Learning to be Indifferent in Complex Decisions: A Coarse Q-learning Model*

Philippe Jehiel† Aviman Satpathy‡

December 15, 2024

Latest version available here

Abstract

We introduce the Coarse Q-learning (CQL) model, which captures reinforcement learning by boundedly rational decision-makers who focus on the aggregate outcomes of choosing among exogenously defined clusters of alternatives (similarity classes), rather than evaluating each alternative individually. Analyzing a smooth approximation of the model, we show that the learning dynamics exhibit steady-states corresponding to smooth Valuation Equilibria (Jehiel and Samet, 2007). We demonstrate the existence of multiple equilibria in decision trees with generic payoffs and establish the local asymptotic stability of pure equilibria when they occur. Conversely, when trivial choices featuring alternatives within the same similarity class yield sufficiently high payoffs, a unique mixed equilibrium emerges, characterized by indifferences between similarity classes, even under acute sensitivity to payoff differences. Finally, we prove that this unique mixed equilibrium is globally asymptotically stable under the CQL dynamics.

Keywords: Reinforcement Learning, Bounded Rationality, Coarse Reasoning
JEL Codes: C62, C73, D83

*We thank Evan Friedman, Jean-Marc Tallon, Olivier Tercieux & Giacomo Weber for valuable comments.
†Paris School of Economics & University College London; jehiel@enpc.fr
‡Paris School of Economics; aviman.satpathy@psemail.eu

1 Introduction

Traditional economic models of decision-making under payoff uncertainty often assume that agents construct a Bayesian representation of the distribution of payoffs as a function of the chosen alternatives. When repeatedly faced with such scenarios, commonly referred to as multi-armed bandit problems, they design optimal strategies that trade off the exploration-exploitation incentives based on the discount factor (rate of impatience) (Gittins, 1979).

We consider a standard decision problem under uncertainty where different alternatives may be available depending on the state of the world. However, we depart from the Bayesian paradigm by assuming that our decision-maker employs a simple learning heuristic to guide her strategy and its adjustments over time as she accumulates experience. Specifically, our decision-maker starts with initial assessments of the values of her alternatives, makes her choice by employing a mixed strategy that assigns higher probabilities to alternatives with higher assessments, and updates her assessments of the chosen alternatives in the direction of the observed payoff. Our learning scheme broadly belongs to the tradition of model-free reinforcement learning (RL), to the extent that the dynamics are driven solely by the observed payoffs. This particular genre of reinforcement learning with bandit feedback has been referred to as the payoff-assessment learning model in the literature (Sarin and Vahid, 1999), although we incorporate a noisy (smooth) version of it that uses a logit choice rule based on the assessments, similar to the one used in Cominetti et al. (2010).[1]

In this paper, we modify the aforementioned smooth payoff-assessment learning model in one key aspect. Rather than assuming that the decision-maker forms a distinct assessment for each individual alternative, we propose that the decision-maker considers a smaller set of categories (equivalence classes) that partition the overall set of alternatives, forming assessments only at the level of categories. When faced with a choice among alternatives belonging to different categories, we assume that the decision-maker implements the heuristic described above, first by selecting a category based on its assessment, and then uniformly randomizing among alternatives within the chosen category.[2] Beyond this modification, our learning model is the same as described above. That is, the decision-maker chooses her strategy as a

[1]The logit (softmax) formulation has been viewed by some researchers (Rustichini et al., 2023) as providing a heuristic approach to model how a decision-maker could handle the exploitation/exploration trade-off, although the learning heuristic considered here is not forward-looking per se. In our study, we treat the parameter that governs the logit formulation as exogenous and focus on its limit where the decision-maker almost surely picks the alternative(s) with the highest assessment(s).

[2]This two-step procedure is conceptually similar to the nested logit model (Hausman and McFadden, 1984), but simplifies it by applying a uniform randomization strategy within each category, following the principle of indifference, as such alternatives are considered indistinguishable to the decision-maker.

smooth function of her profile of assessments, and updates the assessment of a given category based on the observed payoff whenever she picks an alternative in that category. We refer to such a learning model as the Coarse Q-learning (CQL) model, adopting a smooth version where the choice policy is modeled using the logit (softmax) choice rule.

We believe that our modification of the usual payoff-assessment learning model is particularly relevant when the grand set of alternatives is too large to be considered extensively. In such cases, it seems highly plausible that the decision-maker would use coarse categories and implement the assessment device at the level of such categories rather than at the level of individual alternatives. This perspective in terms of categorical reasoning in the face of choice overload is well known in the psychological literature (Rosch and Lloyd, 1978); however, to the best of our knowledge, it has not been explored in mathematical analyses of learning models. This is the main focus of our paper. It should be emphasized from the outset that, in our approach, the categories (referred to as similarity classes) are exogenously defined. One possible interpretation is that alternatives are characterized by vectors of attributes and the agent focuses primarily on those attributes that are salient while ignoring the others (Tversky, 1972; Bordalo et al., 2012, 2013). From this perspective, a given similarity class corresponds to a specific profile of realizations of the salient attributes, potentially encompassing a large subset of alternatives.[3]

Our results are as follows. Firstly, we establish that the set of steady-states of the Coarse Q-learning model is non-empty. Moreover, these steady-states correspond to a smooth version of the Valuation Equilibrium (Jehiel and Samet, 2007). As the sensitivity parameter in the logit formulation increases without bound, the steady-states of the CQL model converge to a refinement of the set of Valuation Equilibria. We illustrate through examples the possibilities that multiple steady-states may emerge or that a unique steady-state arises in which the assessments of several similarity classes are equal in the high-sensitivity limit, where the decision-maker almost surely selects alternatives in the similarity class(es) with the highest perceived assessment(s). The latter case, which can arise even for generic specifications of objective payoffs, necessitates mixing between similarity classes in order to sustain the indifferences in the high-sensitivity limit. These insights demonstrate that, although we are considering decision problems, the steady-states of our learning model are better understood as equilibria rather than as the outputs of a maximization problem.

[3]An alternative perspective is that a third party collects the data and organizes it into predefined categories. While this view aligns with the experimental framework in Jehiel and Singh (2021), it is less relevant to our setting, which focuses on a single learning agent. Other potential mechanisms for aggregation include limited memory, unlabeled data, or prevailing narratives.

Our main contributions focus on analyzing the convergence properties of our learning model, particularly in the high-sensitivity limit. We start with the case of decision-makers equipped with at most two similarity classes where we provide a complete characterization - establishing convergence of the learning dynamics to the set of steady-states in the high-sensitivity limit. Specifically, we demonstrate that there always exists a steady-state that is locally asymptotically stable under the CQL dynamics, meaning that if the initial assessments are close to those at the steady-state, the learning dynamics will converge to it. Furthermore, if the steady-state is unique, whether pure or mixed, it is globally asymptotically stable, meaning the learning dynamics will converge to it regardless of the initial conditions.

We then consider the general case of decision-makers with an arbitrary number of similarity classes and prove the following results. We verify that if a strict Valuation Equilibrium (VE) employing pure strategies exists, the steady-state that arises in its vicinity for a sufficiently large sensitivity parameter is locally asymptotically stable under the CQL dynamics.

Our leading convergence results are characterized by varying the payoffs associated with trivial decision problems where all available alternatives belong to the same similarity class. When such payoffs are sufficiently high,[4] we demonstrate the emergence of a unique mixed steady-state, whose limit as the sensitivity parameter approaches infinity involves an equalization of assessments across at least two similarity classes. We establish that this unique mixed steady-state, whose limit features indifference(s) and a refinement of mixed VE, is globally asymptotically stable in the CQL dynamics. By contrast, when trivial decision problems yield sufficiently low payoffs[5], we show the existence of multiple steady-states, with at least one corresponding to a strict pure valuation equilibrium in the high-sensitivity limit. This steady-state is locally asymptotically stable for a sufficiently large sensitivity parameter, as established by our general local stability result for strict pure steady-states.

1.1 Literature Review

Our paper relates to various branches of literature. In the literature on learning, there is a long-standing practice of studying the stability properties of rest-points, beginning with the literature on fictitious play (Brown, 1951), where Shapley (1964) provided an early example of non-convergence. Convergence results in this domain were obtained for 2×2 games (Brown,

[4]Considering the additional cognitive cost associated with non-trivial choices (choice overload), this assumption is especially germane when differences in material payoffs are small, with the cognitive cost outweighing these differences. We believe this in particular applies to most lab experiments on complex decision-making.

[5]Our results on the stability of mixed equilibria do not address the cases where the payoffs attached to trivial decision problems are at intermediate levels. We leave the exploration of these scenarios for future research.

1951), two-player zero-sum games (Robinson, 1951), dominance solvable games (Nachbar, 1990) and potential games (Monderer and Shapley, 1996b,a). The classical fictitious play model was extended to allow for stochastic (smooth) best-responses using the same logit formulation that we use (Fudenberg and Kreps, 1993; Fudenberg and Levine, 1998; Hofbauer and Sandholm, 2002). The continuous-time long-run approximation of such models, developed by Benaïm (1999), is adapted to and utilized in our learning model. Local stability results are derived by linearizing the obtained "mean-field" differential equations around the rest-points and verifying whether the real parts of the eigenvalues of the corresponding Jacobian matrix are all negative. Global stability results are established by either constructing a strict Lyapunov function that decays along all non-constant trajectories of the learning dynamics or leveraging the properties of cooperative dynamical systems (Smith, 1995).

Our learning model aligns with the tradition of reinforcement learning (RL) models in economics (Roth and Erev, 1995; Börgers and Sarin, 1997; Erev and Roth, 1998). However, our updating scheme involves a weighted average of the observed payoff and the previous assessment, while conventional RL models update propensity by directly adding the observed payoff. Besides the Payoff-Assessment Learning model (Sarin and Vahid, 1999; Cominetti et al., 2010), Q-learning models (Watkins and Dayan, 1992; Sutton and Barto, 2018) also use a similar weighted average updating rule. A key innovation in our learning model is the use of coarse categories instead of treating each alternative separately[6]. In our setting, if alternatives were treated individually, convergence toward a nearly optimal strategy would be trivial (as shown in the smooth learning model in Sarin and Vahid (1999)). Our result—that persistent mixing may exist in a decision problem with generic payoffs, even as the decision-maker becomes extremely sensitive to differences in assessments, and that such behavior is globally stable within the learning dynamics—has no counterpart in the literature.

Regarding categorization and its equilibrium consequences, beyond the Valuation Equilibrium (Jehiel and Samet, 2007), the Analogy-based Expectation Equilibrium (Jehiel, 2005) is also worth mentioning. In this game-theoretic setting, players lump several states together into coarse categories to form aggregated beliefs about opponents' behavior.[7]

In Section 2 of the paper, we present the setup and the learning model, providing a continuous-time approximation as well as demonstrating the formal link between steady-states of the

[6]To an extent, our two-step choice procedure echoes some features of deep learning (Mnih et al., 2015). The first step - selecting a category based on assessments using the logit (softmax) rule, parallels high-level decision-making in neural networks, where abstracted features guide overall choices. The second step - randomly selecting an alternative within the chosen category due to a lack of differentiation, mirrors low-level actions in hierarchical models, where specific actions are executed based on higher-level decisions.

[7]It can be viewed as the fictitious play counterpart of the approach developed in Jehiel and Samet (2007).

CQL model and Valuation Equilibrium. Section 3 provides an example with multiple steady-states and another with a unique steady-state whose limit as the decision-maker almost surely selects the similarity class(es) with the highest assessment(s) involves mixing amid indifference. In Sections 4 and 5, we present our main analytical results for scenarios where the payoffs associated with trivial decision problems in which all alternatives belong to the same similarity class, are either sufficiently high or sufficiently low. Section 6 concludes the paper, highlighting open questions and avenues for future research.

2 Model

The Coarse Q-learning (CQL) model is tailored for complex choice environments characterized by individuals repeatedly tasked with evaluating the potential outcomes of their decisions amid a multitude of options and uncertain states of the world. In such settings, where the vast array of alternatives renders a detailed evaluation of each potential outcome across different states infeasible, decision-makers might organically resort to categorical models that help alleviate the inherent complexity.

2.1 Categorization

A natural approach ingrained in human psychology is that decision-makers bundle several alternatives together into coarse categories based on perceived analogies and focus on learning about their collective payoffs. We refer to these coarse subsets as *similarity classes*. These are assumed to be exogenously pre-defined in this paper. The set of similarity classes partitions the space of the agent's alternatives.

A concrete way to think of similarity classes is as follows. Consider alternatives characterized by a vector of attributes $x = (x_1, x_2, \ldots, x_N)$, where N is a large number representing the total number of attributes, and each attribute $x_i \in \{0, 1\}$. The agent considers only a non-empty proper subset $N^s \subset \{1, 2, \ldots, N\}$ of these attributes as salient. What is deemed salient may stem from cultural or psychological factors. A similarity class is parameterized by $\hat{x}(N^s) = (\hat{x}_i)_{i \in N^s}$, where each $\hat{x}_i \in \{0, 1\}$ for every $i \in N^s$. An alternative $x = (x_1, x_2, \ldots, x_N)$ is assigned to the similarity class $\hat{x}(N^s)$, if $x_i = \hat{x}_i$ for all $i \in N^s$. Our learning model assumes that the agent monitors only the salient attributes of the chosen alternatives and their resulting payoffs. Consequently, she reinforces the valuations of similarity classes instead of those of individual alternatives.

2.2 Setup

We imagine a setting where a myopic agent, Alice, faces a stage decision problem with generic payoffs, repeated infinitely. The stage problem can be described by a decision tree $\mathcal{T}(\mathbf{\Psi}, \mathbf{f}, \mu)$, characterized as follows. At the root, nature draws a state $\psi \in \mathbf{\Psi}$ at random according to a fixed probability mass function $\mathbf{f}(\psi)$. We assume that $\mathbf{\Psi}$ is finite and non-empty. Let $\mathcal{C}_\psi$ denote the finite, non-empty set of alternatives available to Alice in state ψ. $\mathcal{C}$ denotes the grand set of alternatives available to Alice across all states, i.e., $\mathcal{C} = \bigcup_{\psi \in \mathbf{\Psi}} \mathcal{C}_\psi$. Finally, $\mu_\psi : \mathcal{C}_\psi \to \mathbb{R}$ denotes a bounded, real-valued, state-dependent payoff function. Alice receives a payoff of $\mu_\psi(c) \in \mathbb{R}$ for picking an alternative $c \in \mathcal{C}_\psi$ in state ψ.

We define a similarity transformation as an equivalence relation on the grand set of alternatives $\mathcal{C}$. Let $\mathcal{S}$ represent the finite set of similarity (equivalence) classes available to Alice, such that $\mathcal{S}$ forms a partition of $\mathcal{C}$. For each alternative $c \in \mathcal{C}$, let $\Gamma(c)$ denote the similarity class $s \in \mathcal{S}$ that contains c. Thus, $\Gamma : \mathcal{C} \to \mathcal{S}$ is a similarity mapping. By definition, we have $|\mathcal{S}| \leq |\mathcal{C}|$, and let $|\mathcal{S}| = n$ where $n \in \mathbb{N}$. We focus on non-trivial similarity relations, where $1 < n < |\mathcal{C}|$. The set $\mathcal{S}$ of similarity classes and the mapping Γ available to Alice are assumed to be exogenously defined in this paper. How does the introduction of the similarity mapping help Alice simplify her complex choice problem? First, her grand choice set is now reduced to $\mathcal{S}$ instead of the larger set $\mathcal{C}$. Furthermore, in any state $\psi \in \mathbf{\Psi}$, the set of distinct options (similarity classes) available to Alice, denoted by $\mathcal{S}_\psi$, is a non-empty set constructed by applying the similarity mapping Γ to each alternative $c \in \mathcal{C}_\psi$. Specifically, $\mathcal{S}_\psi = \bigcup_{c \in \mathcal{C}_\psi} \Gamma(c)$. Finally, $|\mathcal{S}_\psi| \leq |\mathcal{C}_\psi|$, since each alternative belongs to exactly one similarity class, though multiple alternatives may belong to the same similarity class.

In our learning model, when faced with a state where multiple alternatives from different similarity classes are available, Alice first chooses among the similarity classes spanned by the available alternatives. If the chosen similarity class contains several alternatives, she then picks randomly and uniformly among them.[8] This allows us to simplify the original decision tree $\mathcal{T}(\mathbf{\Psi}, \mathbf{f}, \mu)$ into an equivalent decision tree $\mathcal{T}'(\mathbf{\Omega}, \mathbf{p}, \pi)$ where the state space is composed of all non-empty subsets of $\mathcal{S}$, with probabilities and payoffs redefined accordingly.[9] Formally, let $\omega \in \mathbf{\Omega}$ be a representative state where $\mathbf{\Omega} = \mathcal{P}(\mathcal{S}) \setminus \{\emptyset\}$. $\mathbf{p}$ denotes a probability mass function over $\omega \in \mathbf{\Omega}$. $\pi_\omega(s)$ is the expected payoff associated with choosing a similarity class s available in state ω. $\mathbf{p}$ and π are related to $\mathbf{f}$ and μ, respectively, as follows. Define

[8] By incorporating a stochastic choice policy at the level of similarity classes, our two-stage procedure resembles the nested logit model often used in discrete choice frameworks. It also provides a straightforward resolution to the red/blue bus paradox (Anderson et al., 1992): when an alternative is duplicated, the duplicates are naturally treated as part of the same similarity class, leaving the overall choice unaffected.

[9] A detailed algorithm for transforming $\mathcal{T}(\mathbf{\Psi}, \mathbf{f}, \mu)$ into $\mathcal{T}'(\mathbf{\Omega}, \mathbf{p}, \pi)$ can be found in the Online Appendix.

for each $\omega \in \mathbf{\Omega}$,

$$\mathbf{\Psi}(\omega) = \{\psi \in \mathbf{\Psi} : \Gamma(\mathcal{C}_\psi) = \omega\},$$

where $\Gamma(\mathcal{C}_\psi) = \mathcal{S}_\psi = \{\Gamma(c) : c \in \mathcal{C}_\psi\}$. Essentially, for any given non-empty subset $\omega \subseteq \mathcal{S}$, we consider $\mathbf{\Psi}(\omega)$ that represents all the states $\psi \in \mathbf{\Psi}$ in the original decision tree where the available alternatives $\mathcal{C}_\psi$ span ω. Trivially, $\mathcal{S}_\omega = \omega$ in the decision tree $\mathcal{T}'(\mathbf{\Omega}, \mathbf{p}, \pi)$.

The probability of state ω should then be the probability that $\psi \in \mathbf{\Psi}(\omega)$. Hence,

$$p(\omega) = \sum_{\psi \in \mathbf{\Psi}(\omega)} f(\psi). \tag{1}$$

Regarding the payoff specification in $\mathcal{T}'$, consider a given $\psi \in \mathbf{\Psi}(\omega)$ and an arbitrary $s \in \omega$. There are $|\mathcal{C}_\psi \cap \Gamma^{-1}(s))|$ alternatives in $\mathcal{C}_\psi$ that correspond to the same similarity class s. Since we assume that Alice cannot distinguish among alternatives within the same similarity class, she will randomize uniformly among them when multiple such alternatives are available in a given state. Thus, the expected payoff obtained from choosing an alternative in the similarity class s in the state ψ will simply be $\sum_{c \in \mathcal{C}_\psi \cap \Gamma^{-1}(s)} \mu_\psi(c)/|(\mathcal{C}_\psi \cap \Gamma^{-1}(s))|$. Finally, averaging these payoffs over all the states $\psi \in \mathbf{\Psi}(\omega)$ yields

$$\pi_\omega(s) = \frac{\sum_{\psi \in \mathbf{\Psi}(\omega)} f(\psi) \sum_{c \in \mathcal{C}_\psi \cap \Gamma^{-1}(s)} \mu_\psi(c)/|(\mathcal{C}_\psi \cap \Gamma^{-1}(s))|}{\sum_{\psi \in \mathbf{\Psi}(\omega)} f(\psi)}. \tag{2}$$

To illustrate this transformation, we consider the following example. A consumer, Bob, is confronted with three binary choice problems, involving apples, lemons, and limes, each occurring with equal probability (Fig. 1). However, due to his color-blindness, he cannot distinguish between lemons and limes and groups them together under the category of "citrus fruits". Consequently, the bundling reduces his decision-making process into a binary choice problem between apples and citrus fruits (with probability $\frac{2}{3}$), and a trivial unary choice involving only citrus fruits (with probability $\frac{1}{3}$), as seen in Fig. 2.

2.3 Learning Dynamics

We introduce *valuations*, denoted by v, as real-valued functions defined on the set of similarity classes, i.e., $v : \mathcal{S} \to \mathbb{R}$. They represent Alice's assessment of the expected payoff performance of each similarity class available to her. When called upon to make a choice, Alice identifies each available alternative with the similarity class it belongs to. Alice's valua-

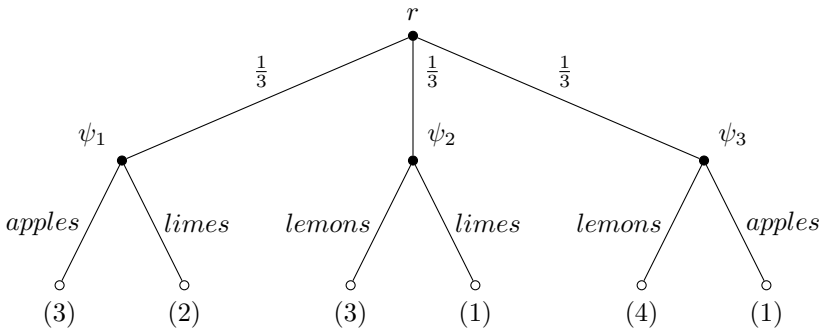

Figure 1: Bob's original decision tree

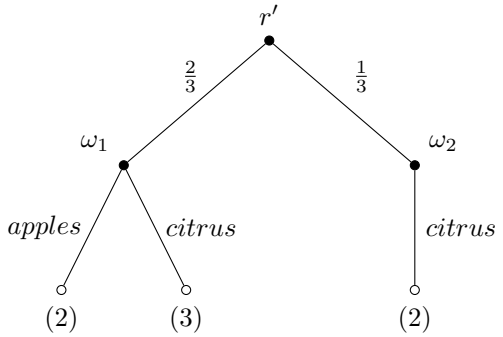

Figure 2: Bob's simplified decision tree

tions determine her strategy in each stage.[10] Once an alternative is chosen, its corresponding payoff is observed, and the valuation of the similarity class containing the chosen alternative is updated based on the observed payoff.

More precisely, Alice encounters a stage decision tree $\mathcal{T}'$ repeated infinitely. At each stage $k \in \mathbb{N} \cup \{0\}$, nature presents Alice with a choice problem $\omega \in \Omega$ with probability $p(\omega)$. The set of similarity classes available to Alice at node ω is denoted by $\mathcal{S}_\omega \subseteq \mathcal{S}$. We refer to this set as the set of her pure strategies at node ω. $\pi_\omega : \mathcal{S}_\omega \to \mathbb{R}$ is a bounded, generic payoff function at node ω. We maintain the same notation to refer to the multi-linear extension of payoffs to the set of her mixed strategies. $\mathbf{\Delta}_\omega$ denotes the set of mixed strategies (probability vectors) over $\mathcal{S}_\omega$. Alice makes her choice at node ω in period k employing a mixed strategy $\delta_{\omega,k} = \sigma_{\omega,k}(\mathbf{v}_k) \in \mathbf{\Delta}_\omega$. Here, $\mathbf{v}_k = (v_k^s)^{s \in \mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$ is a vector of valuations that reflects her assessment of the payoff potential associated with each similarity class at stage k. We use the logit stochastic choice model to characterize the mapping from the space of valuations

[10] Alice perceives her available alternatives coarsely, treating all options within a similarity class as indistinguishable. Thus, following the principle of indifference, she uniformly randomizes among them. During the learning process, she may select different alternatives from the same similarity class across periods, experiencing varying payoffs each time. However, she interprets these payoffs as random fluctuations around the expected payoff of the similarity class and bases her decisions exclusively on this expected value.

to the space of mixed strategies.[11] Thus, given a valuation vector $\mathbf{v}_k$, the probability that Alice chooses an alternative in similarity class $s \in \mathcal{S}_\omega$ at node ω in stage k is given by the real-analytic (C^ω) function:

$$\sigma^s_{\omega,k}(\mathbf{v}_k) = \frac{\exp(\beta v^s_k)}{\sum_{j\in\mathcal{S}_\omega} \exp(\beta v^j_k)}, \tag{3}$$

where $\beta \geq 0$ is a scaling constant determining Alice's sensitivity to differences in valuations.[12] It has a smoothing effect with $\beta = 0$ leading to a trivial uniform random choice, while for $\beta \to \infty$, the probabilities concentrate on the similarity class(es) with the highest valuation(s). Most of the analysis in this paper is conducted in the high-sensitivity limit, i.e., as $\beta \to \infty$. In this limit, the myopic decision-maker, Alice, is guaranteed to almost surely choose an alternative in the similarity class(es) with the highest current valuation(s).

Once Alice makes a choice s based on her mixed strategy $\delta_{\omega,k}$ at node ω in stage k, she observes only the realized payoff $\pi_k = \pi_\omega(s)$. Crucially, she gains no insight into her foregone (counterfactual) payoffs. She then uses this novel information to update her valuation of the similarity class she selected, while leaving the valuations of the remaining strategies unchanged, following a simple iterative weighted averaging scheme:

$$v^s_{k+1} = \begin{cases} (1-\alpha_k)v^s_k + \alpha_k \pi_k & \text{if } s = s_k \\ v^s_k & \text{otherwise.} \end{cases}$$

The reinforcement update rule can also be written in vector notation as

$$\mathbf{v}_{k+1} - \mathbf{v}_k = \alpha_k[\tilde{\mathbf{v}}_k - \mathbf{v}_k] \tag{4}$$

where,

$$\tilde{v}^s_k = \begin{cases} \pi_k & \text{if } s = s_k, \\ v^s_k & \text{otherwise.} \end{cases}$$

Here, $\alpha_k \in (0,1)$ is a sequence of averaging factors such that $\sum_k \alpha_k = \infty$ and $\sum_k \alpha_k^2 < \infty$ (Kushner and Yin, 2003). These conditions are satisfied, for e.g., by setting $\alpha_k = (k+1)^{-1}$, meaning that the valuation v^s_k becomes the simple average of all valuations for similarity

[11] The logit choice model has been widely used in stochastic fictitious play models of learning (Fudenberg and Levine, 1998; Fudenberg and Kreps, 1993; Hofbauer and Sandholm, 2002), QRE models in game theory (McKelvey and Palfrey, 1995; Goeree et al., 2016), and in discrete choice models in the empirical literature (Hausman and McFadden, 1984; Anderson et al., 1992). It's also commonly seen in the literature on reinforcement learning and deep learning, where it's typically referred to as the softmax function.

[12] β can also be interpreted as an inverted noise parameter or as the inverse of the absolute temperature.

class s up to time k. This approach effectively causes Alice's sensitivity to new observations to diminish over time, while ensuring that future observations still exert a non-negligible impact. While Alice chooses her strategy based on her current valuation $\mathbf{v}_k$, the resulting payoffs, and thus the random vector $\tilde{\mathbf{v}}_k$, are influenced by these valuations. Equations (3) and (4) together describe a non-homogeneous Markov process that captures how her valuations evolve over time. This process can be interpreted as Alice exploring various similarity classes to understand their potential rewards and adjusting her future choices exploiting what she has learned. The transition from one valuation $\mathbf{v}_k$ to the next $\mathbf{v}_{k+1}$ involves several steps: starting with her initial valuation, Alice employs a mixed strategy that assigns higher probabilities to alternatives in similarity classes with higher valuations, makes her choice, observes her realized payoff, and then updates her valuation for the relevant similarity class in the direction of the observed payoff. This exercise is iterated indefinitely generating a discrete-time stochastic process.

The discrete-time Coarse Q-learning (CQL) model is fully described by the equations (3) and (4) together with an initial value of the valuation vector $\mathbf{v}_0$. The initial valuations can be interpreted as her prior assessments of the expected payoff for alternatives within each similarity class. Now, upon dividing the reinforcement update rule in Eq. (4) by an infinitesimal α_k, the iterative method in the long-run can be seen as a Finite Difference Euler scheme for an associated system of differential equations. However, there's a twist: the R.H.S. of our equation is not deterministic but a random field. Building on this insight, the theory of stochastic approximation (Benaïm, 1999; Benaïm et al., 2005) has developed techniques that establish the fundamental connections between the asymptotics of the discrete-time random process in Eq. (4) as $k \to \infty$ and the asymptotics of the deterministic continuous-time averaged dynamics as $t \to \infty$, given by:

$$\dot{\mathbf{v}} = E_\sigma(\tilde{\mathbf{v}}|\mathbf{v}) - \mathbf{v} \tag{5}$$

where $E_\sigma(\tilde{\mathbf{v}}|\mathbf{v})$ characterizes the expected payoffs of the similarity classes induced by the mixed strategy probabilities σ (specified by the logit choice rule). These techniques show that one can characterize the limiting behavior of the stochastic discrete-time CQL process in terms of a continuous-time ordinary differential equation defined by the expected motion of the stochastic process. More precisely, any sequence $\mathbf{v}_k$ generated by Eq. (4) must remain bounded, since the payoffs $\pi_\omega(s)$ are bounded. Therefore, using Propositions 4.1 & 4.2 along with the Limit Set Theorem (5.7) in Benaïm (1999), we establish that the ω-limit set[13] of

[13]The ω-limit set of a discrete-time stochastic process $\{\mathbf{X}_k\}_{k=0}^{\infty}$ is the set of all points $\mathbf{x}$ in the state space for which there exists a subsequence $\{k_n\}$ with $k_n \to \infty$ such that $\mathbf{X}_{k_n} \to \mathbf{x}$ almost surely.

any realization of the stochastic discrete-time CQL dynamics in Equations (3) and (4) is almost surely a compact, connected, internally chain transitive set[14] of the deterministic continuous-time averaged CQL dynamics in Eq. (5).

Expanding the expected payoff terms in Eq. (5), $\forall\ s \in \mathcal{S}$, the evolution of the valuation v_s in the continuous-time CQL model is governed by the system of coupled ODEs:

$$\dot{v}_s = f_s(\mathbf{v}) = g_s(\mathbf{v}) - v_s, \tag{6}$$

where,

$$g_s(\mathbf{v}) = \frac{\sum_{\omega\in\Omega: s\in\mathcal{S}_\omega} p(\omega)\sigma^s_\omega(\mathbf{v})\pi_\omega(s)}{\sum_{\omega\in\Omega: s\in\mathcal{S}_\omega} p(\omega)\sigma^s_\omega(\mathbf{v})},$$

$$\sigma^s_\omega(\mathbf{v}) = \frac{\exp(\beta v_s)}{\sum_{j\in\mathcal{S}_\omega}\exp(\beta v_j)}.$$

First, we observe that in light of the continuous-time mean dynamics derived above, our transformation of the decision tree from $\mathcal{T}$ to $\mathcal{T}'$ by collapsing multiple analogous nodes into a single node using Alice's similarity partition is indeed without loss of generality. Second, we note that our choice of the logit rule to model the mixed strategy map leads to the expected payoff terms being non-polynomial and rules out any chance of spelling out explicit, algebraic solutions to the system of ODEs for $\beta > 0$. Of course, the ODE system is explicitly solvable for the trivial case of $\beta = 0$ where Alice uniformly randomizes among all available similarity classes at any node ω.[15] We present a general result on the existence of steady-state solutions of the ODE system in Eq. (6). 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 $\mathbf{v}^*$. We denote the set of steady-state solutions of the system of ODEs in Eq. (6) by $\mathcal{V}$. Applying Brouwer's fixed point theorem, we prove the following existence result in the Appendix.

Theorem 1. *The set $\mathcal{V}$ of steady-state solutions of the CQL dynamics is non-empty.*

Proof. The proof is relegated to Section A.1 of the Appendix. □

[14] The internally chain transitive (ICT) set is a stronger notion of the invariant set for dynamical systems that allows for the possibility of introducing asymptotically vanishing shocks in the dynamics. ICTs may include steady-states, periodic orbits and strange attractors (Conley, 1978).

[15] A trivial case where the ODE system is explicitly solvable, even for large β, occurs when Alice possesses the finest similarity partition, $\mathcal{S} = \mathcal{C}$. That is, she does not cluster her alternatives and instead forms a distinct valuation for each alternative. Consequently, the system of ODEs decouples and simplifies to a linear form with constant $g_s = \mu_\psi(s)$ terms. The solutions exhibit exponential decay in time and asymptotically approach the actual payoffs leading to nearly optimal choices in the long-run. This exercise illustrates that clustering alternatives into coarse similarity classes significantly impacts the learning dynamics.

2.4 Connections with Valuation Equilibrium

The steady-states of the continuous-time CQL model whose existence is guaranteed in Theorem 1 can be interpreted as smooth variants of Valuation Equilibria, first introduced in the context of multi-agent extensive form games by Jehiel and Samet (2007).

Definition 1 (**Smooth Valuation Equilibrium**). *A strategy profile* $\sigma = (\sigma_\omega^s)_{\omega\in\Omega}^{s\in\mathcal{S}}$ *constitutes a smooth valuation equilibrium for* $\mathcal{T}'$ *if there exists a valuation system* $(v_s)_{s\in\mathcal{S}}$ *s.t.*

$$v_s = \frac{\sum_{\omega\in\Omega:s\in\mathcal{S}_\omega} p(\omega)\sigma_\omega^s(v)\pi_\omega(s)}{\sum_{\omega\in\Omega:s\in\mathcal{S}_\omega} p(\omega)\sigma_\omega^s(v)}; \sigma_\omega^s = \frac{\exp(\beta v_s)}{\sum_{j\in\mathcal{S}_\omega}\exp(\beta v_j)}.$$

We readily verify that any steady-state of the CQL model is a Smooth Valuation Equilibrium (SVE), and vice versa. For a finite β, any SVE is fully-mixed, meaning the corresponding vector of valuations lies in the interior of the convex hull of the payoffs. That is, at any node, Alice selects each available similarity class with strictly positive probability, according to the logit choice rule. However, as the logit parameter grows without bound ($\beta \uparrow \infty$), Alice becomes highly sensitive to differences in valuations. In this limit, the logit choice rule almost surely selects the similarity class(es) with the highest valuation(s), $s \in \arg\max_{s\in\mathcal{S}_\omega} v_s$, at any node $\omega \in \Omega$.[16] This property is used to show that as $\beta \uparrow \infty$, the corresponding smooth valuation equilibria lie in an arbitrarily small neighborhood of some valuation equilibrium. The latter requires that in each state ω, Alice chooses similarity class(es) $s \in \arg\max_{s\in\mathcal{S}_\omega} v_s$ and her valuations are consistent in the sense that $v_s = \frac{\sum_{\omega\in\Omega:s\in\mathcal{S}_\omega} p(\omega)\sigma_\omega^s(v)\pi_\omega(s)}{\sum_{\omega\in\Omega:s\in\mathcal{S}_\omega} p(\omega)\sigma_\omega^s(v)}$.

Lemma 2.1. *The smooth valuation equilibria (SVE) of the CQL dynamics converge to valuation equilibria (VE) as the sensitivity parameter* $\beta \to \infty$. *Specifically, for any* $\epsilon > 0$, $\exists$ $\hat{\beta} \in \mathbb{R}_+$ *such that for almost all* $\beta > \hat{\beta}$, *except possibly on a measure zero set, every SVE lies within an* ϵ*-neighborhood of a VE and varies smoothly with* β.

Proof. The proof relegated to Section A.2 of the Appendix. □

It's worth mentioning that while each fixed point of the CQL dynamics in the high-sensitivity

[16] An alternative interpretation of the logit heuristic is that it approximates a noisy (stochastic) choice strategy that Alice employs based on her perceived payoff performance of each similarity class. While this approach allows Alice to make errors in choosing her optimal similarity class, it penalizes these errors in proportion to their severity. Specifically, the penalty incurred for making a sub-optimal choice is exponentially proportional to the payoff loss associated with that choice. As β increases indefinitely, the cost of these mistakes becomes prohibitively high, driving Alice's behavior toward the optimal strategy.

limit is a valuation equilibrium, there exist valuation equilibria in certain decision trees that cannot be characterized as the limiting fixed points of the continuous-time CQL dynamics as $\beta \uparrow \infty$. An example illustrating this point is provided in Sec. 3.1 of the Online Appendix. Consequently, the set of smooth valuation equilibria in the high-sensitivity limit ($\beta \uparrow \infty$) of the CQL model offers a *refinement* of the set of VE in a decision tree.

Our primary focus in this paper is on the asymptotic convergence properties of the CQL dynamics. Eq. (6) defines a real, autonomous, smooth dynamical system on the space of valuations. At any time, $t \in \mathbb{R}$, the state of the dynamical system is given by a vector of valuations, $\mathbf{v}(t) \in \mathbb{R}^{\mathcal{S}}$. The rest-points of the dynamical system are elements $\mathbf{v}^*$ of the set $\mathcal{V}$ such that the time-derivative equals zero at these points. In the following sections, we investigate the asymptotic stability[17] of the rest-points. For local stability, we examine the effects of small perturbations on the long-run behavior of the CQL model in neighborhoods of its rest-points. In particular, we linearize Eq. (6) around the rest-points and analyze the sign of the real parts of the eigenvalues of the corresponding Jacobian matrices to determine the local asymptotic stability of the CQL model at its rest-points. For global stability results of the continuous-time process, we either construct a strict Lyapunov function that decays along all non-constant trajectories of the CQL dynamics or leverage the convergence properties of monotone cooperative dynamical systems (Smith, 1995).

Pemantle (1990) shows that a discrete-time stochastic system as in Eq. (4) has a probability zero of converging to a linearly unstable steady-state of the continuous-time process in Eq. (6), provided that there is a non-negligible amount of noise in the evolution of every component of the system. Benaïm (1999) shows that every locally asymptotically stable steady-state of the continuous-time process has strictly positive probability of being the long-run outcome of the discrete-time process, again provided that there is non-negligible noise in the system. Benaïm (1999) also shows that if the continuous-time process 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 process converges to it almost surely. In light of these results, we focus the remainder of the paper on analyzing the convergence properties of the continuous-time CQL dynamics in Eq. (6).

[17] An equilibrium of a dynamical system is *locally asymptotically stable* if, for any initial condition sufficiently 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 solutions of the dynamical system, regardless of the initial conditions, asymptotically converge to the equilibrium. Refer to Sec. 4.1. of the Online Appendix for a formal treatment of asymptotic stability and the Hartman-Grobman (Linearization) Theorem.

3 Illustrations

We illustrate the dynamics of the CQL model using several examples based on a decision tree shown in Fig. 3 where the decision-maker Alice operates with two similarity classes. At the root r, nature chooses one of three nodes ω_1, ω_2 and ω_3, each with equal probability. At node ω_1, Alice encounters a binary choice between alternatives L_1 and R_1. At nodes ω_2 and ω_3, Alice encounters trivial unary choices, involving L_2 and R_3 respectively. The set of alternatives is partitioned into two similarity classes, $L = \{L_1, L_2\}$ and $R = \{R_1, R_3\}$. We examine three distinct scenarios by altering the payoffs associated with the alternatives at the trivial unary choice nodes, specifically $\mathbf{z_2}$ for L_2 at ω_2 and $\mathbf{z_3}$ for R_3 at ω_3, while keeping the payoffs for the alternatives at the binary choice node ω_1 constant.

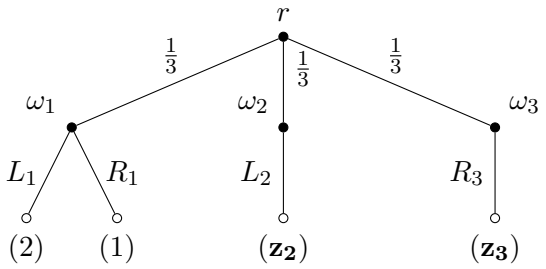

Figure 3: Example of a Decision Tree with Two Similarity Classes

3.1 Example: Multiplicity of SVE

Here, we assume $\mathbf{z_2} = \mathbf{z_3} = 0$. This implies that Alice receives a strictly lower reward at each of the unary choice nodes compared to the binary choice node, regardless of her actions at the latter. As a result, there are three valuation equilibria: two pure and one mixed.

- The pure strategy that selects the alternative in L at each of the nodes ω_1 and ω_2 is a strict pure VE. The corresponding valuation vector is $(v_L = 1, v_R = 0)$ and the strategy is optimal for this valuation. We verify, by direct computation, that the valuation $(1, 0)$ is a limiting steady-state of the CQL model as $\beta \uparrow \infty$. The numerical simulation seen in Fig. 4 points to the same. Moreover, with a large sensitivity parameter ($\beta = 50$), we observe strong evidence of convergence to the steady-state starting from a nearby initial valuation system.

- The pure strategy that selects an alternative in R at each of the nodes ω_1 and ω_3 is a strict pure VE. The corresponding valuation vector is $(v_L = 0, v_R = 0.5)$ and the strategy is optimal for this valuation. We verify, by direct computation, that the valuation $(0, 0.5)$ is a steady-state of the CQL model as $\beta \uparrow \infty$. The numerical

simulation seen in Fig. 5 points to the same. Moreover, with a large sensitivity parameter ($\beta = 50$), we observe strong evidence of convergence to the steady-state starting from a nearby initial valuation system.

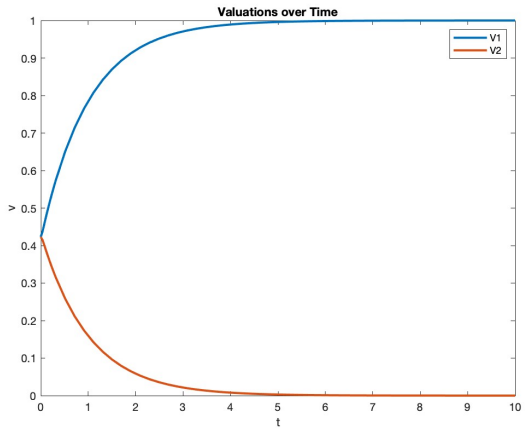

Figure 4: Stable Strict Pure SVE at $(1.0, 0.0)$; $\beta = 50$

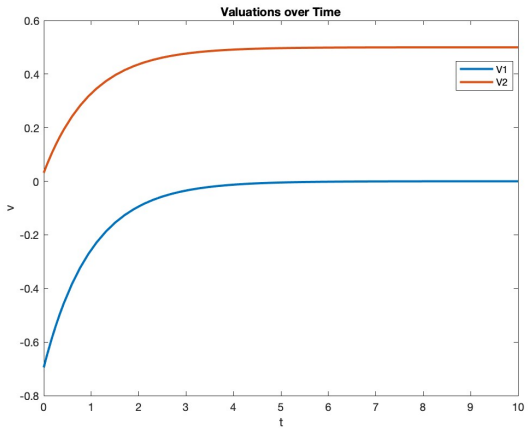

Figure 5: Stable Strict Pure SVE at $(0.0, 0.5)$; $\beta = 50$

- The mixed strategy that selects the alternative in L with probability $2 - \sqrt{3}$ and the alternative in R with probability $\sqrt{3} - 1$ at node ω_1 is a mixed VE. The corresponding valuation is $(v_L = 1 - \frac{1}{\sqrt{3}}, v_R = 1 - \frac{1}{\sqrt{3}})$ and the strategy is optimal for this valuation. We verify, by direct computation, that the valuation $(1 - \frac{1}{\sqrt{3}}, 1 - \frac{1}{\sqrt{3}}) \approx (0.423, 0.423)$ is a steady-state of the CQL model as $\beta \uparrow \infty$. However, from the numerical simulation presented in Fig. 4, it seems that the steady-state is unstable. A small perturbation to an initial valuation of $(0.423, 0.423)$ causes the valuation system to move away and eventually come to rest at one of the strict pure VE.[18]

3.2 Example: Unique Mixed SVE

In this example, we assume $\mathbf{z_2} = \mathbf{z_3} = 3$. That is, Alice receives a strictly higher reward at each of the unary choice nodes compared to the binary choice node, regardless of her actions at the latter. Consequently, the mixed strategy that selects the alternative in L with probability $\sqrt{3} - 1$ and the alternative in R with probability $2 - \sqrt{3}$ at node ω_1 constitutes the unique mixed valuation equilibrium (VE). The corresponding valuation is $(2 + \frac{1}{\sqrt{3}}, 2 + \frac{1}{\sqrt{3}})$ and the strategy is optimal for this valuation. We verify, by direct computation, that the valuation $(2 + \frac{1}{\sqrt{3}}, 2 + \frac{1}{\sqrt{3}}) \approx (2.577, 2.577)$ is a steady-state of the CQL model as $\beta \uparrow \infty$.

[18] Indeed, when we linearize the system around $(0.423, 0.423)$ and compute the eigenvalues of the corresponding Jacobian matrix in the $\beta \uparrow \infty$ limit, we find that one of the eigenvalues is positive. Thus, this mixed VE is an asymptotically unstable rest-point.

Moreover, numerical simulations, as shown in Fig. 6, indicate that this steady-state is asymptotically stable. With a sufficiently large sensitivity parameter ($\beta = 50$), there is strong evidence of convergence to the steady-state from an arbitrary initial valuation system.

3.3 Example: Unique Pure SVE

In this example, we assume $\mathbf{z_2} = 1$ and $\mathbf{z_3} = 0$. Consequently, there exists a unique strict pure valuation equilibrium (VE) where Alice selects an alternative in L at each of the nodes ω_1 and ω_2. The corresponding valuation is $(1.5, 0)$ and the strategy is optimal for this valuation. We verify, by direct computation, that the valuation $(1.5, 0)$ is a steady-state of the CQL model as $\beta \uparrow \infty$. Numerical simulations, as shown in Fig. 7, indicate that this steady-state is asymptotically stable for a large sensitivity parameter ($\beta = 50$).

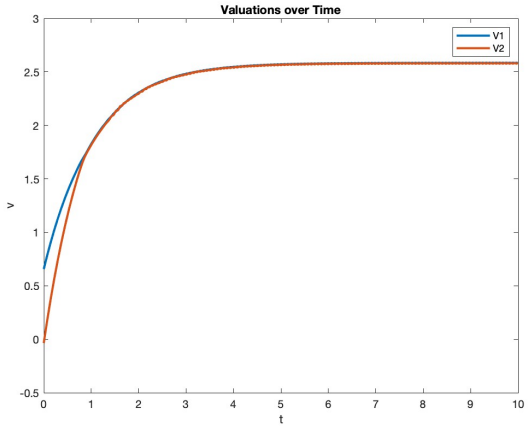

Figure 6: Stable Unique Mixed SVE at $(2.577, 2.577)$; $\beta = 50$

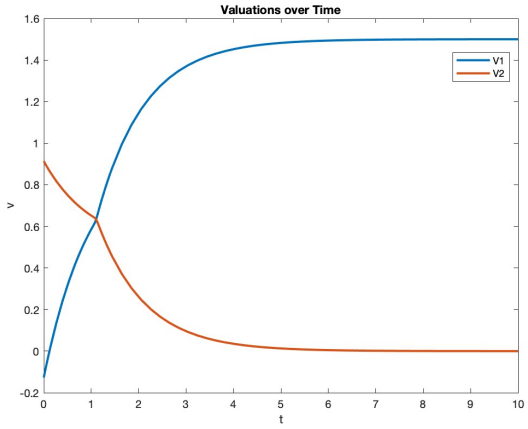

Figure 7: Stable Unique Strict Pure SVE at $(1.5, 0)$; $\beta = 50$

It is important to note that if the decision-maker possesses the finest partition available (i.e., she can distinguish each alternative independently), then multiple equilibria and mixed equilibria are impossible in a decision tree with generic payoffs in our learning model. Thus, the clustering of alternatives has non-trivial consequences on the dynamics of an otherwise standard reinforcement learning model. Finally, the examples suggest that an asymptotically stable equilibrium of the CQL model always exists, and that strict pure equilibria are asymptotically stable whenever they are present. In the next section, we formalize these insights and provide general results on the convergence of CQL dynamics when the DM operates with at most two similarity classes. Later in the paper, we analyze the general case with an arbitrary number of similarity classes.[19]

[19] Supplementary examples showing the workings of the CQL model with more than two similarity classes can be found in Sec. 3 of the Online Appendix.

4 Decision Trees with Two Similarity Classes

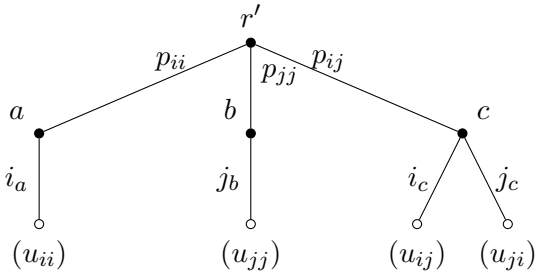

Figure 8: Decision Tree $\mathcal{T}_2^{'}$ with Two Similarity Classes

In this section, we restrict our attention to decision trees with generic payoffs where Alice has two similarity classes available to her.[20] We imagine a general decision tree $\mathcal{T}_2^{'}$ with generic payoffs (depicted in Fig. 8) where at the root r', nature selects one out of three possible nodes a, b and c with strictly positive probabilities p_{ii}, p_{jj} and p_{ij} respectively. At node a, Alice chooses an alternative in the similarity class i and she receives a payoff of $\pi_a(i_a) = u_{ii}$. At node b, she choose an alternative in the similarity class j and receives a payoff of $\pi_b(j_b) = u_{jj}$. Thus, Alice faces trivial unary choices at the nodes a and b. At node c, she faces a binary choice between an alternative in similarity class i (receiving a payoff of $\pi_c(i_c) = u_{ij}$) and an alternative in similarity class j (receiving a payoff of $\pi_c(j_c) = u_{ji}$). The set of alternatives is partitioned into two similarity classes $i = \{i_a, i_c\}$ and $j = \{j_b, j_c\}$.

Theorem 2. *There exists a finite $\hat{\beta}$, such that $\forall\ \beta > \hat{\beta}$, a decision tree $\mathcal{T}'$ with generic payoffs where an agent chooses among alternatives in at most two similarity classes admits a smooth valuation equilibrium that is locally asymptotically stable under the CQL dynamics. Additionally, if the equilibrium is unique, it is globally asymptotically stable.*

Proof. The proof is relegated to Section A.3 of the Appendix. □

5 Decision Trees with More than Two Similarity Classes

We extend our analysis to decision problems with generic payoffs where Alice has an arbitrary number of similarity classes available to her. Specifically, we let $|\mathcal{S}| = n$ where $n \in \mathbb{N} : n > 2$,

[20] The case of the coarsest similarity partition where Alice has only one available similarity class is trivial. She groups all her (indistinguishable) alternatives into a single equivalence class that yields a constant payoff, equal to the simple average of the payoffs of these alternatives across all states. The corresponding smooth dynamical system is linear and admits exponentially decaying solutions that asymptotically converge to the unique fully-mixed valuation equilibrium that involves uniform randomization among all available alternatives in every state of the world.

although we still maintain that n is finite and that at any node ω, $1 \leq |\mathcal{S}_\omega| \leq n$. Once again, we imagine a decision tree where at the root r', nature selects a node $\omega \in \Omega$ with probability $p(\omega)$, where $\Omega = \mathcal{P}(\mathcal{S}) \setminus \{\emptyset\}$. For Theorem 3, we additionally assume that the set of nodes $\{\omega \in \Omega : |\mathcal{S}_\omega| = 1\}$ is included in the support of the probability mass function p. That is, for each similarity class $s \in \mathcal{S}$, the trivial unary choice node featuring it is drawn by nature with a strictly positive probability. We recall that Theorem 1 guarantees the existence of a steady-state of the CQL model and Lemma 2.1 tells us that a steady-state may generically arise in an arbitrarily small neighborhood of either a strict pure VE, a partially-mixed VE or a fully-mixed VE for a sufficiently large sensitivity parameter.

5.1 Strict Pure SVE are Locally Asymptotically Stable

Firstly, we consider the strict pure valuation equilibria where the incentives to follow the corresponding optimal (pure) strategies are strict in every state of the world.

Theorem 3. *There exists a finite $\hat{\beta}$, such that $\forall\ \beta > \hat{\beta}$, in a decision tree $\mathcal{T}'$ with generic payoffs and an arbitrary number of similarity classes, the following holds: If there exists a strict pure valuation equilibrium, then the corresponding smooth valuation equilibrium that arises in its neighborhood is locally asymptotically stable under the CQL dynamics.*

Proof. Let $\mathbf{v}^* = (v_i^*)_{i \in \mathcal{S}}$ be a steady-state of the CQL model corresponding to a strict pure valuation equilibrium. In such an equilibrium, the valuations satisfy a strict total order: for all $i, j \in \mathcal{S}$ with $i \neq j$, we have $v_i^* \neq v_j^*$. To analyze the local asymptotic stability of $\mathbf{v}^*$, we examine the Jacobian matrix $\mathbf{J}$ of the ODE system in Eq. (6) evaluated at $\mathbf{v}^*$:

$$\mathbf{J}_{ij} = \frac{\partial}{\partial v_j}(g_i(\mathbf{v}) - v_i)\bigg|_{\mathbf{v}=\mathbf{v}^*} = \frac{\partial g_i}{\partial v_j}\bigg|_{\mathbf{v}=\mathbf{v}^*} - \delta_{ij},$$

where δ_{ij} is the Kronecker delta (1 if $i = j$, 0 otherwise). Our goal is to show that all eigenvalues of $\mathbf{J}$ have negative real parts for sufficiently large β. We first compute $\frac{\partial g_i}{\partial v_j}$ for all $i, j \in \mathcal{S}$. Since $g_i(\mathbf{v})$ depends on $\sigma_\omega^i(\mathbf{v})$, we need the partial derivatives of $\sigma_\omega^i(\mathbf{v})$ w.r.t. v_j.

For $\omega \in \Omega_i$ and $i \in \mathcal{S}_\omega$, the derivative of $\sigma_\omega^i(\mathbf{v})$ with respect to v_i is:

$$\frac{\partial \sigma_\omega^i}{\partial v_i}(\mathbf{v}) = \beta \sigma_\omega^i(\mathbf{v}) \left(1 - \sigma_\omega^i(\mathbf{v})\right) = \beta \frac{\exp(\beta v_i) \left(\sum_{s \neq i : s \in \mathcal{S}_\omega} \exp(\beta v_s)\right)}{\left(\sum_{s \in \mathcal{S}_\omega} \exp(\beta v_s)\right)^2}.$$

For $\omega \in \Omega_i$ and $i, j \in \mathcal{S}_\omega$, where $j \neq i$, the derivative of $\sigma_\omega^i(\mathbf{v})$ with respect to v_j, is:

$$\frac{\partial \sigma_\omega^i}{\partial v_j}(\mathbf{v}) = \frac{\partial \sigma_\omega^j}{\partial v_i}(\mathbf{v}) = -\beta \sigma_\omega^i(\mathbf{v}) \sigma_\omega^j(\mathbf{v}) = -\beta \frac{\exp(\beta(v_i + v_j))}{\left(\sum_{s \in \mathcal{S}_\omega} \exp(\beta v_s)\right)^2}.$$

As $\beta \to \infty$, the steady-state corresponds to a strict pure valuation equilibrium at $\mathbf{v} = \mathbf{v}^*$ and the choice probabilities $\sigma_\omega^i(\mathbf{v}^*)$ become deterministic. If $v_i^* > v_j^*$ for all $j \in \mathcal{S}_\omega \setminus \{i\}$, then $\sigma_\omega^i(\mathbf{v}^*) \to 1$ and $\sigma_\omega^j(\mathbf{v}^*) \to 0$, as $\beta \uparrow \infty$. If $v_i^* < v_j^*$ for some $j \in \mathcal{S}_\omega$, then $\sigma_\omega^i(\mathbf{v}^*) \to 0$, and $\sigma_\omega^j(\mathbf{v}^*) \to 1$, as $\beta \uparrow \infty$. In either case,

$$\lim_{\beta \to \infty} \frac{\partial \sigma_\omega^i}{\partial v_i}(\mathbf{v}) \bigg|_{\mathbf{v}^*} = \lim_{\beta \to \infty} \beta \sigma_\omega^i(\mathbf{v}^*) \left(1 - \sigma_\omega^i(\mathbf{v}^*)\right) = 0$$

$$\lim_{\beta \to \infty} \frac{\partial \sigma_\omega^i}{\partial v_j}(\mathbf{v}) \bigg|_{\mathbf{v}^*} = \lim_{\beta \to \infty} -\beta \sigma_\omega^i(\mathbf{v}^*) \sigma_\omega^j(\mathbf{v}^*) = 0$$

Since the derivatives of $\sigma_\omega^i(\mathbf{v}^*)$ tend to zero as $\beta \uparrow \infty$ and because $g_i(\mathbf{v})$ is a weighted average of $\pi_\omega(i)$ with weights involving $\sigma_\omega^i(\mathbf{v})$, it follows from the quotient rule and the chain rule:

$$\lim_{\beta \to \infty} \frac{\partial g_i}{\partial v_j} \bigg|_{\mathbf{v} = \mathbf{v}^*} = 0, \quad \forall i, j \in \mathcal{S}.$$

Therefore, as $\beta \to \infty$, the Jacobian matrix $\mathbf{J}$ evaluated at $\mathbf{v} = \mathbf{v}^*$

$$\mathbf{J}^* = \mathbf{J}|_{\mathbf{v} = \mathbf{v}^*} \to -\mathbf{I}_n,$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Since $\mathbf{J}^* \to -\mathbf{I}_n$, all eigenvalues λ_i of $\mathbf{J}^*$ satisfy:

$$\lambda_i \to -1 \quad \text{as} \quad \beta \to \infty.$$

By continuity of eigenvalues with respect to the entries of the Jacobian matrix (that are smooth functions of β), there exists a finite $\hat{\beta} > 0$ such that for all $\beta > \hat{\beta}$, all eigenvalues λ_i of $\mathbf{J}^*$ satisfy: $\mathrm{Re}(\lambda_i) < 0$ implying $\mathbf{J}^*$ is invertible. By the Implicit Function Theorem, we establish that near a strict pure valuation equilibrium, there exists a locally unique steady-state $\mathbf{v}^*(\beta)$ that varies smoothly with β for sufficiently large $\beta > \hat{\beta}$. Moreover, since all eigenvalues of the Jacobian matrix evaluated at the steady-state $\mathbf{v}^*$ have negative real parts for $\beta > \hat{\beta}$, the steady-state is locally asymptotically (exponentially) stable under the CQL dynamics by the Hartman–Grobman (linearization) theorem. □

5.2 When Trivial Choice Payoffs are Large

In this part, we consider the case where the payoffs associated with states that consist of a single similarity class are sufficiently large relative to the payoffs obtained in other states where there is a non-trivial choice to be made between at least two similarity classes. We believe this special case is particularly relevant in applications where the material payoffs are relatively small compared to the cognitive cost associated with the act of making a non-trivial choice. To the extent that deliberation about which choice to make is required only in non-trivial states, it is reasonable to assume that the associated cognitive costs would outweigh the material payoffs. This justifies the relevance of the assumptions outlined below.

Formally, we vary the payoffs associated with the trivial unary choice nodes where Alice chooses among alternatives from within a single similarity class. In particular, we add a large constant $z \in \mathbb{R}_+$ to the unary choice payoffs $\pi_{\omega=\{i\}}(i)$ for all $i \in \mathcal{S}$. Let $\{\pi_{\omega=\{i\}}(i) : i \in \mathcal{S}\}$ denote the set of unary choice payoffs, which are the payoffs Alice receives by choosing an alternative in similarity class i at node ω, such that $\mathcal{S}_\omega = \{i\}$, in the decision tree $\mathcal{T}_n'$, for all $i \in \mathcal{S}$. We add a constant z to every element of this set, where z is above a large positive threshold $\hat{z}$, i.e., $z > \hat{z} > 0$. Meanwhile, the rest of the payoffs, $\{\pi_\omega(i) : i \in \mathcal{S}, \omega \in \Omega, |\mathcal{S}_\omega| > 1\}$, obtained at nodes where Alice faces non-trivial choice problems, remain unchanged. Essentially, for all $i \in \mathcal{S}$, Alice is assumed to receive a substantially higher payoff when she chooses an alternative in a similarity class i at a node where i is the only available similarity class, relative to when she chooses an alternative in i at a node where alternatives from at least one other similarity class are available. Additionally, for the remainder of the paper, we assume that the support of the probability mass function $p(\omega)$, denoted by $\text{supp}(p) = \{\omega \in \Omega : p(\omega) > 0\}$, includes all unary choice nodes and all binary choice nodes. That is, we assume $\{\omega \in \Omega : 1 \leq |\mathcal{S}_\omega| \leq 2\} \subseteq \text{supp}(p)$.

Theorem 4. *There exists a finite $\hat{z} > 0$ such that $\forall\ z > \hat{z}$ and $\forall\ \beta \in \mathbb{R}_+$, a finite decision tree $\mathcal{T}_n'$ with generic payoffs and an arbitrary number of similarity classes always admits a unique smooth valuation equilibrium (SVE). Moreover, there exists a finite $\hat{\beta} > 0$ such that $\forall$ $\beta > \hat{\beta}$, the unique SVE lies in the neighborhood of a mixed valuation equilibrium in which the agent is indifferent between at least two of her similarity classes. Furthermore, the unique SVE is globally asymptotically stable under the CQL dynamics $\forall\ \beta \in \mathbb{R}_+$.*

Proof. The proof involves several steps. We prove that the conditions on the trivial unary choice payoffs that are sufficient for ruling out strict pure VE are also sufficient for establishing the local asymptotic stability of the SVE in the CQL dynamics. Exploiting the local

asymptotic stability result, we prove the uniqueness of such an SVE using the Poincare-Hopf index theorem, and subsequently its global asymptotic stability in the CQL dynamics by leveraging the properties of monotone cooperative dynamical systems.

Non-existence of Strict Pure VE

Assume there exists a strict pure VE for $z > \hat{z}$. Thus, there must be a strict total order relation on the equilibrium valuations. We focus on two similarity classes: the ones with the lowest and the second-lowest equilibrium valuations. Without loss of generality, let these classes be denoted by i and j such that $v_i^* < v_j^*$. Since i has the lowest valuation among all similarity classes, it is selected only at the trivial choice node $\omega_i = \{i\}$. Therefore, by consistency, the equilibrium valuation for i is given by $v_i^* = \pi_{\{i\}}(i) + z$. Now consider j. While j has a higher valuation than i in equilibrium, its valuation is strictly lower than that of every other similarity class besides i. Thus, in equilibrium, j is selected only at the trivial choice node $\omega_j = \{j\}$ and the binary choice node $\omega_{ij} = \{i, j\}$ featuring i and j. Again, by consistency, the equilibrium valuation of j is:

$$v_j^* = \frac{p(\omega_j)(\pi_{\{j\}}(j) + z) + p(\omega_{ij})\pi_{\{i,j\}}(j)}{p(\omega_j) + p(\omega_{ij})}.$$

We observe that the weight on z in the consistent valuation v_j^* of class j is strictly less than the weight on z in the consistent valuation v_i^* of class i, as $p(\omega_{ij}) > 0$. Thus, by making z sufficiently large, we can ensure that $v_i^* \geq v_j^*$, leading to a contradiction. Hence, there exists a finite $\hat{z}$ such that for all $z > \hat{z}$, there does not exist a strict pure VE. As a VE must always exist in a finite decision tree, there exists a mixed valuation equilibrium where the agent is indifferent between at least two of her similarity classes. By Lemma 2.1, for $\beta > \hat{\beta}$, a smooth valuation equilibrium of the CQL dynamics lies in the neighborhood of a mixed VE in which the agent is indifferent between at least two of her similarity classes.

Local Asymptotic Stability of SVE

We aim to show that the Jacobian matrix $\mathbf{J}$ of the CQL dynamical system in Eq.(6), a square matrix of order n, is a stability matrix everywhere in the domain. We begin by computing a typical main-diagonal term of the Jacobian by differentiating $f_s(\mathbf{v})$ with respect to v_s. Let's denote this partial derivative by $\mathbf{J}_{ss}$. First, recall that $\dot{\mathbf{v}} = \mathbf{f}(\mathbf{v})$ where $f_s(\mathbf{v})$ is given as:

$$f_s(\mathbf{v}) = g_s(\mathbf{v}) - v_s = \frac{\sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega)\sigma_\omega^s(\mathbf{v})\pi_\omega(s)}{\sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega)\sigma_\omega^s(\mathbf{v})} - v_s.$$

To compute $\mathbf{J}_{ss} = \frac{\partial f_s(\mathbf{v})}{\partial v_s} = \frac{\partial g_s(\mathbf{v})}{\partial v_s} - 1$, we apply the quotient rule and the chain rule of differentiation. Thus, a typical main-diagonal term of the Jacobian matrix $\mathbf{J}$ is

$$\mathbf{J}_{ss} = \beta \frac{\sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma^s_\omega(\mathbf{v})(1 - \sigma^s_\omega(\mathbf{v}))\left(\pi_\omega(s) - g_s(\mathbf{v})\right)}{\sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma^s_\omega(\mathbf{v})} - 1. \tag{7}$$

We compute a typical off-diagonal term of the Jacobian by differentiating $f_s(\mathbf{v})$ with respect to v_k where $k \neq s$. Let's denote this partial derivative by $\mathbf{J}_{sk}$.

To evaluate $\mathbf{J}_{sk} = \frac{\partial f_s(\mathbf{v})}{\partial v_k} = \frac{\partial g_s(\mathbf{v})}{\partial v_k}$, we again apply the quotient rule and the chain rule of differentiation. Thus, a typical off-diagonal term of the Jacobian matrix $\mathbf{J}$ is:

$$\mathbf{J}_{sk} = \frac{\beta \sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma^s_\omega(\mathbf{v}) \sigma^k_\omega(\mathbf{v}) \left(g_s(\mathbf{v}) - \pi_\omega(s)\right)}{\sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma^s_\omega(\mathbf{v})}. \tag{8}$$

We make the following important observations on $\mathbf{J}_{sk}$. Firstly, for each $s \in \mathcal{S}$, the contribution to $\mathbf{J}_{sk}$ from the trivial choice node involving s - i.e., from $\omega \in \Omega$ where $\mathcal{S}_\omega = \{s\}$ - is 0. This is so because s is the sole available similarity class at this node and the agent selects s with a constant probability $\sigma^s_{\{s\}} = 1$, implying $\sigma^k_{\{s\}} = 0$ for $k \neq s$, independently of $\mathbf{v}$.

Secondly, for each $s \in \mathcal{S}$, by choosing a sufficiently large but finite constant $z > \hat{z}$ added to the payoffs of trivial unary choices, the contribution to $\mathbf{J}_{sk}$ from any non-trivial choice node involving s - i.e., from $\omega \in \Omega$ such that $s \in \mathcal{S}_\omega$ and $1 < |\mathcal{S}_\omega| \leq n$ - can be guaranteed to be positive. Specifically, there exists a threshold $\hat{z} > 0$ such that for all $z > \hat{z}$ and for all $\omega \in \Omega$ with $s \in \mathcal{S}_\omega$ and $1 < |\mathcal{S}_\omega| \leq n$, the following inequality holds: $g_s(\mathbf{v}) - \pi_\omega(s) > 0$.

This ensures that for $z > \hat{z}$, all off-diagonal entries of the Jacobian matrix satisfy $\mathbf{J}_{sk} > 0$ for every $k \in \mathcal{S}$ with $k \neq s$, for all $\mathbf{v} \in \mathbb{R}^{\mathcal{S}}$, and for all $\beta > 0$. Consequently, the absolute values of the off-diagonal terms of the Jacobian matrix are $|\mathbf{J}_{sk}| = \mathbf{J}_{sk} > 0$ for all $k \in \mathcal{S}$ with $k \neq s$. It is important to note that the threshold $\hat{z}$ depends solely on the exogenously specified parameters of the model and is independent of the agent's endogenous valuations.

Finally, we compute the sum of the absolute values of all the off-diagonal terms in the row s of the Jacobian matrix, i.e., we sum the expressions for $\mathbf{J}_{sk}$ for all $k \neq s$. Let's denote this sum by $\mathcal{R}_s = \sum_{k \neq s} |\mathbf{J}_{sk}| = \sum_{k \neq s} \mathbf{J}_{sk} > 0$. Let $D_s(\mathbf{v}) = \sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma^s_\omega(\mathbf{v})$.

$$\mathcal{R}_s = \sum_{k \neq s} \left[\beta \sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma^s_\omega(\mathbf{v}) \sigma^k_\omega(\mathbf{v}) \left(g_s(\mathbf{v}) - \pi_\omega(s)\right) / D_s(\mathbf{v}) \right],$$

Since the summation is over all $k \neq s$ for a fixed s, we can pull the summation over k inside

the summation over ω:

$$\mathcal{R}_s = \beta \sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} \left[p(\omega) \sigma_\omega^s(\mathbf{v}) \left(g_s(\mathbf{v}) - \pi_\omega(s) \right) \sum_{k \neq s} \sigma_\omega^k(\mathbf{v}) \right] / D_s(\mathbf{v}).$$

Next, we use the fact that $\sum_{k \in \mathcal{S}_\omega} \sigma_\omega^k(\mathbf{v}) = 1$:

$$\sum_{k \neq s} \sigma_\omega^k(\mathbf{v}) = 1 - \sigma_\omega^s(\mathbf{v}).$$

Therefore, we can simplify the expression:

$$\mathcal{R}_s = \beta \frac{\sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma_\omega^s(\mathbf{v}) \left(1 - \sigma_\omega^s(\mathbf{v}) \right) \left(g_s(\mathbf{v}) - \pi_\omega(s) \right)}{\sum_{\omega \in \Omega : s \in \mathcal{S}_\omega} p(\omega) \sigma_\omega^s(\mathbf{v})} = -\mathbf{J}_{ss} - 1.$$

Let $\mathcal{D}_s(\mathbf{J}_{ss}, \mathcal{R}_s) \subseteq \mathbb{C}$ be a closed disc in the complex plane centered at $\mathbf{J}_{ss}$ with radius $\mathcal{R}_s$. We refer to such a disc as a Gershgorin disc. Across all the rows of the Jacobian matrix, we define n such discs. Now, by the Gershgorin Circle theorem, every eigenvalue of $\mathbf{J}$ lies within at least one of the Gershgorin discs.[21] As a corollary, all the eigenvalues of $\mathbf{J}$ must lie within the union of the n Gershgorin discs. Finally, we make the following two observations. First, for all rows $s \in \mathcal{S}$, $\mathbf{J}_{ss}$ is real and $\mathbf{J}_{ss} < -1$. To see this, note that $\mathbf{J}_{ss} = -\mathcal{R}_s - 1$ and $\mathcal{R}_s > 0$. Second, $\forall\ s \in \mathcal{S}$, $\mathbf{J}_{ss} + \mathcal{R}_s = -1$. Therefore, for all $\beta > 0$, the real parts of the eigenvalues of the Jacobian matrix have a supremum equal to -1 and $\mathbf{J}$ is a stability matrix everywhere in the domain. By continuity, the real parts of the eigenvalues remain strictly negative as $\beta \uparrow \infty$, where we evaluate $\mathbf{J}$ at a steady-state valuation system $\mathbf{v}^*$ that corresponds to a mixed VE. Therefore, by the linearization theorem, $\forall\ z > \hat{z}$ and $\forall\ \beta > \hat{\beta}$, a smooth valuation equilibrium that corresponds to a mixed valuation equilibrium in the high-sensitivity limit is *locally asymptotically stable* in the CQL dynamics.

Uniqueness of SVE

We use the Poincare-Hopf index theorem to establish the uniqueness of the smooth valuation equilibrium that corresponds to a mixed VE in the high-sensitivity limit.[22]

[21] Refer to Section 1.3 of the Online Appendix for a statement and a proof of the Gershgorin Circle Theorem

[22] The Poincare-Hopf index theorem states: 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:

$$\sum_i \text{index}_{x_i}(\mathbf{h}) = \chi(M),$$

Recall that a smooth valuation equilibrium $\mathbf{v}^* \in \mathbb{R}^n$ is a zero of the smooth vector field $f : \mathbb{R}^n \to \mathbb{R}^n$ where $\mathbf{f}(\mathbf{v}) = \mathbf{g}(\mathbf{v}) - \mathbf{v}$. The zeros of $\mathbf{f}(\mathbf{v})$ are precisely the fixed points of the smooth vector field $\mathbf{g}(\mathbf{v})$. Let $K \subset \mathbb{R}^n$ be the convex hull of the finite set of generic payoffs in $\mathbb{R}^n$. Thus, K is a compact, convex subset of $\mathbb{R}^n$ with non-empty interior and a well-defined boundary ∂K. We've already established that $g(K) \subseteq K$. Therefore, a smooth valuation equilibrium $\mathbf{v}^* \in K$.

The real parts of all eigenvalues of the Jacobian matrix $\mathbf{J_f}$ are strictly negative for all $\mathbf{v} \in \mathbb{R}^n$. This implies that any zero of $\mathbf{f}(\mathbf{v})$ is non-degenerate since $\mathbf{J_f}$ is non-singular everywhere in the domain. Consequently, by the inverse function theorem, the zeroes of $\mathbf{f}(\mathbf{v})$ are also locally isolated. Additionally, there are no zeros of $\mathbf{f}(\mathbf{v})$ on ∂K; all zeros lie in $\text{Int}(K)$. For a finite β, this is trivially true since the corresponding SVE is fully-mixed. More interestingly, it is also true as $\beta \uparrow \infty$. To see this, recall that for all $z > \hat{z}$, there does not exist a smooth valuation equilibrium that corresponds to a strict pure VE as $\beta \uparrow \infty$. Thus, any admissible SVE must necessarily correspond to a mixed VE as $\beta \uparrow \infty$. Moreover, such a mixed VE cannot include a unique strictly dominated similarity class in equilibrium. Therefore, for $z > \hat{z}$, any mixed VE must involve mixing between at least the lowest and the second-lowest similarity classes, where the classes are ranked in ascending order based on their equilibrium valuations. Thus, the zeroes of $\mathbf{f}(\mathbf{v})$ are restricted to the interior of K even in the high-sensitivity limit.

Since each zero is locally isolated, there can only be countably many isolated zeroes in the interior of 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 the smooth vector field $\mathbf{f}(\mathbf{v})$. We define a smooth vector field $\mathbf{h}(\mathbf{v}) : \mathbb{R}^n \to \mathbb{R}^n$ by:

$$\mathbf{h}(\mathbf{v}) = -\mathbf{f}(\mathbf{v}) = \mathbf{v} - \mathbf{g}(\mathbf{v}).$$

The zeros of $\mathbf{h}(\mathbf{v})$ are the same as those of $\mathbf{f}(\mathbf{v})$ and lie in $\text{Int}(K)$. Since $\mathbf{f}(\mathbf{v})$ is smooth, $\mathbf{h}(\mathbf{v})$ is smooth on K. At each zero $\mathbf{v}^*$ of $\mathbf{h}(\mathbf{v})$, the Jacobian $\mathbf{J_h} = -\mathbf{J_f}$ has eigenvalues with strictly positive real parts. Thus, each zero is non-degenerate. The index of a non-degenerate zero of a smooth vector field is determined by the sign of the determinant of the Jacobian matrix at the zero. For $\mathbf{h}(\mathbf{v})$, $\det(\mathbf{J_h}) > 0$, since all the eigenvalues have strictly positive real parts. Therefore, every zero of the smooth vector field $\mathbf{h}(\mathbf{v})$ is a non-degenerate source with an index $\text{ind}_{\mathbf{v}^*}(\mathbf{h}) = +1$.

Let $\mathbf{v} \in \partial K$, and let $\mathbf{n}(\mathbf{v})$ be the outward-pointing unit normal vector to ∂K at $\mathbf{v}$. Since

where the sum is over all isolated zeroes of the vector field $\mathbf{h}$, $\chi(M)$ is the Euler characteristic of M.

$\mathbf{g}(\mathbf{v}) \in K$ and $\mathbf{v} \in \partial K$, the vector $\mathbf{h}(\mathbf{v}) = \mathbf{v} - \mathbf{g}(\mathbf{v})$ satisfies: $\mathbf{h}(\mathbf{v}) = \mathbf{v} - \mathbf{g}(\mathbf{v}) \neq 0$ and $\mathbf{h}(\mathbf{v}) \cdot \mathbf{n}(\mathbf{v}) > 0$. The first part follows from the fact that there are no fixed points of $\mathbf{g}(\mathbf{v})$ on the boundary of K. For the second part, recall that since K is convex, for any $\mathbf{g}(\mathbf{v}) \in K$ and $\mathbf{v} \in \partial K$ with $\mathbf{g}(\mathbf{v}) \neq \mathbf{v}$, the vector $\mathbf{v} - \mathbf{g}(\mathbf{v})$ points outward from K at $\mathbf{v}$. The dot product $\mathbf{h}(\mathbf{v}) \cdot \mathbf{n}(\mathbf{v}) = (\mathbf{v} - \mathbf{g}(\mathbf{v})) \cdot \mathbf{n}(\mathbf{v}) > 0$, since $\mathbf{v}$ is on the boundary and $\mathbf{g}(\mathbf{v})$ is strictly inside K. Therefore, $\mathbf{h}(\mathbf{v})$ points outward at every point $\mathbf{v} \in \partial K$. Correspondingly, $\mathbf{f}(\mathbf{v})$ points inward at every point $\mathbf{v} \in \partial K$. In fact, this ensures that K is a positively invariant set for the CQL dynamics, meaning trajectories do not escape K.

As the convex hull of a finite set of generic payoffs, K is an n-dimensional compact differentiable manifold with a boundary, where $n = |\mathcal{S}|$. The Euler characteristic of any compact, convex subset of $\mathbb{R}^n$ is 1, as such a set is homeomorphic to a closed n-ball in $\mathbb{R}^n$, and therefore, contractible and homotopic to a point. Its fundamental group is trivial. Thus, $\chi(K) = 1$.

Finally, we use the Poincare-Hopf index theorem for $\mathbf{h}(\mathbf{v})$: $\sum_i \text{ind}_{\mathbf{v}^*}(\mathbf{h}) = \chi(K)$. Given $\chi(K) = 1$, and knowing that the index for each zero of $\mathbf{h}(\mathbf{v})$ is $+1$, we have: $N \times (+1) = 1$, where N is the finite number of zeroes of $\mathbf{h}(\mathbf{v})$. Therefore, $N = 1$, i.e., there is a unique zero with index $+1$. Essentially, since the sum of indices must equal 1, and each zero has an index of $+1$, there can only be one such zero. Correspondingly, the smooth vector field $\mathbf{f}(\mathbf{v}) = -\mathbf{h}(\mathbf{v})$ has a unique zero that lies in the interior of K. Hence, there exists a *unique smooth valuation equilibrium* of the CQL dynamics for all $\beta \in \mathbb{R}_+$. The unique SVE lies in the neighborhood of a mixed valuation equilibrium for large but finite $\beta > \hat{\beta}$.

Global Asymptotic Stability of SVE

The aim is to prove that the unique smooth valuation equilibrium that corresponds to a mixed valuation equilibrium in the high-sensitivity limit is globally asymptotically stable in the CQL dynamics for all $\beta \in \mathbb{R}_+$. We begin by establishing that the CQL dynamical system $\dot{\mathbf{v}} = \mathbf{f}(\mathbf{v})$ is a cooperative and irreducible monotone dynamical system.

Firstly, we verify that the CQL dynamical system is cooperative. A system of ODEs is cooperative (competitive) if it generates a monotone semi-flow in the forward (backward) direction. We recall that, $\forall\ z \in (\hat{z}, \infty)$, and $\forall\ \beta \in (0, \infty)$, every off-diagonal term of the Jacobian matrix of the CQL dynamical system is positive everywhere in the domain, i.e.,

$$\mathbf{J}_{sk} = \frac{\partial f_s}{\partial v_k}(\mathbf{v}) > 0, \quad \forall s \neq k, \quad \mathbf{v} \in K.$$

Since K is a convex subset of $\mathbb{R}^n$, it is also p-convex. Therefore, by Condition 3.1.3 in Smith

(1995), $\mathbf{v}$ is of type K in K and $\dot{\mathbf{v}} = \mathbf{f}(\mathbf{v})$ is a cooperative system of ODEs.[23]

Secondly, we verify that the CQL dynamical system is irreducible. A system of ODEs is irreducible if the associated Jacobian matrix is an irreducible matrix at every point in the domain.[24] Clearly, for $\beta \in (0, \infty)$, every off-diagonal term of the Jacobian matrix is strictly positive everywhere in the domain, $\mathbf{J}_{sk} > 0$, where $s \neq k$. Now, we recall from the proof of local asymptotic stability, that every term on the main diagonal of the Jacobian matrix is strictly negative, $\mathbf{J}_{ss} < 0$, at every point of the domain. Therefore, the associated directed graph $\mathbf{G_J}$ is strongly connected and the Jacobian matrix $\mathbf{J}$ is an irreducible matrix.

By Theorem 4.1.1 in Smith (1995), we know that the semi-flow of a cooperative and irreducible system of ODEs is strongly monotone. A strongly monotone semi-flow implies that it is eventually strong monotone and therefore, strongly order preserving (SOP)[25] by Proposition 1.1.1 in Smith (1995). We recall that the unique smooth valuation equilibrium $\mathbf{v}^*$ of the CQL model lies in the interior of K, which is a compact, convex subset of $\mathbb{R}^n$. Therefore, by Theorem 2.3.1 in Smith (1995), $\mathbf{v}^*$ is the unique element of the ω-limit set of every orbit of the CQL dynamical system. Hence, the unique smooth valuation equilibrium that corresponds to a mixed VE in the high-sensitivity limit, is *globally asymptotically stable* in the CQL dynamics for a finite $\beta > \hat{\beta}$. Moreover, Benaïm (1999) shows that if the continuous-time process has a unique steady-state that is globally asymptotically stable, then it is the only element of the internally chain-transitive set and the discrete-time stochastic CQL process in Eq. (4) converges to it almost surely.[26] □

The scope of the indifferences identified in Theorem 4 can be extended to include all similarity classes in the high-sensitivity limit, under additional assumptions on the distribution of states. Notably, if all states are equally likely, as $\beta \uparrow \infty$, the unique SVE corresponds to a fully-mixed VE where Alice is indifferent among all her similarity classes. This in-

[23] The non-negative cone in $\mathbb{R}^n$, denoted by $\mathbb{R}^n_+$, is the set of all n-tuples with non-negative components. It gives rise to a partial order on $\mathbb{R}^n$ by $\mathbf{y} \leq \mathbf{x}$ if $\mathbf{x} - \mathbf{y} \in \mathbb{R}^n_+$. Less formally, this is true if and only if $y_i \leq x_i$ for all i. A vector field $\mathbf{q}$ is said to be of type K in $D \in \mathbb{R}^n$ if for each i, $q_i(\mathbf{a}) \leq q_i(\mathbf{b})$ for any two points $\mathbf{a}$ and $\mathbf{b}$ in D satisfying $\mathbf{a} \leq \mathbf{b}$. The type K condition is most easily identifiable from the sign structure of the Jacobian matrix of the vector field.

[24] A matrix $\mathbf{A}$ can always be associated with a certain directed graph $\mathbf{G_A}$. It has n vertices labeled $1, \ldots, n$, and there is an edge from vertex i to vertex j precisely when $a_{ij} \neq 0$. Then the matrix $\mathbf{A}$ is irreducible if and only if its associated graph $\mathbf{G_A}$ is strongly connected.

[25] A 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 $\phi(t, \mathbf{y})$.

[26] Alternatively, we could also refer to Theorem 3.2 in Hirsch et al. (2001) for an equivalent result.

sight follows as a corollary of a broader result stated in Proposition 1. Specifically, let $\Omega^{|1|} = \{\omega \in \Omega : |S_\omega| = 1\}$ denote the set of trivial unary choice states. Consider Assumption 5.1, which is satisfied, for e.g., under a uniform distribution where all states are equally likely.

Assumption 5.1 (Monotonicity)**.** *For any two states ω, $\omega' \in \Omega$, $\omega \subseteq \omega' \implies p(\omega) \leq p(\omega')$. Additionally, for any two states ω, $\omega' \in \Omega^{|1|}$, $p(\omega) = p(\omega')$.*

Proposition 1. *Let Assumption 5.1 hold. There exists a finite $\hat{z} > 0$ such that for all $z > \hat{z}$ and $\beta \geq 0$, a decision tree $\mathcal{T}_n^{'}$ with generic payoffs and a finite number of similarity classes always admits a unique smooth valuation equilibrium (SVE) that is globally asymptotically stable in the CQL dynamics. Moreover, as $\beta \to \infty$, this unique SVE corresponds to a fully-mixed valuation equilibrium where the agent is indifferent among all her similarity classes.*

Proof. The proof is relegated to Section 2 of the Online Appendix. □

5.3 When Trivial Choice Payoffs are Small

To illustrate the dramatic impact of assuming large trivial choice payoffs, we now consider the opposite scenario where these payoffs are small, leading to markedly different properties for the steady-states of the CQL dynamics. Specifically, we introduce a large, negative constant, z, to the unary choice payoffs $\pi_{\omega=\{i\}}(i)$ for all $i \in \mathcal{S}$. Let $\{\pi_{\omega=\{i\}}(i) : i \in \mathcal{S}\}$ represent the set of unary choice payoffs, corresponding to the payoffs Alice receives when selecting a similarity class i at a node ω where $\mathcal{S}_\omega = \{i\}$. We adjust each payoff in this set by adding the constant z, where z is sufficiently negative, below a specified threshold $\tilde{z} < 0$, i.e., $z < \tilde{z} < 0$. In contrast, the payoffs $\{\pi_\omega(i) : i \in \mathcal{S}, \omega \in \Omega, |\mathcal{S}_\omega| > 1\}$, associated with non-trivial choice nodes (where multiple similarity classes are available), remain unchanged[27].

Theorem 5. *There exists a finite threshold $\tilde{z} < 0$ such that for all $z < \tilde{z}$, in a decision tree $\mathcal{T}_n^{'}$ with generic payoffs and an arbitrary number n of similarity classes, the following holds: There exists a multiplicity of valuation equilibria. For each similarity class $s \in \mathcal{S}$, there exists a valuation equilibrium (VE) where s is the unique strictly dominated similarity class. Moreover, there exists at least one strict pure valuation equilibrium. Correspondingly, there exists a finite $\hat{\beta}$, such that for all $\beta > \hat{\beta}$, the smooth valuation equilibrium that arises in the neighborhood of the strict pure VE is locally asymptotically stable in the CQL dynamics.*

Proof. The proof is relegated to Section A.4 of the Appendix. □

[27] $\forall\ i \in \mathcal{S}$, Alice is assumed to receive significantly lower payoffs when choosing a similarity class i at a node where it is the sole available class, compared to nodes where at least one other similarity class is present.

The multiplicity identified in Theorem 5 can be further strengthened under Assumption 5.1. Notably, if all states 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:

Proposition 2. *Suppose Assumption 5.1 holds. There exists a finite threshold $\tilde{z} < 0$ such that for all $z < \tilde{z}$, in a decision tree $\mathcal{T}_n'$ with generic payoffs and an arbitrary number n of similarity classes, the following holds. Every strict total order on the valuations of the n similarity classes is admissible in a strict pure valuation equilibrium, i.e., there exist $n!$ strict pure VE. Correspondingly, there exists a finite $\hat{\beta} > 0$, such that for all $\beta > \hat{\beta}$, the smooth valuation equilibrium that lies in the neighborhood of each strict pure valuation equilibrium is locally asymptotically stable in the CQL dynamics. Besides the $n!$ strict pure smooth valuation equilibria, there exist at least $n! - 1$ smooth valuation equilibria that correspond to partially-mixed valuation equilibria in the high-sensitivity limit, each of which is asymptotically unstable in the CQL dynamics, for $\beta > \hat{\beta}$.*

Proof. Proof in Section A.5 of the Appendix □

6 Conclusion

In this article, we have introduced the Coarse Q-learning (CQL) model as a novel approach to simplifying decision-making under payoff uncertainty by categorizing alternatives into coarse subsets called similarity classes. This model departs from traditional Bayesian frameworks by employing a heuristic, reinforcement learning-based method in which decision-makers smoothly update their assessments of the categories in the direction of observed payoffs. Our focus is on the convergence properties of such a coarse Q-learning model for predefined categories.

One of our key findings is that persistent mixing can exist in a decision problem with generic payoffs, even as the decision-maker becomes extremely sensitive to differences in assessments. Moreover, such behavior is globally stable in our learning dynamics - a phenomenon that has no counterpart in the existing literature. This result may provide a novel explanation for the presence of indifferences in decision-making, as observed in empirical studies (Iyengar and Lepper, 2000; DellaVigna, 2009). We leave for future research the study of how decision-makers form, adjust, and optimize categories over time, especially when salience does not prescribe which ones to consider.

References

Anderson, S., De Palma, A., and Thisse, J. (1992). *Discrete Choice Theory of Product Differentiation*. Mit Press. MIT Press.

Benaïm, M. (1999). Dynamics of stochastic approximation algorithms. *Séminaire de probabilités de Strasbourg*, 33:1–68.

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

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience Theory of Choice Under Risk. *The Quarterly Journal of Economics*, 127(3):1243–1285.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Brown, G. W. (1951). Iterative solution of games by fictitious play. In Koopmans, T. C., editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, New York.

Börgers, T. and Sarin, R. (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory*, 77(1):1–14.

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.* Regional conference series in mathematics ; no. 38. Published for the Conference Board of the Mathematical Sciences by the American Mathematical Society, Providence.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–72.

Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, 88(4):848–881.

Fudenberg, D. and Kreps, D. M. (1993). Learning mixed equilibria. *Games and Economic Behavior*, 5(3):320–367.

Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*, volume 2. MIT press.

Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):148–164.

Goeree, J. K., HOLT, C. A., and PALFREY, T. R. (2016). *Quantal Response Equilibrium: A Stochastic Theory of Games*. Princeton University Press.

Hausman, J. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, 52(5):1219–1240.

Hirsch, M. W., Smith, H. L., and Zhao, X.-Q. (2001). Chain transitivity, attractivity, and strong repellors for semidynamical systems. *Journal of Dynamics and Differential Equations*, 13:107–131.

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. and Samet, D. (2007). Valuation equilibrium. *Theoretical Economics*, 2(2):163–185.

Jehiel, P. and Singh, J. (2021). Multi-state choices with aggregate feedback on unfamiliar alternatives. *Games and Economic Behavior*, 130:1–24.

Kushner, H. and Yin, G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. Stochastic Modelling and Applied Probability. Springer New York.

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533.

Monderer, D. and Shapley, L. S. (1996a). Fictitious play property for games with identical interests. *Journal of Economic Theory*, 68(1):258–265.

Monderer, D. and Shapley, L. S. (1996b). Potential games. *Games and Economic Behavior*, 14(1):124–143.

Nachbar, J. H. (1990). Evolutionary selection dynamics in games: Convergence and limit properties. *International Journal of Game Theory*, 19(1):59–89.

Pemantle, R. (1990). Nonconvergence to Unstable Points in Urn Models and Stochastic Approximations. *The Annals of Probability*, 18(2):698 – 712.

Robinson, J. (1951). An iterative method of solving a game. *Annals of Mathematics*, 54(2):296–301.

Rosch, E. and Lloyd, B. B. (1978). *Principles of categorization*. MIT press.

Roth, A. E. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, 8(1):164–212.

Rustichini, A., Soukupova, M., and Palminteri, S. (2023). Adaptive coding is optimal in reinforcement learning. *Available at SSRN 4320894*.

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). *1. Some Topics in Two-Person Games*, pages 1–28. Princeton University Press, Princeton.

Smith, H. L. (1995). *Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems*. American Mathematical Soc.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. The MIT Press, second edition.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79:281–299.

Watkins, C. J. C. H. and Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3):279–292.

A Omitted Proofs

A.1 Theorem 1

Proof. To prove that $\mathcal{V}$ is non-empty, it suffices to show that the function $\mathbf{g} : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{S}}$ admits at least one fixed point, i.e., there exists $\mathbf{v}^* = (v_s^*)_{s\in\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$ such that $g_s(\mathbf{v}^*) = v_s^*$ for all $s \in \mathcal{S}$. This function $\mathbf{g}$ is defined component-wise for each strategy $s \in \mathcal{S}$ by

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

where $\mathbf{v} = (v_s)_{s\in\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$, $\Omega_s = \{\omega \in \Omega : s \in \mathcal{S}_\omega\}$ is the set of states where similarity class s is available, $p(\omega)$ is the probability of state ω with $p(\omega) \geq 0$ and $\sum_{\omega\in\Omega} p(\omega) = 1$. By assumption, for each $s \in \mathcal{S}$, there exists at least one $\omega_s \in \Omega_s$ such that $p(\omega_s) > 0$. $\pi_\omega(s)$ is the payoff for similarity class s in state ω, and $\sigma_\omega^s(\mathbf{v})$ represents the logit propensity to choose similarity class s in state ω, which depends continuously on $\mathbf{v}$ and for $\beta \geq 0$, satisfies $0 < \sigma_\omega^s(\mathbf{v}) < 1$, for all $\mathbf{v}$.

Consider the finite set of generic payoffs $\{\pi_\omega(s) : s \in \mathcal{S}, \omega \in \Omega_s\}$. Let K be the convex hull of this finite set. Since $\mathcal{S}$ and Ω are finite, K is a compact and convex subset of $\mathbb{R}^{|\mathcal{S}|}$. For any $\mathbf{v} \in K$, the function $\mathbf{g}$ maps $\mathbf{v}$ to a point in K. Specifically, each $g_s(\mathbf{v})$ is a convex combination of the payoffs $\pi_\omega(s)$ for $\omega \in \Omega_s$, with weights given by $w_\omega^s(\mathbf{v}) = \frac{p(\omega)\sigma_\omega^s(\mathbf{v})}{\sum_{\omega'\in\Omega_s} p(\omega')\sigma_{\omega'}^s(\mathbf{v})}$. These weights satisfy $w_\omega^s(\mathbf{v}) \geq 0$ and $\sum_{\omega\in\Omega_s} w_\omega^s(\mathbf{v}) = 1$ for each $s \in \mathcal{S}$. Therefore, $\mathbf{g}(\mathbf{v})$ is a convex combination of the payoffs, and thus $\mathbf{g}(\mathbf{v}) \in K$. Consequently, $\mathbf{g}(K) \subseteq K$.

The function $\mathbf{g}$ is continuous on K. Each $\sigma_\omega^s(\mathbf{v})$ is continuous in $\mathbf{v}$ and β, and since algebraic operations preserve continuity, each $g_s(\mathbf{v})$ is continuous. Therefore, $\mathbf{g}$ is a continuous function from K to K. Since K is a non-empty, compact, and convex subset of $\mathbb{R}^{|\mathcal{S}|}$, and $\mathbf{g}$ is a continuous endomorphism of K, Brouwer's Fixed-Point Theorem applies. Thus, there exists $\mathbf{v}^* \in K$ such that $\mathbf{g}(\mathbf{v}^*) = \mathbf{v}^*$. This fixed point $\mathbf{v}^*$ satisfies $g_s(\mathbf{v}^*) = v_s^*$ for all $s \in \mathcal{S}$, meaning $\mathbf{v}^*$ is a steady-state solution of the CQL dynamics. Therefore, $\mathcal{V}$, the set of steady-state solutions, is non-empty for all $\beta \geq 0$. □

A.2 Lemma 2.1

Proof. The proof proceeds as follows. Firstly, we establish that the set-valued mapping of smooth valuation equilibria $\mathcal{V}(\beta)$ is upper hemicontinuous in β. Using upper hemicontinuity,

we argue that any limit point of a sequence of SVEs as $\beta \to \infty$ is a VE. Finally, for any $\epsilon > 0$, we show that $\exists\ \hat{\beta}$ such that for all $\beta > \hat{\beta}$, the SVEs are within ϵ of some VE.

Let $\mathcal{V}(\beta)$ denote the set of smooth valuation equilibria at sensitivity level β, i.e., $\mathcal{V}(\beta) = \{\mathbf{v} \in K : \mathbf{g}(\mathbf{v}; \beta) - \mathbf{v} = \mathbf{0}\}$, where $K \subset \mathbb{R}^{|\mathcal{S}|}$ is a compact and convex set defined by the convex hull of the bounded payoffs and $\mathbf{g}(\mathbf{v}; \beta)$ is the expected payoff function that is continuous. A set-valued mapping $\Phi : \Lambda \to 2^X$ is upper hemicontinuous at λ_0 if for every open set U containing $\Phi(\lambda_0)$, there exists a neighborhood V of λ_0 such that for all $\lambda \in V$, $\Phi(\beta) \subseteq U$.

Since $\mathcal{V}(\beta)$ is the set of fixed points in K, and K is compact, $\mathcal{V}(\beta)$ is a closed subset of a compact set, hence compact. Note that $\mathcal{V}(\beta)$ is the pre-image of a closed set $\{\mathbf{0}\}$ under a continuous function $\mathbf{g}(\mathbf{v}; \beta) - \mathbf{v}$, making it closed. Consider sequences $\beta_n \to \beta$ and $\mathbf{v}_n \to \mathbf{v}$ with $\mathbf{v}_n \in \mathcal{V}(\beta_n)$. Since $\mathbf{v}_n = \mathbf{g}(\mathbf{v}_n; \beta_n)$ and $\mathbf{g}$ is continuous in $\mathbf{v}$ and β,

$$\mathbf{v} = \lim_{n\to\infty} \mathbf{v}_n = \lim_{n\to\infty} \mathbf{g}(\mathbf{v}_n; \beta_n) = \mathbf{g}(\mathbf{v}; \beta).$$

Therefore, $\mathbf{v} \in \mathcal{V}(\beta)$, and the graph of $\mathcal{V}$ is closed. Since $\mathcal{V}(\beta)$ is a compact set, the mapping $\mathcal{V}(\beta)$ is upper hemicontinuous in β.

For each β, consider any $\mathbf{v} \in \mathcal{V}(\beta)$. To establish the optimality of the smooth valuation equilibria as $\beta \uparrow \infty$, consider the logit choice policy employed by Alice. For a finite β, this policy assigns exponentially higher probabilities to similarity classes with higher valuations. As $\beta \uparrow \infty$, Alice becomes extremely sensitive to differences in valuations, causing the probabilities to concentrate on the set of similarity classes with maximal limiting valuation, denoted by $\mathcal{S}_\omega^{max} = \arg\max_{s\in\mathcal{S}_\omega} v_s^*$, at each choice node $\omega \in \Omega$. Correspondingly, in any state of the world $\omega \in \Omega$, the probability that Alice chooses an alternative from a dominated similarity class ($s \notin \arg\max_{s\in\mathcal{S}_\omega} v_s^*$) approaches 0 as $\beta \uparrow \infty$. Thus, in the high-sensitivity limit ($\beta \uparrow \infty$), the limiting fixed points satisfy the optimality condition of Valuation Equilibrium.

$$\sigma_\omega^s(\beta) \xrightarrow[\beta\to\infty]{a.s.} \begin{cases} (\sigma_\omega^s)^* & \text{if } s \in \mathcal{S}_\omega^{\max}, \text{ where } 0 < (\sigma_\omega^s)^* \leq 1 \text{ and } \sum_{s\in\mathcal{S}_\omega^{\max}} (\sigma_\omega^s)^* = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The expected payoff function $\mathbf{g}(\mathbf{v}; \beta)$ converges to a limiting function $\mathbf{g}_\infty(\mathbf{v}^*)$, which depends on the maximizers of $\mathbf{v}^*$. Let $\mathbf{v}_n \in \mathcal{V}(\beta_n)$ with $\beta_n \to \infty$. By compactness, $\mathbf{v}_n$ has a convergent subsequence $\mathbf{v}_{n_k} \to \mathbf{v}^*$. Using the closed graph property and continuity, $\mathbf{v}^*$ satisfies:

$$\mathbf{v}^* = \lim_{k\to\infty} \mathbf{g}(\mathbf{v}_{n_k}; \beta_{n_k}) = \mathbf{g}_\infty(\mathbf{v}^*)$$

Thus, $\mathbf{v}^*$ is a fixed point of $\mathbf{g}_\infty$, which characterizes the valuation equilibria. Since $\mathcal{V}(\beta)$ is upper hemicontinuous, for any $\epsilon > 0$, there exists $\hat{\beta}$ such that for all $\beta > \hat{\beta}$:

$$\mathcal{V}(\beta) \subseteq \bigcup_{\mathbf{v}^* \in \mathcal{V}(\infty)} B(\mathbf{v}^*, \epsilon)$$

where $B(\mathbf{v}^*, \epsilon)$ is the open ball of radius ϵ around $\mathbf{v}^*$, and $\mathcal{V}(\infty)$ denotes the set of valuation equilibria. Therefore, for $\beta > \hat{\beta}$, every SVE $\mathbf{v} \in \mathcal{V}(\beta)$ is within ϵ of some VE, $\mathbf{v}^* \in \mathcal{V}(\infty)$.

Moreover, $\mathbf{g}$ being a smooth function from the compact, convex set $K \subset \mathbb{R}^{|\mathcal{S}|}$ to itself, the Morse-Sard Theorem tells us that the set of critical values of $\mathbf{g}$ - which is the image of the set of critical points in K where the Jacobian $D_{\mathbf{v}}\mathbf{g}$ is not surjective - has Lebesgue measure zero in K. Also, since $\mathbf{g}$ is a non-constant real-analytic function from the convex set K to itself, the set of critical points of $\mathbf{g}$ (where the Jacobian is singular) has Lebesgue measure zero in K. This implies that almost all fixed points $\mathbf{v} \in \mathcal{V}(\beta)$ are regular fixed points, i.e., the Jacobian $D_{\mathbf{v}}\mathbf{g}$ is invertible at such fixed points and the Implicit Function Theorem applies.

Therefore, except for a null set of equilibria, the smooth valuation equilibria $\mathbf{v} \in \mathcal{V}(\beta)$ are locally unique, depend smoothly on β, and converge to valuation equilibria as $\beta \to \infty$. □

A.3 Theorem 2

Proof. The existence of a smooth valuation equilibrium (SVE) for the decision tree $\mathcal{T}_2'$ is guaranteed by Theorem 1. Lemma 2.1 guarantees that for a sufficiently large sensitivity parameter $\beta > \hat{\beta}$, each smooth valuation equilibrium arises in the neighborhood of some valuation equilibrium (VE). The decision tree $\mathcal{T}_2'$ with generic payoffs admits at most three valuation equilibria - two strict pure and one mixed. The set of VE characterizes the set $\mathcal{V}$ of stationary points of the CQL model in the high-sensitivity limit.

We denote Alice's valuations of similarity classes i and j at time t by $v_i(t)$ and $v_j(t)$ respectively. $\mathbf{v}(t) = (v_i(t), v_j(t)) \in \mathbb{R}^2$ denotes the valuation vector at time $t \in \mathbb{R}_+ \cup \{0\}$. $\beta > 0$ is her sensitivity parameter. Given her valuations, the mixed strategy map she employs is as follows. At node a, she chooses the alternative i_a in similarity class i with probability $\sigma_{ii} = 1$. At node b, she chooses the alternative j_b in similarity class j with probability $\sigma_{jj} = 1$. At node c, she chooses either the alternative i_c in similarity class i with probability σ_{ij} or the alternative j_c in similarity class j with probability $\sigma_{ji} = 1 - \sigma_{ij}$, where,

$$\sigma_{ij}(\mathbf{v}(t)) = \frac{\exp(\beta v_i(t))}{\exp(\beta v_i(t)) + \exp(\beta v_j(t))}.$$

The expected payoffs for her two similarity class, induced by the mixed strategy map, are:

$$g_i(\mathbf{v}(t)) = \frac{p_{ii}\sigma_{ii}u_{ii} + p_{ij}\sigma_{ij}u_{ij}}{p_{ii}\sigma_{ii} + p_{ij}\sigma_{ij}} = \frac{p_{ii}u_{ii} + p_{ij}\sigma_{ij}u_{ij}}{p_{ii} + p_{ij}\sigma_{ij}},$$
$$g_j(\mathbf{v}(t)) = \frac{p_{jj}\sigma_{jj}u_{jj} + p_{ij}\sigma_{ji}u_{ji}}{p_{jj}\sigma_{jj} + p_{ij}\sigma_{ji}} = \frac{p_{jj}u_{jj} + p_{ij}(1-\sigma_{ij})u_{ji}}{p_{jj} + p_{ij}(1-\sigma_{ij})}.$$

And, $\forall\ s \in \{i, j\}$, the CQL dynamics are governed by the 2-D (planar) system of ODEs:

$$\dot{v}_s = f_s(\mathbf{v}) = g_s(\mathbf{v}) - v_s.$$

We define, without loss of generality, similarity class j as the numeraire and appropriately rescale the valuations by subtracting the valuation of the numeraire j at all times t. The scalars $x(t) = v_i(t) - v_j(t)$ and $y(t) = v_j(t) - v_j(t) = 0$ represent the renormalized valuations for similarity classes i and j, respectively, at time t. This transformation is a translation of the original valuation vector (v_i, v_j) by the vector $(-v_j, -v_j)$, effectively setting the valuation of the numeraire class j to zero and expressing the valuation of class i relative to j. This simplification reduces the two-dimensional planar dynamical system to a one-dimensional flow on the real line. We note that the underlying CQL dynamics are invariant under such a translation of the valuations. To observe this, consider the probability σ_{ij} that similarity class i is chosen over similarity class j at node c, where,

$$\sigma_{ij} = \frac{\exp(\beta v_i(t))}{\exp(\beta v_i(t)) + \exp(\beta v_j(t))}.$$

Upon translating the valuations by $-v_j(t)$, we have:

$$\sigma_{ij} = \frac{\exp(\beta(v_i(t) - v_j(t)))}{\exp(\beta(v_i(t) - v_j(t))) + \exp(\beta(v_j(t) - v_j(t)))}.$$

Since $x(t) = v_i(t) - v_j(t)$, this simplifies to:

$$\sigma_{ij} = \frac{\exp(\beta x(t))}{\exp(\beta x(t)) + \exp(\beta y(t))} = \frac{\exp(\beta x(t))}{\exp(\beta x(t)) + \exp(0)} = \frac{1}{1 + \exp(-\beta x(t))}.$$

Given the logit choice rule, the derivative of σ_{ij} w.r.t. x is:

$$\sigma'_{ij} = \frac{d}{dx}(\sigma_{ij}) = \beta\sigma_{ij}(1 - \sigma_{ij}) = \beta \cdot \frac{\exp(-\beta x)}{(1 + \exp(-\beta x))^2} \geq 0.$$

Correspondingly, at any time t, we translate the vector of expected payoffs $(g_i(v), g_j(v))$ by the vector $(-g_j(v), -g_j(v))$. The scalars $g(x) = g_i(x) - g_j(x)$ and $h(x) = g_j(x) - g_j(x) = 0$

denote the renormalized expected payoffs for the similarity classes i and j at time t. The two-dimensional system of ODEs now reduces to a one-dimensional ODE in $x \in \mathbb{R}$ given by:

$$\dot{x} = f(x) = g(x) - x, \tag{9}$$

Here, $f : \mathbb{R} \to \mathbb{R}$ represents a smooth (C^∞) scalar field and its derivative w.r.t. x is:

$$f'(x) = \frac{d}{dx} f(x) = \frac{d}{dx}(g(x) - x) = \frac{d}{dx}(g(x)) - 1 = \frac{d}{dx}(g_i(x)) - \frac{d}{dx}(g_j(x)) - 1.$$

Plugging in the expressions for g_i and g_j, the derivative $f'(x)$ can be explicitly written as

$$f'(x) = p_{ij}\sigma'_{ij}\left(\frac{p_{ii}(u_{ij} - u_{ii})}{(p_{ii} + p_{ij}\sigma_{ij})^2} + \frac{p_{jj}(u_{ji} - u_{jj})}{(p_{jj} + p_{ij}(1 - \sigma_{ij}))^2}\right) - 1 \tag{10}$$

We observe that the scalar field $f(x)$ is globally Lipschitz continuous for any finite $\beta \geq 0$ since the payoffs are bounded by assumption. By the Picard-Lindelöf theorem, the Lipschitz condition on f guarantees that for any initial condition $x(0) = x_0$, there exists a unique solution $x(t)$ to the differential equation $\dot{x} = f(x)$ for all times $t \in \mathbb{R}_+$. Moreover, this implies that non-trivial periodic solutions such as cycles (oscillations) are impossible. Given that the dynamics are restricted to the real line, the existence of a cycle would contradict the uniqueness of solutions to the initial value problem. Recall that a fixed point $x^* \in \mathbb{R}$ satisfies $f(x^*) = 0$. At these points, the solution remains constant: if $x(0) = x^*$, then $x(t) = x^*$ for all $t \geq 0$. If x is not at a fixed point, i.e. $x \neq x^*$, then the sign of $f(x)$ determines the direction of $x(t)$. If $f(x) > 0$, then $\dot{x} > 0$ and $x(t)$ increases over time. If $f(x) < 0$, then $\dot{x} < 0$ and $x(t)$ decreases over time. Therefore, away from fixed points, the flow is monotonic (i.e., strictly increasing or decreasing). This monotonicity ensures that the trajectory cannot return to a previous state without reversing direction, which would require another fixed point in between where the flow reverses. However, such intermediate fixed points would again either attract or repel the trajectory, preventing cyclic behavior.

Another way of establishing that non-trivial periodic closed orbits are impossible in a one-dimensional real, smooth dynamical system is by observing that such a system always admits a potential function and therefore can be represented as a gradient system. Recall, that a potential function $V : \mathbb{R} \to \mathbb{R}$ for this system is a function such that

$$\dot{x} = f(x) = -\frac{dV}{dx}.$$

To show that a potential function exists, we construct $V(x)$ by integrating $f(x)$ w.r.t. x:

$$V(x) = -\int f(x)\,dx.$$

Since $f(x)$ is smooth, the Fundamental Theorem of Calculus guarantees that the integral $\int f(x)\,dx$ exists and is a smooth function of x. Therefore, we have constructed a potential function $V(x)$ such that the original system can be expressed as a gradient system $\dot{x} = -\frac{dV}{dx}$ with the potential function $V(x)$. Now, we use the potential function $V(x)$ to rule out closed orbits. Along the trajectories of the system, we have:

$$\frac{dV}{dt} = \frac{dV}{dx} \cdot \frac{dx}{dt} = \frac{dV}{dx} \cdot \left(-\frac{dV}{dx}\right) = -\left(\frac{dV}{dx}\right)^2 \leq 0.$$

This indicates that $V(x)$ is non-increasing with time. It either decreases or remains constant. For a trajectory to form a closed orbit, the system must return to its initial state after some finite time, implying $V(x)$ must return to its initial value. However, since $V(x)$ is non-increasing along the trajectory, it cannot return to a previous higher value in finite time unless it remains constant. The only case where $V(x)$ remains constant is when $\frac{dV}{dx} = 0$, which corresponds to equilibrium points ($\dot{x} = 0$). These points are not closed orbits but fixed points where the system is stationary. Hence, the potential function $V(x)$ demonstrates that the system cannot have non-trivial closed orbits. The trajectories cannot loop back to their initial states, ensuring that no closed orbits exist in the 1-D real, smooth CQL dynamics.

Furthermore, we notice that the forward (in time) trajectories in the one-dimensional CQL dynamics are bounded within a positively invariant interval defined by the extreme points of the image of the renormalized $g(x)$ map. This ensures that forward trajectories cannot asymptotically escape to the infinities. To see this, recall that since the payoffs are bounded, there exist $M_1 \geq 0$ and $M_2 \geq 0$ such that $-M_1 \leq g(x) \leq M_2$ for all $x \in \mathbb{R}$. We verify that for $x > M_2$, $f(x) \leq M_2 - x < 0$, and for $x < -M_1$, $f(x) \geq -M_1 - x > 0$. Thus, outside the interval $[-M_1, M_2]$, $f(x)$ directs trajectories inward, ensuring that they do not escape to the infinities and in fact, enter the interval $[-M_1, M_2]$ in finite time. The boundedness of $g(x)$ and the inward direction of $f(x)$ outside $[-M_1, M_2]$ ensure that once trajectories enter this compact interval, they remain confined within it for all future times - positive invariance of $[-M_1, M_2]$. Evidently, any fixed point x^* is contained within the compact interval $[-M_1, M_2]$. Trajectories within the interval are either themselves fixed points or monotonically approach fixed points in the long-run. Thus, the asymptotics of CQL dynamics reduce to analyzing the local stability around rest points in $[-M_1, M_2]$.

We note that, by Lemma 2.1, any SVE must lie in the neighborhood of either a strict pure VE (on either boundary of the interval $[-M_1, M_2]$) or a mixed VE (at the origin) for a sufficiently large β. To examine local stability, we linearize the ODE $\dot{x} = f(x)$ and evaluate $f'(x)$ around a fixed point x^* of the CQL dynamical system. The fixed point may either correspond to a strict pure VE (at $x^* \neq 0$) or a mixed VE (at $x^* = 0$) for $\beta > \hat{\beta}$. Assume there exists an SVE that arises near a strict pure VE for $\beta > \hat{\beta}$, i.e., $x^* \neq 0$. There are potentially two distinct strict pure VE in our setting. Without loss of generality, assume $x^* > 0$. Therefore, similarity class j is dominated by similarity class i in the equilibrium valuation. That is, in equilibrium, the agent at the node c selects the alternative in similarity class i with probability $\sigma_{ij} \to 1$ exponentially as $\beta \to \infty$. This implies,

$$\lim_{\beta\to\infty} \sigma'_{ij}(x)\Big|_{x=x^*>0} = \lim_{\beta\to\infty} \beta \cdot \frac{\exp(-\beta x)}{(1+\exp(-\beta x))^2}\Big|_{x=x^*>0} = 0$$

Thus, we have:

$$\lim_{\beta\to\infty} f'(x)\Big|_{x=x^*>0} = -1$$

Here, $\sigma'_{ij}(x)$ denotes the derivative of the choice probability function $\sigma_{ij}(x)$ with respect to x, and $f'(x)$ represents the derivative of the scalar field $f(x)$ w.r.t. x evaluated at $x = x^* > 0$. By continuity in β, for a sufficiently large but finite $\beta > \hat{\beta}$, the derivative $f'(x)$ remains negative implying a strictly decreasing $f(x)$ in a neighborhood of a strict pure SVE. Therefore, by the Linearization Theorem, if there exists a strict pure valuation equilibrium, a fixed point (SVE) of the CQL model that lies in the vicinity of the strict pure VE, is guaranteed to be locally asymptotically stable for a sufficiently large sensitivity parameter. By the inverse function theorem, such a fixed point is also locally isolated. Naturally, if the strict pure SVE is additionally unique, as in Example 3.3, it is also globally asymptotically stable for a sufficiently large sensitivity parameter. That is, the CQL dynamics asymptotically converge to the unique fixed point x^* from any initial valuation x_0. To see this, recall that the approach of bounded trajectories to the unique attracting fixed point is monotonic in a gradient system with the potential function serving as a Lyapunov function as proved below.

On the other hand, depending on the underlying primitives, it might occur that there are two strict pure valuation equilibria (VE) located at the boundaries, along with one mixed VE situated in the interior. The strict pure VE correspond to similarity class i (at $x_i^* > 0$) and similarity class j (at $x_j^* < 0$) being chosen with probability 1 at node c in the high-sensitivity limit, as illustrated in Example 3.1. We have already established that each strict pure VE is locally asymptotically stable within its distinct basin of attraction, for a sufficiently

large sensitivity parameter. Additionally, there exists a unique mixed VE at the origin (the intersection of the two basins) where Alice randomizes between the two similarity classes due to indifference. The smooth valuation equilibrium (SVE) in the vicinity of the mixed VE at the origin is asymptotically unstable in the CQL dynamics for sufficiently large β. To demonstrate this instability, we exploit the local stability of the pure SVE, which indicates that for sufficiently large β, the smooth function $f(x)$ is decreasing near the pure VE points: $f(x_i^*) = 0$ and $f'(x_i^*) < 0$, as well as $f(x_j^*) = 0$ and $f'(x_j^*) < 0$. Since these fixed points are locally isolated, the intermediate value theorem guarantees the existence of a point $x_{ij}^* \in (x_j, x_i)$ where $f(x_{ij}^*) = 0$. Given that a smooth scalar field f can reverse signs only at fixed points, and there is exactly one mixed VE (high-sensitivity limit of mixed SVEs) at the origin, there exists a unique point in the interval near $x_{ij}^* = 0$ around which $f(x)$ increases from negative to positive, for $\beta > \hat{\beta}$. Thus, $f'(x_{ij}^*) > 0$, indicating that the mixed SVE near $x_{ij}^* = 0$ is repelling or locally asymptotically unstable for a sufficiently large sensitivity parameter. Pemantle (1990) shows that the discrete-time system in Eq. (4) has a probability 0 of converging to a linearly unstable fixed point of the continuous-time process in Eq. (6), provided that there is a non-negligible amount of noise in the system. Given a sufficiently large β, the CQL dynamics converge to the pure VE at $x_i^* > 0$ ($x_j^* < 0$) starting from an arbitrary positive (negative) initial valuation x_0.

We shift our attention to the remaining case where no strict pure VE exists. By the existence result in Jehiel and Samet (2007), we know that there must exist a unique mixed VE where the agent is indifferent between her two similarity classes, i.e., at the origin. In the absence of a strict pure VE, Lemma 2.1 implies that an SVE, x^* s.t. $f(x^*) = 0$, must lie in the vicinity of the unique mixed VE at the origin for a sufficiently large $\beta > \hat{\beta}$. We show that this unique mixed SVE is globally asymptotically stable in the CQL dynamics for sufficiently large β. A geometric proof is as follows. Recall that since the payoffs are bounded, there exist $M_1 \geq 0$ and $M_2 \geq 0$ such that $-M_1 \leq g(x) \leq M_2$ for all $x \in \mathbb{R}$. For $x > M_2$, $f(x) \leq M_2 - x < 0$, and for $x < -M_1$, $f(x) \geq -M_1 - x > 0$. Since a smooth scalar field f cannot reverse sign without encountering a fixed point (by the intermediate value theorem), and there are no fixed points around the boundaries of the interval (by the absence of strict pure VE), by continuity, $f(x) > 0$ for $x < x^*$ and $f(x) < 0$ for $x > x^*$. Therefore, there is a unique, locally isolated fixed point x^* near the origin in whose neighborhood $f(x)$ strictly decreases from positive to negative, i.e. $f'(x^*) < 0$. By the linearization theorem, this fixed point that corresponds to the unique mixed VE in the high-sensitivity limit, is locally asymptotically stable in the CQL dynamics. In fact, since the one-dimensional CQL model is a gradient system, the unique mixed valuation equilibrium is also globally asymptotically stable for a

sufficiently large sensitivity parameter. That is, the CQL dynamics asymptotically converge to the unique fixed point x^* from any initial valuation x_0, for a sufficiently large $\beta > \hat{\beta}$.

To see this, we observe that $(x - x^*)f(x) < 0$ for all $x \neq x^*$, indicating that the function $f(x)$ drives the state x toward the fixed point x^* from both sides. We define a Lyapunov function $V(x)$ that measures the "distance" from the fixed point

$$V(x) = \frac{1}{2}(x - x^*)^2.$$

This function is positive definite and radially unbounded, satisfying $V(x) > 0$ for all $x \neq x^*$ and $V(x^*) = 0$. We calculate the time derivative of $V(x)$ along the solutions of the ODE:

$$\frac{dV}{dt} = V'(x) \cdot \frac{dx}{dt} = (x - x^*) \cdot f(x).$$

Since $(x - x^*)f(x) < 0$ for all $x \neq x^*$, it follows that:

$$\frac{dV}{dt} < 0 \quad \text{for all} \quad x \neq x^*.$$

This implies that $V(x)$ decreases along trajectories, except at the fixed point. Since $V(x)$ is positive definite and its derivative $\frac{dV}{dt}$ is negative definite, Lyapunov's direct method tells us that the fixed point x^* is globally asymptotically stable. Because the Lyapunov function decreases continuously and unboundedly over time, all trajectories starting from any initial condition $x(0) \in \mathbb{R}$ will converge to x^* as $t \to \infty$: $\lim_{t\to\infty} x(t) = x^*$. □

A.4 Theorem 5

Proof. Consider an arbitrary similarity class $s \in \mathcal{S}$. We aim to prove that, for all $z < \tilde{z}$, there exists a valuation equilibrium where s is the unique strictly dominated similarity class. Assume $v_s < v_k$, for all $k \in \mathcal{S} \setminus \{s\}$. We will show that this partial order on valuations can be sustained in equilibrium. Given $v_s < v_k$, for all $k \in \mathcal{S} \setminus \{s\}$, by optimality, the similarity class s is selected exclusively at the trivial unary choice node $\omega_s = \{s\}$, where s is the only available similarity class. Therefore, by consistency, the valuation $v_s = \pi_{\{s\}}(s) + z$. Correspondingly, for an arbitrary class $k \in \mathcal{S} \setminus \{s\}$, by consistency,

$$\underline{v_k} = \inf_{k \in \mathcal{S}\setminus\{s\}} v_k = \frac{p(\omega_k)(\pi_{\{k\}}(k) + z) + p(\omega_{sk})\pi_{\{s,k\}}(k)}{p(\omega_k) + p(\omega_{sk})}.$$

To see this, note that the lowest possible valuation for class $k \in \mathcal{S} \setminus \{s\}$ corresponds to the case where k is the unique strictly dominated similarity class in the set $\mathcal{S} \setminus \{s\}$. In that case, by optimality, class k is selected at the following two nodes - the trivial unary choice node $\omega_k = \{k\}$ and the binary choice node $\omega_{sk} = \{s, k\}$. Given the order of valuations, it follows that the weight associated with the negative constant z in v_s is greater than that in $\underline{v_k}$. By making z sufficiently small (more negative), the difference $\underline{v_k} - v_s$ can be made arbitrarily large and positive. Indeed, it is then optimal for the agent to choose class s exclusively at the trivial choice node $\omega_s = \{s\}$ where s is the only available similarity class. As z is made smaller, the sub-optimality of s relative to $\mathcal{S} \setminus \{s\}$ is reinforced. Therefore, $\exists\ \tilde{z} < 0$, such that $\forall\ z < \tilde{z}$, $v_s^* < v_k^*$ for all $k \in \mathcal{S} \setminus \{s\}$ constitutes a valuation equilibrium. Since class s was arbitrarily chosen and the constant $z < \tilde{z}$ is added to the trivial choice payoff for each similarity class, the argument extends to any similarity class in $\mathcal{S}$. Thus, we've proved that for each similarity class $s \in \mathcal{S}$, there exists a valuation equilibrium (VE) where s is the unique strictly dominated similarity class and that there is a multiplicity of VE for $z < \tilde{z}$.

To rigorously show that the partial order described above can be sustained in equilibrium, we proceed as follows. Let us define the subspace $\mathbf{V}_s$ of valuation vectors where similarity class s has the lowest valuation by at least δ: $\mathbf{V}_s = \{\mathbf{v} \in K : v_s \leq v_k - \delta, \forall k \in \mathcal{S} \setminus \{s\}\}$, where $\delta > 0$ is a fixed positive constant, and K is a compact, convex subset of $\mathbb{R}^{|\mathcal{S}|}$ defined by the convex hull of the bounded payoffs. We need to demonstrate that the mapping $\mathbf{v} \mapsto \mathbf{g}(\mathbf{v})$ maps $\mathbf{V}_s$ to itself; that is, if $\mathbf{v} \in \mathbf{V}_s$, then $\mathbf{g}(\mathbf{v}) \in \mathbf{V}_s$. For similarity class s, at the trivial unary choice node $\omega_s = \{s\}$, the choice probability is: $\sigma^s_{\omega_s}(\mathbf{v}) = 1$. At any non-trivial binary choice node ω where s is available along with another class, since $v_s \leq v_k - \delta$ for all $k \neq s$, we have $v_k - v_s \geq \delta$. Therefore, the choice probability $\sigma^s_\omega(\mathbf{v})$ satisfies:

$$\frac{\sigma^s_\omega(\mathbf{v})}{\sigma^k_\omega(\mathbf{v})} = \frac{\exp(\beta v_s)}{\exp(\beta v_k)} = \exp(\beta(v_s - v_k)) \leq \exp(-\beta\delta) < 1.$$

As $\beta \to \infty$, $\exp(-\beta\delta) \to 0$, so: $\sigma^s_\omega(\mathbf{v}) \leq \exp(-\beta\delta) \to 0$. Thus, the valuation v_s solely depends on the payoff from the unary node ω_s and puts a weight 1 on the negative constant z. For similarity classes $k \neq s$, at their respective unary nodes $\omega_k = \{k\}$, the choice probability is: $\sigma^k_{\omega_k}(\mathbf{v}) = 1$. At non-trivial binary nodes where k competes with s, since $v_k \geq v_s + \delta$, we have $v_k - v_s \geq \delta$. The choice probability $\sigma^k_\omega(\mathbf{v})$ satisfies:

$$\sigma^k_\omega(\mathbf{v}) = \frac{\exp(\beta v_k)}{\sum_{j \in \mathcal{S}_\omega} \exp(\beta v_j)} \geq \frac{\exp(\beta(v_s + \delta))}{\exp(\beta v_s) + \exp(\beta(v_s + \delta))} = \frac{1}{1 + \exp(-\beta\delta)}.$$

As $\beta \to \infty$, $\exp(-\beta\delta) \to 0$, so: $\sigma^k_\omega(\mathbf{v}) \geq \frac{1}{1+\exp(-\beta\delta)} \to 1$. Therefore, the valuations v_k are

influenced by payoffs from multiple nodes (both trivial and non-trivial) and are less affected by the negative constant z. The mapping $\mathbf{g}(\mathbf{v})$ preserves the ordering $v_s \leq v_k - \delta$ for all $k \neq s$, thus mapping $\mathbf{V}_s$ to itself. $\mathbf{V}_s$ is defined by linear inequalities, which are convex constraints. The intersection of convex sets is convex. Since the inequalities are non-strict, $\mathbf{V}_s$ is closed. Being a closed subset of the compact set K, $\mathbf{V}_s$ is compact. Since $\mathbf{V}_s$ is compact, convex, and non-empty, and $\mathbf{g}(\mathbf{v})$ is continuous and maps $\mathbf{V}_s$ to itself, by Brouwer's Fixed Point Theorem, there exists at least one fixed point $\mathbf{v}^* \in \mathbf{V}_s$. This fixed point corresponds to a VE where similarity class s is the unique strictly dominated class in the high-sensitivity limit. We prove the existence of at least one strict pure VE by construction.

Constructing a strict pure VE is equivalent to assigning a strict total order onto the equilibrium valuations. Consider the set of similarity classes $\mathcal{S}$. $\Omega_s = \{\omega \in \Omega : s \in \mathcal{S}_\omega\}$ is the finite set of nodes where similarity class s is available. Let

$$\mathcal{I} = \underset{s \in \mathcal{S}}{\operatorname{argmin}} \frac{p(\omega = \{s\})}{\sum_{\omega \in \Omega_s} p(\omega)}.$$

The set $\mathcal{I}$ is non-empty since $\mathcal{S}$ and Ω_s are finite sets and for each $s \in \mathcal{S}$, there exist at least two distinct $\omega_s \in \Omega_s$ (unary and binary states) such that $p(\omega_s) > 0$, by assumption. We choose a similarity class $i \in \mathcal{I}$ at random and assign it the highest equilibrium valuation such that $v_i^* > v_s^*$ for all $s \in \mathcal{S} \setminus \{i\}$. In fact, by consistency, we know that

$$v_i^* = \frac{\sum_{\omega \in \Omega_i} p(\omega)\pi_\omega(i)}{\sum_{\omega \in \Omega_i} p(\omega)}.$$

Notice that i has maximal equilibrium valuation among all similarity classes precisely because it assigns the lowest possible weight to the arbitrarily small constant z in its consistent equilibrium valuation. Let $\mathcal{S}_{-i} = \mathcal{S} \setminus \{i\}$ and $\Omega^{-i} = \mathcal{P}(\mathcal{S}_{-i}) \setminus \{\emptyset\}$. We define $\Omega_s^{-i} = \{\omega \in \Omega^{-i} : s \in \mathcal{S}_\omega\}$. Let

$$\mathcal{J} = \underset{s \in \mathcal{S}_{-i}}{\operatorname{argmin}} \frac{p(\omega = \{s\})}{\sum_{\omega \in \Omega_s^{-i}} p(\omega)}.$$

Clearly, $\mathcal{J}$ is non-empty. We pick a similarity class, j, from the set $\mathcal{J}$ at random. The similarity class j is assigned the second-highest equilibrium valuation such that $v_j^* < v_i^*$ but $v_j^* > v_s^*$ for all $s \in \mathcal{S} \setminus \{i, j\}$. By consistency, we establish that

$$v_j^* = \frac{\sum_{\omega \in \Omega_j^{-i}} p(\omega)\pi_\omega(j)}{\sum_{\omega \in \Omega_j^{-i}} p(\omega)}.$$

The process is repeated in the same manner until we have arrived at the unique strictly

dominated similarity class n whose equilibrium valuation assigns the highest possible weight, 1, to the arbitrarily small constant z. Indeed, by consistency, $v_n^* = \pi_{\{n\}}(n) + z$.

By construction, the equilibrium valuations of the similarity classes would follow a strict total order: $v_i^* > v_j^* > \cdots > v_n^*$. The construction ensures that higher-valued classes have minimal influence from the negative z, maintaining the strict order. By optimality, at each choice node, the unique similarity class with the highest valuation among the available classes is chosen deterministically. Therefore, for $z < \tilde{z}$, there exists at least one strict pure valuation equilibrium. Correspondingly, by Theorem 3, there exists a $\hat{\beta} > 0$, such that for all $\beta > \hat{\beta}$, the smooth valuation equilibrium that arises in the neighborhood of the strict pure valuation equilibrium is locally unique and locally asymptotically stable in the CQL dynamics. □

A.5 Proposition 2

Proof. Given n similarity classes, we can define $n!$ distinct strict total orders on their valuations. Let's consider an arbitrary strict order $v_1 < \cdots < v_k < v_s < \cdots < v_n$. We consider two arbitrary similarity classes s and k where $k \neq s$ in the set $\mathcal{S}$ such that the valuations satisfy $v_k < v_s$. Recall that z is a small, negative constant added to the trivial unary choice payoffs. Given the order of the valuations, it follows that the weight associated with the negative constant z in v_k is greater than that in v_s. This observation is based on the fact that the class s with a higher valuation is selected at a greater proportion of choice nodes compared to the class k with a lower valuation. For e.g., at the binary choice node featuring both classes, s is strictly preferred over k. By making z sufficiently small, the difference $v_s - v_k$ can be made arbitrarily large and positive. Consequently, it is optimal for the agent to choose class s over class k at a larger number of nodes, including at the binary choice node involving both. As z is made smaller, the optimality of s over k is reinforced.

Thus, the strict order is self-confirming for sufficiently negative z. Given the arbitrary selection of s and k, this reasoning extends to the relative strict ordering between any two similarity classes in $\mathcal{S}$. Therefore, there exists a threshold $\tilde{z} < 0$ such that for all $z < \tilde{z}$, the strict total order $v_1 < \cdots < v_k < v_s < \cdots < v_n$ constitutes a strict pure valuation equilibrium. We note that the consistency condition in VE is also satisfied. The valuations are well-defined since for each similarity class s, there exists at least one node where s is chosen with probability 1, including for the similarity class with the lowest valuation in equilibrium that is selected only at the trivial unary choice node featuring it.

This argument generalizes to every strict total order that can be defined on the set $\{v_s : s \in \mathcal{S}\}$, implying that sufficiently low payoffs at the trivial unary choice nodes result in $n!$

distinct strict pure valuation equilibria. According to Theorem 3, there exists a $\hat{\beta} > 0$, such that for all $\beta > \hat{\beta}$, the smooth valuation equilibrium that arises in the neighborhood of each strict pure valuation equilibrium is locally unique and locally asymptotically stable in the CQL dynamics within its basin of attraction.

At each such equilibrium, the vector field $\mathbf{h}(\mathbf{v}) = -\mathbf{f}(\mathbf{v})$ has an index of $+1$ since all eigenvalues of the Jacobian matrix $\mathbf{J_h} = -\mathbf{J_f}$ are strictly positive for a sufficiently large sensitivity parameter. Consequently, there are $n!$ isolated zeroes with indices $+1$ corresponding to the $n!$ strict pure valuation equilibria, each being locally asymptotically stable in the CQL dynamics. Given $\chi(K) = 1$, by the Poincare-Hopf index theorem, there must be at least $n! - 1$ non-degenerate zeroes of the vector field $\mathbf{h}(\mathbf{v})$ with indices -1 in the interior of K. An index of -1 indicates that at least one eigenvalue of the Jacobian matrix $\mathbf{J_h}$ has a strictly negative real part, which implies that at least one eigenvalue of the Jacobian matrix $\mathbf{J_f} = -\mathbf{J_h}$ has a strictly positive real part. Therefore, by the linearization theorem, each of the at least $n! - 1$ smooth valuation equilibria in the interior (corresponding to partially-mixed VE in the high-sensitivity limit) is asymptotically unstable in the CQL dynamics for $\beta > \hat{\beta}$. $\square$