

Coarse Q-learning: Indifference, Indeterminacy, and Instability*

Philippe Jehiel[†] Aviman Satpathy[‡]

April 29, 2026

[Click here for the latest version](#)

[Click here for the Online Appendix](#)

Abstract

We introduce Coarse Q-learning (CQL), a reinforcement-learning model for bandit problems with stochastically varying menus. Alternatives are exogenously partitioned into similarity classes, and feedback from sampled alternatives is pooled within classes into class-level valuations. Choices follow multinomial logit over class valuations, and valuations update toward realized payoffs as in Q-learning. Using stochastic approximation, we derive the mean-field dynamics and characterize the steady states as smooth analogues of Valuation Equilibria. The model yields novel long-run phenomena in the high payoff-sensitivity limit: depending on the environment, CQL may exhibit multiple stable strict equilibria, a unique globally stable mixed equilibrium with indifference across classes, or no stable equilibrium at all, with valuations and choice probabilities converging instead to a stable limit cycle. These outcomes are driven by coarse aggregation and do not arise in the standard alternative-level benchmark.

Keywords: Reinforcement Learning, Coarse Inference, Misspecified Learning

JEL Codes: C62, C73, D83, D91

*We thank Evan Friedman, Johannes Hörner, Margarita Pavlova, Ariel Rubinstein, Jean-Marc Tallon, Olivier Tercieux, and Giacomo Weber for valuable comments. All remaining errors are our own.

[†]Paris School of Economics & University College London; jehiel@enpc.fr

[‡]Paris School of Economics; aviman.satpathy@psemail.eu

1 Introduction

Consider a consumer, Alice, who has recently moved to Paris and dines out regularly. To discover her taste in wine, she orders a bottle to share with friends on each occasion. Facing a vast and unfamiliar selection across different restaurants, Alice simplifies her choice by relying on coarse categories based on region of origin (e.g., Bordeaux, Chianti) and price range. While she typically favors wines from the category that yields the highest expected net benefit based on her current estimates, she also occasionally explores other categories to broaden her palate. After each bottle, she observes her realized utility and updates her valuation of the chosen category in the direction of the experienced pleasure.

We are interested in understanding how Alice would compare these various categories in the long run. The mechanism just described shares many features with reinforcement learning models of trial-and-error adaptation in humans and animals (Skinner, 1938; Caplin and Dean, 2008; Niv, 2009; Glimcher, 2011). However, our framework introduces a crucial twist: while standard reinforcement learning evaluates each alternative separately, Alice learns at the level of coarse categories. This coarsening affects both the decision stage (as distinct alternatives within a category are treated alike) and the updating stage (since the realized payoff from a single item updates the estimate of the entire category). Our focus is on how this coarsening alters the learning process, specifically the long-run ranking of the categories.

Our main finding is that coarse, category-level reinforcement generates novel long-run phenomena whenever menus exogenously vary across periods. More precisely, we consider environments where not all categories are available in every menu - a natural benchmark in many settings (just as few restaurants offer wines from every region and every price range). Under these conditions, we identify three distinct long-run outcomes. First, we observe *indeterminacy*: despite continuous, systematic exploration, the long-run ranking of categories can remain highly sensitive to initial conditions, as the system may admit multiple strict rankings as distinct long-run attractors. Second, we find *indifference*: for open sets of underlying objective payoff distributions, Alice may inevitably end up evaluating several categories as equally attractive. Finally, the model can generate *instability*, where Alice’s ranking may never settle down, resulting in perpetual fluctuations in her revealed patterns of choice.

Formally, we study a repeated decision problem in which the decision-maker (DM) faces a randomly drawn menu (a subset of alternatives) in each period, generated independently from a stationary distribution. As in the standard multi-armed bandit literature (Robbins, 1952; Gittins, 1979), the DM does not a priori know the payoff distributions of the alternatives and must learn the expected payoffs through sampled experience. We adopt a model-free learning

framework, specifically Q-learning (Watkins, 1989) or, more broadly, temporal difference learning (Sutton and Barto, 2018). Formalizing our earlier intuition, we endow the DM with categories defined as *similarity classes* that partition the grand set of alternatives, with learning occurring at this category level. In each period, the DM assigns valuations to the different categories and selects from those available in the drawn menu using a logit (or softmax) choice rule, which probabilistically favors categories with higher valuations.¹ If the chosen category contains multiple available alternatives, the DM selects one uniformly at random (our analysis trivially extends to any exogenous tie-breaking rule). Upon realizing the payoff from the chosen alternative, the DM updates the valuation of the corresponding category via a convex combination of the prior estimate and the experienced payoff. The weight placed on the new observation diminishes as the number of times a category is sampled grows, mirroring the standard temporal difference updates in Q-learning.

We first note as a benchmark that if each alternative were treated as a separate category in our model, the DM would eventually learn to select the best available alternative in each menu almost surely (assuming a high payoff-sensitivity level where she overwhelmingly favors the highest-valued option). Under this benchmark, the long-run ranking is independent of initial conditions, features no instability, and exhibits no indifference outside of non-generic cases where mean payoffs are exactly identical. This follows as a corollary of standard convergence results of single-agent Q-learning in the context of general finite Markov decision processes (Watkins and Dayan, 1992). By contrast, when the DM relies on coarse categories, the long-run dynamics are fundamentally altered. To characterize these novel long-run outcomes, we apply stochastic approximation to derive the underlying continuous-time mean-field learning dynamics, a framework we refer to as Coarse Q-Learning (CQL).

We establish that the set of steady states under Coarse Q-Learning is non-empty and, generically, finite with odd parity. As the payoff sensitivity parameter approaches infinity, the accumulation points of these steady states select a subset of the Valuation Equilibria (VE) of Jehiel and Samet (2007). We subsequently use different examples to illustrate the long-run dynamics. In settings with two similarity classes, we demonstrate both the possibility of multiple steady states and the existence of a unique steady state where valuations of the two classes are equalized in the high-sensitivity limit (indifference).² In these two-class examples, the system always converges to a steady state, though the specific attractor depends on the

¹The logit formulation has the advantage that it is the unique stochastic choice rule satisfying several desirable properties, such as continuity, independence of irrelevant alternatives, and translation invariance.

²While the observation that multiple (strict) VE can arise in decision problems appears in Jehiel and Samet (2007), the finding that mixed VE may occur, requiring the equalization of valuations, is new to this paper. Moreover, we show that while the set of mixed VE may be infinite in generic decision trees with three or more classes, the corresponding set of limiting CQL steady states for diverging sensitivity remains finite.

initial conditions when multiple exist (indeterminacy). Finally, using a decision problem with three similarity classes, we show that the learning dynamics may never settle down, instead cycling perpetually through all possible rankings (instability).

We extend these insights by allowing for general payoff distributions and an arbitrary (finite) number of similarity classes. We first prove that when a strict valuation equilibrium (with a strict total order on classes) exists, there is a unique steady state in its neighborhood that is locally asymptotically stable under the Coarse Q-Learning dynamics in the high-sensitivity limit. Conversely, when no strict valuation equilibrium exists, it is possible that no steady state is asymptotically stable, leading to the cyclical dynamics observed in the three-class example. To proceed further, we parameterize decision problems by the magnitude of (class-level) expected payoffs obtained in “singleton menus” where all available alternatives belong to the same similarity class. When these singleton menu expected payoffs are uniformly large, we show that there is a unique steady state involving the equalization of valuations over several classes in the high-sensitivity limit; furthermore, the learning dynamics converge to this steady state independent of initial conditions, making it globally asymptotically stable. By contrast, when singleton menu expected payoffs are uniformly small, multiple steady states emerge, with at least one being strict and, hence, locally stable. Finally, we show that instability may become unavoidable when singleton menu payoffs are intermediate.³

The fundamental mechanism driving these novel dynamics is the endogeneity of the learning process: under Coarse Q-Learning, the valuation updates for a given category depend endogenously on how frequently that category is chosen across different menus. Ultimately, a category’s long-run valuation is a weighted average of the payoffs from its constituent alternatives, but crucially, these weights are determined by the DM’s endogenous and evolving strategy.⁴ Consequently, steady states must be analyzed as fixed points of a dynamical system rather than the solutions to an optimization problem, which also explains why convergence is not always guaranteed, as our three-class example demonstrates. The stabilizing effect of large singleton menu payoffs arises because an overvalued category is chosen more frequently in competitive (multi-category) menus where expected payoffs are relatively lower. This disproportionate sampling pulls its valuation downward, creating a mean-reverting pressure. Conversely, when singleton menu payoffs are uniformly small, the highest payoffs are

³Singleton menus are especially tractable because the agent’s class choice is degenerate and valuation-independent there. They provide a clean way to shift payoffs exogenously without directly affecting the interaction among classes in non-singleton menus. For an economic motivation, see the application in 5.1.

⁴If menus were invariant, this weighting would be independent of the choice strategy. This underscores the critical role that menu variation (coupled with coarse categorization) plays in generating these rich dynamics. Bandits with stochastically varying menus or “sleeping bandits” have also been studied in the online regret-minimization literature (Kleinberg et al., 2010) but without the coarse categorization aspect.

obtained in competitive menus, which makes several (initial) strict rankings of valuations self-confirming and naturally generates a multiplicity of locally stable steady states.

The above mechanism relies on the DM pooling heterogeneous alternatives into coarse categories. We motivate such coarse learning as a natural response to the complexity of rich choice environments. When the universe of alternatives is vast (as in the wine example), evaluating and memorizing a distinct valuation for every single item is cognitively infeasible. A DM constrained to tracking at most n valuations must compress the environment into no more than n coarse classes. Indeed, the necessity of categorization for learning has a rich tradition in psychology (Rosch and Lloyd, 1978). In this light, our framework formally bridges two foundational psychological concepts - reinforcement learning and categorization, within a tractable economic model. While we treat the DM’s similarity partition as a primitive, a natural interpretation is that decision-makers group alternatives based on their most salient attributes (Tversky, 1972; Bordalo et al., 2012, 2013). From this perspective, physiological and psychological considerations can inform which attributes become salient in a given context, such as region and price for wine. In market settings, this raises the possibility that strategic actors, such as monopolist online platforms, may exploit limited salience by influencing which attributes consumers attend to, and/or by shaping which products represent each perceived category across menus, as illustrated in an application in Section 5.1.

Sec. 2 presents the learning model in discrete time, and provides a continuous-time ODE approximation of the asymptotic dynamics. Sec. 3 illustrates the novel qualitative phenomena observed in the long-run through various examples. Sec. 4 presents our main analytical results on the equilibrium structure as well as convergence (or lack of) of Coarse Q-learning trajectories to equilibrium. Sec. 5 discusses complementary interpretations for the exogenous similarity partition in our model and provides a misspecified Bayesian foundation for our learning model. Sec. 6 discusses related literature and Sec. 7 concludes the paper by highlighting key takeaways and open questions. All proofs appear in the appendix.

2 Model

2.1 Primitives

Let \mathcal{A} be a finite universal set of alternatives, and let $\Psi := \{\psi \subseteq \mathcal{A} : \psi \neq \emptyset\}$ denote the set of *menus*. Time is discrete, $k \in \mathbb{N}_0$. In each period k , Nature draws a menu $\psi_k \in \Psi$ i.i.d. according to an exogenous probability mass function $f \in \Delta(\Psi)$. If Alice samples alternative $a \in \psi_k$ in period k , she observes a payoff $r_{a,k}$ distributed according to $R_a \in \Delta(\mathbb{R})$. For each $a \in \mathcal{A}$, the sequence $\{r_{a,k}\}_{k \geq 0}$ is i.i.d., and payoff draws are independent across alternatives

and time. Payoff laws are menu-independent: R_a does not depend on ψ_k . We assume a uniform moment bound: there exist $\delta > 0$ and $\kappa < \infty$ such that $\sup_{a \in \mathcal{A}} \mathbb{E}[|r_{a,k}|^{2+\delta}] \leq \kappa$, and define $\mu_a := \mathbb{E}[r_{a,k}]$. The collection $\{R_a\}_{a \in \mathcal{A}}$ is unknown to the risk-neutral Alice who wishes to learn her preferences represented by $\{\mu_a\}_{a \in \mathcal{A}}$ by aggregating sampled payoffs.⁵

We assume that Alice represents the space of alternatives through a coarse “similarity” partition. Formally, let \sim be an exogenously given equivalence relation on \mathcal{A} , interpreted as perceived similarity. Let $\mathcal{S} := \mathcal{A}/\sim = \{[a] : a \in \mathcal{A}\}$ denote the set of similarity classes (quotient space), where $[a] := \{a' \in \mathcal{A} : a' \sim a\}$ is the equivalence class of a . The associated canonical surjection map $\rho : \mathcal{A} \rightarrow \mathcal{S}$, $\rho(a) := [a]$ induces a partition of \mathcal{A} into similarity classes. Throughout, Alice represents an alternative a only via its class $s = \rho(a) \in \mathcal{S}$. Moreover, she forms and updates payoff estimates only at the class level, treating alternatives within a class as having approximately the same expected payoff. We motivate this coarse representation in Sections 5.1–5.2, via complementary interpretations that apply across a wide range of economic contexts: (i) misspecification arising from limited salience of payoff-relevant attributes that characterize alternatives, and (ii) memory and sample complexity constraints when the space of alternatives is very large, in line with categorization principles widely studied in psychology (Anderson, 1991; Rosch and Lloyd, 1978).

2.2 Discrete-time Learning Model

Let $\mathcal{T} = (\mathcal{A}, \Psi, f, r)$ denote the stage decision tree with Nature at the root. In each period k , Nature draws a menu $\psi_k \in \Psi$ i.i.d. with law f . Alice perceives the realized menu only through the induced set of classes $\omega_k := \rho[\psi_k] = \{\rho(a) : a \in \psi_k\} \subseteq \mathcal{S}$. Given her information \mathcal{I}_k ,⁶ she selects a class $s_k \in \omega_k$ according to an \mathcal{I}_k -measurable stochastic choice policy $\sigma_k(\cdot | \omega_k)$, then chooses some alternative $a_k \in s_k \cap \psi_k$ (according to a fixed tie-breaking rule), and receives payoff $r_k = r_{a_k,k} \sim R_{a_k}$. Conditional on a_k , the payoff draw r_k is independent of the history and of the current menu ψ_k . Thus, the stage problem repeats i.i.d. over time.

We introduce *valuations* as real-valued functions on the set of similarity classes, $v : \mathcal{S} \rightarrow \mathbb{R}$. For each period k , let $v_k(s)$ denote Alice’s current estimate of the expected payoff of class $s \in \mathcal{S}$, and let $\mathbf{v}_k := (v_k(s))_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$ be the valuation vector that parametrizes her decision rule. Alice maintains one valuation per class, treating alternatives within a class as payoff-equivalent in expectation and residual within-class heterogeneity as i.i.d. noise.

⁵A classic interpretation is that each alternative a has a deterministic *fundamental* payoff μ_a , and the observed payoff is a noisy signal $r_{a,k} = \mu_a + \varepsilon_{a,k}$, where $\{\varepsilon_{a,k}\}_{k \geq 0}$ are i.i.d. mean-zero shocks. The stochasticity of payoffs may reflect either imperfect monitoring of the fundamentals or idiosyncratic taste shocks.

⁶Alice’s information is summarized by the filtration generated by realized class menus, chosen classes, and realized payoffs. At the start of period k , she observes the current class menu ω_k along with past $(s_t, r_t)_{t < k}$. Formally, let $\mathcal{I}_0 := \sigma\text{-alg}(\omega_0)$, with $\mathcal{I}_{k+1} := \sigma\text{-alg}(\mathcal{I}_k, s_k, r_k, \omega_{k+1})$, so that $\mathcal{I}_k = \sigma\text{-alg}(\omega_0, s_0, r_0, \dots, \omega_k)$.

In period k , faced with menu ψ_k , Alice perceives each $a \in \psi_k$ only through its similarity class $s = \rho(a)$ and therefore randomizes over the available class set $\omega_k := \rho[\psi_k] \subseteq \mathcal{S}$. She selects a class $s_k \in \omega_k$ with probability $\sigma_k(s_k | \omega_k)$, where $\sigma_k(\cdot | \omega_k)$ depends on the current valuations $(v_k(s'))_{s' \in \omega_k}$ of available classes. Conditional on choosing class s_k , she selects an alternative $a \in s_k \cap \psi_k$ uniformly at random.⁷ For each available $\omega \subseteq \mathcal{S}$ and valuation profile $\mathbf{v} \in \mathbb{R}^{\mathcal{S}}$, a stochastic choice rule assigns a *mixed action* $\sigma(\cdot | \omega; \mathbf{v}) \in \Delta(\omega)$. We model $\sigma(\cdot | \omega; \mathbf{v})$ using the multinomial logit (softmax) rule (Luce, 1959), as is standard in classic learning models such as reinforcement learning and stochastic fictitious play (Sutton and Barto, 2018; Fudenberg and Levine, 1998; Hofbauer and Sandholm, 2002).⁸ Specifically, given menu $\omega_k \subseteq \mathcal{S}$ and valuations $\mathbf{v}_k \in \mathbb{R}^{\mathcal{S}}$, class s is chosen in period k with probability

$$\sigma(s | \omega_k; \mathbf{v}_k) = \mathbf{1}\{s \in \omega_k\} \frac{\exp(\beta v_k(s))}{\sum_{j \in \omega_k} \exp(\beta v_k(j))}, \quad \beta \geq 0.$$

The parameter $\beta \in \mathbb{R}_+$ is a scaling constant that determines Alice’s sensitivity to valuation differences. It has a smoothing effect with $\beta = 0$ leading to a uniform choice over all available classes, while for $\beta \rightarrow \infty$, the probabilities concentrate on similarity class(es) with the highest valuation(s). Most of the analysis in this paper is conducted in the high-sensitivity limit ($\beta \rightarrow \infty$) where Alice almost surely chooses an alternative in a similarity class with maximal current valuation. Let $|\cdot|$ denote the cardinality of a set. Under the nested rule (softmax over classes and uniform tie-breaking within the chosen class),⁹ given menu ψ_k and valuations $\mathbf{v}_k \in \mathbb{R}^{\mathcal{S}}$, alternative a is chosen in period k with probability

$$\nu_k(a | \psi_k, \mathbf{v}_k) := \Pr(a_k = a | \psi_k, \mathbf{v}_k) = \frac{\sigma(\rho(a) | \omega_k; \mathbf{v}_k)}{|\rho(a) \cap \psi_k|}, \quad a \in \psi_k. \quad (1)$$

⁷Since Alice’s information \mathcal{I}_k contains no within-class identities, any admissible \mathcal{I}_k -measurable within-class rule cannot condition on labels of alternatives in $s_k \cap \psi_k$. More generally, after Alice selects a class s_k , the realized alternative can be drawn according to an exogenously specified tie-breaking kernel $\Upsilon(\cdot | s_k, \psi_k)$ supported on $s_k \cap \psi_k$. We work with the uniform kernel, $\Upsilon(a | s, \psi_k) = \frac{1}{|s \cap \psi_k|}$, $a \in s \cap \psi_k$, because it is anonymous (label-independent) within each class but our analysis extends to any exogenous kernel.

⁸Building on Luce (1959), Breitmoser (2021) provides an axiomatic characterization of multinomial logit, showing that for a fixed utility scale, it is the unique stochastic choice rule that satisfies continuity, positivity, independence of irrelevant alternatives (IIA), label-independence, and translation invariance.

⁹Alice’s two-step nested choice rule (logit over classes, uniform within a class) is immune to the “duplicates problem” (Debreu, 1960). If an alternative $a \in \psi$ is duplicated by adding a copy a' with $\rho(a') = \rho(a) =: s$, then the induced class menu $\omega = \rho(\psi)$ is unchanged, hence the class-choice probabilities $\sigma(\cdot | \omega; \mathbf{v}_k)$ are unchanged. Only the within-class implementation adjusts: under uniform tie-breaking, each $b \in s \cap \psi$ is chosen with probability $\sigma(s | \omega; \mathbf{v}_k) / |s \cap \psi|$. Consequently, the total probability of selecting class s is invariant to duplication, avoiding the inflation that would arise under a one-step Luce choice rule over alternatives in the presence of identical copies. Finally, because Alice models payoff shocks as i.i.d. across alternatives (hence no correlation across classes), the usual information-based critiques of IIA do not apply.

At the transition from period k to $k+1$, after choosing $a_k \in \psi_k$ according to the alternative-level mixed action $\nu_k(\cdot \mid \psi_k, \mathbf{v}_k)$, Alice observes the realized payoff $r_k := r_{a_k} \sim R_{a_k}$ and the next menu ψ_{k+1} , and receives no feedback about forgone payoffs. At the beginning of period $k+1$ (before choosing a_{k+1}), she updates only the valuation $v_k(s_k)$ of the previously chosen class $s_k := \rho(a_k)$ toward the observed payoff via the temporal-difference (TD) recursion

$$v_{k+1}(s) = v_k(s) + \alpha_k(s) \mathbf{1}\{s = s_k\} \left[r_{a_k} + \gamma \max_{s' \in \omega_{k+1}} v_k(s') - v_k(s) \right], \quad s \in \mathcal{S}, \quad (2)$$

where $\gamma \in [0, 1)$ is her discount factor and $M_{k+1} := \max_{s' \in \omega_{k+1}} v_k(s')$ is her maximum continuation value in period $k+1$ given her period- k valuations \mathbf{v}_k . Alice's update rule is a simple temporal-difference value iteration scheme with a Bellman-style one-step target: between periods k and $k+1$, she revises only the valuation of the chosen class $s_k = \rho(a_k)$ by taking an affine combination of its current valuation $v_k(s_k)$ and the TD target $r_k + \gamma M_{k+1}$, with step-size $\alpha_k(s_k) > 0$ determining the weight placed on new information. In vector form, letting $\mathbf{e}_s \in \mathbb{R}^{\mathcal{S}}$ denote the standard basis vector, the TD update can be written as

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \alpha_k(s_k) \left(r_{a_k} + \gamma M_{k+1} - v_k(s_k) \right) \mathbf{e}_{s_k}.$$

This mirrors the bandit (stateless) variant of canonical Q-learning (Watkins and Dayan, 1992; Tsitsiklis, 1994), except that valuations are indexed by similarity classes rather than individual alternatives, reflecting that Alice treats only classes as payoff-relevant. Our goal is to characterize how this coarsening affects the long-run Q-learning dynamics.

We assume that for each $s \in \mathcal{S}$, $\{\alpha_k(s)\}_{k \geq 0}$ are (possibly stochastic) step-sizes adapted to Alice's information $(\mathcal{I}_k)_{k \geq 0}$, and satisfy the Robbins and Monro (1951) conditions:¹⁰

$$0 < \alpha_k(s) \leq \bar{\alpha} < \infty \quad \text{and} \quad \sum_k \alpha_k(s) = \infty \quad \text{and} \quad \sum_k (\alpha_k(s))^2 < \infty \quad \text{almost surely.}$$

These conditions on the step-sizes imply that Alice's sensitivity to new observations diminishes eventually, while ensuring that future observations still exert a non-negligible impact.¹¹

Alice's choices in period k are determined by her current valuation vector \mathbf{v}_k , while the real-

¹⁰When $\bar{\alpha} = 1$, the valuation update can be interpreted as a convex combination of the old estimate and the new information in each period, with $\alpha_k(s)$ being the weight on the latter.

¹¹If Alice views payoffs as i.i.d. draws from a stationary distribution, it's natural that she eventually down-weights new observations (see Sec. 5.3 for a Bayesian foundation or Stylized Fact 2 in Benjamin (2019)). While Q-learning is sometimes criticized as slow or sample-inefficient as a model of human learning (Gershman and Daw, 2017), we show in Online Appendix (Sec. A) that this assessment is sensitive to the learning-rate design: inverse-propensity-weighted step-sizes that upweight signals from rarely sampled classes accelerate convergence by asymptotically equalizing effective information arrival rates across classes.

ized payoff (and hence the update to \mathbf{v}_{k+1}) depends on \mathbf{v}_k through the induced choice probabilities. Iterating this feedback yields a discrete-time stochastic process $\{\mathbf{v}_k\}_{k \geq 0}$ governed by (1) and (2). The discrete-time Coarse Q-learning model is fully specified by (1)–(2) together with an initial condition \mathbf{v}_0 , interpreted as Alice’s initial assessments of class-level expected payoffs. To characterize long-run behavior, we use stochastic approximation. Under Robbins-Monro step-sizes, (2) can be written as $v_{k+1}(s) - v_k(s) = \alpha_k(s) \left(h_s(\mathbf{v}_k) + \eta_{k+1}(s) \right)$, $s \in \mathcal{S}$, where $h_s(\mathbf{v}_k) := \mathbb{E}[v_{k+1}(s) - v_k(s) \mid \mathcal{I}_k] / \alpha_k(s)$ is the conditional drift (systematic update direction) and $\eta_{k+1}(s)$ is a mean-zero noise term with $\mathbb{E}[\eta_{k+1}(s) \mid \mathcal{I}_k] = 0$ implying $\{\eta_{k+1}(s)\}_{k \geq 0}$ is a martingale-difference sequence.¹² The conditional drift $h_s(\mathbf{v}_k)$ written as $g_s(\mathbf{v}_k) - v_k(s)$ with $g_s(\mathbf{v}) = \mathbb{E}\left[r_k + \gamma \max_{j \in \omega_{k+1}} v(j) \mid s_k = s, \mathbf{v}_k = \mathbf{v}\right]$, represents the expected prediction error for class s given current estimates. Intuitively, the recursion moves $v(s)$ upward when the realized reward for class s exceeds $v(s)$ on average, and downward otherwise. Replacing random increments by their conditional means yields the mean-field ODE $\dot{v}(s) = h_s(\mathbf{v})$, $s \in \mathcal{S}$, which describes the expected motion of the learning process in the long run.

2.3 Continuous-time Asymptotic Approximation

Standard stochastic-approximation results (Benaïm, 1999; Kushner and Yin, 2003) imply that the continuous-time interpolation of $\{\mathbf{v}_k\}_{k \geq 0}$ is an asymptotic pseudo-trajectory of the mean-field ODE (see Online Appendix). In particular, if valuations remain bounded a.s.,¹³ then the ω -limit set of any sample path of (2) is a.s. a compact, connected, internally chain-transitive (ICT) set of the mean-field flow (Benaïm, 1999).¹⁴ Starting with the case of a myopic DM ($\gamma = 0$), the mean-field Coarse Q-learning (CQL) dynamics are given by:

$$\dot{v}_s = g_s(\mathbf{v}) - v_s, \quad s \in \mathcal{S}, \quad (3)$$

where $g_s(\mathbf{v})$ is the expected payoff of class s conditional on choosing s under \mathbf{v} :

$$g_s(\mathbf{v}) := \frac{\sum_{\psi \in \Psi} f(\psi) \sigma(s \mid \rho(\psi); \mathbf{v}) \mu_s(\psi)}{\sum_{\psi \in \Psi} f(\psi) \sigma(s \mid \rho(\psi); \mathbf{v})}, \quad \mu_s(\psi) := \frac{1}{|s \cap \psi|} \sum_{a \in s \cap \psi} \mu_a,$$

and $\mu_a := \mathbb{E}[r_a]$. By translation invariance of softmax, $g(\mathbf{v} + c\mathbf{1}) = g(\mathbf{v})$ for all $c \in \mathbb{R}$.

¹²A formal construction of (h, η) and the relevant regularity conditions for SA appear in the Online Appx.

¹³This holds trivially under bounded payoffs; under unbounded payoffs we work with a projected recursion onto a sufficiently large, compact, positively invariant set K , which is without loss for the mean-field analysis since the ODE vector field points inward on ∂K (see the Online Appendix (Sec. A) for details).

¹⁴The ω -limit set of a stochastic process $\{\mathbf{x}_k\}$ is the set of all points \mathbf{x} in the associated state space such that $\mathbf{x}_{k_n} \rightarrow \mathbf{x}$ along some subsequence $k_n \rightarrow \infty$, almost surely. Internally chain transitive (ICT) sets of the flow generated by the ODE (3) are compact invariant sets that are chain-transitive under arbitrarily small pseudo-orbits (Conley, 1978). They arise as ω -limit sets of asymptotic pseudo-trajectories and may consist of equilibria, periodic orbits, or chaotic attractors.

To build intuition for the ODE approximation, fix a valuation vector \mathbf{v} and consider a short time window over which valuations change negligibly, so behavior is well approximated by the stationary logit policy induced by \mathbf{v} . In each period Nature draws ψ with probability $f(\psi)$, and conditional on ψ Alice selects $s \in \rho(\psi)$ with probability $\sigma(s \mid \rho(\psi); \mathbf{v})$. If s is selected in menu ψ , the expected payoff equals the within-menu average $\mu_s(\psi) = |s \cap \psi|^{-1} \sum_{a \in s \cap \psi} \mu_a$. Averaging over menus and conditioning on s being chosen yields the class- s signal $g_s(\mathbf{v}) = \mathbb{E}[\mu_s(\psi) \mid s \text{ is chosen under the policy induced by } \mathbf{v}]$, which is the ratio defining $g_s(\mathbf{v})$ above. Under the myopic TD update ($\gamma = 0$), the expected increment in v_s is proportional to the prediction error $g_s(\mathbf{v}) - v_s$, leading after time-rescaling to the mean-field dynamics $\dot{v}_s = g_s(\mathbf{v}) - v_s$. Vanishing step-sizes underpin this approximation: valuations evolve slowly relative to the i.i.d. menu and payoff draws, so over long horizons the accumulated noise averages out and the recursion tracks the deterministic drift.¹⁵

Even if each alternative $a \in \mathcal{A}$ has a menu-invariant expected payoff μ_a , the class-level mean $\mu_s(\psi)$ is typically menu-dependent: different menus induce different within-class subsets $s \cap \psi$, and alternatives within a class need not share the same μ_a .¹⁶ Under uniform within-class tie-breaking, choosing class s in menu ψ therefore generates a payoff drawn from the uniform mixture over $s \cap \psi$, so the distribution of class- s signals varies with the realized menu. Moreover, the signal stream relevant for updating class s is endogenously reweighted across menus: conditional on s being chosen under valuations \mathbf{v} , menus are weighted by $f(\psi)\sigma(s \mid \rho(\psi); \mathbf{v})$, and $\sigma(s \mid \rho(\psi); \mathbf{v})$ depends on the set of co-occurring classes $\rho(\psi)$, which varies with ψ . As a result, the empirical signals used to update each class depend on the menu process and endogenously on Alice’s past valuations through her own sampling behavior, even if payoffs are i.i.d. at the alternative-level, leading to *active learning* dynamics.

For a forward-looking DM ($\gamma > 0$), the TD increment in (2) includes the greedy Q-learning continuation term $\max_{j \in \rho(\psi)} v(j)$. More generally, let $\kappa_\psi(\mathbf{v})$ denote the continuation operator

¹⁵For e.g., take $\alpha_k = (k+1)^{-1}$. The errors $\{\eta_{k+1}\}$ form a martingale-difference sequence with $\mathbb{E}[\eta_{k+1} \mid \mathcal{I}_k] = 0$ and, under the uniform moment bound, $\mathbb{E}[\|\eta_{k+1}\|^2 \mid \mathcal{I}_k] = O(1)$. Hence the scaled noise increment $\alpha_k \eta_{k+1}$ has conditional variance $\text{Var}(\alpha_k \eta_{k+1} \mid \mathcal{I}_k) = \alpha_k^2 \text{Var}(\eta_{k+1} \mid \mathcal{I}_k) = O(1/k^2)$, whereas the deterministic drift increment $\alpha_k h(\cdot)$ is of order $O(\alpha_k) = O(1/k)$. With such a deterministic calendar-time step-size, the limiting mean-field dynamics are generally asynchronous: each component is rescaled by the corresponding strictly positive class-selection propensity, so that $\dot{v}_s = \lambda_s(\mathbf{v})(g_s(\mathbf{v}) - v_s)$ for suitable $\lambda_s(\mathbf{v}) > 0$. The synchronous ODE in (3) is obtained by using inverse-propensity-weighted step-sizes, with propensities estimated on a faster timescale, which asymptotically equalize expected updating rates across classes. As discussed in Sec. 5.3, our results extend to the asynchronous case, so this normalization is without loss.

¹⁶Our analysis extends verbatim to menu-dependent primitive payoffs. Allow $r_{a,k} \sim R_{a,\psi_k}$ whenever $a \in \psi_k$, with means $\mu_a(\psi) := \mathbb{E}[r_{a,k} \mid a \in \psi_k = \psi]$, and assume the same conditional independence holds uniformly over (a, ψ) . Then $\mu_s(\psi) := \frac{1}{|s \cap \psi|} \sum_{a \in s \cap \psi} \mu_a(\psi)$ represents the class average in the drift, with no other changes to the mean-field analysis. We impose menu-invariant $\{R_a\}$ for parsimony, to emphasize that menu dependence in our dynamics is induced by coarse perception rather than by primitive payoff externalities.

associated with menu ψ and define $C(\mathbf{v}) := \sum_{\psi \in \Psi} f(\psi) \kappa_{\psi}(\mathbf{v})$. Since menu transitions are action-independent, the expected continuation term does not depend on the chosen class, so it enters the mean drift as a common scalar $\gamma C(\mathbf{v})$ added to every component: $\dot{\mathbf{v}} = g(\mathbf{v}) + \gamma C(\mathbf{v})\mathbf{1} - \mathbf{v}$. If κ_{ψ} is translation-covariant, i.e. $C(\mathbf{v} + c\mathbf{1}) = C(\mathbf{v}) + c$ as $\max_{j \in \rho(\psi)} v(j)$ clearly is, then this common term can be translated out without affecting the induced softmax policy.¹⁷ Accordingly, the relative valuations $\tilde{\mathbf{v}}(t) := \mathbf{v}(t) - \eta(t)\mathbf{1}$ (with $\dot{\eta} = \gamma C(\mathbf{v}) - \eta$) evolve as in the myopic mean-field system $\dot{\tilde{\mathbf{v}}} = g(\tilde{\mathbf{v}}) - \tilde{\mathbf{v}}$ (see Lemma B.9). Thus, the forward-looking system is semi-conjugate to the myopic system via the time-varying shift $\mathbf{v} = \tilde{\mathbf{v}} + \eta\mathbf{1}$; only the level of valuations changes, not their relative evolution. As softmax probabilities depend only on relative valuations, this semi-conjugacy ensures that all qualitative long-run phenomena established for the myopic case carry over exactly to the forward-looking agent.

Reduction of the decision tree: Since the long-run behavior of the discrete-time CQL model in (2) is governed by the mean-field CQL dynamics in (3), it's without loss to collapse the original decision tree \mathcal{T} by aggregating alternatives within each class. We denote the reduced menu space by $\Omega := \{\omega \subseteq \mathcal{S} : \omega \neq \emptyset\}$, and let $\Psi_{\omega} := \{\psi \in \Psi : \rho(\psi) = \omega\}$. The latter aggregates all menus in the original decision tree that feature the same set of available similarity classes. We define a probability mass function on Ω by $p(\omega) = \sum_{\psi \in \Psi_{\omega}} f(\psi)$, $\forall \omega \in \Omega$. For each $\omega \in \Omega$ and $s \in \omega$, define the expected payoff of class s in menu ω by

$$\pi_s(\omega) = \frac{\sum_{\psi \in \Psi_{\omega}} f(\psi) \mu_s(\psi)}{\sum_{\psi \in \Psi_{\omega}} f(\psi)}, \quad \mu_s(\psi) := \frac{1}{|s \cap \psi|} \sum_{a \in s \cap \psi} \mu_a.$$

The reduced tree $\mathcal{T}'(\mathcal{S}, \Omega, p, \pi)$ has menu-space Ω , with menus drawn i.i.d. by Nature with probability $p(\cdot)$, and class-level expected payoffs $\pi_s(\cdot)$, for example, see Fig. 1. Each menu in \mathcal{T}' features a unique set of available classes. Denoting the logit policy in menu ω by

$$\sigma_{\omega}^s(\mathbf{v}) = \mathbf{1}\{s \in \omega\} \frac{\exp(\beta v_s)}{\sum_{j \in \omega} \exp(\beta v_j)},$$

the conditional expected payoff of class s computed on $\mathcal{T}'(\mathcal{S}, \Omega, p, \pi)$ under \mathbf{v}

$$g_s(\mathbf{v}) = \frac{\sum_{\omega \in \Omega} p(\omega) \sigma_{\omega}^s(\mathbf{v}) \pi_s(\omega)}{\sum_{\omega \in \Omega} p(\omega) \sigma_{\omega}^s(\mathbf{v})},$$

¹⁷(2) uses the greedy Q-learning operator $\kappa_{\psi}(\mathbf{v}) = \max_{j \in \rho(\psi)} v(j)$. Two common alternatives are the Log-SumExp operator in dynamic logit $\kappa_{\psi}(\mathbf{v}) = \beta^{-1} \log \sum_{j \in \rho(\psi)} \exp(\beta v(j))$ (Rust, 1987; Steiner et al., 2017) and the SARSA operator $\kappa_{\psi}(\mathbf{v}) = \sum_{j \in \rho(\psi)} \sigma(j | \rho(\psi); \mathbf{v}) v(j)$ (Sutton and Barto, 2018). All are translation-covariant, hence induce the same relative dynamics in our setting with action-independent menu transitions.

is identical to the drift obtained on the original tree $\mathcal{T}(\Psi, f, r)$, since $\pi_s(\omega)$ is exactly the f -average of the within-menu class expected payoffs $\mu_s(\psi)$ over Ψ_ω . Hence the reduction to \mathcal{T}' is without loss for the continuous-time CQL dynamics $\dot{v}_s = g_s(\mathbf{v}) - v_s$ as the continuous-time limit depends only on the expected motion of the discrete-time stochastic CQL process.¹⁸

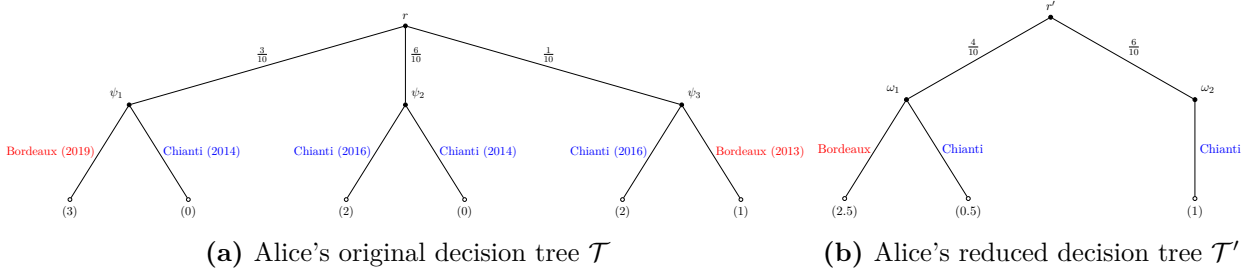


Figure 1: Example of tree-reduction for the mean-field CQL dynamics

Trivial cases: Since $g_s(\mathbf{v})$ involves the softmax $\sigma_\omega^s(\mathbf{v})$, the drift is a transcendental function of \mathbf{v} , and closed-form algebraic solutions of (3) are generally unavailable for $\beta > 0$. We list four regimes in which the drift is linear in \mathbf{v} and the ODE decouples with $\dot{\mathbf{v}}(t) = \bar{\boldsymbol{\pi}} - \mathbf{v}(t) \Rightarrow \mathbf{v}(t) = \bar{\boldsymbol{\pi}} + (\mathbf{v}(0) - \bar{\boldsymbol{\pi}})e^{-t}$, $\forall t \geq 0$ and for some constant $\bar{\boldsymbol{\pi}}$. Hence, as $t \rightarrow \infty$, $\mathbf{v}(t) \rightarrow \bar{\boldsymbol{\pi}}$.

- (i) *Pure exploration* ($\beta = 0$) where Alice is totally insensitive to payoffs. For any $\omega = \rho[\psi]$, $\sigma_\omega^s(\mathbf{v}) = 1/|\omega|$, so the constant $\bar{\pi}_s = \frac{\sum_\psi f(\psi) (|\omega|)^{-1} \mathbf{1}_{\{s \in \mathcal{S}_\psi\}} \mu_s(\psi)}{\sum_\psi f(\psi) (|\omega|)^{-1} \mathbf{1}_{\{s \in \mathcal{S}_\psi\}}}$, the selection-conditional average of the within-menu class-level expected payoffs. The endogenous selection bias channel drops out as Alice's sampling process is independent of her valuations.
- (ii) *Finest partition* ($\mathcal{S} = \mathcal{A}$): Alice maintains a separate valuation for each alternative. Since $\rho(a) = a$, for any menu ψ and $a \in \psi$, the within-menu average $\mu_{\rho(a)}(\psi) = \mu_a$. The mean-field drift decouples coordinate-wise: $\dot{v}_a = \mu_a - v_a$, $a \in \mathcal{A}$, so each valuation converges to μ_a (provided each alternative is sampled infinitely often).
- (iii) *Full availability* ($\rho(\psi) = \mathcal{S}$ for all ψ with $f(\psi) > 0$): every menu contains at least one alternative from each class. The reduced tree has a single class menu $\omega = \mathcal{S}$ with $p(\mathcal{S}) = 1$, and $g_s(\mathbf{v}) = \frac{p(\mathcal{S}) \sigma_{\mathcal{S}}^s(\mathbf{v}) \pi_s(\mathcal{S})}{p(\mathcal{S}) \sigma_{\mathcal{S}}^s(\mathbf{v})} = \pi_s(\mathcal{S})$. Thus, the ODE system decouples and for all s , $v_s \rightarrow \pi_s(\mathcal{S})$. Intuitively, since the set of co-occurring classes is the same in every period (as in a multi-armed bandit), the valuation-dependent sampling probability $\sigma_\omega^s(\mathbf{v})$ does not vary across menus and therefore does not endogenously reweight the menu distribution conditional on selecting s ; the selection-bias channel drops out.

¹⁸To see this, note that for any two menus $\psi, \psi' \in \Psi_\omega$ the choice policy over similarity classes depends only on the set of available classes ω and the current valuations, hence $\sigma_\psi^s(\mathbf{v}) = \sigma_{\psi'}^s(\mathbf{v}) = \sigma_\omega^s(\mathbf{v})$ for all $s \in \omega$, and, conditional on choosing class s , the within-class randomization is uniform in both menus.

(iv) *Correct specification*: Alice is correct in believing that alternatives pooled within a class are payoff-homogeneous in expectation, i.e. for each class $s \in \mathcal{S}$, $\mu_a = \mu_s$ for all $a \in s$. Thus, the within-menu class mean is menu-invariant $\mu_s(\psi) \equiv \mu_a$, and the mean-field drift reduces to $\dot{v}_s = \mu_s - v_s$ and the system decouples.¹⁹

In contrast, when similarity classes pool multiple alternatives with heterogeneous expected payoffs and availability varies across menus, the drift $g_s(\mathbf{v})$ inherits the softmax non-linearity, so the system is coupled, non-algebraic, and endogenous for any $\beta > 0$. Our main objective is to study the long-run consequences of such coarse inference.

Steady states: Smooth Valuation Equilibria

Our focus is on the asymptotic behavior of CQL dynamics in (3). The associated dynamical system is real, autonomous, and smooth on $\mathbb{R}^{\mathcal{S}}$. At any time $t \in \mathbb{R}_+$, the state of the CQL dynamical system is given by a vector of valuations, $\mathbf{v}(t) \in \mathbb{R}^{\mathcal{S}}$. First, we consider the steady state solutions of the CQL system in (3), where a steady state solution is defined as a stationary system of valuations $\mathbf{v}^* \in \mathbb{R}^{\mathcal{S}}$ such that $\dot{\mathbf{v}} = 0$ when evaluated at \mathbf{v}^* . Denoting the set of steady states of the ODE system in (3) by $\mathcal{V}(\beta)$ for a given β , we show in Theorem 1 that $\mathcal{V}(\beta)$ is nonempty for any $\beta \in [0, \infty)$. We refer to a steady state of the CQL ODE as a *smooth valuation equilibrium (SVE)* defined below.²⁰ The nomenclature follows Jehiel and Samet (2007) who introduce a similar solution concept in the context of multi-agent extensive-form games called *valuation equilibrium (VE)*.

Definition 1. A choice profile $\sigma = (\sigma_\omega^s)_{\omega \in \Omega}^{s \in \mathcal{S}}$ constitutes a *smooth valuation equilibrium* for $\mathcal{T}'(\mathcal{S}, \Omega, p, \pi)$ if there exists a valuation system $\mathbf{v}^* = (v_s^*)_{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$ such that, $\forall s \in \mathcal{S}$,

$$v_s^* = \frac{\sum_{\omega \in \Omega} p(\omega) \sigma_\omega^s(\mathbf{v}^*) \pi_s(\omega)}{\sum_{\omega \in \Omega} p(\omega) \sigma_\omega^s(\mathbf{v}^*)}, \quad \sigma_\omega^s(\mathbf{v}^*) = \mathbf{1}\{s \in \omega\} \frac{\exp(\beta v_s^*)}{\sum_{j \in \omega} \exp(\beta v_j^*)}.$$

We emphasize that even though the underlying environment is of a decision problem, the definition of smooth valuation equilibrium (SVE) has an inherently fixed-point structure. The logit rule pins down choice probabilities as a function of valuations, while the valuation consis-

¹⁹If every member of class s is equally good in expectation, then which members of s happen to be available in a given menu is irrelevant: menu variation cannot distort the class-level mean signal, and Alice’s sampling behavior cannot induce an endogenous reweighting of “better” or “worse” within-class alternatives because there are none. The residual randomness is i.i.d. noise, so coarse learning is equivalent to correctly learning the alternative means. Indeed, the same logic applies whenever misspecification is nevertheless ordinally correct at the class level: if one class uniformly dominates another in expected payoff across all underlying alternatives, then menu variation cannot overturn their ranking in the induced class-level payoff signals.

²⁰When $\gamma > 0$, the drift includes an additive continuation term of the form $\gamma C(\mathbf{v})\mathbf{1}$. We translate out this common shift and work with normalized valuations, so steady states characterized by $\mathbf{v} = g(\mathbf{v})$ correspond one-to-one (modulo an additive constant) to steady states of the forward-looking dynamics; see Lemma B.9.

tency equation pins down each class valuation as the expected payoff that is induced by those very choice probabilities across menus. An SVE is therefore a pair (σ, \mathbf{v}^*) in which choices are logit-optimal given valuations and valuations are correct assessments of the outcomes generated by those choices. This mutual consistency is the source of genuinely equilibrium-like phenomena in our framework, which are absent from standard decision-theoretic models that treat valuations as exogenous inputs rather than endogenously determined objects.

Before turning to the examples in the next section, it is useful to describe the high-sensitivity benchmark that will help interpret them. As the logit parameter grows without bound ($\beta \uparrow \infty$), Alice chooses among classes with maximal current valuation in each menu with probability approaching one. This motivates the natural greedy benchmark, the Valuation Equilibrium (VE) of [Jehiel and Samet \(2007\)](#): for some menu-wise tie-breaking rule, Alice chooses among valuation-maximizing classes in each menu, and valuations are self-consistent with the expected payoffs generated by this greedy behavior. Theorem 1 in Sec. 4.1 formalizes this link by showing that every high-sensitivity accumulation point of SVE is a VE.²¹

3 Illustrations

We now illustrate the range of CQL dynamics through a few numerical examples, which the subsequent analysis in Sec. 4 formalizes. The first two are based on a decision tree in Fig. 2 with two similarity classes. At the root r , Nature chooses one of three menus ω_1 , ω_2 and ω_3 , uniformly at random. In ω_1 , Alice encounters a binary choice between alternatives L_1 and R_1 . In menus ω_2 and ω_3 , Alice encounters degenerate singleton choices, involving L_2 and R_3 respectively. The set of alternatives is partitioned into two classes, $L = \{L_1, L_2\}$ and $R = \{R_1, R_3\}$. We examine two distinct scenarios by altering the expected payoffs associated with alternatives in the singleton menus, specifically z_2 for L_2 and z_3 for R_3 , while keeping the expected payoffs for alternatives in the binary menu unchanged.

3.1 Example: Multiplicity of SVE

Here, we assume $z_2 = -0.25$ and $z_3 = 0.25$. This implies that Alice receives a strictly lower expected reward in each of the singleton choice menus compared to the binary choice menu, regardless of her actions at the latter. As a result, there are three VE: two pure and one mixed. For each VE, there’s a corresponding SVE in its neighborhood for large β .

- *Strict pure VE favoring L*: The pure policy that selects L in menus ω_1 and ω_2 , and R only in menu ω_3 , constitutes a *strict* pure VE. The consistent valuation at this equilibrium is $(v_L, v_R) = (0.875, 0.250)$, for which the policy is optimal. Numerically,

²¹See Appendix B.1 for a precise definition of VE.

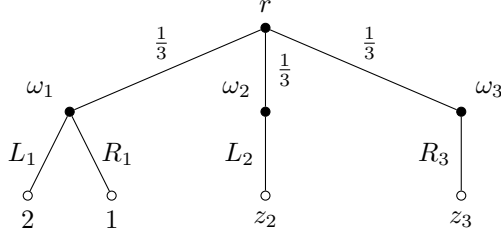


Figure 2: Example of a Reduced Decision Tree with Two Similarity Classes

$(0.875, 0.250)$ arises as a limiting SVE of the CQL dynamics as $\beta \uparrow \infty$; the phase portrait in Fig. 3a is consistent with this. For a large sensitivity ($\beta = 10$), simulations show convergence from nearby initial valuations, providing evidence of *local stability* of this strict pure VE. Intuitively, choosing L in the binary menu ω_1 loads the class- L update on the high expected payoff $\mu_{L_1} = 2$ and largely insulates it from the low singleton $\mu_{L_2} = z_2 < 0$ (which is sampled only when ω_2 occurs). Once $v_L > v_R$, the logit rule further tilts sampling toward L in ω_1 , reinforcing her expectations and generating a locally self-confirming drift toward the L -favoring valuation.

- *Strict pure VE favoring R:* The pure policy that selects R in menus ω_1 and ω_3 , and L only in menu ω_2 , constitutes a *strict* pure VE. The consistent valuation at this equilibrium is $(v_L, v_R) = (-0.250, 0.625)$, for which the policy is optimal. Numerically, $(-0.250, 0.625)$ arises as a limiting SVE of the CQL dynamics as $\beta \uparrow \infty$; the phase portrait in Fig. 3a is consistent with this. For a large sensitivity ($\beta = 10$), simulations show convergence from nearby initial valuations, providing evidence of *local stability* of this strict pure VE. Intuitively, selecting R in ω_1 makes the class- R signal primarily reflect the high expected payoff $\mu_{R_1} = 1$ while the bad singleton $z_3 = 0.25$ is only sampled when ω_3 occurs. On the contrary, if L is rarely selected at ω_1 , its class-level signal is mostly dominated by the low payoff $z_2 = -0.25$ from ω_2 where L is always selected. Thus, when $v_R > v_L$, logit sampling concentrates on R in the only nontrivial menu, so the induced signal stream keeps validating the R -favoring policy. This self-confirming feedback yields local stability of the strict R -favoring equilibrium.
- *Mixed VE with indifference:* In menu ω_1 , the mixed strategy that uniformly randomizes between L_1 and R_1 is a mixed VE. The associated consistent valuation is $(v_L, v_R) = (0.5, 0.5)$, at which this strategy is optimal. Numerically, $(0.5, 0.5)$ also arises as a limiting SVE of the CQL dynamics as $\beta \uparrow \infty$; the phase portrait in Fig. 3a corroborates this. However, it turns out that for a large sensitivity ($\beta = 10$), this mixed VE is unstable: any small perturbation that makes the two valuations unequal leads to the dynamics polarizing toward one of the two strict pure equilibria discussed above.

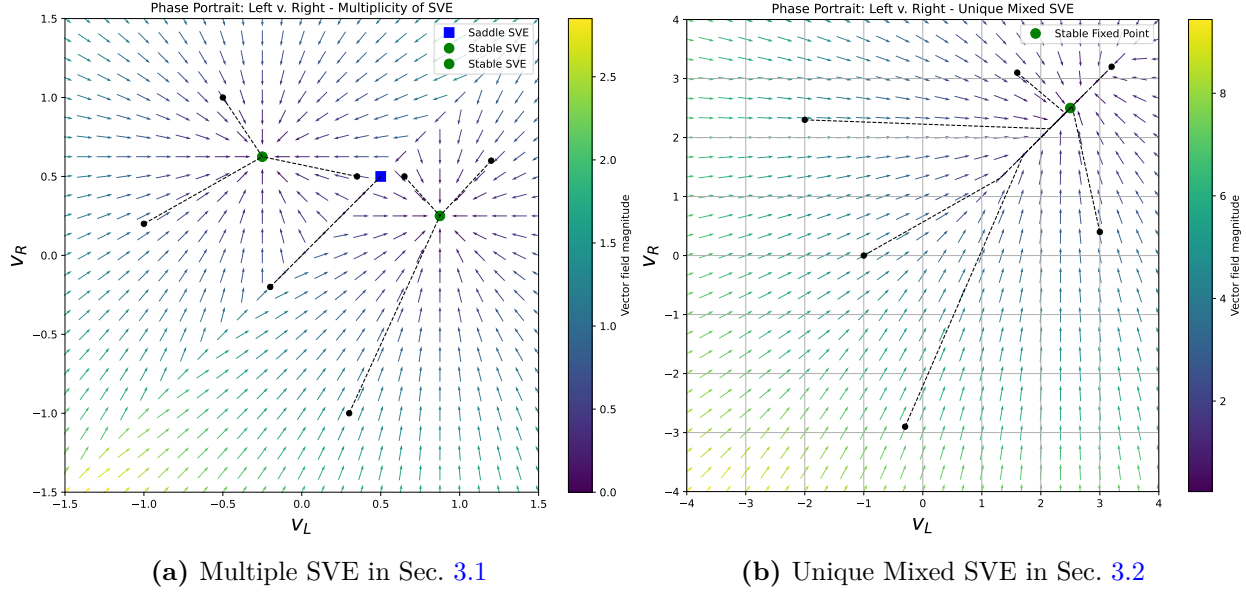


Figure 3: Phase portraits of the CQL dynamics for decision tree in Fig. 2 with $\beta = 10$

3.2 Example: Unique Stable Mixed SVE

Assume now $z_2 = 2.75$ and $z_3 = 3.25$. Thus, Alice receives a strictly higher expected reward in each of the two singleton choice menus compared to her expected reward in the binary choice menu, regardless of her actions at the latter. Consequently, the mixed strategy that uniformly randomizes between L_1 and R_1 in ω_1 constitutes the unique (mixed) VE. The associated equilibrium valuation is $(2.5, 2.5)$, for which the mixed strategy is optimal. Numerically, $(2.5, 2.5)$ also arises as a limiting SVE of the CQL dynamics as $\beta \uparrow \infty$; the phase portrait in Fig. 3b corroborates this. For a large sensitivity ($\beta = 10$), simulations show convergence from all initial valuations, providing evidence of *global stability* of this unique mixed VE. Intuitively, when singleton expected payoffs are sufficiently large, they act as an anchoring force on class valuations. If $v_L > v_R$, logit makes L chosen relatively more often in the only nontrivial menu ω_1 . The singleton-menu choices themselves remain exogenous, but the composition of updates changes: the singleton-menu sampled payoffs now represent a smaller share of L 's total signal stream than of R 's. Thus, the high-valuation class is updated disproportionately from the lower-payoff binary menu, while the low-valuation class is updated disproportionately from the high-payoff singleton menu. This generates mean reversion in the valuation gap, stabilizing the unique mixed equilibrium.

3.3 Example: Unique Unstable Mixed SVE

Consider now an example with three similarity classes based on the decision tree in Fig. 4 with $z_R = -0.4$, $z_P = -0.5$, and $z_S = -0.6$. The twelve alternatives are partitioned into three equivalence classes: $R = \{R_1, R_3, R_4, R_7\}$, $P = \{P_1, P_2, P_5, P_7\}$, and $S = \{S_2, S_3, S_6, S_7\}$.

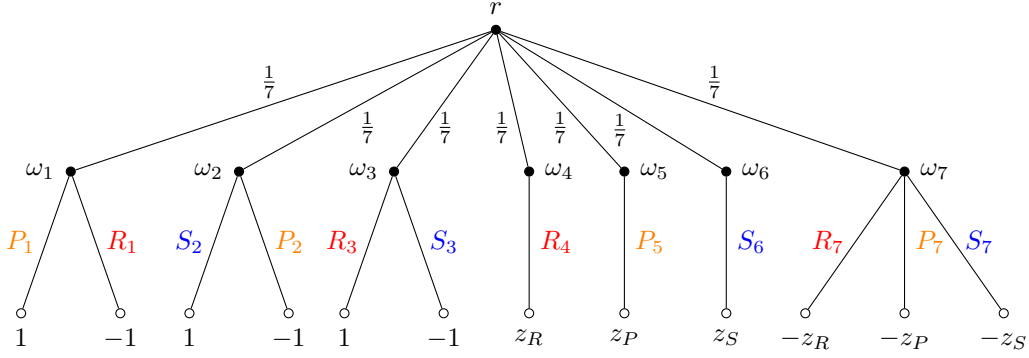


Figure 4: Reduced decision tree for RPS with class-level menu-contingent expected payoffs

This tree admits neither strict pure VE nor partially-mixed VE. Hence any VE must be fully-mixed in the sense that Alice is indifferent across all classes, so $v_R = v_P = v_S$.²² Accordingly, the set of fully-mixed VE forms a continuum, parameterized by menu-wise tie-breaking probabilities.²³ Turning to SVE, for each $\beta \geq 0$, the logit rule pins down a unique tie-breaking in every menu. For a.e. sufficiently large $\beta < \infty$, the corresponding SVE is unique and lies near one particular fully-mixed VE; moreover, as $\beta \uparrow \infty$ it converges to a *unique* selection from the continuum of fully-mixed VE.²⁴

Numerical experiments (see Fig. 5b) confirm that the corresponding valuation \mathbf{v}^* lies near this fully-mixed VE and that $\mathcal{V}(\beta)$ selects it as β grows. However, for large sensitivity ($\beta = 10$), this unique mixed SVE is *unstable*: in the two-dimensional relative coordinates²⁵ $(x, y) = (v_R - v_S, v_P - v_S)$, the fixed point near $(0, 0)$ repels, and nearby trajectories spiral toward a stable hexagonal *limit cycle*; see Fig. 5. Thus, in the high-sensitivity regime the CQL dynamics exhibit persistent cycling of valuations (and induced mixed strategies). The intuition for why cycling is asymptotically inevitable here appears after Theorem 4.

It is worth highlighting that none of the phenomena illustrated above could arise in a standard version of Q-Learning (or generally, TD-learning). If Alice were to perfectly discriminate alternatives, i.e. if she were equipped with the finest partition $\mathcal{S} = \mathcal{A}$, the ODE would decouple and become linear as mentioned above. Then the dynamics would converge to a

²²Recall that VE is defined via menu-wise argmax best responses. When valuations tie, VE imposes no restriction on how Alice mixes among tied classes within a menu; these menu-wise tie-breaking probabilities can be chosen to satisfy indifference even when the singleton terms z_R, z_P, z_S differ.

²³The fully-mixed VE involves five independent tie-breaking probabilities (three binary & two ternary), but only two independent indifference conditions, so the solution set generically has dimension at least three.

²⁴Under logit, (nearly) equal valuations imply (nearly) equal choice weights: if $v_R = v_P = v_S$ then $\sigma_\omega^s(\mathbf{v}) = 1/|\omega|$ for all $s \in \omega$, and if v_R, v_P, v_S are close then the induced menu-wise probabilities are close to uniform.

²⁵For the phase portrait in Fig. 5b where we plot $v_R - v_S$ against $v_P - v_S$, we exploit the translation invariance of the logit choice rule to reduce the learning dynamics to the plane. Essentially, we translate every component of the valuation vector by the negative of the valuation of scissors (v_S) at all times $t \geq 0$.

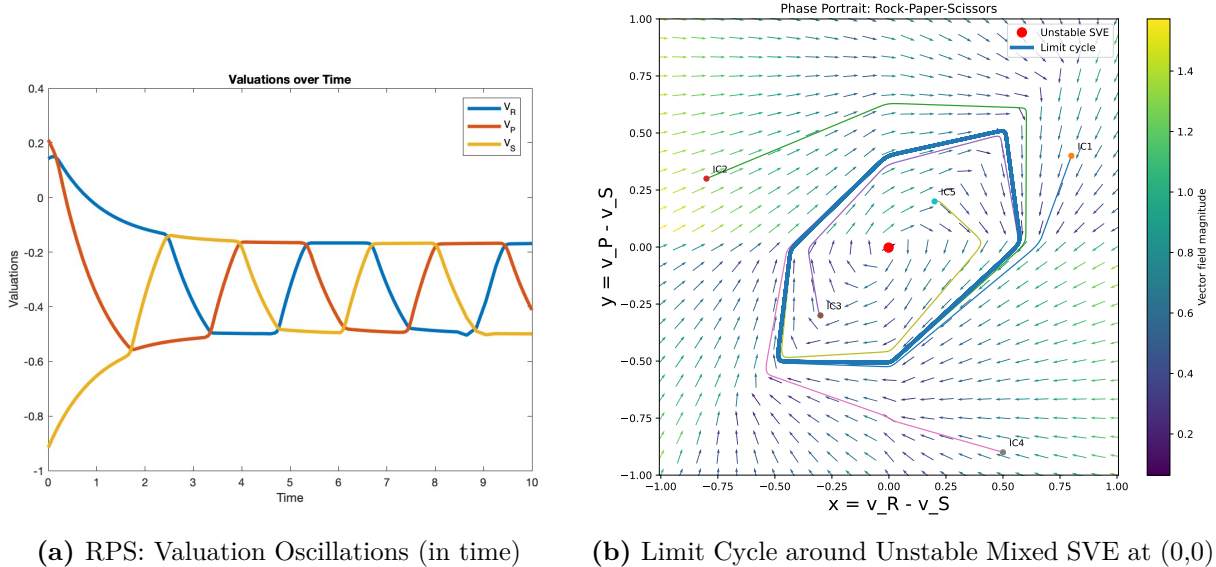


Figure 5: Phase portraits of the CQL dynamics for decision tree in Fig. 4 with $\beta = 10$

unique steady state as in [Watkins and Dayan \(1992\)](#), precluding the possibility of multiple equilibria or mixed equilibria or limit cycles arising in generic trees. Hence, it’s the coarse representation of alternatives across variable menus - rather than reinforcement learning per se - that drives the aforementioned non-trivial phenomena in the long-run.

4 Results

4.1 Smooth Valuation Equilibria

To begin, we analyze the steady states of the CQL ODE in (3) that we refer to as smooth valuation equilibria (SVE). We observe that for any $0 \leq \beta < \infty$, every SVE is fully-mixed, i.e., in any menu ω , Alice selects each available similarity class $s \in \omega$ with positive probability $\sigma_\omega^s(\mathbf{v}) > 0$. Moreover, at an SVE \mathbf{v}^* , each coordinate v_s^* is a strict convex combination of the menu-contingent class expected payoffs on menus in which s is available: $v_s^* \in \left(\min_{\omega \in \Omega_s^+} \pi_s(\omega), \max_{\omega \in \Omega_s^+} \pi_s(\omega) \right)$, where $\Omega_s^+ := \{\omega \in \text{supp}(p) : s \in \omega\}$. We assume each class s is available with positive probability in at least two distinct menus $\omega, \omega' \in \Omega_s^+$ with $\pi_s(\omega) \neq \pi_s(\omega')$. Thus, \mathbf{v}^* lies in the relative interior of the product set of intervals $K := \prod_{s \in \mathcal{S}} \left[\min_{\omega \in \Omega_s^+} \pi_s(\omega), \max_{\omega \in \Omega_s^+} \pi_s(\omega) \right]$. The following result establishes the existence of an SVE for every sensitivity $\beta \in \mathbb{R}_+$ and describes the limiting relation between SVE and the set of VE (denoted by \mathcal{VE}) as sensitivity becomes arbitrarily large.

Theorem 1. *Fix a decision tree $(\mathcal{S}, \Omega, p, \pi)$. For $\beta \geq 0$, let $\mathcal{V}(\beta) := \{\mathbf{v} \in K : \mathbf{v} = g(\mathbf{v}; \beta)\}$ denote the SVE set, and let $\mathcal{V}(\infty)$ denote the set of accumulation points of $\{\mathcal{V}(\beta)\}_{\beta \uparrow \infty}$. Then:*

- (i) For every $\beta \geq 0$, $\mathcal{V}(\beta)$ is nonempty and compact, and the correspondence $\beta \mapsto \mathcal{V}(\beta)$ is upper hemicontinuous and compact-valued on $[0, \infty)$. The SVE graph $\{(\mathbf{v}, \beta) : \mathbf{v} \in \mathcal{V}(\beta)\}$ has finitely many path-connected components. For a generic specification of parameters (p, π) and for a.e. $\beta \geq 0$, $\mathcal{V}(\beta)$ is finite and comprises an odd number of isolated regular equilibria, each lying on a locally unique real-analytic branch in β .²⁶
- (ii) Every accumulation point of SVE as $\beta \rightarrow \infty$ is a VE: $\mathcal{V}(\infty) \subseteq \mathcal{VE}$. Equivalently, for every $\varepsilon > 0$, $\exists \hat{\beta} < \infty$ s.t. $\forall \beta \geq \hat{\beta}$, every $\mathbf{v} \in \mathcal{V}(\beta)$ lies in an ε -neighborhood of some $\mathbf{v}^\infty \in \mathcal{VE}$. For generic (p, π) , $\mathcal{V}(\infty)$ is a nonempty finite subset of VE. Moreover, the path-connected component of the SVE graph containing the unique SVE at $\beta = 0$ projects onto the entire β -axis $[0, \infty)$. For generic (p, π) , this principal component is a smooth embedded one-dimensional curve whose high-sensitivity tail admits a unique limit point $\mathbf{v}^\infty \in \mathcal{V}(\infty)$ as $\beta \rightarrow \infty$, thereby selecting a unique VE.

Part (ii) of Theorem 1 implies that along any sequence $\beta_n \uparrow \infty$, every convergent sequence of SVE $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ with $\mathbf{v}_n \rightarrow \mathbf{v}^\infty$, has a limit $\mathbf{v}^\infty \in \mathcal{VE}$. Importantly, the converse need not hold: not every VE must arise as a high-sensitivity limit of an SVE branch, implying the inclusion may be strict. For instance, in generic decision trees with three or more similarity classes (as in Example 3.3), the set of fully-mixed VEs, when nonempty, forms a continuum, whereas the set of limiting SVE $\mathcal{V}(\infty)$ is finite. A second illustration appears in the [Online Appendix \(Sec. E\)](#): a continuum of strict pure VE fails to be approached by any SVE branch. Since the logit rule is uniquely pinned down by the axioms of continuity, interiority, IIA, and translation invariance in our setting, $\mathcal{V}(\infty)$ provides an axiomatic refinement of VE.

Part (ii) yields a procedure for selecting VE in a.e. decision tree analogous to the “tracing” procedure of [Harsanyi and Selten \(1988\)](#) for selecting Nash equilibria in generic normal-form games. Starting from the unique (uniform) SVE at $\beta = 0$, one can follow the globally connected principal component by continuation. In generic decision trees, this component is a smooth embedded curve whose projection covers $[0, \infty)$, though it may have finitely many fold points. Tracking its high-sensitivity tail selects a unique VE in $\mathcal{V}(\infty)$.²⁷

In the remainder of this section, we investigate the asymptotic stability²⁸ of the steady states.

²⁶Throughout, when we say that a property holds *generically* (or for a *generic decision tree*), we keep fixed the reduced tree structure, namely \mathcal{S} , Ω , and the support of p , and vary only the numerical parameters $(p(\omega), \pi_s(\omega))$ over the associated finite-dimensional parameter space. Thus, genericity means that the property holds on a comeager subset of this parameter space that also has full Lebesgue measure.

²⁷Thus, a VE can be computed by continuation: initialize at the uniform SVE at $\beta = 0$, follow the principal component, and trace its high-sensitivity tail to large β , in the spirit of tracing logit QRE to select Nash equilibria ([McKelvey and Palfrey, 1995](#)). Nonetheless, since the VE set may be infinite even for generic trees, our proof departs substantially from theirs and instead relies on new tools from o-minimal geometry.

²⁸An equilibrium of a dynamical system is *locally asymptotically stable* if, for any initial condition sufficiently

For local stability, we examine the effects of small perturbations on the long-run behavior of the CQL dynamics in neighborhoods of its steady states. In particular, we linearize the RHS of (3), $g(\mathbf{v}) - \mathbf{v}$, around the steady states and analyze the sign of the real parts of the eigenvalues of the corresponding Jacobian matrices to determine the local stability of the CQL dynamics at its steady states. For global stability results, we either construct a strict Lyapunov function that decays along all non-constant trajectories or, when applicable, leverage the convergence properties of monotone dynamical systems (Smith, 1995).

4.2 Dynamics

We present general results on the asymptotics of CQL dynamics. For convenience, we let $\text{supp}(p) = \{\omega \in \Omega : p(\omega) > 0\}$ and define the undirected co-occurrence graph $G = (\mathcal{S}, E)$ by: $\{i, j\} \in E \iff \exists \omega \in \text{supp}(p)$ with $\{i, j\} \subseteq \omega$. We begin by observing that since the softmax function is translation-invariant, it depends on valuations only through differences. Hence, the CQL drift map $g(\mathbf{v})$ is translation-invariant and only *relative* valuations matter. Without loss of generality, we fix a pivot class $q \in \mathcal{S}$ and work with the translated vector of differences $u_s := v_s - v_q$ (so $u_q = 0$). The induced dynamics on u are autonomous in $\mathbb{R}^{|\mathcal{S}|-1}$: they track how each class drifts *relative* to the pivot, with the same softmax probabilities and the same drift as in the original system. In particular, rest points of the full system are in one-to-one correspondence with rest points of the reduced system, and the Jacobian at a rest point has the same nontrivial eigenvalues as the reduced Jacobian, plus the additional eigenvalue $-(1 - \gamma) < 0$ in the common-shift direction (see Lemma B.9 in Supp. Appendix).

If Alice operates with one class, the analysis is trivial as she uniformly randomizes among all available alternatives for all valuation profiles. Considering decision problems where Alice has two similarity classes available to her, we prove the following general convergence result by exploiting the translation-invariance of $g(\mathbf{v})$ to reduce the dynamics to one dimension.

Theorem 2. *Fix a reduced decision tree \mathcal{T}'_2 with at most two similarity classes. There exists $\hat{\beta} \in [0, \infty)$ such that for all $\beta \geq \hat{\beta}$, \mathcal{T}'_2 admits an SVE that is locally asymptotically stable for the CQL dynamics. If the SVE is unique, it is globally asymptotically stable. In general, the CQL dynamics converge to an SVE for every initial condition.*

4.2.1 Strict Pure LSVE are Locally Asymptotically Stable

We now extend the analysis to decision trees with an arbitrary number of similarity classes. Recall that $\mathbf{v}^\infty \in \mathcal{V}(\infty) \subset \mathbb{R}^{\mathcal{S}}$ is a *limiting smooth valuation equilibrium* (LSVE) if there

close to the equilibrium, the solution trajectory remains close (Lyapunov stability) and asymptotically converges to the equilibrium (attractivity). *Global asymptotic stability* refers to the property of an equilibrium where all solution trajectories, regardless of initial conditions, asymptotically converge to the equilibrium.

exist $\beta_n \uparrow \infty$ and $\mathbf{v}^{(\beta_n)} \in \mathcal{V}(\beta_n)$ such that $\mathbf{v}^{(\beta_n)} \rightarrow \mathbf{v}^\infty$. We call an LSVE \mathbf{v}^∞ *strict pure* if, for every menu ω with $p(\omega) > 0$, the valuation-maximizing class $s^*(\omega) := \arg \max_{s \in \omega} v_s^\infty$ is unique and strictly dominates all other available classes, i.e., $v_{s^*(\omega)}^\infty > v_s^\infty, \forall s \in \omega \setminus \{s^*(\omega)\}$. A strict total order on \mathcal{S} implies menu-wise strictness. Conversely, for generic decision trees with a connected co-occurrence graph G , menu-wise strictness induces a strict total order on \mathcal{S} . In either case, $s^*(\omega)$ is unique for every ω , and the induced choice rule is pure and optimal, with the top-ranked available class chosen almost surely.

Theorem 3. *Consider a decision tree \mathcal{T}'_n with arbitrarily many similarity classes. Suppose there exists a strict pure LSVE \mathbf{v}^∞ that induces a strict total order on the set of classes chosen in equilibrium in at least one menu. Then there exists $\hat{\beta} < \infty$ such that, for every $\beta \geq \hat{\beta}$, the unique SVE $\mathbf{v}^{(\beta)}$ that lies in a neighborhood of \mathbf{v}^∞ is locally asymptotically stable.*

Informally, at a strict pure LSVE \mathbf{v}^∞ , every menu ω in $\text{supp}(p)$ has a unique “winner” $s^*(\omega)$ separated from all other available classes by a uniform valuation gap $m > 0$. For large β , the logit rule therefore places probability $1 - O(e^{-\beta m})$ on the winner and only exponentially small probability on every loser, so choice behavior becomes nearly insensitive to small perturbations of \mathbf{v}^* . Since the CQL Jacobian is $J(\mathbf{v}) = Dg(\mathbf{v}) - I$, it is enough to argue that $Dg(\mathbf{v}^{(\beta)})$ is negligible for large β . This follows because $g_i(\mathbf{v})$ is an average payoff for class i across the menus Ω_i in which i is available, with weights proportional to the probability that i is chosen in those menus. In general, a partial-derivative of $g(\mathbf{v})$ can be written as,

$$\frac{\partial g_i}{\partial v_j} = \beta \sum_{\omega \in \Omega_i} w_\omega^i(\mathbf{v}) \left(\mathbf{1}\{i = j\} - \sigma_\omega^j(\mathbf{v}) \right) \left(\pi_i(\omega) - g_i(\mathbf{v}) \right); \quad w_\omega^i(\mathbf{v}) := \frac{p(\omega) \sigma_\omega^i(\mathbf{v})}{\sum_{\omega' \in \Omega_i} p(\omega') \sigma_{\omega'}^i(\mathbf{v})} \in \Delta(\Omega_i).$$

When i is a winner in some menu, the denominator of the partial derivative is bounded away from zero, while the relevant logit derivatives are of the form $\beta \sigma_\omega^i (1 - \sigma_\omega^i)$ for own effects and $-\beta \sigma_\omega^i \sigma_\omega^j$ with $j \neq i$ for cross effects. Under menu-wise strictness, these products are $O(\beta e^{-\beta m})$, and hence exponentially small for large β . When i is never a winner, i is selected only through exponentially rare “trembles”, and the conditional menu distribution (given i is selected) concentrates on the closest-loss menus where i loses by the smallest valuation margin, so the leading dependence on \mathbf{v} cancels between the numerator and denominator.²⁹ In either case, for some $\epsilon > 0$, $\|Dg(\mathbf{v}^{(\beta)})\| = O(\beta e^{-\beta \epsilon}) \rightarrow 0$, hence $J(\mathbf{v}^{(\beta)})$ is a small

²⁹This “closest-loss concentration” is exactly the refinement logic behind strict pure limiting SVE. A (strict) pure VE imposes no restrictions on the valuation levels of classes that are never optimal in any menu, beyond their sub-optimality. In contrast, as $\beta \uparrow \infty$ a never-winner class i is chosen only through exponentially small logit trembles, and conditional on choosing i the menu distribution concentrates on the closest-loss menus where i loses by the smallest valuation gap. Hence v_i must be consistent with the expected payoffs of i in those closest-loss menus, which pins down the otherwise unrestricted “off-path” valuations.

perturbation of $-I$. By continuity of eigenvalues, for large $\beta < \infty$, the SVE $\mathbf{v}^{(\beta)}$ is hyperbolic with all Jacobian eigenvalues in the open left half-plane, yielding local exponential stability of the unique SVE $\mathbf{v}^{(\beta)}$ near the strict pure LSVE \mathbf{v}^∞ by the Hartman-Grobman theorem.

Therefore, if a decision tree admits a strict pure LSVE, the set of asymptotically stable SVE for the CQL dynamics is nonempty. [Benaim \(1999\)](#) shows that every locally asymptotically stable steady state of the continuous-time ODE (3) has strictly positive probability of being the long-run outcome of the discrete-time stochastic CQL process (2), given non-degenerate noise in the system (true for any $\beta < \infty$ in our setting). Thus, asymptotic stability may serve as a dynamic equilibrium selection criterion.

4.2.2 Non-existence of Asymptotically Stable SVE

In the absence of a strict LSVE, asymptotic stability need not obtain. The theorem below shows that, for an open set of decision trees with three similarity classes and sufficiently large payoff sensitivity $\beta < \infty$, the CQL mean-field dynamics admit no asymptotically stable SVE. Since stochastic approximation processes such as the discrete-time CQL process in (2) converge with probability zero to linearly unstable steady states of the mean-field ODE under non-degenerate noise ([Pemantle, 1990](#); [Murooka and Yamamoto, 2025](#)), this result implies that convergence to equilibrium is not guaranteed in general for the CQL dynamics.

Theorem 4. *There exists an open set \mathcal{U} of decision trees with $n \geq 3$ similarity classes such that, for each $\mathcal{T}'_n \in \mathcal{U}$, there exists $\hat{\beta} < \infty$ for which, for all $\beta \geq \hat{\beta}$, the CQL dynamics admit no asymptotically stable SVE.*

To prove the theorem, we consider the RPS family of decision trees with three similarity classes (see Fig. 4 in Example 3.3). We simplify the algebra by choosing symmetric $z_R = z_P = z_S = z$ with $z \in (-1, 0)$ and uniform $p(\omega) = 1/6$, $\forall \omega \in \Omega \setminus \{\omega_7\}$ with $p(\omega_7) = 0$.³⁰ We show that such a tree admits a unique fully-mixed SVE (with index +1) that is linearly unstable for large β and the CQL trajectories spiral into a stable limit cycle in the long-run. Informally, to see why the cycling occurs, it is sufficient to examine optimal choice behavior in the high-sensitivity limit. Consider a generic (strict) ranking of the three valuations. In the essentially noise-free limit ($\beta \rightarrow \infty$), Alice a.s. selects the top-ranked class whenever it's available, whereas the bottom-ranked one is chosen a.s. only in its singleton menu. Hence,

³⁰In the proof, we show that the result is robust to small perturbations of both the mean payoff parameters and the menu probabilities, provided the support of p is held fixed. In the [Online Appendix \(Sec. C\)](#), we show that the result continues to hold when the ternary menu ω_7 is added to $\text{supp}(p)$ as in Fig. 4, although the algebra becomes more cumbersome. The [Online Appendix \(Sec. E\)](#) contains additional three-class examples, including an RPS decision tree with (triangular) cycling obtained by keeping the binary-menu payoffs unchanged, removing all singleton menus from $\text{supp}(p)$ and replacing them with the ternary menu.

the top class accrues a strictly larger expected payoff ($z/3$) than the bottom class does (z) ensuring the top class strictly outperforms the bottom one. The behavior of the middle class is pivotal and depends solely on its binary interaction with the lowest class.

Two mutually exclusive cases obtain: (i) *negative interaction* - if, in their shared binary menu, the expected payoff of the middle class is -1 , it becomes the worst performer overall (since -1 is the minimal feasible expected payoff). It then drops below the former bottom class, swapping their positions in the ranking. (ii) *positive interaction* - if, in that binary menu, the middle class' expected payoff is $+1$, it becomes the best performer overall (since $+1$ is the maximal feasible expected payoff). It then overtakes the former top class, again swapping positions. Each such swap changes which class is “middle”, and the same logic re-applies. Consequently, this decision tree admits no strict pure VE - the “middle” class is always either the best performer or the worst performer in expected payoffs. Therefore, the strict order of valuations rotates cyclically rather than settling, and no fixed point can be stable. The resulting perpetual rotation of rankings generates a hexagonal limit cycle in the valuations (and induced mixed strategies), as illustrated in Fig. 5 for Example 3.3.

4.2.3 Global Asymptotic Stability of a Unique Mixed SVE

Theorem 4 shows that, for large sensitivity, a unique mixed SVE need not be asymptotically stable. This does not exclude the possibility that, under additional structure, a mixed SVE is the global attractor. In this subsection, we identify such a structure. Specifically, we consider variations that uniformly shift the expected payoffs in singleton menus where all available alternatives belong to a single class, while keeping all other menu-contingent class-level expected payoffs unchanged. For each class i , let $\pi_i(\{i\}) \equiv \mathbb{E}[r \mid \omega = \{i\}]$ denote the expected payoff in its singleton menu $\{i\}$. We translate all singleton-menu payoffs by a common scalar $z \in \mathbb{R}$, i.e. $\pi_i(\{i\}) \mapsto \pi_i(\{i\}) + z$, for all $i \in \mathcal{S}$, while leaving $\{\pi_i(\omega) : \omega \in \Omega, |\omega| > 1\}$ unchanged. We focus on singleton menus because Alice's class choice is degenerate there: when $\omega = \{i\}$, she selects class i with probability 1 regardless of her valuations. Singleton shifts therefore perturb class-level feedback without directly changing the logit odds in competitive menus, providing a clean comparative static for stability analysis.³¹

Assumption 4.1. G is connected, and for all $s \in \mathcal{S}$, the singleton menu $\{s\} \in \text{supp}(p)$.

Asm. 4.1 implies that each $s \in \mathcal{S}$ has degree at least one in G . Thus, every class appears in some non-singleton menu in $\text{supp}(p)$ along with its singleton menu. While we allow $p(\{s\})$ to be arbitrarily small, we assume it is strictly positive for each class s . Under Asm. 4.1, we show

³¹A high (low) uniform singleton payoff shift z can be interpreted as parameterizing decision trees where each similarity class's expected payoff maximizers (minimizers) occur only in singleton menus.

that there exists a threshold $\hat{z} \in \mathbb{R}$ (depending on $(\mathcal{S}, \Omega, p, \pi)$) such that whenever $z \geq \hat{z}$, the CQL dynamics admit a unique mixed SVE, and this SVE is globally asymptotically stable. [Benaïm \(1999\)](#) shows that if the continuous-time ODE (3) has a unique steady state that is a global attractor, then it is the unique element of the internally chain-transitive set and the discrete-time CQL stochastic process (2) converges to it almost surely.

Theorem 5. *Fix a reduced decision tree $\mathcal{T}'_n = (\mathcal{S}, \Omega, p, \pi)$ such that [Asm. 4.1](#) holds. There exists a threshold $\hat{z} < \infty$ such that for every $z \geq \hat{z}$ the following hold:*

- (i) *Uniqueness: For every $\beta \in \mathbb{R}_+$, the CQL dynamics admit a unique steady state (SVE).*
- (ii) *High-sensitivity indifference: There exists $\hat{\beta} < \infty$ such that for every $\beta \geq \hat{\beta}$, the unique SVE lies in a neighborhood of a unique mixed VE at which the agent is indifferent between at least two similarity classes.*
- (iii) *Global asymptotic stability: For every $\beta \in \mathbb{R}_+$, the unique SVE is globally asymptotically stable for the CQL dynamics $\dot{\mathbf{v}} = g(\mathbf{v}; \beta) - \mathbf{v}$. In particular, for every initial condition $\mathbf{v}(0)$, the solution $\mathbf{v}(t)$ converges to this SVE as $t \rightarrow \infty$.*

When z is high, singleton menus create a common payoff baseline that makes strict valuation orders self-defeating. To see this, suppose for contradiction that a strict VE exists with lowest-ranked class i . Class i is chosen only at its singleton menu and is assessed entirely by that outcome, so $g_i(\mathbf{v})$ places full weight on the singleton shift z . By contrast, the highest-ranked class j is chosen in at least one non-singleton menu, so its singleton payoff is diluted by its relatively weaker non-singleton payoff(s), implying a smaller weight on z in $g_j(\mathbf{v})$. For high enough z , this difference in exposure to the singleton shift reverses the putative ranking between i and j , contradicting strictness and implying any VE must be mixed.

Local stability stems from a strong negative self-correction induced by singleton payoff dominance. For high z , each class s has a high singleton benchmark relative to any non-singleton-menu payoff. If v_s increases, s is selected more often in non-singleton menus where its realized payoff is below this benchmark, which pulls its conditional assessment $g_s(\mathbf{v})$ downward. This generates a stabilizing self-effect. If another class k becomes more attractive, i.e. v_k increases, it absorbs probability mass in non-singleton menus where it co-occurs with s . This crowds s out of precisely those menus in which its payoff is relatively low, thereby increasing s 's conditional assessment $g_s(\mathbf{v})$. Hence cross-effects are non-negative (cooperativity), and the Jacobian of $F(\mathbf{v}) = g(\mathbf{v}) - \mathbf{v}$ is Metzler on the compact, convex set K . Moreover, the overall sensitivity of g_s to changes in valuations of co-occurring classes is exactly tied to the self-effect by the translation invariance of the drift ($Dg(\mathbf{v}) \mathbf{1} = \mathbf{0}$ for any \mathbf{v}). Thus, the non-

negative cross-effects sum to the magnitude of the negative diagonal term up to the constant -1 implying that each row of the Jacobian is uniformly strictly diagonally dominant with rightmost Gershgorin bound at -1 . Hence, any SVE is locally exponentially stable.

Second, this local stability ensures that any equilibrium is an isolated non-degenerate zero of $-F(\mathbf{v})$ with index $+1$. Since K is compact and convex and $-F(\mathbf{v})$ points strictly outward on its boundary ∂K , the Poincaré-Hopf index theorem yields exactly one equilibrium in $\text{int } K$, that is, a unique SVE for each β . Finally, the Metzler property together with irreducibility of the Jacobian implies that the mean-field dynamics are cooperative and consequently strongly monotone on the compact, positively invariant set K . A strongly monotone semi-flow on such a set with a unique interior equilibrium must converge globally to that equilibrium.

Moreover, since for each $\beta \in \mathbb{R}_+$ the SVE is unique, the equilibrium correspondence is a single-valued map and its graph in (\mathbf{v}, β) -space is a unique globally connected component (the principal branch); by Theorem 1, any sequence $\beta_n \uparrow \infty$ has $\mathbf{v}(\beta_n)$ converging (along a subsequence) to a valuation equilibrium in $\mathcal{V}(\infty)$, so tracing this principal branch to the high-sensitivity limit selects a unique mixed VE even when the VE set forms a continuum.

More generally, beyond the sufficient condition provided above, the same argument yields existence, uniqueness, and global asymptotic stability of a mixed SVE throughout a *cooperative* regime, namely whenever classes behave as complements in the sense that raising the valuation of any class weakly increases the conditional expected payoff of every class with which it co-occurs, uniformly over $\mathbf{v} \in K$. Formally, this means that the Jacobian of the CQL drift $F(\mathbf{v}) = g(\mathbf{v}) - \mathbf{v}$ is Metzler on K , i.e. all off-diagonal entries are nonnegative.

Assumption 4.2 (Monotonicity). *For any two menus $\omega, \omega' \in \Omega$, $\omega \subseteq \omega' \implies p(\omega) \leq p(\omega')$. Additionally, for any two singleton menus $\omega, \omega' \in \{\omega \in \Omega : |\omega| = 1\}$, $p(\omega) = p(\omega')$.*

Theorem 5 guarantees high-sensitivity indifference between at least two classes. With an additional restriction on the menu distribution, this indifference can be strengthened to total indifference across all classes. Given Asm. 4.1, Asm. 4.2 imposes full support over $\Omega = \mathcal{P}(\mathcal{S}) \setminus \{\emptyset\}$, equal likelihood of singleton menus, and monotonicity under set inclusion (richer menus are weakly more likely). For e.g., it holds for the uniform distribution.

Proposition 1. *Let assumptions 4.1-4.2 hold. There exists a $\hat{z}_1 < \infty$ such that for all $z \geq \hat{z}_1$ and $\beta \geq 0$, every reduced decision tree \mathcal{T}'_n admits a unique SVE that is globally asymptotically stable for the CQL dynamics. Moreover, as $\beta \rightarrow \infty$, this unique SVE corresponds to a unique fully-mixed VE where the agent is indifferent among all similarity classes.*

To see the intuition behind this result, recall that the consistency map $g_s(\mathbf{v})$ averages ex-

pected payoffs across menus conditional on s being selected. With equal singleton probabilities, the singleton bonus enters $g_s(\mathbf{v})$ with the same unconditional weight for every class, but with a class-dependent conditional weight that is inversely proportional to the total probability mass of menus in which s is selected. Under the argmax rule, any strict valuation gap $v_i < v_j$ implies that class j is selected in additional menus where i is not, so the selection mass for j is strictly larger. Thus, the singleton component is more diluted for j than for i , and class i places a strictly larger effective weight on the singleton bonus z in its consistency calculation. For sufficiently large z , this amplification overturns any putative strict ranking, so no strict valuation gaps can persist, implying any VE must be fully-mixed.

4.2.4 Multiplicity of SVE with at least one asymptotically stable SVE

We now consider the complementary regime in which singleton menus are uniformly *disadvantaged* relative to non-singleton menus. Concretely, we perturb payoffs by lowering the expected payoff in each singleton menu while leaving all other menu-contingent expected payoffs unchanged. For sufficiently low z shifts, the CQL dynamics exhibit multiple equilibria, and crucially at least one SVE lies in the neighborhood of a strict pure LSVE and is therefore locally asymptotically stable for large sensitivity by Theorem 3.

Theorem 6. *Fix a generic reduced decision tree \mathcal{T}'_n such that Asm. 4.1 is satisfied. Then there exists a threshold $\tilde{z} > -\infty$ such that for every $z < \tilde{z}$ the following statements hold:*

- (i) *There exists at least one strict pure VE. Consequently, there exists $\hat{\beta} < \infty$ such that for all $\beta \geq \hat{\beta}$, the unique SVE that arises in a neighborhood of that strict pure VE is locally asymptotically stable for the CQL dynamics.*
- (ii) *Additionally, if all binary menus are in $\text{supp}(p)$, there exist multiple VE such that for each $s \in \mathcal{S}$, there exists a distinct VE at which s is the unique worst-ranked class.*

The intuition for this theorem is the mirror image of Theorem 5. When a class is ranked low, it is chosen relatively more often in its singleton menu than in non-singleton menus. If the singleton shift z is sufficiently negative, this selection pattern depresses the class's assessed payoff even further, reinforcing its low rank. Hence, for z small enough, strict valuation hierarchies can be self-confirming, yielding at least one strict pure VE and, by the high-sensitivity correspondence, a locally stable nearby SVE. When all binary menus are in support, this reinforcement argument can be run with any designated class as the unique worst class, generating multiple distinct valuation equilibria. The multiplicity identified in Theorem 6 can be further strengthened under stronger assumptions on the menu distribution. Notably, if all menus are equally likely, any strict ordering of the valuations can arise in a valuation equilibrium. This insight emerges as a corollary of the following broader result

that relies solely on the weaker Assumption 4.2.

Proposition 2. *Let Asms. 4.1, 4.2 hold. There exists a threshold $\tilde{z} \in \mathbb{R}$ such that for all $z \leq \tilde{z}$, in a finite decision tree \mathcal{T}'_n , the following holds. Every strict total order on the valuations of the n similarity classes is admissible in some VE, i.e., there exist $n!$ strict pure VE. Correspondingly, there exists a $\hat{\beta} < \infty$, such that for all $\beta \geq \hat{\beta}$, the (unique) SVE that lies in the neighborhood of each strict pure VE is locally asymptotically stable.*

To highlight Theorems 5 & 6 by varying singleton payoffs, it's instructive to revisit the RPS family of decision-trees (Fig. 4) adopting the specification used to illustrate the possibility of non-convergence in Theorem 4: $z_R = z_P = z_S = z$, with $z \in \mathbb{R}$, and $p(\omega) = 1/6$ for all $\omega \in \Omega \setminus \{\omega_7\}$ with $p(\omega_7) = 0$. Recall that, for $z \in (-1, 0)$, there is a unique *unstable* fully-mixed SVE at the origin (indifference across all three classes) for large sensitivity β ; with valuations perpetually oscillating. As z increases, the mixed equilibrium gains stability. In particular, for $z \geq 0$, the fully-mixed SVE is *unique* and *globally stable* for all $\beta \geq 0$ in accordance with Theorem 5. Conversely, as z decreases, multiplicity arises: for $z \leq -1$, the fully-mixed SVE persists but is *unstable*. In addition, multiple distinct SVE emerge near the corresponding strict pure VE, each of which is *locally stable* in line with Theorem 6, while further SVE appear near partially-mixed VE that act as saddle points for large β . Together, these cases display the bifurcation picture in the RPS tree: the unique, globally stable mixed equilibrium for $z \geq 0$ loses stability as z crosses into $(-1, 0)$, and for $z \leq -1$ the system exhibits coexistence of multiple locally stable SVE near the corresponding strict VE.

5 Discussion

5.1 Limited Salience and Induced Misspecification

The exogenous similarity partition in our model can be interpreted through the lens of limited salience. Consider alternatives that are characterized by several payoff-relevant attributes (price, size, etc.), but suppose Alice attends only to a fixed proper subset of them, the salient attributes. This induces a projection from the full attribute vector of an alternative to its salient coordinates. Alice then perceives two alternatives as similar whenever their salient projections coincide, even if they differ along non-salient payoff-relevant dimensions. The resulting similarity classes need not reflect objective payoff homogeneity, because the salient attributes need not coincide with the attributes that are most predictive of payoffs. Rather, they reflect the granularity at which Alice perceives her environment. This interpretation gives a natural foundation for coarsening. When only a subset of attributes are salient, it is reasonable for Alice to organize her options according to those attributes alone and to treat residual within-class heterogeneity as i.i.d. noise. We assume the partition is induced by a

fixed attention mechanism: an attribute is either salient in a given context or it is not.

Limited salience can be viewed as inducing a simple form of misspecification. Since Alice reasons using similarity classes only, she behaves as if all alternatives within a class were governed by the same stationary payoff law. A realized payoff from one alternative is then interpreted as feedback about its class as a whole, and the corresponding class valuation is updated accordingly. In reality, however, alternatives that look identical on salient dimensions may yet differ systematically in expected payoffs because of latent attributes. Thus, the DM unwittingly pools information across possibly heterogeneous options.

This perspective is particularly natural in online retail environments. A consumer shopping from home often sees only a coarse digital representation of a product and cannot inspect dimensions that would be readily observable in person, such as expiry dates, ripeness, or packaging condition. Two products may therefore appear identical on the online platform even though they differ in ways that matter for the consumer’s preferences.³²

Application: Assortment Design on an Online Retail Platform

A potential application of this perspective is assortment design on a monopolistic online platform. Consumers shopping on such a platform typically observe only a limited set of salient product attributes, such as posted price, color, size, or brand label, while other payoff-relevant dimensions remain difficult to inspect at the moment of choice, including sourcing, durability, compliance, or seller reliability. The platform interface thus induces a partition of the product space into similarity classes, and products that appear similar along these salient dimensions are treated by consumers as close substitutes even when they differ systematically in latent quality. Learning at the level of classes rather than individual products is therefore a natural response to the informational constraints of digital shopping.

This structure gives the platform considerable power over how consumers learn. The platform does not merely choose which categories appear together; it can also choose which particular products are used to represent each category in a given menu. In particular, it can expose the consumer to especially attractive products in some menus and more ordinary products in others, while preserving the same visible category labels and prices throughout. The consumer may then attribute the quality of those experiences to the category itself rather than to the hidden composition of products shown within it.

³²This perspective is consistent with evidence from online grocery retail. [Sharib et al. \(2024\)](#) document that mandatory nutrition information is often missing, inaccessible, or not legible on major U.S. online food retail platforms, despite the underlying technical capacity to display it. Consumer evidence from online grocery orders also points to short-dated and even past-date deliveries, together with limited consumer control over exact use-by dates in online fulfillment; see the following [report](#) from Australia.

This creates a natural strategic motive when the platform earns different margins across categories. For e.g., a platform may earn much higher profits on its own in-house products than on third-party products sold through the marketplace yielding a small commission. Assume further that, in ordinary comparisons (non-singleton menus), the in-house products are on average somewhat inferior, and that the platform’s profit margins are similar across its category of in-house products. A correctly-specified consumer would then tend to favor the third-party category. However, the platform can manipulate coarse learning by disproportionately featuring unusually attractive in-house products in singleton menus while showing more ordinary products in menus where categories compete side by side. This corresponds to a configuration in which $\pi_s(\{s\}) > \pi_s(\omega)$ for non-singleton menus $\omega \ni s$, as singleton menus contain especially favorable representatives of the in-house class s . In line with Theorem 5’s indifference insight, we expect such manipulations to induce more sales of the in-house products even when there are products from other brands available, thereby increasing the platform’s profits. Ongoing companion work formalizes this mechanism and develops the broader strategic-platform implications of coarse consumer learning.

5.2 Cognitive Complexity & Choice Overload

A complementary interpretation of CQL is as a deliberate form of stateless Q-learning adopted for cognitive and statistical tractability rather than as a consequence of limited salience. Suppose Alice is correctly specified: the alternative set \mathcal{A} is finite, all payoff-relevant attributes are salient, payoffs may depend on both the chosen alternative and the realized menu, and Ψ consists of all nonempty menus. A canonical correctly specified learner would then maintain a separate valuation $v(a, \psi)$ for each feasible alternative-menu pair (a, ψ) with $a \in \psi$. While this tabular representation accords with the standard logic of Q-learning, it has exponential worst-case dimension: if all nonempty menus are feasible and $n := |\mathcal{A}|$, then it requires $\sum_{\psi \in \Psi} |\psi| = n \cdot 2^{n-1}$ valuation coordinates. This is the familiar curse of dimensionality in dynamic programming (Bellman, 1957). By contrast, CQL compresses this representation to one valuation per similarity class, so the dimension of the valuation state grows only linearly in the number of classes, independently of the number of menus.

The burden is not only mnemonic but also statistical. At such a fine level of description, a given coordinate $v(a, \psi)$ is updated only when the precise pair (a, ψ) is encountered and the alternative a is chosen from menu ψ . When the menu space is large, many such coordinates are visited only rarely, so learning at the fully disaggregated level is based on extremely thin data. Even if exact asymptotic identification remains possible under strong exploration assumptions, estimation at that granularity is slow, noisy, and sample-inefficient. By contrast, a cognitively frugal learner who abstracts from menu dependence and maintains a coarser set

of valuations pools feedback across many observations, thereby alleviating the sparse-data problem that arises under the fully specified representation. This pooling, however, is statistically non-innocuous. By updating one valuation per category, Alice treats realized payoffs as if they were representative draws from a menu-invariant class-level distribution. Yet when alternatives within a category are heterogeneous and the composition of the category varies across menus, selecting that category in different menus amounts to sampling from different conditional distributions. Since the sampling intensity of a class across menus is itself endogenous to current valuations, the pooled data are systematically selected in a way that Alice fails to recognize - leading to the endogenous selection bias in our model.

It's worth noting that Coarse Q-learning extends immediately to environments with infinitely many alternatives, provided the number of similarity classes remains finite. This is useful because classical alternative-level Q-learning is ill-defined on a continuum of alternatives: under any absolutely continuous within-menu choice rule, a given alternative is selected with probability zero at any date and hence is typically never sampled repeatedly, so the notion of learning an alternative's value is not meaningful. By contrast, CQL remains well-defined because learning takes place at the level of classes rather than individual alternatives. Indeed, all of our analysis continues to apply so long as the reduced decision tree $\mathcal{T}'(\mathcal{S}, \Omega, p, \pi)$ is finite, with the only modification being that within-class randomization is defined relative to a normalized reference measure on the feasible part of the chosen class.

We also note that the singleton-payoff shifts above can be interpreted as reduced-form *choice aversion/affinity* (Iyengar and Lepper, 2000; Fudenberg and Strzalecki, 2015). If expected payoffs take the form $u_a(\psi) = \mu_a(\psi) + \xi(\psi)$, where $\xi(\psi)$ is common to all $a \in \psi$, then this menu-level component is irrelevant for within-menu choice under correct specification but affects pooled feedback under coarse learning. Theorems 5 & 6 study the case where $\xi(\psi) = z \cdot \mathbf{1}\{|\psi| = 1\}$ depends only on whether ψ is a singleton: strong choice aversion ($z > 0$) leads to robust indifference, while strong choice affinity ($z < 0$) leads to multiplicity.

5.3 Bayesian learning under conjugate priors

While model-free reinforcement learning is fundamentally non-parametric and does not require the full specification of a likelihood for rewards and a prior, it can nevertheless coincide with Bayesian learning for the *posterior mean* under standard conjugate models. Fix a class s and suppose that, conditional on choosing s , Alice adopts a (misspecified) subjective model in which realized rewards are i.i.d. from a (regular) one-parameter natural exponential family with scalar sufficient statistic $T(\cdot)$ and (unknown) mean parameter of interest $\theta_s := \mathbb{E}[T(r) \mid s \text{ chosen}]$. Assume she uses the corresponding conjugate prior, parameterized

by prior mean $m_{0,s}$ and pseudo-count $n_{0,s} > 0$.³³ After n_s observations r_0, \dots, r_{n_s-1} from class s , the Diaconis and Ylvisaker (1979) conjugate prior yields a posterior expectation of the parameter θ_s that is affine in the sufficient statistic $T(r)$, and can be written as:

$$\mathbb{E}[\theta_s \mid r_0, \dots, r_{n_s-1}] = \frac{n_{0,s}m_{0,s} + \sum_{\ell=0}^{n_s-1} T(r_\ell)}{n_{0,s} + n_s}.$$

Let $N_k(s) := \sum_{t=0}^{k-1} \mathbf{1}\{s_t = s\}$ be the number of times class s has been chosen strictly before date k , and set $v_0(s) = m_{0,s}$. Consider the per-visit CQL update in (2) with $\gamma = 0$,

$$v_{k+1}(s) = v_k(s) + \mathbf{1}\{s_k = s\} \alpha_k(s) (T(r_k) - v_k(s)), \quad \alpha_k(s) := \frac{1}{1 + n_{0,s} + N_k(s)}.$$

A simple induction on the visit count $N_k(s)$ shows that for every date $k \geq 0$,

$$v_k(s) = \frac{n_{0,s}m_{0,s} + \sum_{t < k: s_t = s} T(r_t)}{n_{0,s} + N_k(s)} = \mathbb{E}[\theta_s \mid \text{data from } s \text{ up to (but excluding) date } k].$$

In common specifications (e.g. Bernoulli, Poisson, Normal), one may take $T(r) = r$, so $\theta_s = \mathbb{E}[r \mid s \text{ chosen}]$ and the recursion tracks the posterior mean of the expected reward. Hence, assuming rewards are menu-invariant conditional on s , this myopic Q-learning valuation recursion reproduces the Bayesian posterior mean exactly along the realized path.³⁴

In addition to the Bayesian foundation for our *valuation update rule*, a natural foundation for the *logit choice rule* is entropy-regularized choice. Conditional on current valuations \mathbf{v} and menu ω , suppose Alice chooses a probability vector $\sigma_\omega \in \Delta(\omega)$ to maximize $\sum_{s \in \omega} \sigma_\omega^s v_s - \frac{1}{\beta} \sum_{s \in \omega} \sigma_\omega^s \log \left(\frac{\sigma_\omega^s}{1/|\omega^s|} \right)$, where the second term is the Shannon entropy cost relative to the uniform distribution on ω . The parameter β measures choice precision: larger β places more weight on payoff considerations relative to the cognitive cost of concentrating too sharply on a particular class. The optimal class choice probabilities then are exactly the multinomial logit probabilities (Matějka and McKay, 2015).³⁵

³³For a likelihood function in the natural exponential family, a conjugate prior always exists (Diaconis and Ylvisaker, 1979). This covers, inter alia, Bernoulli–Beta, Binomial–Beta, Multinomial–Dirichlet, Poisson–Gamma, and Normal–Normal (with known variance) when the object of interest is $\mathbb{E}[\mu_s \mid \text{data}]$. See https://en.wikipedia.org/wiki/Conjugate_prior and [lecture notes](#) by Michael I. Jordan.

³⁴For $\gamma \in (0, 1)$, the relative valuations $v_k(s) - v_k(s')$ for a fixed pivot class s' coincide asymptotically with differences in Bayesian posterior means of the corresponding class-level expected rewards along paths where both classes are visited infinitely often. For classes visited infinitely often, posterior uncertainty typically vanishes (or concentrates around the pseudo-true value under misspecification), so the distinction between full posterior tracking and mean tracking is only relevant for finite-sample uncertainty quantification.

³⁵Under this interpretation, the CQL valuation recursion may be viewed as Bayesian learning for the posterior mean under a coarse subjective model, while the logit rule reflects a trade-off between exploiting currently

This Bayesian recursion corresponds to asynchronous CQL with step-sizes indexed by class visit counts: since a class is updated only when selected, classes with lower choice propensities receive fewer updates and therefore evolve more slowly in calendar time. However, the ODE (3) is the synchronous mean-field limit associated with inverse-propensity-weighted step-sizes in discrete-time CQL, under which effective update rates are asymptotically equalized across classes in expectation. While we prefer the synchronous ODE (3) for its analytical tractability, the [Online Appendix \(Sec. B\)](#) derives the corresponding asynchronous mean-field ODE for deterministic calendar-time gains and shows that all our results continue to hold under this specification, except for the *global* convergence with two classes in [Theorem 2](#).

6 Related Literature

In addition to the literature on reinforcement learning (RL) in computer science that we’ve already discussed, our paper also relates to a closely connected literature in economics. Relative to classical dynamic programming ([Bellman, 1957](#)), model-free RL methods such as Q-learning estimate value functions directly from sampled experience rather than by fitting a fully-specified parametric model of the data-generating process. The myopic variant of Q-learning has been studied independently in economics under the label *payoff-assessment learning* in decision-making ([Sarin and Vahid, 1999](#)), and it has also been used in the analysis of routing games using stochastic approximation ([Cominetti et al., 2010](#)) and extensive-form games ([Jehiel and Samet, 2005](#)). Recently, [Moll \(2025\)](#) has advocated for RL as a computational tool in heterogeneous-agent macroeconomics. However, none of these papers have considered the effect of coarse perception on Q-learning, which is our main contribution.

Our analysis also connects to the growing economic literature on Bayesian learning under model misspecification (see [Esponda and Pouzo \(2026\)](#) for a recent survey). A foundational insight, tracing back to [Berk \(1966\)](#)’s consistency results, is that under misspecification Bayesian posteriors concentrate on parameter values (or models) that best approximate the true data-generating process in Kullback–Leibler divergence. Building on this, a growing literature studies *active learning* environments in which the data an agent observes is itself endogenous to her actions, and characterizes long-run behavior via equilibrium notions that impose “best-fit” beliefs on the realized path of play, such as Berk–Nash equilibrium ([Esponda and Pouzo, 2016, 2021](#)) and its refinements ([Fudenberg et al., 2021](#)). Convergence results have been obtained in limited settings ([Heidhues et al., 2018, 2021](#); [Esponda et al., 2021](#)). Our model complements this literature by studying a distinct source of misspecification that arises from limited salience - coarse inference - in a non-parametric learning model.

more attractive classes and avoiding the cognitive cost of excessively sharp discrimination across them.

Moreover, we show in Sec. 5.3, that if the agent’s misspecified subjective class-level payoff model belongs to a natural exponential family with conjugate prior, our valuation recursion with step-sizes indexed by inverse visit counts coincides exactly with Bayesian updating of the posterior mean of the (misspecified) class-mean parameter along the realized path. Under this interpretation, a VE can be viewed as a Berk-Nash equilibrium induced by the agent’s coarse subjective model and the corresponding endogenous data-generating process.³⁶

Regarding coarse categorization and its equilibrium consequences, our closest point of contact is Jehiel and Samet (2007), who introduced valuation equilibrium (VE) for multi-agent extensive-form games. As already noted, the steady states of our learning dynamics coincide with the smooth valuation equilibria (SVE) induced by a logit perturbation of VE. Relative to Jehiel and Samet (2007), we contribute along three dimensions to equilibrium analysis. First, we analyze the learning dynamics that generate SVE as fixed points, deriving stability and selection results. In particular, as sensitivity grows without bound, the finite accumulation points of the SVE correspondence $LSVE$ refine VE by selecting robust limits, even when the set of (mixed) VE is non-isolated. Second, we show that indifference may be unavoidable in equilibrium: even in finite generic decision trees, uncountably many mixed VE may arise, a possibility not highlighted in Jehiel and Samet (2007). To the best of our knowledge, our result on the global stability of a unique mixed equilibrium in generic decision trees with high singleton payoffs, has no parallel in the literature.³⁷ Third, we identify an open set of decision trees for which no SVE is asymptotically stable for sufficiently large sensitivity, so that the learning dynamics admit no equilibrium point that is a long-run attractor.

Beyond VE, also worth mentioning is the game-theoretic concept of Analogy-based Expectation Equilibrium (Jehiel, 2005) where players form coarse expectations about opponents’ behavior.³⁸ Finally, our model is related to case-based decision theory (CBDT) of Gilboa and Schmeidler (1995) where they consider a DM who evaluates actions by aggregating payoffs across similar past cases and they axiomatize a decision rule that chooses a “best” act based on its past performance in similar cases. In our setting, similarity classes induced by salience play the role of “cases” and class-level valuations summarize experience (observed payoffs) pooled across alternatives within each class.³⁹

³⁶VE is conceptually related to Personal Equilibrium (Spiegler, 2016): the DM’s choices generate the data used to fit a misspecified subjective model, and the fitted beliefs (valuations) in turn justify those choices.

³⁷For context, global convergence results are known only for a limited class of games. Hofbauer and Sandholm (2002) show global convergence of stochastic fictitious play in zero-sum, potential, and supermodular games.

³⁸Viewing Nature as a passive player who chooses Alice’s payoff after each chosen alternative, one can view a VE as an ABEE (thus, a Berk-Nash equilibrium) in which Alice bundles all alternatives in a similarity class into an analogy class to form her expectation about the payoffs in each menu (Jehiel, 2026).

³⁹Other articles that use the concept of similarity for individual or evolutionary learning across games and decisions include LiCalzi (1995); Gilboa and Schmeidler (1995); Samuelson (2001); Steiner and Stewart

Our instability and cycling result also relates to a broader set of non-equilibrium learning phenomena. In particular, persistent oscillations have been documented in Bayesian learning under misspecification, where actions affect the information stream: Nyarko (1991) provides an early example in which beliefs and actions in an MDP can cycle on every sample path. Fudenberg et al. (2017) show using the same example that the interaction between misspecification and incentives to experiment can lead to non-convergence for sufficiently patient agents (even when more myopic behavior converges). Cycling is likewise familiar in learning in games, beginning with Shapley’s classic non-convergence example for fictitious play and related constructions such as the Shapley polygon limit sets (Shapley, 1964; Jordan, 1993; Gaunersdorfer and Hofbauer, 1995). Unlike these examples, our cycling arises in a minimal single-agent bandit problem with *exogenous* menu transitions and payoffs that are i.i.d. conditional on the chosen alternative, so the non-convergence is driven purely by categorization and the feedback between valuation-based sampling and misspecified coarse inference.

7 Conclusion

The central message of this paper is that Coarse Q-learning in the long-run generates novel qualitative phenomena in decision problems with stochastic menus and coarse representation - phenomena that are otherwise impossible to obtain in a standard Q-learning setup. When expected payoffs in singleton menus are sufficiently large, the learning dynamics push valuations toward *indifference* across multiple similarity classes. When singleton expected payoffs are sufficiently low, the same feedback logic supports *multiple* self-confirming rankings of similarity classes in the long-run. Outside these polar regimes, the interaction between coarse perception and menu-dependent sampling at the class level can produce endogenous *instability*, including persistent *cycling* in relative valuations and choice frequencies.⁴⁰

More broadly, our results offer a new lens on indifference, heterogeneity, and instability in preferences defined over categories, phenomena that have been documented in field and lab experiments in behavioral economics (DellaVigna, 2009; Lichtenstein and Slovic, 2006). In our framework, preferences over categories are not primitives but outcomes of an active learning process shaped by coarse perception and stochastic availability. This provides a mechanism by which different individuals may converge to different stable valuation orderings, or a single individual may exhibit indifference or preference reversals over time. We leave the endogenous formation and revision of similarity classes for future research, especially in environments where salience does not uniquely pin down the fixed categories.

(2008); Mengel (2012); Mertikopoulos and Sandholm (2024).

⁴⁰Our stability results for mixed SVE do not fully characterize the intermediate singleton payoff regime beyond the negative convergence result; we leave a systematic analysis of these cases for future work.

References

- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological review*, 98(3):409.
- Bellman, R. (1957). *Dynamic Programming*. Princeton University Press.
- Benaïm, M. (1999). Dynamics of stochastic approximation algorithms. *Séminaire de probabilités de Strasbourg*, 33:1–68.
- Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics - Foundations and Applications 2*, pages 69–186. North-Holland.
- Berk, R. H. (1966). Limiting Behavior of Posterior Distributions when the Model is Incorrect. *The Annals of Mathematical Statistics*, 37(1):51 – 58.
- Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Saliency Theory of Choice Under Risk. *The Quarterly Journal of Economics*, 127(3):1243–1285.
- Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Saliency and consumer choice. *Journal of Political Economy*, 121(5):803–843.
- Breitmoser, Y. (2021). An axiomatic foundation of conditional logit. *Economic Theory*, 72(1):245–261.
- Caplin, A. and Dean, M. (2008). Dopamine, reward prediction error, and economics. *The Quarterly Journal of Economics*, 123(2):663–701.
- Cominetti, R., Melo, E., and Sorin, S. (2010). A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior*, 70(1):71–83.
- Conley, C. C. (1978). *Isolated invariant sets and the Morse index / Charles Conley*. Published for the Conference Board of the Mathematical Sciences by the AMS, Providence.
- Debreu, G. (1960). Review of individual choice behavior by rd luce. *The American Economic Review*, 50(1):186–188.
- DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–72.
- Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of Statistics*, 7(2):269–281.

- Dinca, G. and Mawhin, J. (2021). *Continuation, Existence and Bifurcation*, pages 77–133. Springer International Publishing.
- Dries, L. P. D. v. d. (1998). *Tame Topology and O-minimal Structures*. London Mathematical Society Lecture Note Series. Cambridge University Press.
- Esponda, I. and Pouzo, D. (2016). Berk–nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica*, 84(3):1093–1130.
- Esponda, I. and Pouzo, D. (2021). Equilibrium in misspecified markov decision processes. *Theoretical Economics*, 16(2):717–757.
- Esponda, I. and Pouzo, D. (2026). Learning and equilibrium under model misspecification.
- Esponda, I., Pouzo, D., and Yamamoto, Y. (2021). Asymptotic behavior of bayesian learners with misspecified models. *Journal of Economic Theory*, 195:105260.
- Fudenberg, D., Lanzani, G., and Strack, P. (2021). Limit points of endogenous misspecified learning. *Econometrica*, 89(3):1065–1098.
- Fudenberg, D. and Levine, D. (1998). *The Theory of Learning in Games*. MIT Press.
- Fudenberg, D., Romanyuk, G., and Strack, P. (2017). Active learning with a misspecified prior. *Theoretical Economics*, 12(3):1155–1189.
- Fudenberg, D. and Strzalecki, T. (2015). Dynamic logit with choice aversion. *Econometrica*, 83(2):651–691.
- Gaunersdorfer, A. and Hofbauer, J. (1995). Fictitious play, shapley polygons, and the replicator equation. *Games and Economic Behavior*, 11(2):279–303.
- Gershman, S. J. and Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, 68(Volume 68, 2017):101–128.
- Gilboa, I. and Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*, 110(3):605–639.
- Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):148–164.

- Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, 108(3):15647–15654.
- Guillemin, V. and Pollack, A. (2010). *Differential Topology*. AMS Chelsea Publishing. AMS Chelsea Pub.
- Harsanyi, J. and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. MIT Press Classics. MIT Press.
- Heidhues, P., Kőszegi, B., and Strack, P. (2018). Unrealistic expectations and misguided learning. *Econometrica*, 86(4):1159–1214.
- Heidhues, P., Kőszegi, B., and Strack, P. (2021). Convergence in models of misspecified learning. *Theoretical Economics*, 16(1):73–99.
- Hofbauer, J. and Sandholm, W. H. (2002). On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294.
- Iyengar, S. S. and Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995.
- Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic theory*, 123(2):81–104.
- Jehiel, P. (2026). *Analogy-Based Expectation Equilibrium and Related Concepts: Theory, Applications, and Beyond*, page 305–388. Econometric Society Monographs (Cambridge).
- Jehiel, P. and Samet, D. (2005). Learning to play games in extensive form by valuation. *Journal of Economic Theory*, 124(2):129–148.
- Jehiel, P. and Samet, D. (2007). Valuation equilibrium. *Theoretical Economics*, 2(2):163–185.
- Jordan, J. (1993). Three problems in learning mixed-strategy nash equilibria. *Games and Economic Behavior*, 5(3):368–386.
- Kleinberg, R., Niculescu-Mizil, A., and Sharma, Y. (2010). Regret bounds for sleeping experts and bandits. *Machine Learning*, 80(2–3):245–272.
- Kushner, H. and Yin, G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Stochastic Modelling and Applied Probability. Springer New York.
- LiCalzi, M. (1995). Fictitious play by cases. *Games and Economic Behavior*, 11(1):64–89.

- Lichtenstein, S. and Slovic, P. (2006). *The Construction of Preference*. Cambridge University.
- Luce, R. D. (1959). *Individual choice behavior*, volume 4. Wiley New York.
- Marsden, J. and McCracken, M. (2012). *The Hopf Bifurcation and Its Applications*. Applied Mathematical Sciences. Springer New York.
- Matějka, F. and McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98.
- McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38.
- Mengel, F. (2012). Learning across games. *Games and Economic Behavior*, 74(2):601–619.
- Mertikopoulos, P. and Sandholm, W. H. (2024). Nested replicator dynamics, nested logit choice, and similarity-based learning. *Journal of Economic Theory*, 220:105881.
- Milnor, J. (1965). *Topology from the Differentiable Viewpoint*. University Press of Virginia.
- Moll, B. (2025). The trouble with rational expectations in heterogeneous agent models: A challenge for macroeconomics. *The Economic Journal*, page 104.
- Murooka, T. and Yamamoto, Y. (2025). Bayesian learning when players are misspecified about others. ISER Discussion Paper 1284, Osaka University, ISER.
- Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53(3):139–154. Special Issue: Dynamic Decision Making.
- Nyarko, Y. (1991). Learning in mis-specified models and the possibility of cycles. *Journal of Economic Theory*, 55(2):416–427.
- Pemantle, R. (1990). Nonconvergence to Unstable Points in Urn Models and Stochastic Approximations. *The Annals of Probability*, 18(2):698 – 712.
- Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527 – 535.
- Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407.
- Rosch, E. and Lloyd, B. B. (1978). *Principles of categorization*. MIT press.

- Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033.
- Samuelson, L. (2001). Analogies, adaptation, and anomalies. *Journal of Economic Theory*, 97(2):320–366.
- Sarin, R. and Vahid, F. (1999). Payoff assessments without probabilities: A simple dynamic model of choice. *Games and Economic Behavior*, 28(2):294–309.
- Shapley, L. S. (1964). *Some Topics in Two-Person Games*. Princeton University Press.
- Sharib, J. R., Pomeranz, J. L., Mozaffarian, D., and Cash, S. B. (2024). Disclosure of mandatory and voluntary nutrition labelling information across major online food retailers in the usa. *Public Health Nutrition*, 27(1):e203.
- Skinner, B. (1938). *The Behavior of Organisms: An Experimental Analysis*. B.F. Skinner Foundation.
- Smith, H. L. (1995). *Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems*. American Mathematical Soc.
- Spiegler, R. (2016). Bayesian networks and boundedly rational expectations. *The Quarterly Journal of Economics*, 131(3):1243–1290.
- Steiner, J. and Stewart, C. (2008). Contagion through learning. *Theoretical Economics*, 3(4):431–458. online version, listed from 2010 onwards.
- Steiner, J., Stewart, C., and Matějka, F. (2017). Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica*, 85(2):521–553.
- Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. The MIT Press, second edition.
- Tsitsiklis, J. N. (1994). Asynchronous stochastic approximation and q-learning. *Machine Learning*, 16(3):185–202.
- Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79.
- van den Dries, L., Macintyre, A., and Marker, D. (1994). The elementary theory of restricted analytic fields with exponentiation. *Annals of Mathematics*, 140(1):183–205.
- Watkins, C. J. (1989). *Learning from delayed rewards*. King’s College, Cambridge UK.
- Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

A Appendix: Omitted Proofs

A.1 Theorem 4

Proof. Consider the family of decision trees RPS with three similarity classes seen in Fig. 4. We set $z_R = z_P = z_S = z$ with a fixed $z \in (-1, 0)$ and $p(\omega) = 1/6$, for all $\omega \in \Omega \setminus \{\omega_7\}$ with $p(\omega_7) = 0$, in order to simplify the algebra. We verify that any VE in this decision tree must be fully-mixed with Alice being completely indifferent across all three equivalence classes (see Lem. B.3-B.4 in the Supp. Appendix). To see this informally, notice the binary payoffs are such that each class “defeats” one other class in exactly one binary menu and is “defeated” in another. The singleton menus contribute identical negative shifts z symmetrically to the three classes. Hence no class can be a best-reply in *all* menus, ruling out strict-pure VE; by the same symmetry, partially-mixed VE can’t exist. At any binary menu, optimality requires the class with the larger valuation to be chosen; but because the three binary menus cyclically favor different classes and singleton penalties are symmetric, the only way to satisfy all menu-wise optimality conditions is $v_R = v_P = v_S$. Therefore, every VE is fully-mixed and entails indifference across classes; in relative coordinates $x := v_R - v_S = 0$ and $y := v_P - v_S = 0$. All statements in this proof are in relative coordinates $(x, y) = (v_R - v_S, v_P - v_S)$; thus $(0, 0)$ corresponds to the (absolute) SVE $(z/2, z/2, z/2)$, which is an SVE for all $\beta \geq 0$.

Since any fully-mixed VE has relative valuation $(0, 0)$,⁴¹ Theorem 1(ii) implies that any high-sensitivity limit of SVE must have relative valuation $\mathbf{v}^\infty = (0, 0)$. The logit perturbation uniquely selects the uniform element of the fully-mixed VE continuum: $\sigma_\omega^s(\mathbf{v}^\infty) = 1/2$ for every binary menu ω and $s \in \omega$. We show in Section B.4 of the Supplemental Appendix that for sufficiently large β , there’s a unique SVE that lies near $(0, 0)$. Let

$$\begin{cases} \dot{x} := U(x, y) = g_R(x, y) - g_S(x, y) - x, \\ \dot{y} := V(x, y) = g_P(x, y) - g_S(x, y) - y, \end{cases}$$

denote the CQL dynamics (in relative valuations) with parameters $\beta > 0$ and $-1 < z < 0$. The reduction to two-dimensions by translating each of v_R, v_P, v_S by $-v_S$ is w.l.o.g. since the drift map g is translation-invariant (see Lemma B.9). We note that the unique LSVE ($\mathbf{v}_\infty^* = (0, 0)$ and $\sigma_\omega^s(\mathbf{v}_\infty^*) = 0.5$ for all $s \in \mathcal{S}$ and $\{\omega \in \Omega_s : |\Omega_s| > 1\}$) is an SVE for all $\beta \geq 0$. The singleton menus are degenerate with the single available class being chosen with probability 1. Set $\lambda(u) := \frac{1}{1 + e^{-\beta u}}$, with $q := \lambda(x - y)$, $r := \lambda(y)$, $s := \lambda(x)$. At the binary menus, the logit choice probabilities are therefore $\sigma_R^{(RP)} = q$, $\sigma_P^{(PS)} = r$, $\sigma_R^{(RS)} = s$, with

⁴¹Recall that the fully-mixed VE is not unique - in fact, any $q \in (0, 1)$ such that $\sigma_{\omega_1}^P = \sigma_{\omega_2}^S = \sigma_{\omega_3}^R = q$, constitutes a continuum of fully-mixed VE lying on a manifold of dimension 1.

complementary probabilities $1 - q$, $1 - r$, and $1 - s$, respectively. Hence,

$$g_R(x, y) = \frac{-q + s + z}{q + s + 1}, \quad g_P(x, y) = \frac{1 - q - r + z}{2 - q + r}, \quad g_S(x, y) = \frac{s - r + z}{3 - r - s},$$

and the reduced CQL dynamics are $\dot{x} = (g_R - g_S)(x, y) - x$, $\dot{y} = (g_P - g_S)(x, y) - y$. At the fully mixed SVE $(x, y) = (0, 0)$, one has $q = r = s = \frac{1}{2}$, so $g_R(0, 0) = g_P(0, 0) = g_S(0, 0) = \frac{z}{2}$. Moreover, since

$$\lambda(u) = \frac{1}{2} + \frac{\beta}{4}u + o(u) \quad (u \rightarrow 0),$$

$$q = \frac{1}{2} + \frac{\beta}{4}(x - y) + o(\|(x, y)\|), \quad r = \frac{1}{2} + \frac{\beta}{4}y + o(\|(x, y)\|), \quad s = \frac{1}{2} + \frac{\beta}{4}x + o(\|(x, y)\|).$$

Using the Taylor expansions, the Jacobian $A = D(\dot{x}, \dot{y})|_{(0,0)}$ is

$$A = \begin{pmatrix} -1 + \frac{\beta}{16}(-3z - 2) & \frac{\beta}{4} \\ -\frac{\beta}{4} & -1 + \frac{\beta}{16}(-3z + 2) \end{pmatrix}.$$

Hence, $\text{tr}(A) = -2 - \frac{3z}{8}\beta$, and $\det(A) = 1 + \frac{3z}{8}\beta + \frac{9z^2 + 12}{256}\beta^2$, and the discriminant equals $\text{tr}(A)^2 - 4\det(A) = -\frac{3}{16}\beta^2 < 0$, so the eigenvalues are a complex conjugate pair

$$\lambda_{1,2}(\beta, z) = -\frac{3z\beta}{16} - 1 \pm i \frac{\beta\sqrt{3}}{8},$$

with $\Re\lambda = \frac{1}{2}\text{tr}(A) = -1 - \frac{3z}{16}\beta$. Therefore, the fully-mixed SVE $(0, 0)$ with index $+1$ is

$$\begin{cases} \text{a } \textit{stable} \text{ focus} & \text{if } \beta < -\frac{16}{3z}, \\ \textit{non-hyperbolic} & \text{if } \beta = -\frac{16}{3z}, \\ \text{an } \textit{unstable} \text{ focus} & \text{if } \beta > -\frac{16}{3z}. \end{cases}$$

Fix $z \in (-1, 0)$. At $\beta_c = -\frac{16}{3z}$, the real part $\Re\lambda_{1,2} = 0$ with non-zero imaginary part $\frac{\beta\sqrt{3}}{8} > 0$. Moreover, since $\frac{d}{d\beta}\Re\lambda|_{\beta_c} = -\frac{3z}{16} > 0$, the transversality condition for a Hopf bifurcation is satisfied. By strict negativity of the first Lyapunov coefficient $\ell_1(\beta_c) = \frac{7\sqrt{3}}{9z} < 0$, a supercritical Hopf bifurcation (Marsden and McCracken, 2012) occurs at the critical $\beta = \beta_c$ where

a unique curve of periodic solutions bifurcates from the fixed point into the region $\beta > \beta_c$. For all $\beta > \beta_c$, the unique SVE $(0, 0)$ is unstable and the periodic orbit bifurcating from it is an isolated stable limit cycle whose amplitude grows like $\sqrt{|\beta - \beta_c|}$.

Consequently, every bounded trajectory converges to a periodic orbit. Indeed, expected payoffs are finite, so the flow is confined to the compact convex hull $M \subset \mathbb{R}^2$ of the relative expected payoffs, and the vector field points strictly inward on ∂M , making M positively invariant. In the planar system, the only equilibrium is $(0, 0)$, which is an asymptotically unstable focus for $\beta > \beta_c$. By the Poincaré–Bendixson theorem, any ω -limit set contained in M that is not an equilibrium must be a periodic orbit. Hence all non-trivial trajectories are repelled away from $(0, 0)$ when $\beta > \beta_c$, and every bounded trajectory has a periodic orbit as its ω -limit set. Moreover, for β sufficiently close to β_c , the Hopf bifurcation is super-critical, so exactly one stable limit cycle is created; by positive invariance of M and absence of other equilibria, all bounded trajectories spiral into this unique cycle (see Fig. 5). Finally, by Thm. 6.12 & Cor. 6.14 in [Benaïm \(1999\)](#), the ω -limit set of any realization of the discrete-time CQL dynamics (2) is almost surely an internally chain-recurrent set of the flow generated by (3) - thus, it is a periodic orbit (or a cylinder of periodic orbits). Thus, we've demonstrated that the set of asymptotically stable SVE is empty for an open set of decision trees RPS parameterized by $z \in (-1, 0)$. We note that this range of z allows for both competition (substitution) and cooperation (complementarity) among the classes.

The tree used in the proof is chosen to be symmetric only for transparency. The conclusion is not knife-edge as seen in Example 3.3. Endow the parameter space of reduced decision trees with the Euclidean topology in the primitives (p, π) (with p in the relative interior of the simplex). At the symmetric specification, the relative CQL vector field undergoes a *supercritical* Hopf bifurcation at $\beta = \beta_c$, i.e. the equilibrium has a simple pair of imaginary eigenvalues with transversal crossing and $\ell_1(\beta_c) < 0$. These conditions are open. Hence there exists an open neighborhood \mathcal{U} of this specification and an interval $(\beta_c, \beta_c + \varepsilon)$ such that, for every tree in \mathcal{U} and every $\beta \in (\beta_c, \beta_c + \varepsilon)$, the corresponding mean-field dynamics admit a nearby unstable SVE and a locally attracting periodic orbit. In particular, on \mathcal{U} the set of asymptotically stable SVE is empty for all $\beta > \beta_c$ sufficiently close to β_c . \square

A.2 Theorem 5

Proof. Fix a finite reduced decision tree $\mathcal{T}'_n = (\mathcal{S}, \Omega, p, \pi)$ with connected co-occurrence graph G and full support of singleton menus, and boost all singleton expected payoffs by $z > \hat{z}$ as in the theorem. Write $\dot{\mathbf{v}} = F(\mathbf{v})$, where $F(\mathbf{v}) := g(\mathbf{v}; \beta) - \mathbf{v}$, with $g_s(\mathbf{v}) = \frac{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v}) \pi_s(\omega)}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v})}$, and $\sigma_\omega^s(\mathbf{v}) = \mathbf{1}\{s \in \omega\} \frac{\exp(\beta v_s)}{\sum_{j \in \omega} \exp(\beta v_j)}$.

Assumption A.1. *There exists $\hat{z} < \infty$ such that for all $z > \hat{z}$ and all $s \in \mathcal{S}$,*

$$\frac{p(\{s\})}{\sum_{\omega \in \Omega_s} p(\omega)} (\pi_s(\{s\}) + z) + \left(1 - \frac{p(\{s\})}{\sum_{\omega \in \Omega_s} p(\omega)}\right) \min_{\omega \in \Omega_s: |\omega| \geq 2} \pi_s(\omega) > \max_{\omega \in \Omega_s: |\omega| \geq 2} \pi_s(\omega).$$

We assume the following conservative bound on z that is sufficient for the proof.

$$\hat{z} \geq \max_{s \in \mathcal{S}} \left\{ \left(\frac{\sum_{\omega \in \Omega_s} p(\omega)}{p(\{s\})} \right) \left(\max_{\omega \in \Omega_s: |\omega| \geq 2} \pi_s(\omega) - \min_{\omega \in \Omega_s: |\omega| \geq 2} \pi_s(\omega) \right) + \min_{\omega \in \Omega_s: |\omega| \geq 2} \pi_s(\omega) - \pi_s(\{s\}) \right\}.$$

Throughout this proof, once z is fixed we absorb the singleton boost into the payoff array for notational simplicity, i.e. we write $\pi_s(\{s\})$ for the z -shifted payoff. For a fixed $z \geq \hat{z}$, let $K = \prod_{s \in \mathcal{S}} [m_s, M_s]$ where $m_s = \min_{\omega \in \Omega_s} \pi_s(\omega)$ and $M_s = \max_{\omega \in \Omega_s} \pi_s(\omega)$ with $m_s < M_s$ by assumption. Then K is compact, convex, and $g(K) \subseteq K$, hence every SVE lies in K .

Step 1 - No Strict Pure VE for large z : Suppose, toward a contradiction, that for some large z there exists a strict pure VE with a strict total order $v_{i_1}^* < v_{i_2}^* < \dots < v_{i_n}^*$. Let $i := i_1$ be the lowest-ranked class. Since the co-occurrence graph G is connected, there exists a mixed (non-degenerate) menu $\omega \in \text{supp}(p)$ with $i \in \omega$ and $|\omega| \geq 2$. In a strict pure VE the unique maximizer is chosen at each menu, so at ω some $k \in \omega \setminus \{i\}$ is chosen. For $s \in \mathcal{S}$, let $\mathcal{M}_s := \{\omega \in \text{supp}(p) : s \in \omega, s = \arg \max_{j \in \omega} v_j^*\}$ and $D_s := \sum_{\omega \in \mathcal{M}_s} p(\omega)$. Valuation consistency gives $v_s^* = (\sum_{\omega \in \mathcal{M}_s} p(\omega) \pi_s(\omega)) / D_s$. Since i is never maximal in any non-singleton menu, $\mathcal{M}_i = \{\{i\}\}$ and hence $v_i^* = \pi_i(\{i\}) + z$, so the coefficient of z in v_i^* equals 1. For k as above we have $\{k\}, \omega \in \mathcal{M}_k$, so $D_k \geq p(\{k\}) + p(\omega) > p(\{k\})$ and

$$v_k^* = \underbrace{\frac{p(\{k\})}{D_k}}_{\alpha_k \in (0,1)} (z + \pi_k(\{k\})) + \frac{\sum_{\omega' \in \mathcal{M}_k \setminus \{k\}} p(\omega') \pi_k(\omega')}{D_k}.$$

Since $\alpha_k \in (0, 1)$, one has $v_i^*(z) - v_k^*(z) = (1 - \alpha_k)z + O(1) \rightarrow +\infty$ as $z \rightarrow +\infty$, contradicting $v_i^*(z) < v_k^*(z)$. Thus, by continuity, there exists $\hat{z} < \infty$ such that for all $z > \hat{z}$ no strict pure VE exists. Because VE exist in finite trees (Lemma B.5), at least one mixed VE exists. By Thm. 1, for $\beta \geq \hat{\beta}$ any SVE lies arbitrarily close to some mixed VE.

Step 2 - Local Asymptotic Stability of any SVE ($\forall z > \hat{z}$): Differentiating f_s yields

$$J_{ss}(\mathbf{v}) = \beta \frac{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s (1 - \sigma_\omega^s) (\pi_s(\omega) - g_s(\mathbf{v}))}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s} - 1, \quad (4)$$

$$J_{sk}(\mathbf{v}) = \beta \frac{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s \sigma_\omega^k (g_s(\mathbf{v}) - \pi_s(\omega))}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s}, \quad k \neq s. \quad (5)$$

$\forall s \in \mathcal{S}$, at $\omega = \{s\}$ one has $\sigma_\omega^s \equiv 1$ and $\sigma_\omega^k \equiv 0$ and $1 - \sigma_\omega^s \equiv 0$, so singleton menus contribute 0 to all J_{ss} and J_{sk} with $k \neq s$. By Asm. A.1 with $z > \hat{z}$, for every $s \in \mathcal{S}$ and every non-singleton menu $\omega \in \Omega_s$ s.t. $|\omega| \geq 2$ we have $g_s(\mathbf{v}) > \pi_s(\omega)$ for all $\mathbf{v} \in K$. To see this, fix $s \in \mathcal{S}$ and write $P_s := \sum_{\omega \in \Omega_s} p(\omega)$. Since $\sigma_{\{s\}}^s(\mathbf{v}) \equiv 1$ and $\sigma_\omega^s(\mathbf{v}) \leq 1$ for all ω , we have $\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v}) \leq \sum_{\omega \in \Omega_s} p(\omega) = P_s$. Hence the conditional weight placed on the singleton menu $\{s\}$ inside $g_s(\mathbf{v})$ satisfies $\frac{p(\{s\}) \sigma_{\{s\}}^s(\mathbf{v})}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v})} \geq \frac{p(\{s\})}{P_s}$, $\forall \mathbf{v} \in K$. As the residual conditional mass is supported on non-singleton menus, we obtain the uniform bound

$$g_s(\mathbf{v}) \geq \frac{p(\{s\})}{P_s} (\pi_s(\{s\}) + z) + \left(1 - \frac{p(\{s\})}{P_s}\right) \min_{\omega \in \Omega_s: |\omega| \geq 2} \pi_s(\omega), \quad \forall \mathbf{v} \in K.$$

By Assumption A.1, the right-hand side exceeds $\max_{\omega \in \Omega_s: |\omega| \geq 2} \pi_s(\omega)$, and therefore $g_s(\mathbf{v}) > \pi_s(\omega)$ for all $\mathbf{v} \in K$ and all $\omega \in \Omega_s$ with $|\omega| \geq 2$. Hence, for any pair $\{s, k\} \in E(G)$ there exists a non-singleton menu $\omega \in \text{supp}(p)$ with $\{s, k\} \subseteq \omega$ and $|\omega| \geq 2$, and for that menu $p(\omega) \sigma_\omega^s(\mathbf{v}) \sigma_\omega^k(\mathbf{v}) (g_s(\mathbf{v}) - \pi_s(\omega)) > 0$ for all $\mathbf{v} \in K$, so that the corresponding Jacobian entry satisfies $J_{sk}(\mathbf{v}) > 0$ whenever $\{s, k\} \in E(G)$, $s \neq k$, whereas $J_{sk}(\mathbf{v}) \equiv 0$ whenever $\{s, k\} \notin E(G)$ (since s and k never co-occur in a menu). In particular, for every $\mathbf{v} \in K$ and $\beta < \infty$, the Jacobian $J(\mathbf{v})$ is Metzler, $J_{sk}(\mathbf{v}) \geq 0$ for all $s \neq k$. Summing (5) over $k \neq s$,

$$\underbrace{\sum_{k \neq s} |J_{sk}(\mathbf{v})|}_{\mathcal{R}_s(\mathbf{v})} = \beta \frac{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s (1 - \sigma_\omega^s) (g_s(\mathbf{v}) - \pi_s(\omega))}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s} = -J_{ss}(\mathbf{v}) - 1.$$

Thus for each row s , $J_{ss}(\mathbf{v}) = -\mathcal{R}_s(\mathbf{v}) - 1 < -1$, and the s -th Gershgorin disc⁴² is the closed disc centered at J_{ss} with radius \mathcal{R}_s , whose rightmost point is $J_{ss} + \mathcal{R}_s = -1$. By the Gershgorin circle theorem, every eigenvalue λ of $J(\mathbf{v})$ satisfies $\Re \lambda \leq -1 < 0$. In particular, $\forall \beta \geq 0$, at any SVE \mathbf{v}^* the Jacobian is a Hurwitz (stability) matrix, so the SVE is hyperbolic and locally exponentially (asymptotically) stable by the Hartman-Grobman theorem.

Step 3 - Uniqueness of SVE: Consider $h(\mathbf{v}) := \mathbf{v} - g(\mathbf{v})$ on K . We have $h(\mathbf{v}) \neq 0$ on the boundary⁴³ ∂K and $h(\mathbf{v}) \cdot n(\mathbf{v}) > 0$ for the outward unit normal $n(\mathbf{v})$, since $g(K) \subset \text{int } K$; hence h points strictly outward⁴⁴ on ∂K . At any zero \mathbf{v}^* of h , $Dh(\mathbf{v}^*) = I - Dg(\mathbf{v}^*) = -J(\mathbf{v}^*)$

⁴²Let $\mathcal{D}_s(J_{ss}, \mathcal{R}_s) \subset \mathbb{C}$ be a closed disc in the complex plane centered at J_{ss} with radius \mathcal{R}_s . We refer to such a disc as a Gershgorin disc. Across the n rows of the Jacobian matrix, we define n such discs. By the Gershgorin Circle theorem, every eigenvalue of J lies within at least one of the Gershgorin discs. Equivalently, all eigenvalues of J lie within the union of the n Gershgorin discs.

⁴³For $\beta < \infty$, any SVE is interior. Additionally, for $z > \hat{z}$, any limiting SVE as $\beta \uparrow \infty$ features indifference at least at the bottom (no unique strictly dominated class) - ensuring that \mathbf{v}^* is interior even in the limit.

⁴⁴Because K is convex and $g(K) \subset \text{int } K$, for any $\mathbf{v} \in \partial K$ we have $g(\mathbf{v}) \in \text{int } K$ and $g(\mathbf{v}) \neq \mathbf{v}$. Hence the vector $\mathbf{v} - g(\mathbf{v})$ points strictly outward at \mathbf{v} : with $n(\mathbf{v})$ the outward unit normal, $h(\mathbf{v}) \cdot n(\mathbf{v}) =$

has eigenvalues with strictly positive real parts, so each \mathbf{v}^* is an isolated non-degenerate zero⁴⁵ with index +1. The Poincaré–Hopf index theorem⁴⁶ on the compact convex manifold K (Euler characteristic $\chi(K) = 1$) yields that the algebraic sum of indices of all isolated zeros equals 1. Therefore there is exactly one zero, i.e. exactly one SVE in K for every $\beta \geq 0$. Moreover, by Thm. 1, for $\beta \geq \hat{\beta}$ this unique SVE lies in the neighborhood of a mixed VE.

Step 4 - Global Asymptotic Stability: From Step 2, we know for every $\mathbf{v} \in K$ the Jacobian $J(\mathbf{v})$ is Metzler, $J_{sk}(\mathbf{v}) \geq 0$ for all $s \neq k$, and its off-diagonal sign pattern coincides with the adjacency structure of the co-occurrence graph G . Because G is connected and every undirected edge $\{s, k\} \in E(G)$ generates strictly positive entries in both directions, $J_{sk}(\mathbf{v}) > 0$ and $J_{ks}(\mathbf{v}) > 0$, the directed graph of positive off-diagonals of $J(\mathbf{v})$ is strongly connected. Thus $J(\mathbf{v})$ is an irreducible Metzler matrix for every $\mathbf{v} \in K$. It follows that the ODE $\dot{\mathbf{v}} = F(\mathbf{v})$ is *cooperative* on K and, by Thm. 4.1.1 in Smith (1995), the associated semi-flow on K is *strongly monotone* (equivalently, strongly order preserving⁴⁷ in the sense of Prop. 1.1.1 in Smith, 1995). By Step 3, the SVE \mathbf{v}^* is the unique equilibrium of the flow in the compact, convex, positively invariant set K and lies in $\text{int } K$. Strong monotonicity on a compact order interval with a unique equilibrium implies global convergence. By Thm. 2.3.1 in Smith (1995), every trajectory starting in K converges to \mathbf{v}^* . Hence the unique SVE corresponding to a mixed VE in the high-sensitivity limit is a global attractor of the mean-field dynamics for every $\beta \in \mathbb{R}_+$. Finally, by Benaim (1999, Theorem 5.7), when the mean-field dynamic (3) admits a unique globally asymptotically stable equilibrium, the discrete-time stochastic recursion in (2) converges to this equilibrium almost surely. \square

The remaining proofs appear in the Supplemental Appendix.

$(\mathbf{v} - g(\mathbf{v})) \cdot n(\mathbf{v}) > 0$. Equivalently, $F(\mathbf{v}) = -h(\mathbf{v})$ points strictly inward on ∂K . It follows that K is positively invariant for the CQL flow $\dot{\mathbf{v}} = F(\mathbf{v})$: trajectories starting in K cannot exit K .

⁴⁵Since each zero is isolated (by the inverse function theorem), there can only be countably many isolated zeroes in $\text{int } K$. In fact, since K is a compact set in \mathbb{R}^n , by the Heine-Borel theorem, every open cover of K has a finite sub-cover. Consequently, there can only be finitely many isolated zeroes of $F(\mathbf{v})$.

⁴⁶Poincaré-Hopf index theorem (Milnor, 1965): Let M be a compact differentiable manifold. Let \mathbf{h} be a vector field on M with isolated zeroes. If M has a boundary, then we insist that \mathbf{h} be pointing in the outward normal direction along the boundary. Then we have the formula: $\sum_i \text{index}_{x_i}(\mathbf{h}) = \chi(M)$, where the sum is over all isolated zeroes of the vector field \mathbf{h} , $\chi(M)$ is the Euler characteristic of M . $\chi(K) = 1$ since the convex set K has trivial fundamental group - it is contractible and homotopic to a point.

⁴⁷A semi-flow $\phi(t, \mathbf{x})$ is a mapping from $\mathbb{R}_+ \times \mathbb{R}^n$ to \mathbb{R}^n describing the evolution of the system state \mathbf{x} over time t . A semi-flow $\phi(t, \mathbf{x})$ is order preserving if for any two initial conditions $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with $\mathbf{x} \leq \mathbf{y}$, it holds that $\phi(t, \mathbf{x}) \leq \phi(t, \mathbf{y})$ for all $t \geq 0$. A semi-flow $\phi(t, \mathbf{x})$ is strongly order preserving if it is order preserving and, additionally, for any $\mathbf{x} < \mathbf{y}$, $\phi(t, \mathbf{x}) \ll \phi(t, \mathbf{y})$ for $t > 0$, where \ll denotes the strong ordering, i.e. each component of $\phi(t, \mathbf{x})$ is strictly less than the corresponding component of $\phi(t, \mathbf{y})$.

B Supplemental Appendix: Omitted Proofs

Useful definitions

For $\beta \geq 0$ and $\omega \in \Omega$, the logit policy is $\sigma_\omega^s(\mathbf{v}; \beta) := \mathbf{1}\{s \in \omega\} \frac{\exp(\beta v_s)}{\sum_{j \in \omega} \exp(\beta v_j)} \in \Delta(\omega)$. For $s \in \mathcal{S}$, let $\Omega_s := \{\omega \in \Omega : s \in \omega\}$ and assume $\Omega_s \cap \text{supp}(p) \neq \emptyset$ (otherwise s can be removed from \mathcal{S}). Define the map $g(\cdot; \beta) : \mathbb{R}^{\mathcal{S}} \rightarrow \mathbb{R}^{\mathcal{S}}$ by $g_s(\mathbf{v}; \beta) := \frac{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v}; \beta) \pi_s(\omega)}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v}; \beta)}$, $s \in \mathcal{S}$. For each s , set $m_s := \min_{\omega \in \Omega_s \cap \text{supp}(p)} \pi_s(\omega)$, $M_s := \max_{\omega \in \Omega_s \cap \text{supp}(p)} \pi_s(\omega)$, $K := \prod_{s \in \mathcal{S}} [m_s, M_s]$. Thus, K is a nonempty, compact and convex subset of $\mathbb{R}^{\mathcal{S}}$.

Define the SVE set $\mathcal{V}(\beta) := \{\mathbf{v} \in K : \mathbf{v} = g(\mathbf{v}; \beta)\}$. $\mathcal{V}(\infty)$ denotes the set of accumulation points of the correspondence $\beta \mapsto \mathcal{V}(\beta)$: $\mathcal{V}(\infty) := \{\mathbf{v} \in K : \exists \beta_n \uparrow \infty, \exists \mathbf{v}_n \in \mathcal{V}(\beta_n) \text{ with } \mathbf{v}_n \rightarrow \mathbf{v}\}$. Since $\mathcal{V}(\beta) \neq \emptyset$ for every $\beta < \infty$ (proved below) and K is compact, any sequence $\beta_n \uparrow \infty$ admits $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ with a convergent subsequence whose limit lies in $\mathcal{V}(\infty)$. Thus, $\mathcal{V}(\infty) \neq \emptyset$. For $\mathbf{v} \in \mathbb{R}^{\mathcal{S}}$ and $\omega \in \Omega$, let $\text{argmax}_\omega(\mathbf{v}) := \arg \max_{j \in \omega} v_j$. A greedy choice profile is a (menu-wise) collection of mixed actions $\sigma = (\sigma_\omega)_{\omega \in \Omega}$ with $\sigma_\omega \in \Delta(\omega)$ such that $\text{supp}(\sigma_\omega) \subseteq \text{argmax}_\omega(\mathbf{v})$, $\forall \omega \in \Omega$. Given such a σ , define for each s the overall selection intensity $D_s(\sigma) := \sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega(s)$, and, when $D_s(\sigma) > 0$, the induced conditional menu weights $w_\omega^s(\sigma) := \frac{p(\omega) \sigma_\omega(s)}{D_s(\sigma)} \in \Delta(\Omega_s)$. We say that (σ, \mathbf{v}) is a *valuation equilibrium (VE)* if $\text{supp}(\sigma_\omega) \subseteq \text{argmax}_\omega(\mathbf{v})$ for all ω and, for every s with $D_s(\sigma) > 0$,

$$v_s = \sum_{\omega \in \Omega_s} w_\omega^s(\sigma) \pi_s(\omega) = \frac{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega(s) \pi_s(\omega)}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega(s)}.$$

If $D_s(\sigma) = 0$, VE imposes no restriction on v_s . Let \mathcal{VE} denote the set of valuation vectors \mathbf{v} for which there exists a σ such that (σ, \mathbf{v}) is a VE. For the closest-loss refinement of VE, define the loss gap and minimum loss gap $\Delta_s(\omega; \mathbf{v}) := \max_{j \in \omega} v_j - v_s$, and $\Delta_s^{\min}(\mathbf{v}) := \min_{\omega \in \Omega_s \cap \text{supp}(p)} \Delta_s(\omega; \mathbf{v})$, and the minimizing set $\Omega_s^*(\mathbf{v}) := \{\omega \in \Omega_s \cap \text{supp}(p) : \Delta_s(\omega; \mathbf{v}) = \Delta_s^{\min}(\mathbf{v})\}$. Define the set-valued closest-loss correspondence $g_\infty : K \rightrightarrows K$ by $g_\infty(\mathbf{v}) := \prod_{s \in \mathcal{S}} \text{co}\{\pi_s(\omega) : \omega \in \Omega_s^*(\mathbf{v})\}$, and denote its fixed-point set by $\hat{\mathcal{V}}(\infty) := \{\mathbf{v} \in K : \mathbf{v} \in g_\infty(\mathbf{v})\}$. We will show that $\mathcal{V}(\infty) \subseteq \hat{\mathcal{V}}(\infty)$ in general; the inclusion can be strict because g_∞ allows arbitrary convex combinations over $\Omega_s^*(\mathbf{v})$, whereas actual limits of the endogenous weights are constrained by equilibrium (menu-wise) consistency.

Fix a reduced decision tree structure (\mathcal{S}, Ω) and a support set $\Omega^+ \subseteq \Omega$, which we interpret as the fixed support of the menu distribution p . Let $\Delta(\Omega^+) := \{p \in \mathbb{R}^\Omega : p(\omega) > 0 \forall \omega \in \Omega^+, p(\omega) = 0 \forall \omega \notin \Omega^+, \sum_{\omega \in \Omega} p(\omega) = 1\}$ denote the corresponding relative interior of the face of the simplex of menu distributions with support Ω^+ , and let $\pi^+ := \{\pi_s(\omega) : \omega \in$

Ω^+ , $s \in \omega$ denote the expected payoff array on supported class menus, and define $\Pi^+ := \mathbb{R}^{\{(s,\omega): \omega \in \Omega^+, s \in \omega\}}$, with $d := \sum_{\omega \in \Omega^+} |\omega| = \sum_{s \in \mathcal{S}} |\Omega_s \cap \Omega^+|$. We say that a property $Q(p, \pi)$ holds for *generic decision trees* if there exists a residual (comeager) subset of $\Delta(\Omega^+) \times \Pi^+$, also of full Lebesgue measure in its natural affine dimension, such that $Q(p, \pi)$ holds for all (p, π) in that subset. Thus, throughout, whenever we say “for generic decision trees”, we mean for all (p, π) in some fixed residual, full-measure subset of $\Delta(\Omega^+) \times \Pi^+$. More generally, the admissible set of menu distributions can be partitioned into finitely many support strata, and the same argument applies on each stratum. Since there are only finitely many support patterns, one can combine the conclusions across strata.

B.1 Theorem 1

Theorem. (a) For every $\beta < \infty$, $\mathcal{V}(\beta)$ is nonempty and compact, and the correspondence $\beta \mapsto \mathcal{V}(\beta)$ is upper hemicontinuous and compact-valued on $[0, \infty)$.

(b) If $\beta_n \uparrow \infty$ and $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ with $\mathbf{v}_n \rightarrow \mathbf{v}^*$, then $\mathbf{v}^* \in \mathcal{V}\mathcal{E}$ and $\mathbf{v}^* \in \widehat{\mathcal{V}}(\infty)$, implying $\mathcal{V}(\infty) \subseteq \mathcal{V}\mathcal{E} \cap \widehat{\mathcal{V}}(\infty)$ and $\mathcal{V}(\infty)$ is compact. For every $\varepsilon > 0$ there exists $\widehat{\beta} < \infty$ such that

$$\mathcal{V}(\beta) \subseteq \bigcup_{\mathbf{u} \in \mathcal{V}(\infty)} B(\mathbf{u}, \varepsilon) \quad \forall \beta \geq \widehat{\beta}.$$

(c) Let $h(\mathbf{v}, \beta) := \mathbf{v} - g(\mathbf{v}; \beta)$. For generic decision trees and a.e. $\beta \in [0, \infty)$, every SVE is regular: $\det(I - D_{\mathbf{v}}g(\mathbf{v}; \beta)) \neq 0$, $\forall \mathbf{v} \in \mathcal{V}(\beta)$. Hence, for a.e. β , the set $\mathcal{V}(\beta)$ is finite; each SVE is isolated and lies on a locally unique real-analytic branch in β .

(d) Assume for each $s \in \mathcal{S}$ there exist distinct $\omega, \omega' \in \Omega_s \cap \text{supp}(p)$ with $\pi_s(\omega) \neq \pi_s(\omega')$. Let $h(\mathbf{v}; \beta) := \mathbf{v} - g(\mathbf{v}; \beta)$. Then, for every $\beta \geq 0$, $g(K) \subset \text{int } K$ and $\deg(h(\cdot; \beta), K, 0) = 1$. Hence, the sum of the local indices of all SVE in K equals +1; in particular, whenever all SVE are regular, their number is odd. Moreover, for any compact path-connected parameter interval Θ_0 such that $\theta \mapsto g(\cdot; \theta)$ is continuous, the SVE graph $\mathcal{G}_{\Theta_0} := \{(\mathbf{v}, \theta) \in K \times \Theta_0 : \mathbf{v} = g(\mathbf{v}; \theta)\}$ contains a connected component Γ_{ess} whose projection onto Θ_0 is surjective.

For a fixed decision tree and $\theta = \beta \in [0, \infty)$, the global SVE graph $\mathcal{G} := \{(\mathbf{v}, \beta) \in K \times [0, \infty) : \mathbf{v} = g(\mathbf{v}; \beta)\}$ is definable in \mathbb{R}_{exp} and therefore has finitely many definable path-connected components. At $\beta = 0$ there is a unique SVE \mathbf{v}_0 ; the connected component $\Gamma_{\text{prin}} \subset \mathcal{G}$ containing $(\mathbf{v}_0, 0)$ is the unique component whose projection onto $[0, \infty)$ is surjective. For generic decision trees, \mathcal{G} is a one-dimensional definable real-analytic embedded submanifold of $K \times [0, \infty)$, the projection $\xi : \mathcal{G} \rightarrow [0, \infty)$, $(\mathbf{v}, \beta) \mapsto \beta$, is a Morse function, and each connected component of \mathcal{G} is a smooth embedded curve that, over any bounded interval $[0, \bar{\beta}]$, is a finite union of real-analytic graphs separated by finitely many fold points.

(e) For generic decision trees, $\mathcal{V}(\infty) = \{\mathbf{v} \in K : \exists \beta_n \uparrow \infty, \exists \mathbf{v}_n \in \mathcal{V}(\beta_n), \mathbf{v}_n \rightarrow \mathbf{v}\}$ is a nonempty finite subset of \mathcal{VE} . More generally, let $\mathcal{C} \subset \mathcal{G}$ be any connected component whose projection is unbounded above. Then there exist $B < \infty$ and a definable real-analytic map $\varphi : (B, \infty) \rightarrow K$ such that $\Gamma := \{(\varphi(\beta), \beta) : \beta > B\} \subset \mathcal{C}$, $\xi(\Gamma) = (B, \infty)$, and the limit $\lim_{\beta \rightarrow \infty} \varphi(\beta) =: \mathbf{v}^*$ exists and is unique, with $\mathbf{v}^* \in \mathcal{V}(\infty) \subseteq \mathcal{VE} \cap \widehat{\mathcal{V}}(\infty)$. In particular, the principal branch Γ_{prin} converges to a unique valuation equilibrium $\mathbf{v}^* \in \mathcal{V}(\infty)$ as $\beta \rightarrow \infty$.

Proof. (a) Non-emptiness, compactness and upper hemicontinuity: Fix $s \in \mathcal{S}$ and pick $\omega' \in \Omega_s \cap \text{supp}(p)$. For $\beta < \infty$, $\sigma_{\omega'}^s(\mathbf{v}; \beta) > 0$ for all \mathbf{v} , hence $\sum_{\omega \in \Omega_s} p(\omega) \sigma_{\omega}^s(\mathbf{v}; \beta) \geq p(\omega') \sigma_{\omega'}^s(\mathbf{v}; \beta) > 0$, so $g_s(\cdot; \beta)$ is well-defined and continuous (indeed real-analytic in (\mathbf{v}, β)). Moreover, for any \mathbf{v} , the weights $w_{\omega}^s(\mathbf{v}; \beta) := \frac{p(\omega) \sigma_{\omega}^s(\mathbf{v}; \beta)}{\sum_{\omega' \in \Omega_s} p(\omega') \sigma_{\omega'}^s(\mathbf{v}; \beta)}$ form a probability vector on $\Omega_s \cap \text{supp}(p)$, and $g_s(\mathbf{v}; \beta) = \sum_{\omega \in \Omega_s} w_{\omega}^s(\mathbf{v}; \beta) \pi_s(\omega) \in [m_s, M_s]$. Thus $g(\cdot; \beta)$ maps K into itself. Since K is nonempty, compact, and convex and $g(\cdot; \beta) : K \rightarrow K$ is continuous, Brouwer's theorem yields a fixed point $\mathbf{v}^* \in K$ with $\mathbf{v}^* = g(\mathbf{v}^*; \beta)$, so $\mathcal{V}(\beta) \neq \emptyset$. Because $\mathcal{V}(\beta)$ is the zero set of the continuous map $\mathbf{v} \mapsto \mathbf{v} - g(\mathbf{v}; \beta)$ on compact K , it is compact. Take $\beta_n \rightarrow \beta$ and $\mathbf{v}_n \rightarrow \mathbf{v}$ with $\mathbf{v}_n \in \mathcal{V}(\beta_n)$. Continuity of g in (\mathbf{v}, β) implies $\mathbf{v} = \lim_n \mathbf{v}_n = \lim_n g(\mathbf{v}_n; \beta_n) = g(\mathbf{v}; \beta)$, hence $\mathbf{v} \in \mathcal{V}(\beta)$ and the graph of $\beta \mapsto \mathcal{V}(\beta)$ is closed. Since values are compact, the correspondence is upper hemicontinuous and compact-valued on $[0, \infty)$.

(b) High-sensitivity limit ($\mathcal{V}(\infty) \subseteq \mathcal{VE} \cap \widehat{\mathcal{V}}(\infty)$): Fix $\beta_n \uparrow \infty$ and $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ with $\mathbf{v}_n \rightarrow \mathbf{v}^* \in K$. For each $\omega \in \Omega$, define the menu-wise logit distribution $\sigma_{\omega}^n(\cdot) := \sigma(\cdot \mid \omega; \mathbf{v}_n, \beta_n) \in \Delta(\omega)$. Since Ω is finite and each $\Delta(\omega)$ is compact, by a diagonal subsequence argument we may assume $\sigma_{\omega}^n \rightarrow \sigma_{\omega}^*$ for every $\omega \in \Omega$, for some $\sigma_{\omega}^* \in \Delta(\omega)$.

Step 1 ($\mathbf{v}^ \in \mathcal{VE}$):* We claim that $\text{supp}(\sigma_{\omega}^*) \subseteq \text{argmax}_{\omega}(\mathbf{v}^*)$ for every ω . Fix ω and $s \in \omega$ such that $v_s^* < \max_{j \in \omega} v_j^*$. Let $\delta := \max_{j \in \omega} v_j^* - v_s^* > 0$. For n large, $\max_{j \in \omega} v_{n,j} - v_{n,s} \geq \delta/2$, hence $\sigma_{\omega}^n(s) = \frac{\exp(\beta_n v_{n,s})}{\sum_{j \in \omega} \exp(\beta_n v_{n,j})} \leq \exp(-\beta_n \delta/2) \rightarrow 0$. Therefore $\sigma_{\omega}^*(s) = 0$, proving the claim. Fix $s \in \mathcal{S}$ and define $D_s^n := \sum_{\omega \in \Omega_s} p(\omega) \sigma_{\omega}^n(s)$, and $N_s^n := \sum_{\omega \in \Omega_s} p(\omega) \sigma_{\omega}^n(s) \pi_s(\omega)$. By definition of g and the fixed-point identity $\mathbf{v}_n = g(\mathbf{v}_n; \beta_n)$, $v_{n,s} = N_s^n / D_s^n$, $\forall n$. By convergence of $\sigma_{\omega}^n(s)$ for each ω and finiteness of Ω_s , we have $D_s^n \rightarrow D_s^* := \sum_{\omega \in \Omega_s} p(\omega) \sigma_{\omega}^*(s)$ and $N_s^n \rightarrow N_s^* := \sum_{\omega \in \Omega_s} p(\omega) \sigma_{\omega}^*(s) \pi_s(\omega)$. If $D_s^* > 0$, then $D_s^n > 0$ eventually and taking limits in $v_{n,s} = N_s^n / D_s^n$ yields $v_s^* = \frac{N_s^*}{D_s^*} = \frac{\sum_{\omega \in \Omega_s} p(\omega) \sigma_{\omega}^*(s) \pi_s(\omega)}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_{\omega}^*(s)}$. Therefore, (σ^*, \mathbf{v}^*) satisfies the VE conditions: menu-wise optimality and valuation consistency for all s with $D_s(\sigma^*) > 0$. Hence $\mathbf{v}^* \in \mathcal{VE}$.

Step 2 ($\mathbf{v}^ \in \widehat{\mathcal{V}}(\infty)$):* Fix $s \in \mathcal{S}$. For each n and $\omega \in \Omega_s \cap \text{supp}(p)$, write the conditional

menu weights using $v_{n,s} = g_s(\mathbf{v}_n; \beta_n) = \sum_{\omega \in \Omega_s \cap \text{supp}(p)} w_\omega^{s,n} \pi_s(\omega)$, as

$$w_\omega^{s,n} := \frac{p(\omega) \sigma_\omega^s(\mathbf{v}_n; \beta_n)}{\sum_{\omega' \in \Omega_s \cap \text{supp}(p)} p(\omega') \sigma_{\omega'}^s(\mathbf{v}_n; \beta_n)} \in \Delta(\Omega_s \cap \text{supp}(p)).$$

Let $\Delta_s(\omega; \mathbf{v}) := \max_{j \in \omega} v_j - v_s$ and $\Delta_s^{\min}(\mathbf{v}) := \min_{\omega \in \Omega_s \cap \text{supp}(p)} \Delta_s(\omega; \mathbf{v})$, and define the minimizing set $\Omega_s^*(\mathbf{v}) := \{\omega \in \Omega_s \cap \text{supp}(p) : \Delta_s(\omega; \mathbf{v}) = \Delta_s^{\min}(\mathbf{v})\}$. Using the logit formula,

$$\sigma_\omega^s(\mathbf{v}; \beta) = \frac{\exp\{\beta v_s\}}{\sum_{j \in \omega} \exp\{\beta v_j\}} = \exp\{-\beta \Delta_s(\omega; \mathbf{v})\} \cdot \left(\sum_{j \in \omega} \exp\{-\beta(\max_{i \in \omega} v_i - v_j)\} \right)^{-1},$$

and since $1 \leq \sum_{j \in \omega} \exp\{-\beta(\max_{i \in \omega} v_i - v_j)\} \leq |\omega|$, the parenthesized factor lies in $[|\omega|^{-1}, 1]$ for all β and all $\mathbf{v} \in K$. Hence $\sigma_\omega^s(\mathbf{v}_n; \beta_n) = \exp\{-\beta_n \Delta_s(\omega; \mathbf{v}_n)\} \cdot \Theta_{s,\omega}^{(n)}$ with $0 < \underline{c} \leq \Theta_{s,\omega}^{(n)} \leq \bar{c} < \infty$, for some constants \underline{c}, \bar{c} independent of n . Factoring out $\exp\{-\beta_n \Delta_s^{\min}(\mathbf{v}_n)\}$ from the denominator gives the Laplace factorization of $D_s(\mathbf{v}_n; \beta_n)$ into

$$\sum_{\omega \in \Omega_s \cap \text{supp}(p)} p(\omega) \sigma_\omega^s(\mathbf{v}_n; \beta_n) = \exp\{-\beta_n \Delta_s^{\min}(\mathbf{v}_n)\} \sum_{\omega \in \Omega_s \cap \text{supp}(p)} p(\omega) e^{-\beta_n(\Delta_s(\omega; \mathbf{v}_n) - \Delta_s^{\min}(\mathbf{v}_n))} \Theta_{s,\omega}^{(n)}.$$

Since Ω is finite, either $\Omega_s^*(\mathbf{v}^*) = \Omega_s \cap \text{supp}(p)$, or else $\delta := \min_{\omega \notin \Omega_s^*(\mathbf{v}^*)} (\Delta_s(\omega; \mathbf{v}^*) - \Delta_s^{\min}(\mathbf{v}^*)) > 0$. By Lipschitz continuity of $\Delta_s(\omega; \cdot)$ and $\Delta_s^{\min}(\cdot)$, for all sufficiently large n we have $\Delta_s(\omega; \mathbf{v}_n) - \Delta_s^{\min}(\mathbf{v}_n) \geq \delta/2$, $\forall \omega \notin \Omega_s^*(\mathbf{v}^*)$, and $\Delta_s(\omega; \mathbf{v}_n) - \Delta_s^{\min}(\mathbf{v}_n) \leq \delta/4$, $\forall \omega \in \Omega_s^*(\mathbf{v}^*)$. Since $\Omega_s^*(\mathbf{v}^*) \neq \emptyset$ and $\Theta_{s,\omega}^{(n)} \geq \underline{c} > 0$, the contribution of minimizing menus to the bracketed sum is bounded below uniformly for all large n , whereas the contribution of non-minimizers is $O(e^{-\beta_n \delta/2})$. Hence, the total weight assigned to non-minimizers is exponentially small: $\sum_{\omega \notin \Omega_s^*(\mathbf{v}^*)} w_\omega^{s,n} = O(e^{-\beta_n \eta})$ for some $\eta > 0$. Thus, any cluster point $w^{s,*}$ of the sequence $(w^{s,n})_n$ is supported on $\Omega_s^*(\mathbf{v}^*)$, and along any subsequence along which $w^{s,n} \rightarrow w^{s,*}$ we obtain $v_s^* = \lim_{n \rightarrow \infty} v_{n,s} = \lim_{n \rightarrow \infty} \sum_{\omega} w_\omega^{s,n} \pi_s(\omega) = \sum_{\omega \in \Omega_s^*(\mathbf{v}^*)} w_\omega^{s,*} \pi_s(\omega) \in \text{co}\{\pi_s(\omega) : \omega \in \Omega_s^*(\mathbf{v}^*)\}$. Since this holds for every s , we conclude $\mathbf{v}^* \in g_\infty(\mathbf{v}^*)$, i.e. $\mathbf{v}^* \in \widehat{\mathcal{V}}(\infty)$. Combining Steps 1-2, we obtain $\mathbf{v}^* \in \widehat{\mathcal{V}}(\infty) \cap \mathcal{V}\mathcal{E}$ implying $\mathcal{V}(\infty) \subseteq \mathcal{V}\mathcal{E} \cap \widehat{\mathcal{V}}(\infty)$. Finally, since $\mathcal{V}(\infty) = \limsup_{\beta \rightarrow \infty} \mathcal{V}(\beta)$ is the Kuratowski outer limit of a family of subsets of the compact set K , it is closed in K , and hence compact.

Step 3: Fix $\varepsilon > 0$. If the stated neighborhood inclusion failed, there would exist $\beta_n \uparrow \infty$ and $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ such that $\text{dist}(\mathbf{v}_n, \mathcal{V}(\infty)) \geq \varepsilon$ for all n . By compactness of K , a subsequence converges to some $\mathbf{v}^* \in K$. Thus, $\mathbf{v}^* \in \mathcal{V}(\infty)$, contradicting $\text{dist}(\mathbf{v}_n, \mathcal{V}(\infty)) \geq \varepsilon$ for large n .

(c) Generic regularity and finiteness for a.e. β : Fix a reduced tree structure (S, Ω) and a support set $\Omega^+ \subseteq \Omega$, interpreted as the fixed support of the menu distribution. Let

$\Delta(\Omega^+) := \{p \in \mathbb{R}^\Omega : p(\omega) > 0 \forall \omega \in \Omega^+, p(\omega) = 0 \forall \omega \notin \Omega^+, \sum_{\omega \in \Omega} p(\omega) = 1\}$, and let $\Pi^+ := \mathbb{R}^{\{(s,\omega) : \omega \in \Omega_s \cap \Omega^+\}}$ denote the space of supported class-menu payoff arrays. Define $\tilde{F} : U \times (0, \infty) \times \Delta(\Omega^+) \times \Pi^+ \rightarrow \mathbb{R}^S$, $\tilde{F}(\mathbf{v}, \beta, p, \pi) := \mathbf{v} - g(\mathbf{v}; \beta, p, \pi)$, where $U \subset \mathbb{R}^S$ is an open neighborhood of K . Since all denominators $D_s(\mathbf{v}; \beta, p) = \sum_{\omega \in \Omega_s \cap \Omega^+} p(\omega) \sigma_\omega^s(\mathbf{v}; \beta)$ are strictly positive for every $\beta < \infty$, every $p \in \Delta(\Omega^+)$, and every $s \in S$, the map \tilde{F} is C^∞ . For each $s \in S$, $g_s(\mathbf{v}; \beta, p, \pi) = \sum_{\omega \in \Omega_s \cap \Omega^+} w_\omega^s(\mathbf{v}, \beta, p) \pi_s(\omega)$, with $w_\omega^s(\mathbf{v}, \beta, p) := \frac{p(\omega) \sigma_\omega^s(\mathbf{v}; \beta)}{\sum_{\omega' \in \Omega_s \cap \Omega^+} p(\omega') \sigma_{\omega'}^s(\mathbf{v}; \beta)} > 0$. Hence, $\frac{\partial \tilde{F}_s}{\partial \pi_{s'}(\omega)} = -\mathbf{1}\{s = s'\} w_\omega^s(\mathbf{v}, \beta, p)$.

For each coordinate s , varying any one payoff entry $\pi_s(\omega)$, with $\omega \in \Omega_s \cap \Omega^+$, changes only \tilde{F}_s with nonzero derivative. Thus, $D_\pi \tilde{F}(\mathbf{v}, \beta, p, \pi)$ is surjective at every point. Consequently \tilde{F} is a submersion, and in particular $\tilde{F} \pitchfork \{0\}$. By Thom's parametric transversality theorem, there is a residual (hence, dense) set of parameters $(p, \pi) \in \Delta(\Omega^+) \times \Pi^+$ such that the slice map $\tilde{F}_{p,\pi} : U \times (0, \infty) \rightarrow \mathbb{R}^S$, $(\mathbf{v}, \beta) \mapsto \tilde{F}(\mathbf{v}, \beta, p, \pi)$, is transverse to 0. Moreover, the same conclusion holds for a full-measure set of parameters. Indeed, since \tilde{F} is a submersion, $M := \tilde{F}^{-1}(0)$ is a smooth manifold. Consider the projection $P : M \rightarrow \Delta(\Omega^+) \times \Pi^+$, $P(\mathbf{v}, \beta, p, \pi) = (p, \pi)$. A parameter pair (p, π) is a regular value of P if and only if $\tilde{F}_{p,\pi} \pitchfork 0$. By Sard's theorem, the set of critical values of P has Lebesgue measure zero. Thus, the set of parameter pairs for which $\tilde{F}_{p,\pi} \pitchfork 0$ is both residual and of full Lebesgue measure, after intersecting the residual set from Thom's theorem with the full-measure set from Sard's theorem.

Fix such a (p, π) , and set $Z := \tilde{F}_{p,\pi}^{-1}(0) \subset U \times (0, \infty)$. Then Z is a \mathcal{C}^1 submanifold of dimension 1, and the projection $\text{pr}_\beta : Z \rightarrow (0, \infty)$ is \mathcal{C}^1 . By Sard's theorem, the set of critical values of pr_β has Lebesgue measure zero. For any β which is a regular value of pr_β , every point $(\mathbf{v}, \beta) \in Z$ satisfies $D_{\mathbf{v}} \tilde{F}(\mathbf{v}, \beta, p, \pi)$ invertible, i.e. 0 is a regular value of the map $\mathbf{v} \mapsto \tilde{F}(\mathbf{v}, \beta, p, \pi)$. Hence all SVE at that β are isolated. Indeed, if $D_{\mathbf{v}} \tilde{F}(\mathbf{v}, \beta, p, \pi)$ were singular at some $(\mathbf{v}, \beta) \in Z$, then there would exist a nonzero vertical tangent vector $(u, 0) \in T_{(\mathbf{v}, \beta)} Z$, forcing $d(\text{pr}_\beta) = 0$ at (\mathbf{v}, β) , contrary to the regularity of β . Finally, since $\mathcal{V}(\beta) \subset K$ is compact and each SVE is isolated for a.e. β , $\mathcal{V}(\beta)$ is finite for a.e. β .

(d1-d2) Degree one and generic odd parity: For each s and each \mathbf{v} , the weights $w_\omega^s(\mathbf{v}) := \frac{p(\omega) \sigma_\omega^s(\mathbf{v}; \beta)}{\sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v}; \beta)}$ are strictly positive on at least two menus ω, ω' with $\pi_s(\omega) \neq \pi_s(\omega')$. Hence $g_s(\mathbf{v})$ is a *strict* convex combination of $\{\pi_s(\omega) : \omega \in \Omega_s\}$, so $m_s < g_s(\mathbf{v}) < M_s$. Thus $g(K) \subset \text{int } K$. By convexity of K , for $\mathbf{v} \in \partial K$, we have $h(\mathbf{v}) = \mathbf{v} - g(\mathbf{v}) \neq 0$ and $h(\mathbf{v}) \cdot n(\mathbf{v}) > 0$ with $n(\mathbf{v})$ the outward unit normal (by a supporting-hyperplane argument). Fix $\mathbf{c} \in \text{int } K$ and consider the homotopy $H_t(\mathbf{v}) := \mathbf{v} - ((1-t)g(\mathbf{v}) + t\mathbf{c})$, $t \in [0, 1]$. Because $g(K) \subset \text{int } K$ and $\mathbf{c} \in \text{int } K$, $(1-t)g(\mathbf{v}) + t\mathbf{c} \in \text{int } K$ for all $\mathbf{v} \in \partial K$, hence $H_t(\mathbf{v}) \neq 0$ on ∂K . By homotopy invariance of degree, $\deg(h, K, 0) = \deg(H_0, K, 0) = \deg(H_1, K, 0)$. But

$H_1(\mathbf{v}) = \mathbf{v} - \mathbf{c}$ has a unique zero $\mathbf{v} = \mathbf{c}$ with index $+1$, so $\deg(h, K, 0) = 1$. For (d2), degree equals the sum of local indices of zeros of h in $\text{int } K$; thus the sum of indices is $+1$. For a.e. β , all zeros are non-degenerate, each has index ± 1 , hence there are an odd number of SVE.

(d3) Essential component: Let $\mathcal{G} := \{(\mathbf{v}, \theta) \in K \times \Theta_0 : h(\mathbf{v}; \theta) = 0\}$, and $\xi : \mathcal{G} \rightarrow \Theta_0$, $\xi(\mathbf{v}, \theta) = \theta$. Since h is continuous, \mathcal{G} is closed in $K \times \Theta_0$. As K is compact, for every compact $C \subset \Theta_0$ we have $\xi^{-1}(C) = \mathcal{G} \cap (K \times C)$ compact; hence ξ is proper. By (d1), for every $\theta \in \Theta_0$, $\deg(h(\cdot; \theta), \mathcal{K}, 0) = 1$, where $\mathcal{K} := \text{int } K$, so $h(\cdot; \theta)$ has at least one zero in \mathcal{K} and ξ is surjective. Fix $\theta_a, \theta_b \in \Theta_0$. Since Θ_0 is path-connected, there exists a continuous path $\gamma : [a, b] \rightarrow \Theta_0$ with $\gamma(a) = \theta_a$ and $\gamma(b) = \theta_b$. Define $F : K \times [a, b] \rightarrow \mathbb{R}^{|\mathcal{S}|}$, $F(\mathbf{v}, t) := h(\mathbf{v}; \gamma(t))$. By (d1), $g(K; \theta) \subset \mathcal{K}$ for every $\theta \in \Theta_0$, hence $h(\mathbf{v}; \theta) = \mathbf{v} - g(\mathbf{v}; \theta) \neq 0$, $\forall \mathbf{v} \in \partial K = \partial \mathcal{K}$, $\forall \theta \in \Theta_0$, equivalently, $0 \notin F(\partial \mathcal{K} \times [a, b])$. Thus, the Brouwer $\deg(F(\cdot, t), \mathcal{K}, 0)$ is well-defined and constant in $t \in [a, b]$ (by extended homotopy invariance), and by (d1) equals 1 for all t . Let $S_\gamma := \{(\mathbf{v}, t) \in \mathcal{K} \times [a, b] : F(\mathbf{v}, t) = 0\}$, and $S_{\gamma, t} := \{\mathbf{v} \in \mathcal{K} : F(\mathbf{v}, t) = 0\}$.

By the Leray-Schauder continuation theorem (Dinca and Mawhin, 2021, Thm. 2.1.2), there exists a connected component C_γ of S_γ meeting both $S_{\gamma, a} \times \{a\}$ and $S_{\gamma, b} \times \{b\}$. Mapping C_γ into \mathcal{G} via $\phi(\mathbf{v}, t) := (\mathbf{v}, \gamma(t))$ yields a connected subset $\phi(C_\gamma) \subset \mathcal{G}$ meeting both endpoint fibers $\xi^{-1}(\theta_a)$ and $\xi^{-1}(\theta_b)$. Since C_γ is connected and meets $\mathcal{K} \times \{a\}$ and $\mathcal{K} \times \{b\}$, its projection satisfies $\pi_t(C_\gamma) = [a, b]$, where $\pi_t(\mathbf{v}, t) := t$. Noting that $\xi \circ \phi = \gamma \circ \pi_t$, we obtain $\xi(\phi(C_\gamma)) = \gamma([a, b])$. Hence if Γ_γ denotes the connected component of \mathcal{G} containing $\phi(C_\gamma)$, then $\gamma([a, b]) \subset \xi(\Gamma_\gamma)$. In particular, when $\Theta_0 = [a, b]$ and $\gamma = \text{id}$, the corresponding component is *essential* with surjective projection, $\Gamma_{\text{ess}} := \Gamma_{\text{id}} \subset \mathcal{G}$ satisfies $\xi(\Gamma_{\text{ess}}) = [a, b]$. Finally, at any non-degenerate $(\mathbf{v}^*, \theta^*) \in \mathcal{G}$, the implicit function theorem yields a locally unique real-analytic branch of the solution set as a graph $\theta \mapsto \mathbf{v}(\theta)$ in a neighborhood of θ^* .

(d4) Structure of SVE graph: Fix $(\mathcal{S}, \Omega, p, \pi)$ and consider the map $F(\mathbf{v}, \beta) = \mathbf{v} - g(\mathbf{v}; \beta)$ on $K \times [0, \infty)$. Since g is obtained from finitely many algebraic operations and exponentials, F is definable in the o-minimal structure \mathbb{R}_{exp} , and so is its zero set $\mathcal{G} := \{(\mathbf{v}, \beta) \in K \times [0, \infty) : F(\mathbf{v}, \beta) = 0\}$.⁴⁸ In particular, by cell decomposition (Dries, 1998), \mathcal{G} has only finitely many connected components, each of which is definable and path-connected. For generic (p, π) , part (c) implies that the generic fiber of the projection $\xi : \mathcal{G} \rightarrow [0, \infty)$ is finite; since $\xi(\mathcal{G}) = [0, \infty)$, the fiber-dimension theorem yields $\dim(\mathcal{G}) = 1$. Consequently, each positive-dimensional connected component of \mathcal{G} is a definably path-connected one-dimensional set, and admits a finite decomposition into definable \mathcal{C}^1 arcs and points.

⁴⁸van den Dries et al. (1994) show that the field of real numbers with exponentiation \mathbb{R}_{exp} is o-minimal. In o-minimal expansions of the real ordered field, the definable subsets of \mathbb{R}^n share many of the nice structural properties of semi-algebraic sets. For e.g., definable subsets have only finitely many connected components, definable sets can be stratified and triangulated, and continuous definable maps are piecewise trivial.

(d5) Principal component: At $\beta = 0$, the choice probabilities are uniform on each menu, so $g(\cdot; 0)$ is constant and there is a unique SVE \mathbf{v}_0 . Moreover, $D_{\mathbf{v}}h(\mathbf{v}_0; 0) = I$, so $(\mathbf{v}_0, 0)$ is a regular point of \mathcal{G} . By continuity, there exists $\beta^* > 0$ such that $g(\cdot; \beta)$ is a strict contraction on K for all $\beta \in [0, \beta^*]$. Hence, on $[0, \beta^*]$, the SVE is unique and regular, and the component of \mathcal{G} containing $(\mathbf{v}_0, 0)$ is a single real-analytic graph. Let Γ_{prin} denote the connected component of the global SVE graph $\mathcal{G} := \{(\mathbf{v}, \beta) \in K \times [0, \infty) : \mathbf{v} = g(\mathbf{v}; \beta)\}$ that contains $(\mathbf{v}_0, 0)$. We claim that $\xi(\Gamma_{\text{prin}}) = [0, \infty)$. Fix any $\bar{\beta} < \infty$ and set $\mathcal{G}_{\bar{\beta}} := \mathcal{G} \cap (K \times [0, \bar{\beta}])$. By the continuation result from (d3), $\mathcal{G}_{\bar{\beta}}$ contains a connected component $\Gamma_{\bar{\beta}}$ whose projection onto $[0, \bar{\beta}]$ is surjective. In particular, $\Gamma_{\bar{\beta}}$ intersects the fiber $K \times \{0\}$. Since $\mathcal{V}(0) = \{\mathbf{v}_0\}$, this intersection must contain $(\mathbf{v}_0, 0)$. Therefore $\Gamma_{\bar{\beta}} \subseteq \Gamma_{\text{prin}}$, because Γ_{prin} is the connected component of the global graph containing $(\mathbf{v}_0, 0)$. It follows that $[0, \bar{\beta}] \subseteq \xi(\Gamma_{\text{prin}})$. Since $\bar{\beta} < \infty$ was arbitrary, we obtain $\xi(\Gamma_{\text{prin}}) = [0, \infty)$. If another connected component Γ of \mathcal{G} had $\xi(\Gamma) = [0, \infty)$, then it would intersect the fiber $K \times \{0\}$. Since the SVE at $\beta = 0$ is unique, this would force $(\mathbf{v}_0, 0) \in \Gamma$, and hence $\Gamma = \Gamma_{\text{prin}}$. Thus Γ_{prin} is the unique connected component of \mathcal{G} whose projection onto the β -axis is surjective.

(d6) Generic smoothness and Morse projection: As in (c), fix a support set $\Omega^+ \subseteq \Omega$ and regard the parameter pair (p, π) as ranging over the finite-dimensional parameter space $\Theta := \Delta(\Omega^+) \times \Pi^+$. Define $\tilde{F}(\mathbf{v}, \beta, p, \pi) := \mathbf{v} - g(\mathbf{v}; \beta, p, \pi)$, which is real-analytic in $(\mathbf{v}, \beta, p, \pi)$. Let $\mathcal{M} := \tilde{F}^{-1}(0) \subset U \times (0, \infty) \times \Theta$, $P : \mathcal{M} \rightarrow \Theta$, $(\mathbf{v}, \beta, p, \pi) \mapsto (p, \pi)$, and $\Xi : \mathcal{M} \rightarrow \mathbb{R}$, $(\mathbf{v}, \beta, p, \pi) \mapsto \beta$. By part (c), there exists a comeager full-measure subset $\Theta_{\text{reg}} \subset \Theta$ such that, for every $\theta \in \Theta_{\text{reg}}$, the fiber $\mathcal{G}_{\theta} = P^{-1}(\theta) = \{(\mathbf{v}, \beta) \in K \times (0, \infty) : \tilde{F}(\mathbf{v}, \beta, \theta) = 0\}$ is a one-dimensional embedded real-analytic submanifold. Restricting P to $P^{-1}(\Theta_{\text{reg}})$, it follows that $P : P^{-1}(\Theta_{\text{reg}}) \rightarrow \Theta_{\text{reg}}$ is a submersion. Applying the parametric transversality theorem to this restricted submersion yields a residual subset $\Theta_{\text{Morse}} \subset \Theta_{\text{reg}}$ such that, for every $\theta \in \Theta_{\text{Morse}}$, the projection $\xi_{\theta} : \mathcal{G}_{\theta} \rightarrow (0, \infty)$, $(\mathbf{v}, \beta) \mapsto \beta$, is a Morse function (see [Guillemin and Pollack \(2010\)](#), p. 68 and [Online Appendix \(Sec. D\)](#) for details). Hence, for generic decision trees on the fixed support stratum $\Theta_{\text{gen}} := \Theta_{\text{reg}} \cap \Theta_{\text{Morse}}$, every critical point of ξ is a nondegenerate fold. Since \mathcal{G}_{θ} is definable and one-dimensional, the critical set of ξ is a definable 0-dimensional subset of \mathcal{G}_{θ} , hence finite; therefore the set of critical values of ξ is finite. In particular, every connected component of \mathcal{G}_{θ} is a smooth embedded curve, and over any bounded interval in β it is a finite union of real-analytic graphs separated by finitely many fold points. In particular, the principal branch $\Gamma_{\text{prin}}(p, \pi)$ is such a curve.

(e1) Finite limit set of \mathcal{G} : For each $\beta \in [0, \infty)$, the SVE set $\mathcal{V}(\beta)$ is nonempty, so the graph \mathcal{G} is unbounded in the β -direction. Pick any sequence $\beta_n \uparrow \infty$ and choose $\mathbf{v}_n \in \mathcal{V}(\beta_n)$, so $(\mathbf{v}_n, \beta_n) \in \mathcal{G}$. Because K is compact, there exists a subsequence with $\mathbf{v}_n \rightarrow \mathbf{v}^* \in K$.

By part (b), any such limit \mathbf{v}^* satisfies $\mathbf{v}^* \in \mathcal{VE} \cap \widehat{\mathcal{V}}(\infty)$. Hence $\mathcal{V}(\infty)$ is nonempty and $\mathcal{V}(\infty) \subseteq \mathcal{VE} \cap \widehat{\mathcal{V}}(\infty)$. Set $t := 1/(1+\beta) \in (0, 1]$ and define $\widehat{\mathcal{G}} := \{(\mathbf{v}, t) \in K \times (0, 1] : (\mathbf{v}, \beta) \in \mathcal{G}, t = 1/(1+\beta)\}$. The map $(\mathbf{v}, \beta) \mapsto (\mathbf{v}, 1/(1+\beta))$ is a definable homeomorphism between \mathcal{G} and $\widehat{\mathcal{G}}$. In generic trees, since \mathcal{G} is one-dimensional and definable, so is $\widehat{\mathcal{G}}$. Let $\overline{\widehat{\mathcal{G}}}$ denote the closure of $\widehat{\mathcal{G}}$ in the compact set $K \times [0, 1]$. In an o-minimal structure, a one-dimensional definable subset of a compact set has only finitely many connected components, each a finite union of C^1 arcs and points and each component has finitely many boundary points.

In particular, the intersection with the boundary $\mathcal{L} := \{(\mathbf{v}, 0) \in K \times \{0\} : (\mathbf{v}, 0) \in \overline{\widehat{\mathcal{G}}}\}$ is a definable set of dimension 0, hence a finite set of points.⁴⁹ By construction, $\mathbf{v}^* \in \mathcal{V}(\infty)$ if and only if $(\mathbf{v}^*, 0) \in \mathcal{L}$. Thus the global limit set of SVE as $\beta \uparrow \infty$, $\mathcal{V}(\infty)$, is a nonempty finite subset of \mathcal{VE} , even if \mathcal{VE} may be infinite in some trees (see the example in 3.3). As a corollary, defining $\mathcal{V}_{\text{prin}}(\infty) := \{\mathbf{v}^* \in K : \exists \beta_n \uparrow \infty, (\mathbf{v}_n, \beta_n) \in \Gamma_{\text{prin}}, \mathbf{v}_n \rightarrow \mathbf{v}^*\}$ as the limit set of the principal SVE branch, we get $\mathcal{V}_{\text{prin}}(\infty) \subseteq \mathcal{V}(\infty)$.

(e2) Unique LSVE along unbounded SVE tail branches: Fix a generic parameter array $\theta \in \Theta_{\text{gen}}$ so that the conclusions of (d6) hold: $\mathcal{G} = \mathcal{G}_\pi$ is a one-dimensional real-analytic embedded submanifold of $K \times [0, \infty)$ and the projection $\xi : \mathcal{G} \rightarrow [0, \infty)$ is a Morse function. Let \mathcal{C} be a connected component of \mathcal{G} such that $\xi(\mathcal{C})$ is unbounded above. As \mathcal{C} is connected and ξ is continuous and proper, $\xi(\mathcal{C})$ is a connected subset of $[0, \infty)$ and therefore an interval. Since the set of critical values of ξ is finite, there exists $B < \infty$ such that ξ has no critical values in (B, ∞) and $(B, \infty) \subseteq \xi(\mathcal{C})$. Let Γ be any connected component of $\mathcal{C} \cap (K \times (B, \infty))$. Then $\xi|_\Gamma : \Gamma \rightarrow (B, \infty)$ is a proper local diffeomorphism onto its image. Hence it is a covering map onto its image; since Γ is connected, its image is a nonempty connected open-and-closed subset of the connected interval (B, ∞) , so necessarily $\xi(\Gamma) = (B, \infty)$. Because (B, ∞) is simply connected, the covering $\xi|_\Gamma$ is in fact a global real-analytic diffeomorphism.

Therefore there exists a unique definable real-analytic map $\varphi_\Gamma : (B, \infty) \rightarrow K$ such that $\Gamma = \{(\varphi_\Gamma(\beta), \beta) : \beta > B\}$. Since Γ is definable, so is φ_Γ . Each coordinate of φ_Γ is bounded (as K is compact) and definable; by the o-minimal monotonicity theorem, each coordinate is eventually monotone (or constant), hence convergent. Thus the limit $\mathbf{v}_\Gamma^* := \lim_{\beta \rightarrow \infty} \varphi_\Gamma(\beta)$ exists and is unique. For any sequence $\beta_n \uparrow \infty$, setting $\mathbf{v}_n := \varphi_\Gamma(\beta_n)$ yields $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ and $\mathbf{v}_n \rightarrow \mathbf{v}_\Gamma^*$, so by part (b) $\mathbf{v}_\Gamma^* \in \mathcal{V}(\infty) \subseteq \mathcal{VE}$. Therefore every unbounded tail branch of every

⁴⁹The set $\widehat{\mathcal{G}}$ is a one-dimensional definable subset of $K \times (0, 1]$, so its closure $\overline{\widehat{\mathcal{G}}}$ is definable and $\dim(\overline{\widehat{\mathcal{G}}}) = \dim(\widehat{\mathcal{G}}) = 1$. Consider the frontier $\text{Fr}(\widehat{\mathcal{G}}) := \overline{\widehat{\mathcal{G}}} \setminus \widehat{\mathcal{G}}$. It is definable and, by o-minimality, satisfies $\dim(\text{Fr}(\widehat{\mathcal{G}})) < \dim(\widehat{\mathcal{G}}) = 1$. Thus $\dim(\text{Fr}(\widehat{\mathcal{G}})) \leq 0$. By construction, $\mathcal{L} = \{(\mathbf{v}, 0) \in K \times \{0\} : (\mathbf{v}, 0) \in \overline{\widehat{\mathcal{G}}}\} \subseteq \text{Fr}(\widehat{\mathcal{G}})$, so \mathcal{L} is a definable set of dimension at most 0. Since \mathcal{L} lies in the compact set $K \times \{0\}$, $\dim(\mathcal{L}) = 0$ implies that \mathcal{L} is a finite union of points. Hence \mathcal{L} is finite.

connected component \mathcal{C} with unbounded projection converges to a unique point of $\mathcal{V}(\infty)$.

By part (d1), every SVE lies in $\text{int } K$. Hence $\Gamma_{\text{prin}} \subset \text{int } K \times [0, \infty)$ is a connected one-dimensional embedded manifold with boundary, and any boundary point of Γ_{prin} must lie in the fiber $\{\beta = 0\}$. Since $\Gamma_{\text{prin}} \cap \xi^{-1}(0) = \{(\mathbf{v}_0, 0)\}$, the only boundary point of Γ_{prin} is $(\mathbf{v}_0, 0)$. By the classification of connected one-dimensional manifolds with boundary, Γ_{prin} is homeomorphic to $[0, \infty)$. Since $\xi(\Gamma_{\text{prin}}) = [0, \infty)$ and ξ is proper on Γ_{prin} , the projection tends to infinity along the unique non-compact end of Γ_{prin} . Moreover, by (d6), the set of critical values of $\xi|_{\Gamma_{\text{prin}}}$ is finite. Therefore there exists $B < \infty$, chosen above all critical values and above all earlier returns of the curve, such that $\Gamma_{\text{prin}} \cap (K \times (B, \infty))$ is connected and contains no critical point of ξ . For such a B , the restriction of ξ to this tail is a proper local diffeomorphism onto (B, ∞) . Since (B, ∞) is simply connected, this restriction is a global real-analytic diffeomorphism. Thus the principal component is eventually a single real-analytic graph over (B, ∞) and therefore converges to a unique limit in $\mathcal{V}(\infty)$. \square

B.2 Theorem 2

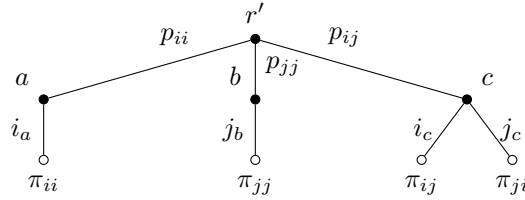


Figure 6: Reduced Decision Tree \mathcal{T}'_2 with Two Similarity Classes

Proof. The alternatives are partitioned into two similarity classes $i = \{i_a, i_c\}$, $j = \{j_b, j_c\}$. By Lem. B.9, it is without loss to work with the relative valuation $x := v_i - v_j$.⁵⁰ Denoting $\sigma(x) = \exp(\beta x)/(1 + \exp(\beta x))$ (logit choice probability of class i at the unique binary menu),

$$g_i(x) = \frac{p_{ii}\pi_{ii} + p_{ij}\sigma(x)\pi_{ij}}{p_{ii} + p_{ij}\sigma(x)}, \quad g_j(x) = \frac{p_{jj}\pi_{jj} + p_{ij}(1 - \sigma(x))\pi_{ji}}{p_{jj} + p_{ij}(1 - \sigma(x))}.$$

The relative valuation satisfies the scalar ODE $\dot{x} = f_\beta(x) := (g_i(x) - g_j(x)) - x$, and the mean valuation coordinate satisfies $\dot{y} = \frac{1}{2}(g_i(x) + g_j(x)) - y$. In particular, once $x(t)$ converges, the forcing term $\frac{1}{2}(g_i(x(t)) + g_j(x(t)))$ converges and hence $y(t)$ converges as well, yielding convergence in the full space. Since there are finitely many finite payoffs there exist M_1, M_2 with $-M_1 \leq g_i(x) - g_j(x) \leq M_2$, hence outside $[-M_1, M_2]$ the flow points inward; thus all forward trajectories enter and remain in a compact interval. In one-dimensional smooth autonomous dynamical system, there are no periodic orbits; flow is monotone and bounded

⁵⁰Writing $y := \frac{1}{2}(v_i + v_j)$, the change of variables $(v_i, v_j) \leftrightarrow (x, y)$ is invertible and triangularizes the system.

within $[-M_1, M_2]$. Thus, ω -limit sets are equilibria x^* with $f_\beta(x^*) = 0$. By the analytic identity theorem, the set of equilibria of the one-dimensional (relative) CQL system is finite.

By Theorem 1, for sufficiently large β every SVE lies near a VE. With two classes there are at most three VE: two strict pure (if they exist) and possibly one mixed (indifference). Thus, for $\beta \geq \hat{\beta}$ the equilibria of f_β lie in small neighborhoods of these VE. Differentiate f_β :

$$f'_\beta(x) = \sigma'(x) \cdot p_{ij} \left(\frac{p_{ii}(\pi_{ij} - \pi_{ii})}{(p_{ii} + p_{ij}\sigma(x))^2} + \frac{p_{jj}(\pi_{ji} - \pi_{jj})}{(p_{jj} + p_{ij}(1 - \sigma(x)))^2} \right) - 1, \quad \sigma'(x) = \beta \sigma(x)(1 - \sigma(x)).$$

If a strict pure VE exists at some $x^\infty \neq 0$, then by Thm. 1 the corresponding SVE $x^*(\beta)$ satisfies $x^*(\beta) \rightarrow x^\infty$ as $\beta \rightarrow \infty$; hence $|x^*(\beta)| \geq \eta > 0$ for all β large enough, which implies $\sigma'(x^*(\beta)) = \beta \sigma(x^*(\beta))(1 - \sigma(x^*(\beta))) \rightarrow 0$. Therefore $f'_\beta(x^*(\beta)) \rightarrow -1$, and for all $\beta \geq \hat{\beta}$ sufficiently large, $f'_\beta(x^*(\beta)) < 0$, so the corresponding SVE is locally asymptotically stable.

If both strict pure VE exist (one with $x^\infty > 0$, one with $x^\infty < 0$), then for all $\beta \geq \hat{\beta}$ there are two corresponding locally asymptotically stable SVE $x_i^*(\beta) > 0$ and $x_j^*(\beta) < 0$ as above. Moreover, by Thm. 1 there is a unique additional equilibrium $x_m^*(\beta)$ in a small neighborhood of the mixed VE. In a one-dimensional smooth autonomous ODE, the sign of f_β must alternate across successive simple zeros; hence the middle equilibrium $x_m^*(\beta)$ is unstable (equivalently $f'_\beta(x_m^*(\beta)) > 0$). The basins of attraction of $x_j^*(\beta)$ and $x_i^*(\beta)$ are therefore separated by $x_m^*(\beta)$, and every trajectory converges to one of the equilibria.

If no strict pure VE exists, there is a unique mixed VE at $x^\infty = 0$; for $\beta \geq \hat{\beta}$ there is a unique nearby SVE $x^*(\beta)$ with $x^*(\beta) \rightarrow 0$ as $\beta \rightarrow \infty$. Since f_β is continuous and points inward outside a compact interval containing the unique root near the origin, it must be that $f_\beta(x) > 0$ for $x < x^*$ and $f_\beta(x) < 0$ for $x > x^*$. Thus $f'_\beta(x^*) < 0$ and the SVE is *locally* asymptotically stable. If the SVE is unique, define $V(x) = \frac{1}{2}(x - x^*)^2$. Then $\dot{V} = (x - x^*)f_\beta(x) < 0$ for $x \neq x^*$ by the sign pattern shown above, so x^* is *globally* asymptotically stable. In all cases, every trajectory monotonically converges to an equilibrium. \square

B.3 Theorem 3

Proof. Let \mathbf{v}^∞ be a *strict pure* LSVE, i.e. there exist $\beta_n \uparrow \infty$ and $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ with $\mathbf{v}_n \rightarrow \mathbf{v}^\infty$, and for every $\omega \in \Omega$ the restriction of \mathbf{v}^∞ to ω has a unique maximizer $s^*(\omega) := \arg \max_{s \in \omega} v_s^\infty$, so the greedy policy induced by \mathbf{v}^∞ is pure. For a never-winning class i , the closest-loss winner is unique as $\arg \min_{\omega \in \Omega_i \cap \text{supp}(p)} (v_{s^*(\omega)}^\infty - v_i^\infty) = \arg \min_{\omega \in \Omega_i \cap \text{supp}(p)} v_{s^*(\omega)}^\infty$, and the winner classes are strictly ordered. Denote this closest-loss winner of class i by t_i . Indeed, this is a generic property when the co-occurrence graph G is connected as

we've assumed.⁵¹ For a class j that is maximal in some menu, i.e. $j = s^*(\omega)$ for some $\omega \in \Omega_j \cap \text{supp}(p)$, we define $t_j = j$. Fix such a strict pure LSVE $\mathbf{v}^\infty \in \mathcal{V}(\infty)$. For $\omega \in \Omega$ and $s \in \omega$, define the loss gap $\Delta_s(\omega) := v_{s^*(\omega)}^\infty - v_s^\infty \geq 0$, such that $\Delta_s(\omega) = 0 \iff s = s^*(\omega)$. For $s \in \mathcal{S}$, recall the following useful expressions:

$$D_s(\mathbf{v}) := \sum_{\omega \in \Omega_s} p(\omega) \sigma_\omega^s(\mathbf{v}), \quad w_\omega^s(\mathbf{v}) := \frac{p(\omega) \sigma_\omega^s(\mathbf{v})}{D_s(\mathbf{v})} \in \Delta(\Omega_s), \quad g_s(\mathbf{v}) = \sum_{\omega \in \Omega_s} w_\omega^s(\mathbf{v}) \pi_s(\omega).$$

For each class i , define $\Omega_i^* := \{\omega \in \Omega_i \cap \text{supp}(p) : s^*(\omega) = t_i\}$, and $\Delta_i := v_{t_i}^\infty - v_i^\infty$. The exact high-sensitivity limit map associated with this strict pure policy is

$$\bar{g}_i(\mathbf{v}^\infty) := \frac{\sum_{\omega \in \Omega_i^*} p(\omega) \pi_i(\omega)}{\sum_{\omega \in \Omega_i^*} p(\omega)}, \quad i \in \mathcal{S}.$$

We first identify what strict pure LSVE means as a refinement of strict VE. Since \mathbf{v}^∞ is an LSVE, there exist $\beta_n \uparrow \infty$ and $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ such that $\mathbf{v}_n \rightarrow \mathbf{v}^\infty$. By the finite-sum Laplace principle used in the proof of Theorem 1(b), for each class i the conditional menu weights concentrate on Ω_i^* , and because generically all menus in Ω_i^* have the same winner class t_i , the asymptotic weights inside Ω_i^* are pinned down by p :

$$w_\omega^i(\mathbf{v}_n; \beta_n) \longrightarrow \bar{w}_\omega^i := \frac{p(\omega)}{\sum_{\omega' \in \Omega_i^*} p(\omega')} \mathbf{1}\{\omega \in \Omega_i^*\}.$$

Passing to the limit in $v_{n,i} = g_i(\mathbf{v}_n; \beta_n) = \sum_{\omega \in \Omega_i} w_\omega^i(\mathbf{v}_n; \beta_n) \pi_i(\omega)$ gives $v_i^\infty = \bar{g}_i(\mathbf{v}^\infty)$, $\forall i \in \mathcal{S}$. Hence \mathbf{v}^∞ is a strict VE satisfying the closest-loss consistency condition $\mathbf{v}^\infty = \bar{g}(\mathbf{v}^\infty)$.

Lemma B.1. *Fix $s \in \mathcal{S}$ and a sequence $\mathbf{v}^{(\beta)} \rightarrow \mathbf{v}^\infty$ as $\beta \rightarrow \infty$. For each $\omega \in \Omega_s = \{\omega \in \Omega : s \in \omega\}$, define the associated loss gap $\Delta_s(\omega) = v_{s^*(\omega)}^\infty - v_s^\infty \geq 0$, and let $\Delta_s := \min_{\omega \in \Omega_s \cap \text{supp}(p)} \Delta_s(\omega)$, and $\Omega_s^* := \{\omega \in \Omega_s : \Delta_s(\omega) = \Delta_s\}$ be the set of closest-loss menus. For sufficiently large β , there exist constants $0 < \underline{c} \leq \bar{c} < \infty$ (independent of β) s.t.,*

- (a) $\sigma_\omega^s(\mathbf{v}^{(\beta)}) = \exp(-\beta \Delta_s(\omega)) \cdot \Theta_\omega^{(\beta)}, \quad \underline{c} \leq \Theta_\omega^{(\beta)} \leq \bar{c},$
- (b) $D_s(\mathbf{v}^{(\beta)}) = \exp(-\beta \Delta_s) (C_s^{(\beta)} + o(1)), \quad \underline{c} \leq C_s^{(\beta)} \leq \bar{c},$
- (c) $\sum_{\omega \notin \Omega_s^*} w_\omega^s(\mathbf{v}^{(\beta)}) = O(\exp(-\beta \eta)) \quad \text{for some } \eta > 0.$

In particular, if $\Omega_s^ = \{\bar{\omega}\}$ is a singleton, then $w_{\bar{\omega}}^s(\mathbf{v}^{(\beta)}) \rightarrow 1$ and $g_s(\mathbf{v}^{(\beta)}) \rightarrow \pi_s(\bar{\omega})$.*

⁵¹Given strict choices in every menu, a strict total order among winners holds generically for a connected G .

Proof. Identical to Step 1 of Thm. 1(b) (see the “high-sensitivity limit” part of B.1). \square

By Lemma B.1, there exist $\eta_i > 0$ and $C_i < \infty$ such that, as $\beta \rightarrow \infty$, $\sum_{\omega \notin \Omega_i^*} w_\omega^i(\mathbf{v}^\infty; \beta) = O(e^{-\beta\eta_i})$, and $w_\omega^i(\mathbf{v}^\infty; \beta) = \frac{p(\omega)}{\sum_{\omega' \in \Omega_i^*} p(\omega')} + O(e^{-\beta\eta_i})$ for $\omega \in \Omega_i^*$. Hence,

$$g_i(\mathbf{v}^\infty; \beta) = \sum_{\omega \in \Omega_i^*} \frac{p(\omega)}{\sum_{\omega' \in \Omega_i^*} p(\omega')} \pi_i(\omega) + O(e^{-\beta\eta_i}) = \bar{g}_i(\mathbf{v}^\infty) + O(e^{-\beta\eta_i}) = v_i^\infty + O(e^{-\beta\eta_i}).$$

Taking $\eta := \min_i \eta_i$, we obtain $\|g(\mathbf{v}^\infty; \beta) - \mathbf{v}^\infty\| \leq C e^{-\beta\eta}$ for all sufficiently large β .

Differentiate $g_i = N_i/D_i$ with $N_i(\mathbf{v}) = \sum_{\omega \in \Omega_i} p(\omega) \sigma_\omega^i(\mathbf{v}) \pi_i(\omega)$ and $D_i(\mathbf{v}) = \sum_{\omega \in \Omega_i} p(\omega) \sigma_\omega^i(\mathbf{v})$. Using $\partial \sigma_\omega^i / \partial v_j = \beta \sigma_\omega^i(\mathbf{v}) (\mathbf{1}\{i = j\} - \sigma_\omega^j(\mathbf{v}))$ and rearranging yields the centered-weights:

$$\frac{\partial g_i}{\partial v_j}(\mathbf{v}) = \beta \sum_{\omega \in \Omega_i} w_\omega^i(\mathbf{v}) (\mathbf{1}\{i = j\} - \sigma_\omega^j(\mathbf{v})) (\pi_i(\omega) - g_i(\mathbf{v})). \quad (6)$$

In particular, the Jacobian of the CQL vector field $F(\mathbf{v}) = g(\mathbf{v}) - \mathbf{v}$ has entries

$$J_{ii}(\mathbf{v}) = \beta \sum_{\omega \in \Omega_i} w_\omega^i(\mathbf{v}) (1 - \sigma_\omega^i(\mathbf{v})) (\pi_i(\omega) - g_i(\mathbf{v})) - 1, \quad (7)$$

$$J_{ik}(\mathbf{v}) = -\beta \sum_{\omega \in \Omega_i \cap \Omega_k} w_\omega^i(\mathbf{v}) \sigma_\omega^k(\mathbf{v}) (\pi_i(\omega) - g_i(\mathbf{v})), \quad k \neq i. \quad (8)$$

Lemma B.2. *In a strict pure LSVE, the following strict gap condition holds in the limit:*

$$\exists m > 0 \text{ s.t. } \forall \omega \in \Omega \text{ with } p(\omega) > 0, \forall j \in \omega \setminus \{s^*(\omega)\}: v_{s^*(\omega)}^\infty - v_j^\infty \geq m. \quad (9)$$

Thus, along any SVE sequence $\mathbf{v}^{(\beta)} \rightarrow \mathbf{v}^\infty$, there exist constants $C < \infty$ and $\eta > 0$, independent of β , such that $\|Dg(\mathbf{v}^{(\beta)})\| \leq C \beta e^{-\beta\eta} \rightarrow 0$ ($\beta \rightarrow \infty$).

Proof. Since Ω is finite and $s^*(\omega)$ is a strict maximizer for every ω with $p(\omega) > 0$, the set $\{v_{s^*(\omega)}^\infty - v_j^\infty : p(\omega) > 0, j \in \omega \setminus \{s^*(\omega)\}\}$ is a finite set of strictly positive numbers; hence its minimum m exists and is strictly positive. Fix a sequence of SVE $\mathbf{v}^{(\beta)} \rightarrow \mathbf{v}^\infty$ and choose β large enough that the ranking in each menu is the same as at the strict pure limit, with a uniform gap. Since $\Omega \cap \text{supp}(p)$ is finite and \mathbf{v}^∞ is strict pure, the uniform gap $m'' := \min_{\substack{\omega \in \Omega \cap \text{supp}(p) \\ j \in \omega \setminus \{s^*(\omega)\}}} (v_{s^*(\omega)}^\infty - v_j^\infty) > 0$ is well-defined, where $s^*(\omega)$ denotes the unique maximizer of \mathbf{v}^∞ on ω . Since $\mathbf{v}^{(\beta)} \rightarrow \mathbf{v}^\infty$ as $\beta \rightarrow \infty$, choose β_0 such that $\|\mathbf{v}^{(\beta)} - \mathbf{v}^\infty\|_\infty \leq m''/4$ for all $\beta \geq \beta_0$. Then for every $\omega \in \Omega \cap \text{supp}(p)$ and every

$j \in \omega \setminus \{s^*(\omega)\}$, $v_{s^*(\omega)}^{(\beta)} - v_j^{(\beta)} \geq (v_{s^*(\omega)}^\infty - v_j^\infty) - 2\|\mathbf{v}^{(\beta)} - \mathbf{v}^\infty\|_\infty \geq m''/2$. Hence all strict-maximizer gaps at \mathbf{v}^∞ persist uniformly for all $\beta \geq \beta_0$. We will show that for each pair (i, j) there is a constant C_{ij} and $\eta_{ij} > 0$ such that $|\partial g_i / \partial v_j(\mathbf{v}^{(\beta)})| \leq C_{ij} \beta e^{-\beta \eta_{ij}}$ for all sufficiently large β . Taking $\eta := \min_{i,j} \eta_{ij} > 0$ and $C := \max_{i,j} C_{ij}$ yields the desired bound.

Step 1: Bounds on softmax derivatives. Fix $\beta \geq \beta_0$ and ω with $p(\omega) > 0$. Let $t := s^*(\omega)$ denote the unique maximizer in ω . By the uniform gap condition, for all $j \in \omega \setminus \{t\}$, $v_t^{(\beta)} - v_j^{(\beta)} \geq m''/2 := m'$. Write

$$\sigma_\omega^j(\mathbf{v}) = \frac{\exp(\beta v_j)}{\sum_{\ell \in \omega} \exp(\beta v_\ell)} = \frac{\exp\{-\beta(v_t - v_j)\}}{1 + \sum_{\ell \in \omega \setminus \{t\}} \exp\{-\beta(v_t - v_\ell)\}}.$$

Hence, for every $j \in \omega \setminus \{t\}$, $\sigma_\omega^j(\mathbf{v}^{(\beta)}) \leq \exp(-\beta m')$. Summing over $j \neq t$ gives $1 - \sigma_\omega^t(\mathbf{v}^{(\beta)}) = \sum_{j \in \omega \setminus \{t\}} \sigma_\omega^j(\mathbf{v}^{(\beta)}) \leq (|\omega| - 1) \exp(-\beta m')$. Since Ω is finite, define the uniform menu-size constant $B := \max_{\omega: p(\omega) > 0} (|\omega| - 1) < \infty$. Then, for every ω with $p(\omega) > 0$, $\sigma_\omega^j(\mathbf{v}^{(\beta)}) \leq e^{-\beta m'}$ ($j \neq t$), and $1 - \sigma_\omega^t(\mathbf{v}^{(\beta)}) \leq B e^{-\beta m'}$.

Now use $\frac{\partial \sigma_\omega^i}{\partial v_j} = \beta \sigma_\omega^i(\mathbf{v}) (\mathbf{1}\{i = j\} - \sigma_\omega^j(\mathbf{v}))$. We bound $|\mathbf{1}\{i = j\} - \sigma_\omega^j(\mathbf{v}^{(\beta)})| \leq 1$ and distinguish whether i is maximal in ω . If $i \neq t$, then $\sigma_\omega^i(\mathbf{v}^{(\beta)}) \leq e^{-\beta m'}$, so $\left| \frac{\partial \sigma_\omega^i}{\partial v_j}(\mathbf{v}^{(\beta)}) \right| \leq \beta e^{-\beta m'}$. If $i = t$, then for $j = t$ we have $\left| \frac{\partial \sigma_\omega^t}{\partial v_t}(\mathbf{v}^{(\beta)}) \right| = \beta \sigma_\omega^t(\mathbf{v}^{(\beta)}) (1 - \sigma_\omega^t(\mathbf{v}^{(\beta)})) \leq \beta B e^{-\beta m'}$, while for $j \neq t$, $\left| \frac{\partial \sigma_\omega^t}{\partial v_j}(\mathbf{v}^{(\beta)}) \right| = \beta \sigma_\omega^t(\mathbf{v}^{(\beta)}) \sigma_\omega^j(\mathbf{v}^{(\beta)}) \leq \beta e^{-\beta m'}$. Combining these bounds and enlarging constants if needed, we obtain the uniform estimate: for all i, j and all ω ,

$$\left| \frac{\partial \sigma_\omega^i}{\partial v_j}(\mathbf{v}^{(\beta)}) \right| \leq C_2 \beta e^{-\beta m'}, \quad \text{with } C_2 := \max\{1, B\}. \quad (10)$$

Step 2: Classes that are maximal in some menu. Fix i such that $i = s^*(\omega)$ for some $\omega \in \Omega_i$ with $p(\omega) > 0$. Then for all large β , $D_i(\mathbf{v}^{(\beta)}) = \sum_{\omega' \in \Omega_i} p(\omega') \sigma_{\omega'}^i(\mathbf{v}^{(\beta)}) \geq p(\omega) \sigma_\omega^i(\mathbf{v}^{(\beta)}) \geq c_i > 0$ for some constant c_i (say, $c_i := p(\omega)/2$) independent of β since $\sigma_\omega^i(\mathbf{v}^{(\beta)}) \rightarrow 1$. Recall

$$\frac{\partial g_i}{\partial v_j}(\mathbf{v}) = \frac{1}{D_i(\mathbf{v})} \sum_{\omega \in \Omega_i} p(\omega) \frac{\partial \sigma_\omega^i}{\partial v_j}(\mathbf{v}) (\pi_i(\omega) - g_i(\mathbf{v})).$$

Since $|\pi_i(\omega) - g_i(\mathbf{v})|$ is uniformly bounded on the compact set K , say by C_π , for each $\beta \geq \beta_0$:

$$\left| \frac{\partial g_i}{\partial v_j}(\mathbf{v}^{(\beta)}) \right| \leq \frac{C_\pi}{c_i} \sum_{\omega \in \Omega_i} p(\omega) \left| \frac{\partial \sigma_\omega^i}{\partial v_j}(\mathbf{v}^{(\beta)}) \right| \leq C_{ij} \beta e^{-\beta m'},$$

where we used (10) and absorbed c_i , C_π , and the finite sum over ω into C_{ij} . This already gives the desired exponential decay for any class i that is maximal in some menu.

Step 3: Classes that are never maximal in any menu. Fix a class i such that $i \neq s^*(\omega)$ for all $\omega \in \Omega_i$ with $p(\omega) > 0$. Define the *closest-loss winner* class $t_i := s^*(\omega_i^*)$ where $\omega_i^* \in \arg \min_{\omega \in \Omega_i \cap \text{supp}(p)} v_{s^*(\omega)}^\infty$. By assumption, t_i is unique. Let $\Omega_i^* := \{\omega \in \Omega_i \cap \text{supp}(p) : s^*(\omega) = t_i\}$, and $\Delta_i := v_{t_i}^\infty - v_i^\infty$. We define

$$\eta_i := \begin{cases} \min_{\omega \in \Omega_i \cap \text{supp}(p) \setminus \Omega_i^*} (v_{s^*(\omega)}^\infty - v_{t_i}^\infty), & \text{if } \Omega_i \cap \text{supp}(p) \setminus \Omega_i^* \neq \emptyset, \\ +\infty, & \text{if } \Omega_i^* = \Omega_i \cap \text{supp}(p). \end{cases}$$

where $\eta_i > 0$ by uniqueness of t_i . As a convention, we work with $\eta_i \in (0, \infty]$ and interpret $e^{-\beta \cdot \infty} = 0$ for $\beta > 0$. By Lemma B.1 applied to class i , the propensity weights concentrate on Ω_i^* : $\sum_{\omega \notin \Omega_i^*} w_\omega^i(\mathbf{v}^{(\beta)}) = O(e^{-\beta \eta_i})$. Now use the centered-weights representation (6) at $\mathbf{v}^{(\beta)}$:

$$\frac{\partial g_i}{\partial v_j}(\mathbf{v}^{(\beta)}) = \beta \sum_{\omega \in \Omega_i} w_\omega^i(\mathbf{v}^{(\beta)}) (\mathbf{1}\{i = j\} - \sigma_\omega^j(\mathbf{v}^{(\beta)})) (\pi_i(\omega) - g_i(\mathbf{v}^{(\beta)})).$$

Split the sum over Ω_i^* and its complement. Since $|\pi_i(\omega) - g_i(\mathbf{v}^{(\beta)})|$ is uniformly bounded on K , and $\sum_{\omega \notin \Omega_i^*} w_\omega^i(\mathbf{v}^{(\beta)}) = O(e^{-\beta \eta_i})$ when $0 < \eta_i < \infty$ (with the convention $e^{-\beta \eta_i} = 0$ when $\eta_i = +\infty$), the contribution of $\Omega_i \setminus \Omega_i^*$ is $O(\beta e^{-\beta \eta_i})$, with this term vanishing when $\eta_i = +\infty$. On Ω_i^* , the winner is t_i , so by the uniform within-menu gap in Eq. (10) there exists $m' > 0$ and $C < \infty$ such that for all $\omega \in \Omega_i^*$ and all large β , $1 - \sigma_\omega^{t_i}(\mathbf{v}^{(\beta)}) \leq C e^{-\beta m'}$, and $\sigma_\omega^j(\mathbf{v}^{(\beta)}) \leq C e^{-\beta m'}$ for all $j \in \omega \setminus \{t_i\}$. We consider three cases.

Case 3.1: $j = t_i$. Then $\mathbf{1}\{i = j\} - \sigma_\omega^{t_i} = -\sigma_\omega^{t_i} = -1 + (1 - \sigma_\omega^{t_i})$. Using $\sum_{\omega \in \Omega_i} w_\omega^i(\pi_i(\omega) - g_i) = 0$ (by definition of g_i) and splitting over Ω_i^* and its complement, we get

$$\sum_{\omega \in \Omega_i^*} w_\omega^i(-1)(\pi_i(\omega) - g_i) = - \sum_{\omega \in \Omega_i^*} w_\omega^i(\pi_i(\omega) - g_i) = \sum_{\omega \notin \Omega_i^*} w_\omega^i(\pi_i(\omega) - g_i),$$

whose magnitude is $O(e^{-\beta \eta_i})$. The remaining part involves $1 - \sigma_\omega^{t_i} = O(e^{-\beta m'})$. Hence the Ω_i^* contribution is $O(\beta e^{-\beta \eta_i}) + O(\beta e^{-\beta m'})$.

Case 3.2: $j = i$. Here $\mathbf{1}\{i = j\} - \sigma_\omega^i = 1 - \sigma_\omega^i$. Decompose $1 - \sigma_\omega^i = 1 + r_\omega$, where $|r_\omega| \leq C e^{-\beta m'}$ on Ω_i^* (because i is not the winner on Ω_i^*). Then

$$\sum_{\omega \in \Omega_i^*} w_\omega^i(1 - \sigma_\omega^i)(\pi_i(\omega) - g_i) = \sum_{\omega \in \Omega_i^*} w_\omega^i(\pi_i(\omega) - g_i) + \sum_{\omega \in \Omega_i^*} w_\omega^i r_\omega (\pi_i(\omega) - g_i).$$

For the first term, use centering on Ω_i : $\sum_{\omega \in \Omega_i^*} w_\omega^i (\pi_i(\omega) - g_i) = -\sum_{\omega \notin \Omega_i^*} w_\omega^i (\pi_i(\omega) - g_i)$, whose magnitude is $O(e^{-\beta\eta_i})$ since the mass outside Ω_i^* is $O(e^{-\beta\eta_i})$ and $|\pi_i(\omega) - g_i|$ is uniformly bounded on K . For the second term, $|r_\omega| = O(e^{-\beta m'})$ and the weights sum to at most 1, hence it is $O(e^{-\beta m'})$. Multiplying by the pre-factor β yields an $O(\beta e^{-\beta \min\{\eta_i, m'\}})$ bound.

Case 3.3: $j \notin \{i, t_i\}$. Then $\mathbf{1}\{i = j\} - \sigma_\omega^j = -\sigma_\omega^j$, and on Ω_i^* we have $\sigma_\omega^j = O(e^{-\beta m'})$ if $j \in \omega$, and $\sigma_\omega^j \equiv 0$ if $j \notin \omega$. Hence the Ω_i^* contribution is $O(\beta e^{-\beta m'})$. Combining the complement bound $O(\beta e^{-\beta\eta_i})$ with the Ω_i^* bounds yields

$$\left| \frac{\partial g_i}{\partial v_j}(\mathbf{v}^{(\beta)}) \right| \leq C_{ij} \beta \left(e^{-\beta\eta_i} + e^{-\beta m'} \right) \leq \tilde{C}_{ij} \beta e^{-\beta \tilde{\eta}_i},$$

where $\tilde{\eta}_i := \min\{\eta_i, m'\} > 0$. This gives the desired exponential decay for all j . Let $\eta := \min_{i \in \mathcal{S}} \tilde{\eta}_i > 0$. Combining Steps 2 and 3 over all classes i and coordinates j gives the existence of constants $C < \infty$ and $\eta > 0$ such that $\|Dg(\mathbf{v}^{(\beta)})\| \leq C \beta e^{-\beta\eta} \xrightarrow{\beta \rightarrow \infty} 0$. \square

Next, choose $r > 0$ small enough that for every $\mathbf{v} \in B_r(\mathbf{v}^\infty)$: (i) the unique winner in each menu is still $s^*(\omega)$, and (ii) for each class i , the closest-loss winner class is still t_i . This is possible because both the menu-wise winner gaps $m := \min_{\omega: p(\omega) > 0} \min_{j \in \omega \setminus \{s^*(\omega)\}} (v_{s^*(\omega)}^\infty - v_j^\infty) > 0$ and the closest-loss-winner gaps $\delta_i := \min_{\omega \in \Omega_i \cap \text{supp}(p)} \min_{j \in \omega \setminus \{t_i\}} (v_{s^*(\omega)}^\infty - v_{t_i}^\infty) > 0$ are strictly positive. We claim that there exist constants $C < \infty$, $\eta > 0$, and $\hat{\beta}_0 < \infty$ such that for all $\beta \geq \hat{\beta}_0$ and all $\mathbf{v} \in B_r(\mathbf{v}^\infty)$, $\|D_{\mathbf{v}}g(\mathbf{v}; \beta)\| \leq C \beta e^{-\beta\eta}$, and $\|g(\mathbf{v}^\infty; \beta) - \mathbf{v}^\infty\| \leq C e^{-\beta\eta}$. By continuity, shrinking r if necessary, we may ensure that for every $\mathbf{v} \in B_r(\mathbf{v}^\infty)$ the menu-wise winner identities $s^*(\omega)$ and, for every never-winning class i , the closest-loss winner class t_i are unchanged. Hence all relevant valuation gaps are bounded below uniformly on $B_r(\mathbf{v}^\infty)$ by positive constants. Therefore the same finite-sum Laplace estimates, centered-weights cancellations, and softmax-derivative bounds used in Lemma B.2 hold uniformly for all $\mathbf{v} \in B_r(\mathbf{v}^\infty)$, yielding constants $C < \infty$, $\eta > 0$, and $\hat{\beta}_0 < \infty$ such that, for all $\beta \geq \hat{\beta}_0$,

$$\sup_{\mathbf{v} \in B_r(\mathbf{v}^\infty)} \|D_{\mathbf{v}}g(\mathbf{v}; \beta)\| \leq C \beta e^{-\beta\eta}, \quad \|g(\mathbf{v}^\infty; \beta) - \mathbf{v}^\infty\| \leq C e^{-\beta\eta}. \quad (11)$$

Choose $\hat{\beta} \geq \hat{\beta}_0$ so large that $q_\beta := C \beta e^{-\beta\eta} < \frac{1}{2}$, and $\varepsilon_\beta := C e^{-\beta\eta} < \frac{r}{2}$ for all $\beta \geq \hat{\beta}$. Then for every $\beta \geq \hat{\beta}$ and every $\mathbf{v} \in B_r(\mathbf{v}^\infty)$, $\|g(\mathbf{v}; \beta) - \mathbf{v}^\infty\| \leq \|g(\mathbf{v}; \beta) - g(\mathbf{v}^\infty; \beta)\| + \|g(\mathbf{v}^\infty; \beta) - \mathbf{v}^\infty\| \leq q_\beta \|\mathbf{v} - \mathbf{v}^\infty\| + \varepsilon_\beta \leq \frac{1}{2}r + \frac{1}{2}r = r$. Hence $g(\cdot; \beta)$ maps the closed ball $\overline{B_r(\mathbf{v}^\infty)}$ into itself. Moreover, by (11), $\|g(\mathbf{v}; \beta) - g(\mathbf{u}; \beta)\| \leq q_\beta \|\mathbf{v} - \mathbf{u}\|$, $\forall \mathbf{u}, \mathbf{v} \in \overline{B_r(\mathbf{v}^\infty)}$, so $g(\cdot; \beta)$ is a contraction on $\overline{B_r(\mathbf{v}^\infty)}$. By Banach's fixed-point theorem, for every $\beta \geq \hat{\beta}$ there exists a unique $\mathbf{v}^{(\beta)} \in \overline{B_r(\mathbf{v}^\infty)}$ such that $\mathbf{v}^{(\beta)} = g(\mathbf{v}^{(\beta)}; \beta)$. Thus, for every sufficiently large β , there exists a unique SVE in a neighborhood of the strict pure LSVE. Finally, let

$J(\mathbf{v}^{(\beta)}) = D_{\mathbf{v}}g(\mathbf{v}^{(\beta)}; \beta) - I$ be the Jacobian of the CQL vector field $F(\mathbf{v}) = g(\mathbf{v}; \beta) - \mathbf{v}$ at this SVE. By (11), $\|D_{\mathbf{v}}g(\mathbf{v}^{(\beta)}; \beta)\| \leq q_{\beta} < \frac{1}{2}$. Hence every eigenvalue λ of $D_{\mathbf{v}}g(\mathbf{v}^{(\beta)}; \beta)$ satisfies $|\lambda| < 1/2$, so every eigenvalue of $J(\mathbf{v}^{(\beta)}) = D_{\mathbf{v}}g(\mathbf{v}^{(\beta)}; \beta) - I$ has real part at most $-1/2$. Therefore $J(\mathbf{v}^{(\beta)})$ is Hurwitz. By the linearization theorem for \mathcal{C}^1 vector fields, $\mathbf{v}^{(\beta)}$ is a hyperbolic rest point and is locally exponentially, hence locally asymptotically, stable. \square

B.4 Theorem 4: Remainder of the proof

Lemma B.3. *Consider the RPS decision tree with classes $\mathcal{S} = \{R, P, S\}$, binary menus $\omega_1 = \{R, P\}$, $\omega_2 = \{P, S\}$, $\omega_3 = \{R, S\}$, unary menus $\omega_4 = \{R\}$, $\omega_5 = \{P\}$, $\omega_6 = \{S\}$, uniform probabilities $p(\omega) = 1/6$ for $\omega \in \{\omega_1, \dots, \omega_6\}$, and expected payoffs $\pi_R(\omega_1) = -1$, $\pi_P(\omega_1) = 1$, $\pi_P(\omega_2) = -1$, $\pi_S(\omega_2) = 1$, $\pi_R(\omega_3) = 1$, $\pi_S(\omega_3) = -1$, $\pi_R(\omega_4) = \pi_P(\omega_5) = \pi_S(\omega_6) = z$, $z \in (-1, 0)$. Then, every VE valuation profile equalizes valuations across classes: if $\mathbf{v} \in \mathcal{VE}$, then $v_R = v_P = v_S$. Equivalently, in relative coordinates $x := v_R - v_S$, $y := v_P - v_S$, the only VE valuation profile is $(x, y) = (0, 0)$. Moreover, the set of LSVE $\mathcal{V}(\infty) = \{(0, 0)\}$.*

Proof. We work in relative coordinates $x := v_R - v_S$ and $y := v_P - v_S$.

Step 1: no VE lies in a strict order cone. There are six strict order cones, corresponding to the six strict rankings of (v_R, v_P, v_S) . By cyclic symmetry it suffices to treat two.

If $R \succ P \succ S$ (equivalently $x > y > 0$), then the binary best replies are (R, P, R) , so

$$g_R = \frac{-1 + 1 + z}{3} = \frac{z}{3}, \quad g_P = \frac{-1 + z}{2}, \quad g_S = z.$$

Hence $g_R - g_S = -\frac{2z}{3} > 0$, and $g_P - g_S = -\frac{1+z}{2} < 0$, contradicting $y > 0$.

If $P \succ R \succ S$ (equivalently $y > x > 0$), then the binary best replies are (P, P, R) , so

$$g_R = \frac{1+z}{2}, \quad g_P = \frac{z}{3}, \quad g_S = z.$$

Hence $g_R - g_S = \frac{1-z}{2} > 0$, and $g_P - g_S = -\frac{2z}{3} > 0$, but $(g_R - g_S) - (g_P - g_S) = \frac{3+z}{6} > 0$, contradicting $y > x$. By cyclic relabeling, all six strict cones are excluded.

Step 2: no VE lies on a tie face away from the origin. It remains to rule out $\{x = 0\} \setminus \{(0, 0)\}$, $\{y = 0\} \setminus \{(0, 0)\}$, and $\{x = y\} \setminus \{(0, 0)\}$. By symmetry it suffices to treat $x = 0$, $y \neq 0$.

Suppose first that $x = 0$ and $y > 0$. Then ω_1 and ω_2 have unique best reply P , while ω_3 is

a tie menu. Let $\alpha \in [0, 1]$ denote the probability with which R is chosen in ω_3 . Then

$$g_R = \frac{\alpha + z}{1 + \alpha}, \quad g_S = \frac{z - 1 + \alpha}{2 - \alpha},$$

so $g_R - g_S = \frac{1 + z + 2\alpha(1 - z - \alpha)}{(1 + \alpha)(2 - \alpha)}$. Since $z \in (-1, 0)$ and $\alpha \in [0, 1]$, the numerator is strictly positive, hence $g_R - g_S > 0$, contradicting $x = 0$.

Now suppose $x = 0$ and $y < 0$. Then ω_1 has unique best reply R , ω_2 has unique best reply S , and ω_3 is again a tie menu. With the same notation,

$$g_R = \frac{z - 1 + \alpha}{2 + \alpha}, \quad g_S = \frac{z + \alpha}{3 - \alpha},$$

hence $g_R - g_S = \frac{z - 3 + 2\alpha(1 - z - \alpha)}{(2 + \alpha)(3 - \alpha)}$. Using $2\alpha(1 - z - \alpha) \leq \frac{(1 - z)^2}{2}$, we obtain

$$z - 3 + 2\alpha(1 - z - \alpha) \leq z - 3 + \frac{(1 - z)^2}{2} = \frac{z^2 - 5}{2} < 0,$$

so $g_R - g_S < 0$, again contradicting $x = 0$. Thus no VE lies on $\{x = 0\} \setminus \{(0, 0)\}$. By cyclic relabeling, the same holds on $\{y = 0\} \setminus \{(0, 0)\}$ and $\{x = y\} \setminus \{(0, 0)\}$, so the only VE valuation profile is $(x, y) = (0, 0)$. Finally, $\mathcal{V}(\infty) \subseteq \mathcal{VE}$ by Thm. 1(b). Since the only VE valuation profile in reduced coordinates is $(0, 0)$, it follows that $\mathcal{V}(\infty) = \{(0, 0)\}$. \square

Lemma B.4 (Eventual uniqueness of SVE). *Fix $z \in (-1, 0)$. Then there exist constants $\rho > 0$ and $\bar{\beta} < \infty$ such that $\mathcal{V}(\beta) \subset B_{\rho/\beta}(0)$ and $\mathcal{V}(\beta) = \{(0, 0)\}, \forall \beta \geq \bar{\beta}$.*

Proof. Write

$$F_\beta(x, y) := \begin{pmatrix} (g_R - g_S)(x, y; \beta) - x \\ (g_P - g_S)(x, y; \beta) - y \end{pmatrix}, \quad h_\beta := -F_\beta.$$

Let $(x_n, y_n) \in \mathcal{V}(\beta_n)$ with $\beta_n \uparrow \infty$. Since every accumulation point of (x_n, y_n) is an LSVE, Lemma B.3 implies $(x_n, y_n) \rightarrow (0, 0)$. Define scaled variables $X_n := \beta_n x_n$, $Y_n := \beta_n y_n$, and

$$a_n := \ell(X_n - Y_n), \quad b_n := \ell(Y_n), \quad c_n := \ell(X_n), \quad \ell(t) := \frac{1}{1 + e^{-t}}.$$

Then the equilibrium conditions are equivalent to $x_n = \Phi_1(X_n, Y_n)$, $y_n = \Phi_2(X_n, Y_n)$, where

$$\Phi_1(X, Y) := \frac{-\ell(X - Y) + \ell(X) + z}{\ell(X - Y) + \ell(X) + 1} - \frac{\ell(X) - \ell(Y) + z}{3 - \ell(Y) - \ell(X)},$$

$$\Phi_2(X, Y) := \frac{1 - \ell(X - Y) - \ell(Y) + z}{2 - \ell(X - Y) + \ell(Y)} - \frac{\ell(X) - \ell(Y) + z}{3 - \ell(Y) - \ell(X)}.$$

We claim that (X_n, Y_n) must remain bounded. Suppose not. Passing to a subsequence, each of X_n , Y_n , and $X_n - Y_n$ converges in $[-\infty, \infty]$, with at least one infinite limit. If all three quantities have eventually fixed nonzero signs, then the limiting binary best-reply pattern lies in one of the six strict order cones. Consequently $(\Phi_1(X_n, Y_n), \Phi_2(X_n, Y_n)) \rightarrow (\bar{\Phi}_1, \bar{\Phi}_2) \neq (0, 0)$ by Step 1 of Lemma B.3, contradicting $(x_n, y_n) = (\Phi_1(X_n, Y_n), \Phi_2(X_n, Y_n)) \rightarrow (0, 0)$. Otherwise, exactly one of X_n , Y_n , and $X_n - Y_n$ has a finite limit; this corresponds to one of the three tie faces. Let $\alpha \in [0, 1]$ denote the limiting tie-breaking probability induced by that finite limit. Then Step 2 of Lemma B.3 shows that the corresponding limit of (Φ_1, Φ_2) cannot equal $(0, 0)$, again a contradiction. Hence (X_n, Y_n) is bounded for every sequence of SVE with $\beta_n \uparrow \infty$. Equivalently, there exist $\rho > 0$ and $\beta_0 < \infty$ such that $\mathcal{V}(\beta) \subset B_{\rho/\beta}(0)$, $\forall \beta \geq \beta_0$.

We now prove eventual uniqueness. Since each $g_i(\cdot; \beta)$ is a convex combination of the payoffs in $\{-1, 1, z\}$, we have $|g_i| \leq 1$ and therefore $\|(g_R - g_S, g_P - g_S)\|_\infty \leq 2$, $\forall (x, y)$, $\forall \beta \geq 0$. Choose $R > 2\sqrt{2}$. For $u \in \partial B_R(0)$, define $H_t(u) := u - t(g_R - g_S, g_P - g_S)(u; \beta)$, $t \in [0, 1]$. Then $\|u\|_2 = R$ while $\|t(g_R - g_S, g_P - g_S)\|_2 \leq 2\sqrt{2} < R$, so $H_t(u) \neq 0$ on $\partial B_R(0)$. Hence, by homotopy invariance, $\deg(h_\beta, B_R(0), 0) = \deg(\text{id}, B_R(0), 0) = 1$, $\forall \beta \geq 0$. Next, by linearization at the origin (see Section A.1),

$$Dh_\beta(0, 0) = -A(\beta, z), \quad \det Dh_\beta(0, 0) = \det A(\beta, z) = 1 + \frac{3z}{8}\beta + \frac{9z^2 + 12}{256}\beta^2.$$

Since $z \in (-1, 0)$, there exist $\kappa > 0$ and $\beta_1 < \infty$ such that $\det Dh_\beta(0, 0) \geq \kappa\beta^2$, $\forall \beta \geq \beta_1$. Also, because the components of h_β are rational combinations of binary logits and all denominators are uniformly bounded away from zero, there exists $C_1 < \infty$ such that $\sup_{u \in \mathbb{R}^2} \|D^2 h_\beta(u)\| \leq C_1\beta^2$, $\forall \beta \geq 1$. Hence, for $u \in B_{\rho/\beta}(0)$, $\|Dh_\beta(u) - Dh_\beta(0)\| \leq C_1\beta^2\|u\| \leq C_1\rho\beta$. Since the determinant is locally Lipschitz on bounded sets of 2×2 matrices, there exists $C_2 < \infty$ such that $|\det Dh_\beta(u) - \det Dh_\beta(0)| \leq C_2\rho\beta^2$, $\forall u \in B_{\rho/\beta}(0)$. Choose $\rho > 0$ smaller if necessary so that $C_2\rho < \kappa/2$. Then $\det Dh_\beta(u) \geq \frac{\kappa}{2}\beta^2 > 0$, $\forall u \in B_{\rho/\beta}(0)$, $\forall \beta \geq \beta_1$. Thus every zero of h_β in $B_{\rho/\beta}(0)$ is regular and has local index $+1$.

Fix $\beta \geq \bar{\beta} := \max\{\beta_0, \beta_1\}$. Since all zeros lie in $B_{\rho/\beta}(0)$, excision gives $\deg(h_\beta, B_{\rho/\beta}(0), 0) = \deg(h_\beta, B_R(0), 0) = 1$. But every zero in $B_{\rho/\beta}(0)$ has local index $+1$, so

$$1 = \deg(h_\beta, B_{\rho/\beta}(0), 0) = \sum_{u \in \mathcal{V}(\beta)} \text{ind}_u(h_\beta) = \#\mathcal{V}(\beta).$$

Since $(0, 0)$ is always an SVE, it follows that $\mathcal{V}(\beta) = \{(0, 0)\}$, $\forall \beta \geq \bar{\beta}$. □

Lemma B.5 (Existence of VE). *Let $K := \prod_{s \in \mathcal{S}} [m_s, M_s]$, where $m_s = \min_{\omega \in \Omega_s} \pi_s(\omega)$ and $M_s = \max_{\omega \in \Omega_s} \pi_s(\omega)$ with $\Omega_s = \{\omega \in \Omega : s \in \omega\}$. Define the set-valued map $G_\infty : K \rightrightarrows K$ as $[G_\infty(\mathbf{v})]_s = \text{co}\{\pi_s(\omega) : \omega \in \Omega_s, \omega \in \text{supp}(p), s \in \arg \max_{j \in \omega} v_j\}$. If Asm. 4.1 holds, then there exists $\mathbf{v}^* \in K$ with $\mathbf{v}^* \in G_\infty(\mathbf{v}^*)$; i.e. the set of valuation equilibria \mathcal{VE} is nonempty.*

Proof. For each $s \in \mathcal{S}$, $G_s(\mathbf{v}; \infty)$ is a nonempty, compact, convex subset of $[m_s, M_s]$: nonempty because the arg max set in each ω is nonempty and by Asm. 4.1 there exists some $\omega \in \Omega_s$ with $p(\omega) > 0$ such that $s \in \arg \max_{j \in \omega} v_j$; compact and convex because we take the convex hull of a finite set of expected payoffs. Upper-hemicontinuity of G_∞ follows from upper hemicontinuity of the arg max correspondence and continuity of the payoff array $\{\pi_s(\omega)\}$ and probabilities p . Thus $G_\infty : K \rightrightarrows K$ is nonempty, convex, compact-valued and is upper hemicontinuous. Kakutani's fixed-point theorem yields $\mathbf{v}^* \in K$ with $\mathbf{v}^* \in G_\infty(\mathbf{v}^*)$. \square

B.5 Theorem 6

Proof. **(i) Existence of strict pure LSVE:** Let $K := \prod_{s \in \mathcal{S}} [m_s(z), M_s(z)] \subset \mathbb{R}^{\mathcal{S}}$, $\Omega^* := \{\omega \in \Omega : p(\omega) > 0\}$, and $(\pi_s(\{s\}))_{s \in \mathcal{S}}$ denote singleton expected payoffs (before the z -shift).

Definition 2. *Define inductively an ordered list (s_1, \dots, s_n) and a selector $\phi : \Omega^* \rightarrow \mathcal{S}$ as follows. Set $\mathcal{S}_0 := \emptyset$ and $\Omega^{(1)} := \Omega^*$. For each stage $r = 1, \dots, n$ and each class $k \in \mathcal{S} \setminus \mathcal{S}_{r-1}$ define the residual mixed mass and the residual singleton share respectively by,*

$$D_k^{(r)} := \sum_{\omega \in \Omega^{(r)}: k \in \omega, |\omega| \geq 2} p(\omega), \quad \text{and} \quad \lambda_k^{(r)} := \frac{p(\{k\})}{p(\{k\}) + D_k^{(r)}} \in (0, 1].$$

Choose s_r as a minimizer of $\lambda_k^{(r)}$ over $k \in \mathcal{S} \setminus \mathcal{S}_{r-1}$; breaking ties using any fixed deterministic rule. Define $\mathcal{S}_r := \mathcal{S}_{r-1} \cup \{s_r\}$ and update $\Omega^{(r+1)} := \{\omega \in \Omega^{(r)} : s_r \notin \omega\}$. Given the resulting order (s_1, \dots, s_n) , define $\phi(\omega) := s_r$, where $r = \min\{t \in \{1, \dots, n\} : s_t \in \omega\}$, $\omega \in \Omega^$.*

Assumption B.1. *The probability vector p lies outside a finite union of algebraic hypersurfaces in the simplex; equivalently, along the adaptive construction, whenever at some stage r two distinct remaining classes $i, j \in \mathcal{S} \setminus \mathcal{S}_{r-1}$ both satisfy $D_i^{(r)} > 0$ and $D_j^{(r)} > 0$, their residual singleton shares satisfy $\lambda_i^{(r)} \neq \lambda_j^{(r)}$.⁵²*

⁵²If $D_k^{(r)} = 0$ then $\lambda_k^{(r)} = 1$, so equal residual shares can occur structurally only among classes with zero residual mixed mass; such ties are resolved by the fixed tie-break rule. When $D_i^{(r)}, D_j^{(r)} > 0$, the identity $\lambda_i^{(r)} = \lambda_j^{(r)}$ is equivalent to the polynomial equality $p(\{i\})D_j^{(r)} = p(\{j\})D_i^{(r)}$, excluded under Asm. B.1.

Define the ϕ -selection set of non-singleton menus and the induced singleton share, $\forall k \in \mathcal{S}$,

$$\Omega_k^\phi := \{\omega \in \Omega^* : k \in \omega, |\omega| \geq 2, \phi(\omega) = k\}, \quad \lambda_k^\phi := \frac{p(\{k\})}{p(\{k\}) + \sum_{\omega \in \Omega_k^\phi} p(\omega)} \in (0, 1].$$

Definition 3 (Frozen- ϕ drift). Fix $z \in \mathbb{R}$ and define the frozen-selector drift $G_\infty^\phi : K \rightarrow K$,

$$\left[G_\infty^\phi(\mathbf{v}) \right]_k := \frac{p(\{k\}) (\pi_k(\{k\}) + z) + \sum_{\omega \in \Omega_k^\phi} p(\omega) \pi_k(\omega)}{p(\{k\}) + \sum_{\omega \in \Omega_k^\phi} p(\omega)}.$$

Lemma B.6. For any $z \in \mathbb{R}$, the map G_∞^ϕ is constant on K and has a unique fixed point $\tilde{\mathbf{v}}^\phi(z) \in K$ with coordinates affine in z ,

$$\tilde{v}_k^\phi(z) = a_k^\phi + \lambda_k^\phi z, \quad a_k^\phi := \lambda_k^\phi \pi_k(\{k\}) + \frac{\sum_{\omega \in \Omega_k^\phi} p(\omega) \pi_k(\omega)}{p(\{k\}) + \sum_{\omega \in \Omega_k^\phi} p(\omega)}.$$

Lemma B.7. Under Asm. B.1, for every $\omega \in \Omega^*$ and every $s \in \omega \setminus \{\phi(\omega)\}$, $\lambda_{\phi(\omega)}^\phi < \lambda_s^\phi$.

Proof. Fix $\omega \in \Omega^*$ and write $\phi(\omega) = s_r$, i.e. r is the smallest index with $s_r \in \omega$. Then $\omega \in \Omega^{(r)}$ by construction, and every $s \in \omega \setminus \{s_r\}$ satisfies $s \notin \mathcal{S}_{r-1}$. First, note that by the priority definition, s_r is selected by ϕ in every non-singleton menu of $\Omega^{(r)}$ that contains s_r . Hence the non-singleton menus contributing to s_r under ϕ are exactly those counted in $D_{s_r}^{(r)}$, so $\lambda_{s_r}^\phi = \lambda_{s_r}^{(r)}$. Next, for any $s \notin \mathcal{S}_{r-1}$, ϕ selects s only on a subset of the non-singleton menus in $\Omega^{(r)}$ that contain s , since some of those menus may also contain classes with higher priority than s . Therefore, the total mixed mass contributing to s under ϕ is weakly smaller than $D_s^{(r)}$, implying $\lambda_s^\phi \geq \lambda_s^{(r)}$. Finally, fix $s \in \omega \setminus \{s_r\}$. Since $\omega \in \Omega^{(r)}$ and contains both s_r and s , it follows that ω is a residual non-singleton menu for each of these classes at stage r . In particular, $D_{s_r}^{(r)} \geq p(\omega) > 0$ and $D_s^{(r)} \geq p(\omega) > 0$. By construction of the adaptive order, s_r minimizes $\lambda_k^{(r)}$ over $k \in \mathcal{S} \setminus \mathcal{S}_{r-1}$, so $\lambda_{s_r}^{(r)} \leq \lambda_s^{(r)}$. Since $D_{s_r}^{(r)} > 0$ and $D_s^{(r)} > 0$, Asm. B.1 rules out equality, hence $\lambda_{s_r}^{(r)} < \lambda_s^{(r)}$. Together, $\lambda_{s_r}^\phi = \lambda_{s_r}^{(r)} < \lambda_s^{(r)} \leq \lambda_s^\phi$, which proves $\lambda_{\phi(\omega)}^\phi < \lambda_s^\phi$ for all $s \in \omega \setminus \{\phi(\omega)\}$. \square

Lemma B.8. Under Asm. B.1, there exists $\tilde{z}^\phi \in \mathbb{R}$ such that for every $z < \tilde{z}^\phi$ and every $\omega \in \Omega^*$, $\tilde{v}_{\phi(\omega)}^\phi(z) > \tilde{v}_s^\phi(z)$, $\forall s \in \omega \setminus \{\phi(\omega)\}$. In particular, $\phi(\omega)$ is the unique maximizer of $\tilde{\mathbf{v}}^\phi(z)$ in every supported menu $\omega \in \Omega^*$.

Proof. Fix $\omega \in \Omega^*$ and $s \in \omega \setminus \{\phi(\omega)\}$. Since $\tilde{v}_k^\phi(z) = a_k^\phi + \lambda_k^\phi z$, $\tilde{v}_{\phi(\omega)}^\phi(z) - \tilde{v}_s^\phi(z) = (a_{\phi(\omega)}^\phi - a_s^\phi) + (\lambda_{\phi(\omega)}^\phi - \lambda_s^\phi) z$. By Lem. B.7, $\lambda_{\phi(\omega)}^\phi - \lambda_s^\phi < 0$, hence the right-hand side tends to $+\infty$ as $z \rightarrow -\infty$. Therefore there exists a finite cutoff $z_{\omega,s} \in \mathbb{R}$ such that the strict inequality holds for all $z < z_{\omega,s}$. Taking $\tilde{z}^\phi := \min_{\omega \in \Omega^*} \min_{s \in \omega \setminus \{\phi(\omega)\}} z_{\omega,s}$ yields the claim. \square

Fix any $z < \tilde{z}^\phi$ and define a pure selector $\sigma^\phi = (\sigma_\omega^\phi)_{\omega \in \Omega^*}$ by $\sigma_\omega^\phi(k) := \mathbf{1}\{k = \phi(\omega)\}$, $\omega \in \Omega^*$, $k \in \mathcal{S}$. By Lemma B.8, for every supported menu $\omega \in \Omega^*$, $\phi(\omega)$ is the unique maximizer of $\tilde{\mathbf{v}}^\phi(z)$ on ω . Hence σ^ϕ is a greedy policy at $\tilde{\mathbf{v}}^\phi(z)$. Moreover, for each class $k \in \mathcal{S}$, the set of supported menus in which k is chosen under σ^ϕ is exactly $\{\{k\}\} \cup \Omega_k^\phi$. Therefore,

$$\tilde{v}_k^\phi(z) = \frac{p(\{k\}) (\pi_k(\{k\}) + z) + \sum_{\omega \in \Omega_k^\phi} p(\omega) \pi_k(\omega)}{p(\{k\}) + \sum_{\omega \in \Omega_k^\phi} p(\omega)} = \frac{\sum_{\omega \in \Omega^*} p(\omega) \sigma_\omega^\phi(k) \pi_k(\omega; z)}{\sum_{\omega \in \Omega^*} p(\omega) \sigma_\omega^\phi(k)}.$$

Thus $(\sigma^\phi, \tilde{\mathbf{v}}^\phi(z))$ is a valuation equilibrium. Since every supported menu has a unique maximizer under $\tilde{\mathbf{v}}^\phi(z)$, this VE is strict pure. It remains to show that $\tilde{\mathbf{v}}^\phi(z)$ is a strict pure LSVE. Since $\tilde{\mathbf{v}}^\phi(z)$ is strict pure and Ω^* is finite, there exist $m > 0$ and $r > 0$ such that on the closed ball $B_r := \overline{B_r(\tilde{\mathbf{v}}^\phi(z))} \subset K$ the unique maximizer in each supported menu remains $\phi(\omega)$ and the corresponding valuation gap is at least m . Hence, uniformly for $\mathbf{v} \in B_r$, every non-maximizer $j \in \omega \setminus \{\phi(\omega)\}$ has logit share $\sigma_\omega^j(\mathbf{v}; \beta) = O(e^{-\beta m})$, and therefore all softmax derivatives satisfy $\left| \frac{\partial \sigma_\omega^i}{\partial v_j}(\mathbf{v}; \beta) \right| = O(\beta e^{-\beta m})$. Since every singleton menu is in support, for each class s and every $\mathbf{v} \in B_r$, $D_s(\mathbf{v}; \beta) \geq p(\{s\}) > 0$. Using the centered-weights formula (6) and boundedness of π on K , it follows that $\sup_{\mathbf{v} \in B_r} \|D_{\mathbf{v}} g(\mathbf{v}; \beta, z)\| \leq C \beta e^{-\beta m}$ for some constant $C < \infty$. In particular, for all sufficiently large β , the map $g(\cdot; \beta, z)$ is a contraction on B_r . Let $\bar{G}_{\infty, z}^\phi$ denote the strict-pure consistency map induced by ϕ , i.e.

$$[\bar{G}_{\infty, z}^\phi(\mathbf{v})]_k := \frac{p(\{k\}) (\pi_k(\{k\}) + z) + \sum_{\omega \in \Omega_k^\phi} p(\omega) \pi_k(\omega)}{p(\{k\}) + \sum_{\omega \in \Omega_k^\phi} p(\omega)}.$$

This map is constant and satisfies $\bar{G}_{\infty, z}^\phi(\tilde{\mathbf{v}}^\phi(z)) = \tilde{\mathbf{v}}^\phi(z)$. Because the maximizing class in each supported menu is fixed throughout B_r , the conditional menu weights induced by logit converge uniformly on B_r to the corresponding p -weights normalized on $\{\{k\}\} \cup \Omega_k^\phi$. Hence $\sup_{\mathbf{v} \in B_r} \|g(\mathbf{v}; \beta, z) - \bar{G}_{\infty, z}^\phi(\mathbf{v})\| \rightarrow 0$ as $\beta \rightarrow \infty$. Therefore, after possibly shrinking r , for all sufficiently large β we have $g(B_r; \beta, z) \subseteq B_r$. By Banach's fixed-point theorem, there exists a unique SVE $\mathbf{v}^{(\beta)} \in B_r$ satisfying $\mathbf{v}^{(\beta)} = G(\mathbf{v}^{(\beta)}; \beta, z)$, and necessarily $\mathbf{v}^{(\beta)} \rightarrow \tilde{\mathbf{v}}^\phi(z)$ as $\beta \rightarrow \infty$. Thus $\tilde{\mathbf{v}}^\phi(z)$ is a strict pure LSVE. The local asymptotic stability claim now follows directly from Thm. 3: there exists $\hat{\beta} < \infty$ such that, for every $\beta \geq \hat{\beta}$, the unique SVE in a neighborhood of $\tilde{\mathbf{v}}^\phi(z)$ is locally asymptotically stable for the CQL dynamics.

(ii) Multiplicity of VE: Assume in addition that all binary menus are in support, i.e. $\{\omega \in \Omega : |\omega| = 2\} \subset \text{supp}(p)$. Fix $s \in \mathcal{S}$. We will construct, for all sufficiently small z , a VE at which s is the unique worst-ranked class. Let $K^{(-s)}(z) := \prod_{k \neq s} [m_k(z), M_k(z)]$ and define

the frozen- s best response correspondence $G_{\infty,z}^{(-s)} : K^{(-s)}(z) \rightrightarrows K^{(-s)}(z)$ by

$$G_{k,\infty,z}^{(-s)}(\mathbf{v}^{-s}) \in \text{co} \left\{ \pi_k(\omega; z) : \omega \in \Omega, \omega \in \text{supp}(p), k \in \omega, k \in \arg \max_{j \in \omega \setminus \{s\}} v_j \right\}, \quad k \neq s,$$

where $\pi_k(\{k\}; z) := \pi_k(\{k\}) + z$ and $\pi_k(\omega; z) := \pi_k(\omega)$ for $|\omega| \geq 2$. Thus z shifts all singleton payoffs uniformly, and in menus containing s we ignore s in the maximization. Because the co-occurrence graph is complete and all singleton menus are in support, the reduced $(n-1)$ -class problem (in which s is ignored) satisfies the same standing assumptions as the original tree. Hence the frozen-priority (strict-pure) construction from part (i) applies verbatim to this frozen- s problem:⁵³ there exist $\bar{z}^{(-s)} \in \mathbb{R}$ and a *fixed* strict selector $\varphi^{(-s)}$ (independent of z) such that, for every $z < \bar{z}^{(-s)}$, the correspondence $G_{\infty,z}^{(-s)}$ admits a fixed point $\tilde{\mathbf{v}}^{(-s)}(z) \in K^{(-s)}(z)$ implementing $\varphi^{(-s)}$. Conditional on $\varphi^{(-s)}$, Lemma B.6 implies that $\tilde{\mathbf{v}}^{(-s)}(z)$ is affine in z : $\tilde{v}_k^{(-s)}(z) = a_k^{(-s)} + \lambda_k^{(-s)} z$, for $k \neq s$, with $\lambda_k^{(-s)} \in (0, 1]$ equal to the conditional weight of singleton menus in k 's consistency calculation under $\varphi^{(-s)}$. Moreover, $\lambda_k^{(-s)} < 1$ for every $k \neq s$. Indeed, since all binary menus are in support, $\{s, k\} \in \text{supp}(p)$; as s is ignored in the maximization, the selector must choose k in the non-singleton menu $\{s, k\}$, so k is selected in at least one non-singleton menu under $\varphi^{(-s)}$, which forces $\lambda_k^{(-s)} < 1$.

Define $\mathbf{v}^{(s)}(z) \in K$ by $v_s^{(s)}(z) := \pi_s(\{s\}) + z$, and $v_k^{(s)}(z) := \tilde{v}_k^{(-s)}(z) = a_k^{(-s)} + \lambda_k^{(-s)} z$, $k \neq s$. For each $k \neq s$, $v_k^{(s)}(z) - v_s^{(s)}(z) = (a_k^{(-s)} - \pi_s(\{s\})) + (\lambda_k^{(-s)} - 1)z$, and since $\lambda_k^{(-s)} - 1 < 0$, the right-hand side tends to $+\infty$ as $z \rightarrow -\infty$. Therefore there exists a finite threshold $\tilde{z}_s := \min_{k \neq s} \frac{a_k^{(-s)} - \pi_s(\{s\})}{1 - \lambda_k^{(-s)}} \in \mathbb{R}$ such that $v_s^{(s)}(z) < v_k^{(s)}(z)$ for all $k \neq s$ whenever $z < \tilde{z}_s$.

Fix $z < \min\{\bar{z}^{(-s)}, \tilde{z}_s\}$, so that s is uniquely the least-valued class at $\mathbf{v}^{(s)}(z)$. We now verify that $\mathbf{v}^{(s)}(z) \in G_{\infty,z}(\mathbf{v}^{(s)}(z))$, i.e. $\mathbf{v}^{(s)}(z)$ is a VE. For any menu ω with $s \notin \omega$, the best-response sets are the same as in the frozen- s problem. For any non-singleton menu $\omega \ni s$, since s is strictly least-valued, the best response lies in $\omega \setminus \{s\}$, so again the best-response set coincides with that in the frozen- s problem. Hence for each $k \neq s$, $g_{k,\infty,z}(\mathbf{v}^{(s)}(z)) = G_{k,\infty,z}^{(-s)}(\tilde{\mathbf{v}}^{(-s)}(z))$, and $v_k^{(s)}(z) = \tilde{v}_k^{(-s)}(z) \in G_{k,\infty,z}^{(-s)}(\tilde{\mathbf{v}}^{(-s)}(z))$. Therefore $v_k^{(s)}(z) \in g_{k,\infty,z}(\mathbf{v}^{(s)}(z))$ for all $k \neq s$. Finally, since $\{s\} \in \text{supp}(p)$ and s is not a best response in any non-singleton menu at $\mathbf{v}^{(s)}(z)$, the s -coordinate of the best-response correspondence reduces to its singleton payoff: $g_{s,\infty,z}(\mathbf{v}^{(s)}(z)) = \pi_s(\{s\}) + z = v_s^{(s)}(z)$. Thus $\mathbf{v}^{(s)}(z) \in G_{\infty,z}(\mathbf{v}^{(s)}(z))$, and $\mathbf{v}^{(s)}(z)$ is a VE with

⁵³Strictly speaking, one must also exclude the exceptional parameter values for which the generic separation condition in Asm. B.1 fails in one of the frozen- s reduced problems. For each fixed s , the equal-residual-share conditions arising in the frozen- s construction define a finite union of algebraic hypersurfaces in the simplex of menu probabilities. Since there are only finitely many classes s , their union over $s \in \mathcal{S}$ is still a finite union of algebraic hypersurfaces. Hence, after excluding this enlarged exceptional set, the generic separation property holds simultaneously for the original problem and for every frozen- s reduced problem.

s uniquely worst-ranked. Since $s \in \mathcal{S}$ was arbitrary, letting $\tilde{z}' := \min_{s \in \mathcal{S}} \min\{\bar{z}^{(-s)}, \tilde{z}_s\}$, we conclude that for all $z < \tilde{z}'$ there are at least $|\mathcal{S}|$ distinct valuation equilibria, each with a different uniquely worst-ranked class. \square

B.6 Proposition 1

Proof. By Theorem 5, under Assumption 4.1 there exists $\hat{z} < \infty$ such that, for every $z \geq \hat{z}$ and every $\beta \geq 0$, the CQL dynamics admit a unique SVE, and this SVE is globally asymptotically stable. Moreover, by Theorem 1, every accumulation point of this unique SVE branch as $\beta \rightarrow \infty$ is a mixed VE. It therefore remains only to show that, under Assumption 4.2, every VE must in fact be *fully mixed*, i.e. must equalize valuations across all classes. Fix a VE (σ^*, v^*) of the z -shifted tree. By Assumption 4.2, all singleton menus have the same probability; write $q := p(\{s\}) > 0$, $s \in \mathcal{S}$. For each class $s \in \mathcal{S}$, let $D_s := \sum_{\omega \ni s} p(\omega) \sigma_\omega^*(s)$ denote the total selection mass of class s under the VE. Since $\sigma_{\{s\}}^*(s) = 1$, we have $D_s \geq q$. Writing the singleton payoff as $\pi_s(\{s\}) + z$, valuation consistency gives

$$v_s^* = \frac{q(\pi_s(\{s\}) + z) + \sum_{\omega \ni s, |\omega| > 1} p(\omega) \sigma_\omega^*(s) \pi_s(\omega)}{D_s} = \lambda_s z + b_s,$$

where $\lambda_s := \frac{q}{D_s} \in (0, 1]$ and $b_s := \frac{q \pi_s(\{s\}) + \sum_{\omega \ni s, |\omega| > 1} p(\omega) \sigma_\omega^*(s) \pi_s(\omega)}{D_s}$. Because the reduced tree is finite, there exists $M < \infty$ such that $|b_s| \leq M$ for every $s \in \mathcal{S}$. We claim that if $v_i^* < v_j^*$ for some $i, j \in \mathcal{S}$, then $D_j \geq D_i + q$, and hence

$$\lambda_i - \lambda_j = q \left(\frac{1}{D_i} - \frac{1}{D_j} \right) = \frac{q(D_j - D_i)}{D_i D_j} \geq q^2.$$

To prove the claim, let $\mathcal{M}_i := \{\omega \in \Omega : i \in \omega, \sigma_\omega^*(i) > 0\}$. If $\omega \in \mathcal{M}_i$, then i is a best reply in ω , so every class in ω has valuation at most v_i^* . Since $v_j^* > v_i^*$, necessarily $j \notin \omega$. Hence in the enlarged menu $\omega \cup \{j\}$, class j is the unique best reply, so $\sigma_{\omega \cup \{j\}}^*(j) = 1$. Moreover, the map $\omega \mapsto \omega \cup \{j\}$ is injective on \mathcal{M}_i , and Assumption 4.2 implies $p(\omega \cup \{j\}) \geq p(\omega)$. Therefore, $D_j \geq q + \sum_{\omega \in \mathcal{M}_i} p(\omega \cup \{j\}) \geq q + \sum_{\omega \in \mathcal{M}_i} p(\omega) \geq q + \sum_{\omega \in \mathcal{M}_i} p(\omega) \sigma_\omega^*(i) = q + D_i$, which proves the claim. Now suppose, toward a contradiction, that $v_i^* < v_j^*$ for some i, j . Then

$$v_i^* - v_j^* = (\lambda_i - \lambda_j)z + (b_i - b_j) \geq q^2 z - 2M.$$

Hence, if $z > \bar{z} := \frac{2M}{q^2}$, we obtain $v_i^* - v_j^* > 0$, a contradiction. Therefore, for every $z \geq \bar{z}$, no strict valuation gap can exist in any VE. It follows that every VE satisfies $v_i^* = v_j^* \forall i, j \in \mathcal{S}$, so every VE is fully mixed in the sense of total indifference among all similarity classes. Finally, let $\hat{z}_1 := \max\{\hat{z}, \bar{z}\}$. For every $z \geq \hat{z}_1$ and every $\beta \geq 0$, existence, uniqueness, and

global asymptotic stability of the SVE follow from Theorem 5. Since the SVE is unique for every β , its high-sensitivity limit selects a unique VE along the principal branch, and the previous argument shows that this VE must be fully mixed. This proves the proposition. \square

B.7 Proposition 2

Proof. Fix an arbitrary strict total order on \mathcal{S} , written as $\pi(1) \prec \pi(2) \prec \dots \prec \pi(n)$, where $\pi(1)$ is the lowest-ranked class and $\pi(n)$ the highest-ranked class. Under Assumptions 4.1 and 4.2, every nonempty menu is in support and all singleton menus have the same probability; write $q := p(\{s\}) > 0$, $s \in \mathcal{S}$. For this order, define the pure greedy selector σ^π by choosing in each menu the π -highest available class: $\sigma_\omega^\pi(\pi(r)) = 1 \iff \pi(r) \in \omega$ and $\omega \cap \{\pi(r+1), \dots, \pi(n)\} = \emptyset$. Equivalently, class $\pi(r)$ is selected exactly on the menu family $\Omega_r^\pi := \left\{ \omega \in \Omega : \pi(r) \in \omega, \omega \cap \{\pi(r+1), \dots, \pi(n)\} = \emptyset \right\}$. Let $D_r^\pi := \sum_{\omega \in \Omega_r^\pi} p(\omega)$, $\lambda_r^\pi := \frac{q}{D_r^\pi} \in (0, 1]$, and

$$b_r^\pi := \frac{q \pi_{\pi(r)}(\{\pi(r)\}) + \sum_{\omega \in \Omega_r^\pi, |\omega| > 1} p(\omega) \pi_{\pi(r)}(\omega)}{D_r^\pi}.$$

Then the valuation-consistency equation induced by σ^π is $v_{\pi(r)}^\pi(z) = b_r^\pi + \lambda_r^\pi z$, $r = 1, \dots, n$. We claim that $\lambda_1^\pi > \lambda_2^\pi > \dots > \lambda_n^\pi$. Fix $r < n$. If $\omega \in \Omega_r^\pi$, then $\pi(r+1) \notin \omega$ and $T_r(\omega) := \omega \cup \{\pi(r+1)\} \in \Omega_{r+1}^\pi$. The map $T_r : \Omega_r^\pi \rightarrow \Omega_{r+1}^\pi$ is injective, and by Assumption 4.2, $p(T_r(\omega)) \geq p(\omega)$. Since $\{\pi(r+1)\} \in \Omega_{r+1}^\pi$ and is not in the image of T_r , we obtain $D_{r+1}^\pi \geq q + \sum_{\omega \in \Omega_r^\pi} p(T_r(\omega)) \geq q + \sum_{\omega \in \Omega_r^\pi} p(\omega) = q + D_r^\pi$. Hence

$$\lambda_r^\pi - \lambda_{r+1}^\pi = q \left(\frac{1}{D_r^\pi} - \frac{1}{D_{r+1}^\pi} \right) = \frac{q(D_{r+1}^\pi - D_r^\pi)}{D_r^\pi D_{r+1}^\pi} \geq q^2,$$

because $D_r^\pi, D_{r+1}^\pi \leq 1$. Thus the singleton coefficient on z is strictly larger for lower-ranked classes. Now let $M := \max_{s \in \mathcal{S}, \omega \ni s} |\pi_s(\omega)| < \infty$. Since each b_r^π is a convex combination of primitive payoffs, $|b_r^\pi| \leq M$ for every r and every π . Therefore, for $r < n$, $v_{\pi(r+1)}^\pi(z) - v_{\pi(r)}^\pi(z) = (\lambda_{r+1}^\pi - \lambda_r^\pi)z + (b_{r+1}^\pi - b_r^\pi) \geq q^2|z| - 2M$. Choose $\tilde{z} := -\frac{2M+1}{q^2}$. Then for every $z \leq \tilde{z}$ and every $r < n$, $v_{\pi(r+1)}^\pi(z) - v_{\pi(r)}^\pi(z) \geq 1 > 0$. Hence $v_{\pi(1)}^\pi(z) < v_{\pi(2)}^\pi(z) < \dots < v_{\pi(n)}^\pi(z)$.

It follows that, for every menu ω , the unique maximizer of $v^\pi(z)$ on ω is the π -highest element of ω , i.e. exactly the class selected by σ^π . Since the coordinates of $v^\pi(z)$ were defined by the consistency equations induced by σ^π , the pair $(\sigma^\pi, v^\pi(z))$ is a strict pure VE. Because π was arbitrary, every strict total order is admissible in some VE; hence there are $n!$ strict pure VE. Finally, fix $z \leq \tilde{z}$. The last step in the proof of Theorem 6 applies verbatim to each strict pure VE constructed above: since every class is selected at least at its singleton menu,

each such strict pure VE is a strict pure LSVE, and therefore, by Theorem 3, for each order π there exists $\hat{\beta}_\pi(z) < \infty$ such that for all $\beta \geq \hat{\beta}_\pi(z)$, the unique SVE in a neighborhood of $v^\pi(z)$ is locally asymptotically stable. Because there are only finitely many permutations, setting $\hat{\beta}(z) := \max_\pi \hat{\beta}_\pi(z) < \infty$ completes the proof. \square

B.8 Equivalence of the Myopic and the Forward-Looking CQL dynamics (modulo translations)

Lemma B.9. *Let $\mathcal{S} = \{1, \dots, n\}$ index similarity classes and consider the CQL ODE*

$$\dot{\mathbf{v}} = \hat{f}(\mathbf{v}) := g(\mathbf{v}) + \gamma C(\mathbf{v}) \mathbf{1} - \mathbf{v}, \quad \sigma_\omega^s(\mathbf{v}) = \frac{\exp(\beta v_s)}{\sum_{j \in \omega} \exp(\beta v_j)}.$$

Assume that the continuation term satisfies the shift property $C(\mathbf{v} + c\mathbf{1}) = C(\mathbf{v}) + c$, $\forall c \in \mathbb{R}$, which is satisfied by the max (Q-learning), LogSumExp (dynamic logit), and on-policy (SARSA) continuation operators. Then the following statements hold for $\gamma \in [0, 1)$:

- (i) *The softmax is translation-invariant, hence $g(\mathbf{v} + c\mathbf{1}) = g(\mathbf{v})$ for all $c \in \mathbb{R}$. Consider the forward-looking CQL ODE: $\dot{\mathbf{v}} = g(\mathbf{v}) + \gamma C(\mathbf{v})\mathbf{1} - \mathbf{v}$, where $g(\mathbf{v} + c\mathbf{1}) = g(\mathbf{v})$ and $C(\mathbf{v} + c\mathbf{1}) = C(\mathbf{v}) + c$ for all $c \in \mathbb{R}$. Let $\eta(t)$ solve $\dot{\eta}(t) = \gamma C(\mathbf{v}(t)) - \eta(t)$, $\eta(0) \in \mathbb{R}$, and define $\tilde{\mathbf{v}}(t) := \mathbf{v}(t) - \eta(t)\mathbf{1}$. Then $\tilde{\mathbf{v}}$ solves the myopic mean-field ODE $\dot{\tilde{\mathbf{v}}} = g(\tilde{\mathbf{v}}) - \tilde{\mathbf{v}}$.*
- (ii) *Fix a pivot class $q \in \mathcal{S}$ and define $u_s = v_s - v_q, \forall s \in \mathcal{S}$, and $U = \{u \in \mathbb{R}^{\mathcal{S}} : u_q = 0\} \cong \mathbb{R}^{n-1}$. The relative valuations \mathbf{u} evolve according to the $(n-1)$ -dimensional ODE on U : $\dot{u}_s = \tilde{g}_s(\mathbf{u}) - u_s, s \in \mathcal{S} \setminus \{q\}, u_q \equiv 0$, where $\tilde{g}_s(\mathbf{u}) := g_s(\mathbf{u} + c\mathbf{1}) - g_q(\mathbf{u} + c\mathbf{1})$ is well-defined and independent of c by translation invariance.*
- (iii) *If \mathbf{v}^* is a rest point of the full ODE system \hat{f} , then $\mathbf{u}^* = \mathbf{v}^* - v_q^*\mathbf{1} \in U$ solves $\tilde{g}(\mathbf{u}^*) = \mathbf{u}^*$. Conversely, if $\mathbf{u}^\dagger \in U$ satisfies $\tilde{g}(\mathbf{u}^\dagger) = \mathbf{u}^\dagger$, then $c^\dagger := \frac{g_q(\mathbf{u}^\dagger) + \gamma C(\mathbf{u}^\dagger)}{1 - \gamma}$ is well-defined and $\mathbf{v}^\dagger = \mathbf{u}^\dagger + c^\dagger\mathbf{1}$ is a rest point of the full system.*
- (iv) *For any \mathbf{v} , $Dg(\mathbf{v})\mathbf{1} = \mathbf{0}$. The shift property of C implies $DC(\mathbf{v})\mathbf{1} = \mathbf{1}$, hence $D\hat{f}(\mathbf{v})\mathbf{1} = -(1 - \gamma)\mathbf{1}$. Fix a pivot q and consider the linear change of variables $(\mathbf{u}, c) = \Phi(\mathbf{v})$ given by $u_s = v_s - v_q$ for $s \neq q$, $u_q = 0$, and $c = v_q$ (so $\mathbf{v} = \Psi(\mathbf{u}, c) = \mathbf{u} + c\mathbf{1}$). In these coordinates the reduced dynamics for \mathbf{u} are autonomous and the c -direction satisfies $\dot{c} = \dot{v}_q = \hat{f}_q(\mathbf{u} + c\mathbf{1})$. Thus, the Jacobian of the full system at a rest point is similar to a block lower-triangular matrix whose diagonal blocks are the Jacobian of the reduced system on U and the scalar $-(1 - \gamma)$. Consequently, the eigenvalues of the reduced Jacobian coincide with the eigenvalues of $D\hat{f}(\mathbf{v})$ other than $-(1 - \gamma)$, and local asymptotic stability of hyperbolic rest points is preserved under reduction.*

Proof. (i) For any $c \in \mathbb{R}$, $g(\mathbf{v} + c\mathbf{1}) = g(\mathbf{v})$ because for any $\omega \in \Omega$,

$$\sigma_\omega^s(\mathbf{v} + c\mathbf{1}) = \frac{\exp(\beta(v_s + c))}{\sum_{j \in \omega} \exp(\beta(v_j + c))} = \sigma_\omega^s(\mathbf{v}).$$

Differentiate $\tilde{\mathbf{v}} = \mathbf{v} - \eta\mathbf{1}$: $\dot{\tilde{\mathbf{v}}} = \dot{\mathbf{v}} - \dot{\eta}\mathbf{1} = (g(\mathbf{v}) + \gamma C(\mathbf{v})\mathbf{1} - \mathbf{v}) - (\gamma C(\mathbf{v}) - \eta)\mathbf{1} = g(\mathbf{v}) - (\mathbf{v} - \eta\mathbf{1})$. By translation invariance of g , $g(\mathbf{v}) = g(\tilde{\mathbf{v}})$, hence $\dot{\tilde{\mathbf{v}}} = g(\tilde{\mathbf{v}}) - \tilde{\mathbf{v}}$.

(ii) Differentiating $u_s = v_s - v_q$ gives $\dot{u}_s = \dot{v}_s - \dot{v}_q = (g_s(\mathbf{v}) + \gamma C(\mathbf{v}) - v_s) - (g_q(\mathbf{v}) + \gamma C(\mathbf{v}) - v_q) = (g_s(\mathbf{v}) - g_q(\mathbf{v})) - u_s$. Since $g_s(\mathbf{v}) - g_q(\mathbf{v})$ is unchanged by adding $c\mathbf{1}$ to \mathbf{v} , we may write it as $\tilde{g}_s(\mathbf{u})$, yielding the closed ODE on U , $\dot{u}_s = \tilde{g}_s(\mathbf{u}) - u_s, \forall s \in \mathcal{S} \setminus \{q\}$ with $u_q \equiv 0$.

(iii) If $\hat{f}(\mathbf{v}^*) = \mathbf{0}$, then $0 = \hat{f}_s(\mathbf{v}^*) - \hat{f}_q(\mathbf{v}^*) = (g_s(\mathbf{v}^*) - g_q(\mathbf{v}^*)) - (v_s^* - v_q^*)$, so $\tilde{g}(\mathbf{u}^*) = \mathbf{u}^*$. Conversely, suppose $\tilde{g}(\mathbf{u}^\dagger) = \mathbf{u}^\dagger$. For any c , translation invariance implies $g_q(\mathbf{u}^\dagger + c\mathbf{1}) = g_q(\mathbf{u}^\dagger)$, and the shift property gives $C(\mathbf{u}^\dagger + c\mathbf{1}) = C(\mathbf{u}^\dagger) + c$. Setting $c^\dagger = \frac{g_q(\mathbf{u}^\dagger) + \gamma C(\mathbf{u}^\dagger)}{1 - \gamma}$ yields $g(\mathbf{u}^\dagger + c^\dagger\mathbf{1}) + \gamma C(\mathbf{u}^\dagger + c^\dagger\mathbf{1})\mathbf{1} = \mathbf{u}^\dagger + c^\dagger\mathbf{1}$, so $\mathbf{v}^\dagger := \mathbf{u}^\dagger + c^\dagger\mathbf{1}$ is a rest point.

(iv) To compare spectra with the reduced system, use the linear coordinates from (ii)–(iii). Let $\Phi : \mathbb{R}^{\mathcal{S}} \rightarrow U \times \mathbb{R}$ be given by $u_s = v_s - v_q$ for $s \neq q$, $u_q = 0$, and $c = v_q$, so that $\Psi(\mathbf{u}, c) := \mathbf{u} + c\mathbf{1}$ satisfies $\Psi \circ \Phi = \text{id}$. By (ii), the \mathbf{u} -dynamics are autonomous: $\dot{\mathbf{u}} = F(\mathbf{u}) := \tilde{g}(\mathbf{u}) - \mathbf{u}$. Moreover, along $\mathbf{v} = \Psi(\mathbf{u}, c)$ we have $\dot{c} = \dot{v}_q = g_q(\mathbf{u} + c\mathbf{1}) + \gamma C(\mathbf{u} + c\mathbf{1}) - c = g_q(\mathbf{u}) + \gamma C(\mathbf{u}) + (\gamma - 1)c$, where we used translation invariance of g and the shift property of C . Hence, in (\mathbf{u}, c) coordinates the vector field takes the triangular form

$$\begin{pmatrix} \dot{\mathbf{u}} \\ \dot{c} \end{pmatrix} = \begin{pmatrix} F(\mathbf{u}) \\ h(\mathbf{u}) + (\gamma - 1)c \end{pmatrix}, \quad h(\mathbf{u}) := g_q(\mathbf{u}) + \gamma C(\mathbf{u}).$$

At a rest point (\mathbf{u}^*, c^*) , the Jacobian therefore has the block lower-triangular form

$$D(\Phi \circ \hat{f} \circ \Psi)(\mathbf{u}^*, c^*) = \begin{pmatrix} DF(\mathbf{u}^*) & 0 \\ Dh(\mathbf{u}^*) & \gamma - 1 \end{pmatrix}.$$

Since Φ is invertible, this matrix is similar to $D\hat{f}(\mathbf{v}^*)$. Thus the spectrum of $D\hat{f}(\mathbf{v}^*)$ consists of the eigenvalues of the reduced Jacobian $DF(\mathbf{u}^*)$ together with the strictly negative additional eigenvalue $\gamma - 1 = -(1 - \gamma) < 0$, proving the claim. \square