long run. Then, (iii) implies that complexity aversion cuts off these cycles by selecting the safe arm in the long run; the agent becomes more conservative, but in this environment conservatism is beneficial: it prevents them from over–reacting to good runs of the risky arm that are only weakly supported by the data. Thus, even though $\bar{\mu} > 0$ makes the agent more pessimistic about worst-case payoffs in each period, it can raise their *actual* long–run payoff by improving the stability of their behavior.

D.3.1 Persistent complexity aversion

We now study the welfare implications of allowing complexity aversion to persist in the long run ($\mu_t \to \mu_\infty > 0$) even when misspecification concerns vanish ($\lambda_t \to 0$) under correct specification. Proposition 13 shows this can be *welfare-reducing*.

Lemma 7. *Suppose Assumptions 1 and 3 hold. Let σ be a policy such that, at each history h_t, $\sigma_t(h_t) \in \arg\max_{a\in A} V_{\lambda_t,\mu_t}(a;\pi_t)$, where $\mu_t \geq 0$, $\mu_t < 1/\lambda_t$ eventually, and $\mu_t \to \mu_\infty \in [0,\infty)$ a.s. Let α^Λ be any Λ–limit frequency of σ. Then, for every $a \in A$ with $\alpha^\Lambda(a) > 0$,*

$$V_{\lambda_t,\mu_t}(a;\pi_t) \longrightarrow \bar{u}^\star(a) + \mu_\infty H(p_a^\star) \qquad a.s.$$

The next result shows that if complexity aversion does not vanish after misspecification concerns vanish, then it can be strictly welfare-reducing under correct specification.

Proposition 13. *Fix any $\mu_\infty > 0$. There exists a correctly specified environment such that $\bar{u}^\star(s) > \bar{u}^\star(r)$, but every policy that is optimal for $V_{\lambda_t,\mu_t}(\cdot;\pi_t)$ and satisfies $\mu_t \to \mu_\infty$ has $\alpha^\Lambda(r) = 1$ for every Λ–limit frequency α^Λ. Thus, $U^\star(\alpha^\Lambda) = \bar{u}^\star(r) < \bar{u}^\star(s)$.*

Proposition 13 identifies the pathology ruled out by our specification $\mu_t = \bar{\mu}\lambda_t$. Under correct specification, misspecification concerns vanish on path. If μ_t does not vanish with them, then the agent does not converge to Bayesian expected utility. Instead, by Lemma 7, the limiting criterion is $\bar{u}^\star(a) + \mu_\infty H(p_a^\star)$, so simplicity continues to distort choice even after misspecification concerns have disappeared, which then causes welfare losses. The restriction $\mu_t = \bar{\mu}\lambda_t$ therefore removes this residual wedge to prevent potential welfare losses.

D.3.2 Proof of Proposition 12

Step 1: Structure of the $\bar{\mu} = 0$ equilibrium. By assumption (ii), there exists a 0–optimal policy with Λ–limit frequency α^0 satisfying $\alpha^0(r) \in (0,1)$ and $\alpha^0(s) > 0$. This means the agent cycles between safe and risky actions in the long run under $\bar{\mu} = 0$.

Step 2: Structure of the $\bar{\mu} \geq \bar{\mu}^*$ equilibrium. By assumption (iii) and Theorem 1, for all $\bar{\mu} \geq \bar{\mu}^*$, every mixed c–robust equilibrium has $\alpha^{\bar{\mu}}(s) = 1$. By Proposition 2, every Λ–limit frequency of a $\bar{\mu}$–optimal policy satisfies $\alpha^{\bar{\mu}}(s) = 1$.

Step 3: Comparison of long–run payoffs. The objective expected payoff of a mixed action α is: $U^\star(\alpha) = \alpha(r)\bar{u}^\star(r) + \alpha(s)\bar{u}^\star(s)$. For α^0 with $\alpha^0(r) > 0$:

$$U^\star(\alpha^0) = \alpha^0(r)\bar{u}^\star(r) + (1 - \alpha^0(r))\bar{u}^\star(s) = \bar{u}^\star(s) - \alpha^0(r)(\bar{u}^\star(s) - \bar{u}^\star(r)).$$

By assumption (i), $\bar{u}^\star(s) > \bar{u}^\star(r)$, so $\bar{u}^\star(s) - \bar{u}^\star(r) > 0$. Since $\alpha^0(r) > 0$: $U^\star(\alpha^0) < \bar{u}^\star(s)$. For $\alpha^{\bar{\mu}}$ with $\alpha^{\bar{\mu}}(s) = 1$: $U^\star(\alpha^{\bar{\mu}}) = \bar{u}^\star(s)$. Therefore, $U^\star(\alpha^{\bar{\mu}}) = \bar{u}^\star(s) > U^\star(\alpha^0)$. □

D.3.3 Proof of Lemma 7

The proof is nearly identical to that of Proposition 3. Under correct specification, the same argument gives $\lambda_t \to 0$ a.s. and posterior concentration on models that coincide with $p^\star$ on played actions. For every such model q and every played action a, Theorem 2.3 implies

$$v_{\lambda_t,\mu_t}(a; q) \longrightarrow \mathbb{E}_{p_a^\star}[u(a, Y)] + \mu_\infty H(p_a^\star),$$

because $q_a = p_a^\star$, $\lambda_t \to 0$, and $\mu_t \to \mu_\infty$. Averaging over the posterior yields the claim. □

D.3.4 Proof of Proposition 13

Fix $\mu_\infty > 0$. Let $A = \{r, s\}$, $Y = \{g, b\}$, $Q = \{q^\star\}$ and $q_a^\star = p_a^\star$ for all a, so the agent is correctly specified by construction. Pick any $\delta \in (0, 1/2)$ and then choose $\varepsilon > 0$ such that $0 < \varepsilon < \mu_\infty\big(\log 2 + (1 - \delta)\log(1 - \delta) + \delta \log \delta\big)$. Define payoffs and the true DGP by

$$u(s, g) = u(s, b) = 1, \qquad p_s^\star(g) = 1 - \delta, \quad p_s^\star(b) = \delta,$$

and

$$u(r, g) = 1, \qquad u(r, b) = 1 - 2\varepsilon, \qquad p_r^\star(g) = p_r^\star(b) = \frac{1}{2}.$$

Since $Q = \{q^\star\}$ and $q_a^\star = p_a^\star$ for both $a \in \{r, s\}$, Assumption 3 holds. The true expected payoffs are $\bar{u}^\star(s) = 1$ and $\bar{u}^\star(r) = \frac{1}{2} \cdot 1 + \frac{1}{2}(1 - 2\varepsilon) = 1 - \varepsilon$, so $\bar{u}^\star(s) > \bar{u}^\star(r)$. The corresponding entropies are $H(p_r^\star) = \log 2$ and $H(p_s^\star) = -(1 - \delta)\log(1 - \delta) - \delta \log \delta$. Hence, by the choice of ε, $\big(\bar{u}^\star(r) + \mu_\infty H(p_r^\star)\big) - \big(\bar{u}^\star(s) + \mu_\infty H(p_s^\star)\big) = -\varepsilon + \mu_\infty\big(H(p_r^\star) - H(p_s^\star)\big) > 0$.

Now, let σ be any policy that is optimal for $V_{\lambda_t,\mu_t}(\cdot;\pi_t)$ and satisfies $\mu_t \to \mu_\infty$. Since $Q = \{q^\star\}$ is a singleton, $\pi_t = \delta_{q^\star}$ for every t, so $V_{\lambda_t,\mu_t}(a;\pi_t) = v_{\lambda_t,\mu_t}(a;q^\star)$ for each $a \in \{r,s\}$. Moreover, under correct specification, Lemma 7 gives $\lambda_t \to 0$ a.s. Since $q^\star_a = p^\star_a$ for both $a \in \{r,s\}$, Theorem 2.3 applies action by action and yields $V_{\lambda_t,\mu_t}(a;\pi_t) \to \bar{u}^\star(a) + \mu_\infty H(p^\star_a)$ a.s. for each $a \in \{r,s\}$. Hence, $V_{\lambda_t,\mu_t}(r;\pi_t) - V_{\lambda_t,\mu_t}(s;\pi_t) \to \bar{u}^\star(r) + \mu_\infty H(p^\star_r) - \bar{u}^\star(s) - \mu_\infty H(p^\star_s) > 0$ a.s. Therefore, there exists an a.s. finite T such that for all $t \geq T$, $V_{\lambda_t,\mu_t}(r;\pi_t) > V_{\lambda_t,\mu_t}(s;\pi_t)$. Thus, r is the unique best reply from date T onward, so every optimal policy chooses r at every date $t \geq T$. It follows that $\alpha_t(r) \to 1$, hence every Λ–limit frequency satisfies $\alpha^\Lambda(r) = 1$. Consequently, $U^\star(\alpha^\Lambda) = \bar{u}^\star(r) < \bar{u}^\star(s)$. □

D.4 Omitted Proofs

D.4.1 Proof of Theorem 2

Proof of Lemma 5. Fix $q \in \Delta(Y)$ and an act f. If $p \not\ll q$ then $R(p\|q) = +\infty$ and such p cannot be optimal. Hence, restrict attention to $p \in \Delta(Y)$ with $\mathrm{supp}(p) \subseteq \mathrm{supp}(q)$. Then,

$$\begin{aligned}\mathbb{E}_p[u(f)] + \frac{1}{\lambda}R(p\|q) + \mu H(p) &= \sum_{y\in Y} p(y)u(f(y)) + \Big(\frac{1}{\lambda} - \mu\Big)\sum_{y\in Y} p(y)\log p(y) - \frac{1}{\lambda}\sum_{y\in Y} p(y)\log q(y)\\ &= \sum_{y\in Y} p(y)\Big(u(f(y)) - \frac{1}{\lambda}\log q(y)\Big) + \kappa\sum_{y\in Y} p(y)\log p(y).\end{aligned} \tag{D.14}$$

Step 1. Since $p(y) \in [0,1]$ for each y, one has $\log p(y) \leq 0$ and hence $\sum_y p(y)\log p(y) \leq 0$. Since $\kappa = \frac{1}{\lambda} - \mu \leq 0$, it follows that

$$\kappa \sum_{y\in Y} p(y)\log p(y) \geq 0. \tag{D.15}$$

Moreover, $\sum_y p(y)\log p(y) = 0$ holds if and only if $p(y) \in \{0,1\}$ for all y, i.e., if and only if p is a Dirac measure.

Step 2: minimizer is a Dirac measure. Let $a^f_q(y) := u(f(y)) - \frac{1}{\lambda}\log q(y) \in \mathbb{R}\cup\{+\infty\}$, where $a^f_q(y) = +\infty$ when $q(y) = 0$ (consistent with $R(\delta_y\|q) = +\infty$). For every feasible p we have

$$\sum_{y\in Y} p(y)a^f_q(y) \geq \min_{y\in\mathrm{supp}(q)} a^f_q(y), \tag{D.16}$$

with equality if $p = \delta_{y^*}$ for any minimizer y^* of a_q^f on $\text{supp}(q)$.

Step 3: combine all steps. Combining (D.14)–(D.16), for every feasible p,

$$\mathbb{E}_p[u(f)] + \frac{1}{\lambda}R(p\|q) + \mu H(p) \geq \sum_y p(y)a_q^f(y) \geq \min_{y\in\text{supp}(q)} a_q^f(y).$$

Choose $p = \delta_{y^*}$ where y^* minimizes a_q^f over $\text{supp}(q)$. Then, $\sum_y p(y)\log p(y) = 0$ so the first inequality holds with equality, and (D.16) holds with equality as well. Hence the minimum over p equals $\min_{y\in\text{supp}(q)} a_q^f(y)$, which coincides with (C.1). □

Proof of Theorem 2.2. When $\mu \geq 1/\lambda$, the preference is not ARC because $\kappa \leq 0$ in (5.4). We aim to prove that there exist a constant $C \in \mathbb{R}$ independent of acts and a grounded, convex, lower semicontinuous cost function $\hat{c} : \Delta(Y) \to [0, +\infty]$ such that for all f,

$$V_{\lambda,\mu}(f;\pi) = C + \min_{p\in\Delta(Y)} \Big\{\mathbb{E}_p[u(f)] + \hat{c}(p)\Big\}. \tag{D.17}$$

Let $n := |Y|$ and identify utility acts with vectors in $\mathbb{R}^n$. For each act f define its utility vector $\xi_f \in \mathbb{R}^n$ by $\xi_f(y) := u(f(y))$. Define a functional I on the convex set $\mathcal{D} := u(X)^Y \subset \mathbb{R}^n$ by

$$I(\xi) := \sum_{q\in Q} \pi(q) \min_{y\in\text{supp}(q)} \Big\{\xi(y) + \frac{1}{\lambda}\log\frac{1}{q(y)}\Big\}. \tag{D.18}$$

Then, by Lemma 5, $V_{\lambda,\mu}(f;\pi) = I(\xi_f)$ for all acts f.

Step 1: basic properties of I. Let $b_q(y) := \frac{1}{\lambda}\log(1/q(y))$, then each map $\xi \mapsto \min_{y\in\text{supp}(q)}\{\xi(y)+b_q(y)\}$ is the pointwise minimum of finitely many affine functionals in ξ, hence is concave and continuous on $\mathbb{R}^n$. A finite weighted sum of concave (respectively, continuous) functions is concave (respectively, continuous), so I is concave and continuous on $\mathbb{R}^n$ and hence on $\mathcal{D}$. Moreover, I is monotone (if $\xi \leq \eta$ coordinatewise then $I(\xi) \leq I(\eta)$).

Step 2: supergradients exist and are probabilities. Fix $\xi \in \mathbb{R}^n$. For each $q \in Q$, let $A_q(\xi) := \arg\min_{y\in\text{supp}(q)}\{\xi(y)+b_q(y)\} \neq \emptyset$. Define the (concave) function $m_q(\xi) := \min_{y\in\text{supp}(q)}\{\xi(y)+b_q(y)\}$. Since m_q is the minimum of finitely many affine maps, its superdifferential at ξ contains the convex hull of the gradients of the active affine pieces, i.e.

$$\text{co}\{\delta_y : y \in A_q(\xi)\} \subseteq \partial m_q(\xi) \subseteq \Delta(Y), \tag{D.19}$$

where $\delta_y \in \Delta(Y)$ denotes the Dirac mass on y and co denotes the convex hull. Indeed, for any $y \in A_q(\xi)$ we have $m_q(\eta) \leq \eta(y) + b_q(y) = m_q(\xi) + \delta_y \cdot (\eta - \xi)$ for all η, and taking convex combinations preserves the supergradient inequality.

Since $I = \sum_q \pi(q) m_q$, the sum of supergradients is a supergradient of the sum: if $p_q \in \partial m_q(\xi)$ for each q, then $\sum_{q \in Q} \pi(q) p_q \in \partial I(\xi)$. In particular, $\partial I(\xi) \neq \emptyset$, because each $\partial m_q(\xi)$ is nonempty by (D.19). Moreover, every vector of the form $\sum_q \pi(q) p_q$ (with $p_q \in \partial m_q(\xi)$) belongs to $\Delta(Y)$ because $\Delta(Y)$ is convex and closed under finite convex combinations.

Step 3: construct a convex cost function. Define $c : \Delta(Y) \to (-\infty, +\infty]$ by

$$c(p) := \sup_{\eta \in \mathcal{D}} \big\{ I(\eta) - p \cdot \eta \big\}. \tag{D.20}$$

Since $p \mapsto -p \cdot \eta$ is affine for each fixed η and a supremum of affine functions is convex and lower semicontinuous, c is convex and lower semicontinuous.

Step 4: the variational representation. First, for any $p \in \Delta(Y)$ and any $\xi \in \mathcal{D}$, the definition (D.20) implies

$$c(p) \geq I(\xi) - p \cdot \xi \implies p \cdot \xi + c(p) \geq I(\xi). \tag{D.21}$$

Taking the infimum over $p \in \Delta(Y)$ yields

$$\inf_{p \in \Delta(Y)} \{p \cdot \xi + c(p)\} \geq I(\xi). \tag{D.22}$$

For the reverse inequality, fix $\xi \in \mathcal{D}$ and choose any $p^\xi \in \partial I(\xi)$, which exists by Step 2. By definition of superdifferential for a concave function, $p^\xi \in \partial I(\xi)$ means that

$$I(\eta) \leq I(\xi) + p^\xi \cdot (\eta - \xi) \qquad \forall \eta \in \mathcal{D}.$$

Rearranging gives $I(\eta) - p^\xi \cdot \eta \leq I(\xi) - p^\xi \cdot \xi$ for all $\eta \in \mathcal{D}$. Taking supremum over η and using (D.20) yields

$$c(p^\xi) = \sup_{\eta \in \mathcal{D}} \{I(\eta) - p^\xi \cdot \eta\} \leq I(\xi) - p^\xi \cdot \xi. \tag{D.23}$$

On the other hand, applying (D.21) at $p = p^\xi$ gives $c(p^\xi) \geq I(\xi) - p^\xi \cdot \xi$. Therefore equality holds in (D.23) and hence

$$I(\xi) = p^\xi \cdot \xi + c(p^\xi) \geq \inf_{p \in \Delta(Y)} \{p \cdot \xi + c(p)\}. \tag{D.24}$$

Combining (D.22) and (D.24), we conclude

$$I(\xi) = \inf_{p\in\Delta(Y)} \{p \cdot \xi + c(p)\} \qquad \forall \xi \in \mathcal{D}. \tag{D.25}$$

Step 5: ground the cost. By Step 2, $\partial I(\xi)$ is nonempty for every ξ; pick some $\xi_0 \in \mathcal{D}$ and $p^{\xi_0} \in \partial I(\xi_0)$. Then by (D.24), $c(p^{\xi_0}) = I(\xi_0) - p^{\xi_0} \cdot \xi_0 < +\infty$, so c is proper on $\Delta(Y)$. Let $m := \inf_{p\in\Delta(Y)} c(p) \in \mathbb{R}$, where finiteness follows because c is proper, lower semicontinuous, and $\Delta(Y)$ is compact. Define the grounded cost $\hat{c}(p) := c(p) - m \geq 0$ and the constant $C := m$. Then, (D.25) implies for all $\xi \in \mathcal{D}$, $I(\xi) = C + \inf_{p\in\Delta(Y)}\{p \cdot \xi + \hat{c}(p)\}$. Substituting $\xi = \xi_f$ yields (D.17) with $\mathbb{E}_p[u(f)] = p \cdot \xi_f$. □

D.4.2 Proof of Proposition 6

Proof of Proposition 6. Let $\kappa := \frac{1}{\lambda} - \mu > 0$ and define the max–min objective (dropping constants independent of ψ)

$$\mathcal{J}(\psi, m) := \sum_{\omega\in\Omega} m(\omega) \sum_{a\in A} v(a, \omega)\psi(a \mid \omega) - \xi \sum_{\omega\in\Omega} g(\omega) \sum_{a\in A} \psi(a \mid \omega) \log \frac{\psi(a \mid \omega)}{\bar{\psi}(a)},$$

where $\bar{\psi}(a) = \sum_\omega g(\omega)\psi(a \mid \omega)$.

Step 1. By Observation 1, for each ψ the inner minimization over m is well-posed and has a unique minimizer $m^\star_{\lambda,\mu}(\cdot; \psi)$ with full support (since g has full support and (4.3) is strictly positive). Moreover, the sets $\Delta(A)^\Omega$ and $\Delta(\Omega)$ are compact convex. For each fixed m, the mapping $\psi \mapsto \mathcal{J}(\psi, m)$ is concave on $\Delta(A)^\Omega$: the payoff term is linear in ψ, and the Shannon term is convex in ψ. To see convexity, write $I(\psi; g) := \sum_\omega g(\omega) \, R\big(\psi(\cdot \mid \omega) \,\|\, \bar{\psi}\big)$, and note that $R(\cdot\|\cdot)$ is jointly convex and both $\psi(\cdot \mid \omega)$ and $\bar{\psi}$ are affine in ψ; thus $I(\psi; g)$ is convex and $-\xi I(\psi; g)$ is concave. For each fixed ψ, the mapping $m \mapsto \mathcal{J}(\psi, m)$ is linear in m. By Sion's minimax theorem, since the objective is concave in Ψ and linear in m over compact convex sets, there exists a saddle point $(\psi^\star, m^\star)$ for the max–min problem (4.2), satisfying $\mathcal{J}(\psi^\star, m) \leq \mathcal{J}(\psi^\star, m^\star) \leq \mathcal{J}(\psi, m^\star)$ for all $(\psi, m) \in \Delta(A)^\Omega \times \Delta(\Omega)$. In particular, $\psi^\star$ maximizes $\psi \mapsto \mathcal{J}(\psi, m^\star)$ over $\Delta(A)^\Omega$, and $m^\star = m^\star_{\lambda,\mu}(\cdot; \psi^\star)$.

Step 2. Fix $a \in A$ and assume $\bar{\psi}^\star(a) > 0$. We claim that then $\psi^\star(a \mid \omega) > 0$ for every ω. Suppose instead that $\psi^\star(a \mid \omega_0) = 0$ for some ω_0. Choose any $b \in A$ with $\psi^\star(b \mid \omega_0) > 0$ and consider, for $\varepsilon > 0$ small, a feasible perturbation ψ^ε that increases $\psi(a \mid \omega_0)$ by ε and decreases $\psi(b \mid \omega_0)$ by ε, leaving all other components unchanged. Since $m^\star(\omega_0) > 0$

and payoffs are bounded, the change in the payoff term in $\mathcal{J}(\cdot, m^\star)$ is $O(\varepsilon)$. However, the Shannon term produces a strictly first-order gain: for $\varepsilon \downarrow 0$,

$$\frac{1}{\varepsilon}\left[-\xi\, g(\omega_0)\, \varepsilon \log \frac{\varepsilon}{\bar{\psi}^\star(a) + O(\varepsilon)}\right] = -\xi\, g(\omega_0) \log \varepsilon + O(1) \;\longrightarrow\; +\infty,$$

so $\mathcal{J}(\psi^\varepsilon, m^\star) > \mathcal{J}(\psi^\star, m^\star)$ for ε small enough, contradicting optimality. Hence $\psi^\star(a \mid \omega) > 0$ for all ω whenever $\bar{\psi}^\star(a) > 0$. Under the maintained assumption that $\bar{\psi}^\star(a) > 0$ for all a, it follows that $\psi^\star(a \mid \omega) > 0$ for all (a, ω), and all derivatives below are finite.

Step 3. Consider the concave maximization problem $\max_{\psi \in \Delta(A)^\Omega} \mathcal{J}(\psi, m^\star)$ and form the Lagrangian $\mathcal{L}(\psi, \zeta) := \mathcal{J}(\psi, m^\star) + \sum_{\omega \in \Omega} \zeta(\omega)\big(\sum_{a \in A} \psi(a \mid \omega) - 1\big)$, with multipliers $\zeta(\omega) \in \mathbb{R}$. For each (a, ω), stationarity requires

$$0 = \frac{\partial \mathcal{L}}{\partial \psi(a \mid \omega)}(\psi^\star, \zeta) = m^\star(\omega)\, v(a, \omega) - \xi\, \frac{\partial I(\psi; g)}{\partial \psi(a \mid \omega)}\Big|_{\psi = \psi^\star} + \zeta(\omega). \tag{D.26}$$

We now compute $\partial I / \partial \psi(a \mid \omega)$ explicitly (including the dependence of $\bar{\psi}$ on ψ). Write

$$I(\psi; g) = \sum_\omega g(\omega) \sum_a \psi(a \mid \omega) \log \psi(a \mid \omega) - \sum_a \bar{\psi}(a) \log \bar{\psi}(a), \qquad \bar{\psi}(a) = \sum_\omega g(\omega) \psi(a \mid \omega).$$

Hence, for $\psi(a \mid \omega) > 0$ and $\bar{\psi}(a) > 0$,

$$\frac{\partial}{\partial \psi(a \mid \omega)} \sum_{\omega'} g(\omega') \sum_b \psi(b \mid \omega') \log \psi(b \mid \omega') = g(\omega)\big(1 + \log \psi(a \mid \omega)\big),$$

and by the chain rule,

$$\frac{\partial}{\partial \psi(a \mid \omega)} \sum_b \bar{\psi}(b) \log \bar{\psi}(b) = (1 + \log \bar{\psi}(a)) \cdot \frac{\partial \bar{\psi}(a)}{\partial \psi(a \mid \omega)} = g(\omega)\big(1 + \log \bar{\psi}(a)\big).$$

Subtracting yields the identity

$$\frac{\partial I(\psi; g)}{\partial \psi(a \mid \omega)} = g(\omega) \log \frac{\psi(a \mid \omega)}{\bar{\psi}(a)}. \tag{D.27}$$

Substituting (D.27) into (D.26) gives, for all (a, ω),

$$\log \frac{\psi^\star(a \mid \omega)}{\bar{\psi}^\star(a)} = \frac{m^\star(\omega)}{\xi g(\omega)}\, v(a, \omega) + \tilde{\zeta}(\omega), \qquad \tilde{\zeta}(\omega) := \frac{\zeta(\omega)}{\xi g(\omega)}.$$

Exponentiating and using $\sum_a \psi^\star(a \mid \omega) = 1$ to normalize across a yields

$$\psi^\star(a \mid \omega) = \frac{\bar{\psi}^\star(a) \exp\left\{\frac{m^\star(\omega)}{\xi g(\omega)} v(a,\omega)\right\}}{\sum_{b \in A} \bar{\psi}^\star(b) \exp\left\{\frac{m^\star(\omega)}{\xi g(\omega)} v(b,\omega)\right\}} = \frac{\bar{\psi}^\star(a) \exp\left\{\frac{v(a,\omega)}{\xi g(\omega)/m^\star(\omega)}\right\}}{\sum_{b \in A} \bar{\psi}^\star(b) \exp\left\{\frac{v(b,\omega)}{\xi g(\omega)/m^\star(\omega)}\right\}}.$$

Finally, by definition (4.4), $\xi_\omega(\psi^\star) = \xi\, g(\omega)/m^\star(\omega)$, which gives (4.5). □

D.4.3 Proof of Proposition 7

Proof of Proposition 7. Fix $\lambda > 0$ and $\mu \in [0, 1/\lambda)$ and define $\kappa := \frac{1}{\lambda} - \mu > 0$ and $\beta := \frac{1}{1-\lambda\mu} > 0$ as in Lemma 1. For $a \in \{d, f\}$ define the normalizing constant $\mathcal{Z}_a := \sum_{y \in Y} \exp\{-u(a,y)/\kappa\}\, q_a(y)^\beta$. By Lemma 1, the unique minimizer of (2.1) is $\hat{p}_{\lambda,\mu}(a;q)(y) = \frac{\exp\{-u(a,y)/\kappa\} q_a(y)^\beta}{\mathcal{Z}_a}$ for $y \in Y$. We first show the value identity

$$v_{\lambda,\mu}(a;q) = -\kappa \log \mathcal{Z}_a, \qquad a \in \{d, f\}. \tag{D.28}$$

For any $p \in \Delta(Y)$ with $p \ll q_a$, expand (2.1) using $H(p) = -\sum_y p(y) \log p(y)$:

$$\sum_y u(a,y)p(y) + \frac{1}{\lambda} R(p \| q_a) + \mu H(p) = \sum_y u(a,y)p(y) + \kappa \sum_y p(y) \log p(y) - \frac{1}{\lambda} \sum_y p(y) \log q_a(y).$$

Evaluating at $p = \hat{p}_{\lambda,\mu}(a;q)$ and using $\log \hat{p}_{\lambda,\mu}(a;q)(y) = -\frac{u(a,y)}{\kappa} + \beta \log q_a(y) - \log \mathcal{Z}_a$ yields

$$\kappa \sum_y \hat{p}_{\lambda,\mu}(a;q)(y) \log \hat{p}_{\lambda,\mu}(a;q)(y) = -\sum_y \hat{p}_{\lambda,\mu}(a;q)(y) u(a,y) + \kappa\beta \sum_y \hat{p}_{\lambda,\mu}(a;q)(y) \log q_a(y) - \kappa \log \mathcal{Z}_a.$$

Substituting into the objective at $\hat{p}_{\lambda,\mu}(a;q)$ yields

$$v_{\lambda,\mu}(a;q) = \left(\kappa\beta - \frac{1}{\lambda}\right) \sum_y \hat{p}_{\lambda,\mu}(a;q)(y) \log q_a(y) - \kappa \log \mathcal{Z}_a.$$

Since $\kappa = \frac{1-\lambda\mu}{\lambda}$ and $\beta = \frac{1}{1-\lambda\mu}$, we have $\kappa\beta = \frac{1}{\lambda}$, so the first term vanishes and (D.28) follows.

Under the binary payoff specification,

$$\mathcal{Z}_a = e^{-1/\kappa}\, q_a(y^*)^\beta + \sum_{i=1}^N q_a(y_i)^\beta = e^{-1/\kappa}(1-\delta)^\beta + \sum_{i=1}^N q_a(y_i)^\beta.$$

If $\mu = 0$, then $\beta = 1$ and $\sum_{i=1}^N q_d(y_i) = \sum_{i=1}^N q_f(y_i) = \delta$, hence $\mathcal{Z}_d = \mathcal{Z}_f$ and therefore

$v_{\lambda,0}(d;q) = v_{\lambda,0}(f;q)$ by (D.28).

If $\mu \in (0, 1/\lambda)$, then $\beta > 1$ and $x \mapsto x^\beta$ is strictly convex on $(0,\infty)$. Since $(q_f(y_1),\ldots,q_f(y_N))$ is not constant and $\frac{1}{N}\sum_{i=1}^N q_f(y_i) = \delta/N$, Jensen's inequality implies $\frac{1}{N}\sum_{i=1}^N q_f(y_i)^\beta > \left(\frac{1}{N}\sum_{i=1}^N q_f(y_i)\right)^\beta = \left(\frac{\delta}{N}\right)^\beta$, so $\sum_{i=1}^N q_f(y_i)^\beta > N\left(\frac{\delta}{N}\right)^\beta = \sum_{i=1}^N q_d(y_i)^\beta$. Hence, $\mathcal{Z}_f > \mathcal{Z}_d$, and (D.28) with $\kappa > 0$ gives $v_{\lambda,\mu}(d;q) = -\kappa \log \mathcal{Z}_d > -\kappa \log \mathcal{Z}_f = v_{\lambda,\mu}(f;q)$. □

D.4.4 Proof of Proposition 8

Given a model $p \in \Delta(Y)$, the growth-optimal portfolio problem in Robson et al. (2023, eq. (7)) is

$$V_0(p) := \max_{\alpha\in\Delta(A)} \mathbb{E}_p\Big[\log\Big(\mathbb{E}_\alpha e^{u(a,y)}\Big)\Big] = \max_{\alpha\in\Delta(A)} \mathbb{E}_p\Big[\log\Big(\sum_{a\in A}\alpha(a)e^{u(a,y)}\Big)\Big]. \tag{D.29}$$

Given an optimizer $\alpha_0^*(p)$ of (D.29), define the sampled choice rule

$$q_0^*(a \mid y) := \frac{\alpha_0^*(p)(a)e^{u(a,y)}}{\sum_{b\in A}\alpha_0^*(p)(b)e^{u(b,y)}}, \qquad q_0^*(a) := \sum_{y\in Y} p(y)q_0^*(a \mid y), \tag{D.30}$$

and, whenever $q_0^*(a) > 0$, the associated posterior

$$q_0^*(y \mid a) := \frac{p(y)q_0^*(a \mid y)}{q_0^*(a)}. \tag{D.31}$$

The regularity condition used in Robson et al. (2023, Proposition 4) is stated below.

Assumption 10 (Regularity Condition 2)**.** The misspecified model $p' \in \Delta(Y)$ lies in the convex hull of $\{q_0^*(\cdot \mid a) : a \in A, q_0^*(a) > 0\}$.

Lemma 8. *For every $\mu \in [0,1)$, $\alpha \in \Delta(A)$ and $y \in Y$,*

$$G_\mu(\alpha, y) = (1-\mu)\log\Big(\sum_{a\in A}\alpha(a)^{\frac{1}{1-\mu}}e^{\frac{u(a,y)}{1-\mu}}\Big). \tag{D.32}$$

In particular, $G_0(\alpha,y) = \log\left(\sum_{a\in A}\alpha(a)e^{u(a,y)}\right)$, so $V_0(\cdot)$ coincides with (D.29).

Proof of Lemma 8. Fix $\mu \in [0,1)$ and $y \in Y$. Write $s := \frac{1}{1-\mu} > 1$ and define $v_y(a) := u(a,y) + \log\alpha(a)$. The objective becomes

$$\sum_a q(a)u(a,y) - \sum_a q(a)\log\frac{q(a)}{\alpha(a)} - \mu\Big(-\sum_a q(a)\log q(a)\Big) = \sum_a q(a)v_y(a) - (1-\mu)\sum_a q(a)\log q(a).$$

Thus, for fixed α, y, the maximization problem is

$$\max_{q\in\Delta(A)} \Big\{ \sum_a q(a)v_y(a) - (1-\mu)\sum_a q(a)\log q(a) \Big\}.$$

The objective function is strictly concave in q (since $1-\mu > 0$) and the Lagrangian first-order conditions yield the unique maximizer $q^*(a) \propto \exp(v_y(a)/(1-\mu)) = \alpha(a)^s e^{su(a,y)}$. Substituting this optimizer back gives the standard log-sum-exp value

$$G_\mu(\alpha, y) = (1-\mu)\log\Big(\sum_a \exp\Big(\frac{v_y(a)}{1-\mu}\Big)\Big) = (1-\mu)\log\Big(\sum_a \alpha(a)^{\frac{1}{1-\mu}} e^{\frac{u(a,y)}{1-\mu}}\Big),$$

which is (D.32). Setting $\mu = 0$ yields $G_0(\alpha, y) = \log(\sum_a \alpha(a)e^{u(a,y)})$ and hence $V_0(\cdot)$ coincides with (D.29). $\square$

Lemma 9. *Fix $\mu \in (0,1)$, $\alpha \in \Delta(A)$ and $y \in Y$. Let $x_a := \alpha(a)e^{u(a,y)}$ for $a \in A$. Then*

$$G_\mu(\alpha, y) = \log\Big(\big(\sum_{a\in A} x_a^{\frac{1}{1-\mu}}\big)^{1-\mu}\Big) \le \log\Big(\sum_{a\in A} x_a\Big) = G_0(\alpha, y),$$

with strict inequality whenever $x_a > 0$ for at least two actions.

Proof of Lemma 9. Let $s := \frac{1}{1-\mu} > 1$. By Lemma 8, $G_\mu(\alpha, y) = \frac{1}{s}\log(\sum_a x_a^s) = \log((\sum_a x_a^s)^{1/s})$ and $G_0(\alpha, y) = \log(\sum_a x_a)$. For nonnegative $(x_a)_a$, the ℓ^s-norm is dominated by the ℓ^1-norm: $(\sum_a x_a^s)^{1/s} \le \sum_a x_a$. To verify this directly when at least two x_a are positive, pick any two indices with $x_i, x_j > 0$ and set $t := x_j/x_i > 0$. Then $\left(\frac{x_i^s + x_j^s}{x_i^s}\right)^{1/s} = (1+t^s)^{1/s} < 1 + t = \frac{x_i + x_j}{x_i}$, where the strict inequality follows from $1 + t^s < (1+t)^s$ for $t > 0$ and $s > 1$ since $f(t) := (1+t)^s - t^s$ satisfies $f'(t) = s[(1+t)^{s-1} - t^{s-1}] > 0$ and $f(0) = 1$. Multiplying by x_i and adding the remaining nonnegative components yields $(\sum_a x_a^s)^{1/s} < \sum_a x_a$. $\square$

Proof of Proposition 8. We construct an explicit two-state example satisfying Assumption 10 and show that $L_\mu(p, p') < L_0(p, p')$ for every $\mu \in (0,1)$.

Step 1: primitives and Assumption 10. Let $Y = \{y_1, y_2\}$ and $A = \{1, 2\}$. Let the gross returns be $e^{u(1,y_1)} = 3$, $e^{u(2,y_1)} = 1$, $e^{u(1,y_2)} = 1$, $e^{u(2,y_2)} = 4$, and let the true prior be $p(y_1) = p(y_2) = \frac{1}{2}$. For $\mu = 0$, the benchmark objective (D.29) becomes, writing $\alpha := \alpha(1) \in [0,1]$,

$$\mathbb{E}_p \log\Big(\sum_a \alpha(a)e^{u(a,y)}\Big) = \frac{1}{2}\log\big(1 + 2\alpha\big) + \frac{1}{2}\log\big(4 - 3\alpha\big).$$

This is strictly concave on $(0,1)$ since

$$\frac{d^2}{d\alpha^2}\Big[\tfrac{1}{2}\log(1+2\alpha)+\tfrac{1}{2}\log(4-3\alpha)\Big] = -\frac{2}{(1+2\alpha)^2} - \frac{9}{2(4-3\alpha)^2} < 0,$$

hence it has a unique maximizer $\alpha_0^*(p)(1) = \frac{5}{12} \in (0,1)$. Using (D.30)–(D.31) and $p(y_1) = p(y_2) = \frac{1}{2}$, $q_0^*(1 \mid y_1) = \frac{\frac{5}{12}\cdot 3}{\frac{5}{12}\cdot 3 + \frac{7}{12}\cdot 1} = \frac{15}{22}$ and $q_0^*(1 \mid y_2) = \frac{\frac{5}{12}\cdot 1}{\frac{5}{12}\cdot 1 + \frac{7}{12}\cdot 4} = \frac{5}{33}$. Therefore, by Bayes rule (D.31), $q_0^*(y_1 \mid 1) = \frac{\frac{1}{2}\cdot\frac{15}{22}}{\frac{1}{2}\cdot\frac{15}{22}+\frac{1}{2}\cdot\frac{5}{33}} = \frac{\frac{15}{22}}{\frac{15}{22}+\frac{5}{33}} = \frac{9}{11}$. Define the misspecified prior $p' \in \Delta(Y)$ by $p'(y_1) = \frac{9}{11}$ and $p'(y_2) = \frac{2}{11}$. Then, $p' = q_0^*(\cdot \mid 1)$ is one of the posteriors induced by the benchmark optimizer under p, hence p' lies in the convex hull of $\{q_0^*(\cdot \mid a) : q_0^*(a) > 0\}$. Thus, Assumption 10 holds.

Step 2: benchmark loss equals Kullback–Leibler divergence. By Lemma 8, $V_0(\cdot)$ coincides with the benchmark problem in (D.29) or Robson et al. (2023, eq. (7)). Since Assumption 10 holds for (p, p') by Step 1, Robson et al. (2023, Proposition 4) applies and yields

$$L_0(p, p') = R(p\|p') = \frac{1}{2}\log\frac{1/2}{9/11} + \frac{1}{2}\log\frac{1/2}{2/11} = \frac{1}{2}\log\frac{121}{72}. \tag{D.33}$$

Step 3: for $\mu > 0$, loss is strictly smaller than benchmark. Fix $\mu \in (0,1)$. By Lemma 9, for every portfolio α and both states y_1, y_2, $G_\mu(\alpha, y) \le G_0(\alpha, y)$, with equality only for degenerate portfolios (since all returns are strictly positive). Taking expectations under p and maximizing over α gives

$$V_\mu(p) = \max_\alpha \mathbb{E}_p[G_\mu(\alpha, y)] \le \max_\alpha \mathbb{E}_p[G_0(\alpha, y)] = V_0(p). \tag{D.34}$$

We claim that the inequality in (D.34) is strict. Toward contradiction, suppose $V_\mu(p) = V_0(p)$ held. Let $\widehat{\alpha} \in \arg\max_\alpha \mathbb{E}_p[G_\mu(\alpha, y)]$ be an optimizer at μ. Then,

$$V_0(p) = V_\mu(p) = \mathbb{E}_p[G_\mu(\widehat{\alpha}, y)] \le \mathbb{E}_p[G_0(\widehat{\alpha}, y)] \le V_0(p),$$

so all inequalities bind. Binding of the first inequality forces $G_\mu(\widehat{\alpha}, y) = G_0(\widehat{\alpha}, y)$ for both y_1 and y_2, which, by Lemma 9 and strict positivity of returns for both actions in both states, implies that $\widehat{\alpha}$ is degenerate. However, for any degenerate portfolio $\alpha = \delta_a$, Lemma 8 implies $G_\mu(\delta_a, y) = u(a, y) = G_0(\delta_a, y)$, so $\mathbb{E}_p[G_0(\delta_1, y)] = \frac{1}{2}\log 3$ and $\mathbb{E}_p[G_0(\delta_2, y)] = \frac{1}{2}\log 4$. On the other hand, the feasible diversified portfolio $\alpha(1) = \alpha(2) = \frac{1}{2}$ delivers

$$\mathbb{E}_p[G_0(\alpha, y)] = \frac{1}{2}\log\left(\frac{3+1}{2}\right) + \frac{1}{2}\log\left(\frac{1+4}{2}\right) = \frac{1}{2}\log 2 + \frac{1}{2}\log\frac{5}{2} = \frac{1}{2}\log 5 > \frac{1}{2}\log 4,$$

so no degenerate portfolio can maximize $V_0(p)$. This contradicts the conclusion that $\widehat{\alpha}$ must be degenerate. Hence, $V_\mu(p) < V_0(p)$. Next, we show that the optimal portfolio under p' is degenerate for all $\mu \in (0,1)$. Since $p'(y_1) = \frac{9}{11}$ is sufficiently tilted toward y_1, the benchmark optimizer at $\mu = 0$ is $\alpha_0^*(p') = \delta_1$. To see this, given $p' = (9/11, 2/11)$, the objective is $\frac{9}{11}\log(1+2\alpha) + \frac{2}{11}\log(4-3\alpha)$, which is increasing in $\alpha \in [0,1]$ and strictly increasing for $\alpha < 1$, so $\alpha = 1$ is the unique maximizer. Now fix any $\mu \in (0,1)$. By Lemma 9, $\mathbb{E}_{p'}[G_\mu(\alpha, y)] \leq \mathbb{E}_{p'}[G_0(\alpha, y)]$ for all α, while $\mathbb{E}_{p'}[G_\mu(\delta_1, y)] = \mathbb{E}_{p'}[G_0(\delta_1, y)]$ because δ_1 is degenerate. Therefore, $\mathbb{E}_{p'}[G_\mu(\alpha, y)] \leq \mathbb{E}_{p'}[G_0(\alpha, y)] \leq \mathbb{E}_{p'}[G_0(\delta_1, y)] = \mathbb{E}_{p'}[G_\mu(\delta_1, y)]$, so $\alpha_\mu^*(p') = \delta_1$ is optimal for every $\mu \in (0,1)$. Combining, for every $\mu \in (0,1)$, we have

$$L_\mu(p, p') = \mathbb{E}_p[G_\mu(\alpha_\mu^*(p), y)] - \mathbb{E}_p[G_\mu(\delta_1, y)] = V_\mu(p) - \tfrac{1}{2}\log 3 < V_0(p) - \tfrac{1}{2}\log 3 = L_0(p, p').$$

Using (D.33), this implies $L_\mu(p, p') < R(p\|p')$ for every $\mu \in (0,1)$. □

D.4.5 Proof of Proposition 9

By Theorem 2.2, for $Y = \{0,1\}$ and $Q = \{q\}$,

$$v_{\lambda,\mu}(f_b; q) = \min\left\{f_b(0) + \frac{1}{\lambda}\log\frac{1}{q(0)}, f_b(1) + \frac{1}{\lambda}\log\frac{1}{q(1)}\right\}.$$

Under symmetric binary q, $\log\frac{1}{q(0)} = \log\frac{1}{q(1)} = \log 2$, so

$$v_{\lambda,\mu}(f_b; q) = \frac{1}{\lambda}\log 2 + \min\{0, -\gamma|b|\} = \frac{1}{\lambda}\log 2 - \gamma|b|.$$

Substituting into V^* yields $V^*(b) = \left(V(0) + \frac{1}{\lambda}\log 2\right) + 2b^r b - (2-m)b^2 - \gamma|b|$. Dropping the additive constant and imposing $\gamma = 2\xi\sigma_a(1-m)$ gives (5.5), and therefore, the maximizers coincide. □