Coarse Q-learning in Complex Decisions

Indifference v. Indeterminacy v. Instability

Philippe Jehiel (PSE & UCL) Aviman Satpathy (PSE)

EEA Congress 2025
Bordeaux School of Economics

August 25, 2025

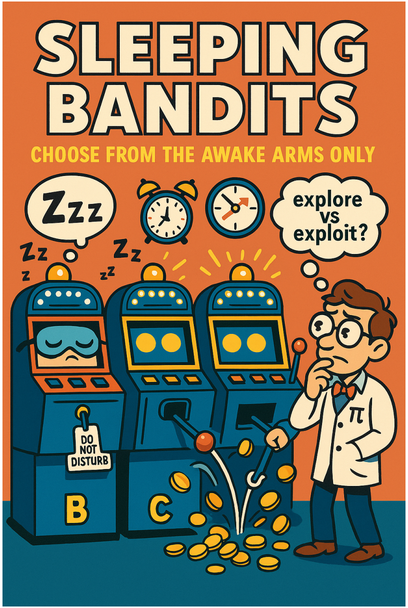


Figure: Sleeping Bandits generalize the classic multi-armed bandits problem allowing for the menu of available arms to vary i.i.d. across periods

Introduction: Setting

- A decision-maker (DM) repeatedly encounters a complex choice problem with:

Introduction: Setting

- A decision-maker (DM) repeatedly encounters a complex choice problem with:
  - an extensive finite grand set of alternatives denoted by $\mathcal{A}$

Introduction: Setting

- A decision-maker (DM) repeatedly encounters a complex choice problem with:
  - an extensive finite grand set of alternatives denoted by $\mathcal{A}$
  - a finite state space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives

Introduction: Setting

- A decision-maker (DM) repeatedly encounters a complex choice problem with:
  - an extensive finite grand set of alternatives denoted by $\mathcal{A}$
  - a finite state space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\emptyset\}$, i.e., set of all non-empty menus of alternatives
  - each alternative $a \in \mathcal{A}$ is described by a high-dimensional feature vector

Introduction: Setting

- A decision-maker (DM) repeatedly encounters a complex choice problem with:
  - an extensive finite grand set of alternatives denoted by $\mathcal{A}$
  - a finite state space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives
  - each alternative $a \in \mathcal{A}$ is described by a high-dimensional feature vector
  - DM only observes a fixed, lower-dimensional salient projection of this vector, with many features remaining unobserved

Introduction: Setting

- A decision-maker (DM) repeatedly encounters a complex choice problem with:
  - an extensive finite grand set of alternatives denoted by $\mathcal{A}$
  - a finite state space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives
  - each alternative $a \in \mathcal{A}$ is described by a high-dimensional feature vector
  - DM only observes a fixed, lower-dimensional salient projection of this vector, with many features remaining unobserved
  - stochastic (state-independent) payoffs r_a drawn i.i.d. from distributions unknown to the DM

Introduction: Setting

- A decision-maker (DM) repeatedly encounters a complex choice problem with:
  - an extensive finite grand set of alternatives denoted by $\mathcal{A}$
  - a finite state space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives
  - each alternative $a \in \mathcal{A}$ is described by a high-dimensional feature vector
  - DM only observes a fixed, lower-dimensional salient projection of this vector, with many features remaining unobserved
  - stochastic (state-independent) payoffs r_a drawn i.i.d. from distributions unknown to the DM
- **Motivation:** In such a setting, individually assessing each alternative becomes impractical and the literature on psychology (Rosch & Lloyd, 1978) suggests that the DM may naturally resort to categorical models of reasoning for simplification

  $\implies$ need for developing models of coarse (aggregated) learning

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)
- We introduce the **Coarse Q-learning (CQL)** Model:

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)
- We introduce the **Coarse Q-learning (CQL)** Model:
  - salient projection of the feature vector partitions the grand set of alternatives $\mathcal{A}$ into equivalence classes called *similarity classes*, $s \in \mathcal{S}$

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)
- We introduce the **Coarse Q-learning (CQL)** Model:
  - salient projection of the feature vector partitions the grand set of alternatives $\mathcal{A}$ into equivalence classes called *similarity classes*, $s \in \mathcal{S}$
  - two distinct alternatives deemed similar by the DM iff they share identical salient projections, even if they may disagree on one or more non-salient features

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)
- We introduce the **Coarse Q-learning (CQL)** Model:
  - salient projection of the feature vector partitions the grand set of alternatives $\mathcal{A}$ into equivalence classes called *similarity classes*, $s \in \mathcal{S}$
  - two distinct alternatives deemed similar by the DM iff they share identical salient projections, even if they may disagree on one or more non-salient features
  - misspecified DM erroneously believes that non-salient features are payoff-irrelevant and reinforces estimates (valuations) of the expected payoffs at the level of similarity classes

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)
- We introduce the **Coarse Q-learning (CQL)** Model:
  - salient projection of the feature vector partitions the grand set of alternatives $\mathcal{A}$ into equivalence classes called *similarity classes*, $s \in \mathcal{S}$
  - two distinct alternatives deemed similar by the DM iff they share identical salient projections, even if they may disagree on one or more non-salient features
  - misspecified DM erroneously believes that non-salient features are payoff-irrelevant and reinforces estimates (valuations) of the expected payoffs at the level of similarity classes
  - decisions are based on the aggregate payoff consequences of the clusters

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)
- We introduce the **Coarse Q-learning (CQL)** Model:
  - salient projection of the feature vector partitions the grand set of alternatives $\mathcal{A}$ into equivalence classes called *similarity classes*, $s \in \mathcal{S}$
  - two distinct alternatives deemed similar by the DM iff they share identical salient projections, even if they may disagree on one or more non-salient features
  - misspecified DM erroneously believes that non-salient features are payoff-irrelevant and reinforces estimates (valuations) of the expected payoffs at the level of similarity classes
  - decisions are based on the aggregate payoff consequences of the clusters
  - beyond salience: other drivers for use of similarity classes in learning may include memory constraints, aggregated feedback, popular narratives, etc.

Introduction: Model

- **Focus**: prior-free setting where a DM with coarse perception of her alternatives employs a simple adaptive reinforcement learning heuristic: Q-learning
  - Watkins & Dayan (1992) prove almost sure convergence of Q-learning to nearly-optimal policies in standard (finite) stationary Markov Decision Processes (MDPs)
- We introduce the **Coarse Q-learning (CQL)** Model:
  - salient projection of the feature vector partitions the grand set of alternatives $\mathcal{A}$ into equivalence classes called *similarity classes*, $s \in \mathcal{S}$
  - two distinct alternatives deemed similar by the DM iff they share identical salient projections, even if they may disagree on one or more non-salient features
  - misspecified DM erroneously believes that non-salient features are payoff-irrelevant and reinforces estimates (valuations) of the expected payoffs at the level of similarity classes
  - decisions are based on the aggregate payoff consequences of the clusters
  - beyond salience: other drivers for use of similarity classes in learning may include memory constraints, aggregated feedback, popular narratives, etc.
- Research Question: How does coarse inference influence Q-learning dynamics — specifically, the nature of steady-states and convergence in the long-run?

Introduction: Results

- Key Findings: Coarse Q-learning generates rich long-run dynamics:

Introduction: Results

- Key Findings: Coarse Q-learning generates rich long-run dynamics:
    - if a strict pure equilibrium exists, it is **locally asymptotically stable** in the CQL dynamics.

Introduction: Results

- Key Findings: Coarse Q-learning generates rich long-run dynamics:
  - if a strict pure equilibrium exists, it is **locally asymptotically stable** in the CQL dynamics.
  - in certain decision trees with generic expected payoffs, a **unique mixed equilibrium** exists that is **globally asymptotically stable** in the CQL dynamics (#*Indifference*)

Introduction: Results

- Key Findings: Coarse Q-learning generates rich long-run dynamics:
  - if a strict pure equilibrium exists, it is **locally asymptotically stable** in the CQL dynamics.
  - in certain decision trees with generic expected payoffs, a **unique mixed equilibrium** exists that is **globally asymptotically stable** in the CQL dynamics (#*Indifference*)
  - in other decision trees, we observe **multiple equilibria**, with each **strict pure** equilibrium exhibiting **local asymptotic stability** (#*Indeterminacy*)

Introduction: Results

- Key Findings: Coarse Q-learning generates rich long-run dynamics:
  - if a strict pure equilibrium exists, it is **locally asymptotically stable** in the CQL dynamics.
  - in certain decision trees with generic expected payoffs, a **unique mixed equilibrium** exists that is **globally asymptotically stable** in the CQL dynamics (#*Indifference*)
  - in other decision trees, we observe **multiple equilibria**, with each **strict pure** equilibrium exhibiting **local asymptotic stability** (#*Indeterminacy*)
  - in yet other cases, there exists **no asymptotically stable equilibrium**, and the long-run CQL dynamics converge to a **stable limit cycle** (#*Instability*)

Model: Alternatives & Menus

- $\mathcal{A}$ denotes a finite grand set of alternatives, with each $a \in \mathcal{A}$ characterized by an N-tuple of features
$$x(a) = \Big(x_1(a), \ldots, x_N(a)\Big) \in X_1 \times \cdots \times X_N,$$
where each X_i is standard Borel. For e.g. - color, size, shape, etc.

Model: Alternatives & Menus

- $\mathcal{A}$ denotes a finite grand set of alternatives, with each $a \in \mathcal{A}$ characterized by an N-tuple of features
$$x(a) = \Big(x_1(a), \ldots, x_N(a)\Big) \in X_1 \times \cdots \times X_N,$$
where each X_i is standard Borel. For e.g. - color, size, shape, etc.

- State space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives

Model: Alternatives & Menus

- $\mathcal{A}$ denotes a finite grand set of alternatives, with each $a \in \mathcal{A}$ characterized by an N-tuple of features
$$x(a) = \Big(x_1(a), \ldots, x_N(a)\Big) \in X_1 \times \cdots \times X_N,$$
where each X_i is standard Borel. For e.g. - color, size, shape, etc.
- State space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives
- For state $\psi \in \Psi$ and available $a \in \psi$, the realized payoff r_a is drawn from a distribution
$$F_a \ = \ G\Big(x(a)\Big) \in \Delta(\mathbb{R}),$$
where $G : \prod_{i=1}^{N} X_i \to \Delta(\mathbb{R})$ is measurable, $\mu_a = \mathbb{E}[r_a] < \infty$, and $\sigma_a^2 = \mathrm{Var}(r_a) < \infty$

Model: Alternatives & Menus

- $\mathcal{A}$ denotes a finite grand set of alternatives, with each $a \in \mathcal{A}$ characterized by an N-tuple of features

$$x(a) = \Big(x_1(a), \ldots, x_N(a)\Big) \in X_1 \times \cdots \times X_N,$$

  where each X_i is standard Borel. For e.g. - color, size, shape, etc.

- State space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives

- For state $\psi \in \Psi$ and available $a \in \psi$, the realized payoff r_a is drawn from a distribution

$$F_a \;=\; G\Big(x(a)\Big) \in \Delta(\mathbb{R}),$$

  where $G : \prod_{i=1}^{N} X_i \to \Delta(\mathbb{R})$ is measurable, $\mu_a = \mathbb{E}[r_a] < \infty$, and $\sigma_a^2 = \mathrm{Var}(r_a) < \infty$

- Genericity: no ties in expectations, $a \neq b \;\Rightarrow\; \mu_a \neq \mu_b$

Model: Alternatives & Menus

- $\mathcal{A}$ denotes a finite grand set of alternatives, with each $a \in \mathcal{A}$ characterized by an N-tuple of features
$$x(a) = \Big(x_1(a), \ldots, x_N(a)\Big) \in X_1 \times \cdots \times X_N,$$
where each X_i is standard Borel. For e.g. - color, size, shape, etc.
- State space: $\Psi = \mathcal{P}(\mathcal{A}) \setminus \{\varnothing\}$, i.e., set of all non-empty menus of alternatives
- For state $\psi \in \Psi$ and available $a \in \psi$, the realized payoff r_a is drawn from a distribution
$$F_a \;=\; G\Big(x(a)\Big) \in \Delta(\mathbb{R}),$$
where $G : \prod_{i=1}^{N} X_i \to \Delta(\mathbb{R})$ is measurable, $\mu_a = \mathbb{E}[r_a] < \infty$, and $\sigma_a^2 = \mathrm{Var}(r_a) < \infty$
- Genericity: no ties in expectations, $a \neq b \implies \mu_a \neq \mu_b$
- The functional form G (hence each F_a) is unknown to the DM; she holds no prior and attempts to learn μ_a solely through sampling

Model: Similarity classes via salient features

- DM not privy to the entire feature vector; instead she has a coarse perception of alternatives observing only a fixed, non-empty, proper subset of indices and inadvertently grouping alternatives into *similarity classes* based on her salient indices

$$N^s \subset \{1, \dots, N\}, \qquad 1 \leq |N^s| < N.$$

Model: Similarity classes via salient features

- DM not privy to the entire feature vector; instead she has a coarse perception of alternatives observing only a fixed, non-empty, proper subset of indices and inadvertently grouping alternatives into *similarity classes* based on her salient indices

$$N^s \subset \{1, \dots, N\}, \qquad 1 \leq |N^s| < N.$$

- Salient product space and projection:

$$X^s = \prod_{i \in N^s} X_i, \qquad x^s(a) = \big(x_i(a)\big)_{i \in N^s} \in X^s.$$

Model: Similarity classes via salient features

- DM not privy to the entire feature vector; instead she has a coarse perception of alternatives observing only a fixed, non-empty, proper subset of indices and inadvertently grouping alternatives into *similarity classes* based on her salient indices

$$N^s \subset \{1, \ldots, N\}, \qquad 1 \leq |N^s| < N.$$

- Salient product space and projection:

$$X^s = \prod_{i \in N^s} X_i, \qquad x^s(a) = \big(x_i(a)\big)_{i \in N^s} \in X^s.$$

- Alternatives $a, a' \in \mathcal{A}$ are *similar*, written $a \sim_s a'$, iff $x^s(a) = x^s(a')$.

Model: Similarity classes via salient features

- DM not privy to the entire feature vector; instead she has a coarse perception of alternatives observing only a fixed, non-empty, proper subset of indices and inadvertently grouping alternatives into *similarity classes* based on her salient indices

$$N^s \subset \{1, \dots, N\}, \qquad 1 \leq |N^s| < N.$$

- Salient product space and projection:

$$X^s = \prod_{i \in N^s} X_i, \qquad x^s(a) = \big(x_i(a)\big)_{i \in N^s} \in X^s.$$

- Alternatives $a, a' \in \mathcal{A}$ are *similar*, written $a \sim_s a'$, iff $x^s(a) = x^s(a')$.
- The induced partition $\mathcal{S} = \mathcal{A}/\sim_s = \{[a] : a \in \mathcal{A}\}$, where $[a] = \{a' \in \mathcal{A} : x^s(a') = x^s(a)\}$.

Model: Similarity classes via salient features

- DM not privy to the entire feature vector; instead she has a coarse perception of alternatives observing only a fixed, non-empty, proper subset of indices and inadvertently grouping alternatives into *similarity classes* based on her salient indices

$$N^s \subset \{1, \dots, N\}, \qquad 1 \leq |N^s| < N.$$

- Salient product space and projection:

$$X^s = \prod_{i \in N^s} X_i, \qquad x^s(a) = \big(x_i(a)\big)_{i \in N^s} \in X^s.$$

- Alternatives $a, a' \in \mathcal{A}$ are *similar*, written $a \sim_s a'$, iff $x^s(a) = x^s(a')$.
- The induced partition $\mathcal{S} = \mathcal{A}/\sim_s = \{[a] : a \in \mathcal{A}\}$, where $[a] = \{a' \in \mathcal{A} : x^s(a') = x^s(a)\}$.
- Similarity projection map:

$$\rho : \mathcal{A} \to \mathcal{S}, \quad \rho(a) = [a].$$

It is surjective and satisfies $x^s(a) = x^s(a') \iff \rho(a) = \rho(a')$.

Model: Similarity classes via salient features

- DM not privy to the entire feature vector; instead she has a coarse perception of alternatives observing only a fixed, non-empty, proper subset of indices and inadvertently grouping alternatives into *similarity classes* based on her salient indices

$$N^s \subset \{1, \ldots, N\}, \qquad 1 \le |N^s| < N.$$

- Salient product space and projection:

$$X^s = \prod_{i \in N^s} X_i, \qquad x^s(a) = \big(x_i(a)\big)_{i \in N^s} \in X^s.$$

- Alternatives $a, a' \in \mathcal{A}$ are *similar*, written $a \sim_s a'$, iff $x^s(a) = x^s(a')$.
- The induced partition $\mathcal{S} = \mathcal{A}/\sim_s = \{[a] : a \in \mathcal{A}\}$, where $[a] = \{a' \in \mathcal{A} : x^s(a') = x^s(a)\}$.
- Similarity projection map:

$$\rho : \mathcal{A} \to \mathcal{S}, \quad \rho(a) = [a].$$

  It is surjective and satisfies $x^s(a) = x^s(a') \iff \rho(a) = \rho(a')$.
- Non-triviality (informational coarseness): each class has size at least 2, $2 \le |s| \le |\mathcal{A}|$ for all $s \in \mathcal{S}$, so ρ is not injective

Model: Timeline and Stage Decision Problem

- Periods: $k = 0, 1, 2, \ldots$; depth-one stage decision tree $\mathcal{T}(\Psi, f, r)$ repeated indefinitely

Model: Timeline and Stage Decision Problem

- Periods: $k = 0, 1, 2, \ldots$; depth-one stage decision tree $\mathcal{T}(\Psi, f, r)$ repeated indefinitely
- Each period k: Nature draws a state (menu) $\psi_k \in \Psi$ i.i.d. with pmf f

Model: Timeline and Stage Decision Problem

- Periods: $k = 0, 1, 2, \ldots$; depth-one stage decision tree $\mathcal{T}(\Psi, f, r)$ repeated indefinitely
- Each period k: Nature draws a state (menu) $\psi_k \in \Psi$ i.i.d. with pmf f
- Stage problem repeats i.i.d. each date - special case of a stationary MDP with action-independent transition

Model: Timeline and Stage Decision Problem

- Periods: $k = 0, 1, 2, \ldots$; depth-one stage decision tree $\mathcal{T}(\Psi, f, r)$ repeated indefinitely
- Each period k: Nature draws a state (menu) $\psi_k \in \Psi$ i.i.d. with pmf f
- Stage problem repeats i.i.d. each date - special case of a stationary MDP with action-independent transition
- DM observes only the salient coordinates of each $a \in \psi_k$, chooses $a_k \in \psi_k$, then observes a corresponding reward $r_k = r_{a_k}$, drawn i.i.d. from $F_{a_k} = G\Big(x(a_k)\Big)$

Model: Timeline and Stage Decision Problem

- Periods: $k = 0, 1, 2, \ldots$; depth-one stage decision tree $\mathcal{T}(\Psi, f, r)$ repeated indefinitely
- Each period k: Nature draws a state (menu) $\psi_k \in \Psi$ i.i.d. with pmf f
- Stage problem repeats i.i.d. each date - special case of a stationary MDP with action-independent transition
- DM observes only the salient coordinates of each $a \in \psi_k$, chooses $a_k \in \psi_k$, then observes a corresponding reward $r_k = r_{a_k}$, drawn i.i.d. from $F_{a_k} = G\Big(x(a_k)\Big)$
- Model Misspecification: DM ignores the payoff-relevance of non-salient features and (erroneously) treats alternatives sharing salient projection as payoff-identical (considered to be i.i.d. within class).

Model: Stochastic Choice Policy

- Valuations $v : \mathcal{S} \to \mathbb{R}$; in period k, $\mathbf{v}_k = (v_k^s)_{s \in \mathcal{S}}$.

Model: Stochastic Choice Policy

- Valuations $v : \mathcal{S} \to \mathbb{R}$; in period k, $\mathbf{v}_k = (v_k^s)_{s \in \mathcal{S}}$.
- Interpretation: v_k^s is DM's current estimate of the expected payoff for class s.

Model: Stochastic Choice Policy

- Valuations $v : \mathcal{S} \rightarrow \mathbb{R}$; in period k, $\mathbf{v}_k = (v_k^s)_{s \in \mathcal{S}}$.
- Interpretation: v_k^s is DM's current estimate of the expected payoff for class s.
- Given state ψ: randomize over classes in $\omega = \mathcal{S}_\psi = \rho(\psi)$ via multinomial logit:

$$\sigma_{\psi,k}^s(\mathbf{v}_k) = \frac{\exp(\beta v_k^s)}{\sum_{j \in \omega} \exp(\beta v_k^j)}, \qquad s \in \omega.$$

Model: Stochastic Choice Policy

- Valuations $v : \mathcal{S} \to \mathbb{R}$; in period k, $\mathbf{v}_k = (v_k^s)_{s \in \mathcal{S}}$.
- Interpretation: v_k^s is DM's current estimate of the expected payoff for class s.
- Given state ψ: randomize over classes in $\omega = \mathcal{S}_\psi = \rho(\psi)$ via multinomial logit:

$$\sigma^s_{\psi,k}(\mathbf{v}_k) = \frac{\exp(\beta v_k^s)}{\sum_{j \in \omega} \exp(\beta v_k^j)}, \qquad s \in \omega.$$

- Within-class tie-breaking: having chosen s, select $a \in s \cap \psi$ uniformly,

$$\Pr(a \mid s, \psi) = \frac{1}{|s \cap \psi|}, \quad a \in s \cap \psi, \quad \Rightarrow \quad \delta^a_{\psi,k} = \frac{1}{|s \cap \psi|} \sigma^s_{\psi,k}.$$

Model: Stochastic Choice Policy

- Valuations $v : \mathcal{S} \to \mathbb{R}$; in period k, $\mathbf{v}_k = (v_k^s)_{s \in \mathcal{S}}$.
- Interpretation: v_k^s is DM's current estimate of the expected payoff for class s.
- Given state ψ: randomize over classes in $\omega = \mathcal{S}_\psi = \rho(\psi)$ via multinomial logit:

$$\sigma_{\psi,k}^s(\mathbf{v}_k) = \frac{\exp(\beta v_k^s)}{\sum_{j \in \omega} \exp(\beta v_k^j)}, \qquad s \in \omega.$$

- Within-class tie-breaking: having chosen s, select $a \in s \cap \psi$ uniformly,

$$\Pr(a \mid s, \psi) = \frac{1}{|s \cap \psi|}, \quad a \in s \cap \psi, \quad \Rightarrow \quad \delta_{\psi,k}^a = \frac{1}{|s \cap \psi|} \sigma_{\psi,k}^s.$$

- Sensitivity (inverse noise) parameter $\beta \geq 0$:

Model: Stochastic Choice Policy

- Valuations $v : \mathcal{S} \to \mathbb{R}$; in period k, $\mathbf{v}_k = (v_k^s)_{s \in \mathcal{S}}$.
- Interpretation: v_k^s is DM's current estimate of the expected payoff for class s.
- Given state ψ: randomize over classes in $\omega = \mathcal{S}_\psi = \rho(\psi)$ via multinomial logit:

$$\sigma_{\psi,k}^s(\mathbf{v}_k) = \frac{\exp(\beta v_k^s)}{\sum_{j \in \omega} \exp(\beta v_k^j)}, \qquad s \in \omega.$$

- Within-class tie-breaking: having chosen s, select $a \in s \cap \psi$ uniformly,

$$\Pr(a \mid s, \psi) = \frac{1}{|s \cap \psi|}, \quad a \in s \cap \psi, \quad \Rightarrow \quad \delta_{\psi,k}^a = \frac{1}{|s \cap \psi|} \sigma_{\psi,k}^s.$$

- Sensitivity (inverse noise) parameter $\beta \geq 0$:
  - $\beta = 0$: uniform random choice

Model: Stochastic Choice Policy

- Valuations $v : \mathcal{S} \to \mathbb{R}$; in period k, $\mathbf{v}_k = (v_k^s)_{s\in\mathcal{S}}$.
- Interpretation: v_k^s is DM's current estimate of the expected payoff for class s.
- Given state ψ: randomize over classes in $\omega = \mathcal{S}_\psi = \rho(\psi)$ via multinomial logit:

$$\sigma^s_{\psi,k}(\mathbf{v}_k) = \frac{\exp(\beta v_k^s)}{\sum_{j\in\omega}\exp(\beta v_k^j)}, \qquad s \in \omega.$$

- Within-class tie-breaking: having chosen s, select $a \in s \cap \psi$ uniformly,

$$\Pr(a \mid s, \psi) = \frac{1}{|s\cap\psi|}, \quad a \in s\cap\psi, \quad \Rightarrow \quad \delta^a_{\psi,k} = \frac{1}{|s\cap\psi|}\,\sigma^s_{\psi,k}.$$

- Sensitivity (inverse noise) parameter $\beta \geq 0$:
  - $\beta = 0$: uniform random choice
  - $\beta \to \infty$: probability mass concentrates on class(es) with maximal v_k^s (almost-sure best response with uniform tie-breaking) - our focus in this paper

Model: Update Rule

- After observing $r_k = r_{a_k}$, DM updates only the chosen similarity class $s = \rho(a_k)$:

$$v_{k+1}(s') = v_k(s') + \alpha_k(s')\,\mathbf{1}_{\{s'=\rho(a_k)\}}\Big[r_k - v_k(s')\Big], \quad \forall s' \in \mathcal{S}.$$

Model: Update Rule

- After observing $r_k = r_{a_k}$, DM updates only the chosen similarity class $s = \rho(a_k)$:

$$v_{k+1}(s') = v_k(s') + \alpha_k(s')\,\mathbf{1}_{\{s'=\rho(a_k)\}}\Big[r_k - v_k(s')\Big], \quad \forall s' \in \mathcal{S}.$$

- Equivalently, in vector notation:

$$\mathbf{v}_{k+1} - \mathbf{v}_k = \alpha_{\mathbf{k}}\Big[\tilde{\mathbf{v}}_k - \mathbf{v}_k\Big], \quad \tilde{v}_k^s = \begin{cases} r_k, & s = \rho(a_k), \\ v_k^s, & \text{otherwise.} \end{cases}$$

Model: Update Rule

- After observing $r_k = r_{a_k}$, DM updates only the chosen similarity class $s = \rho(a_k)$:

$$v_{k+1}(s') = v_k(s') + \alpha_k(s')\,\mathbf{1}_{\{s'=\rho(a_k)\}}\Big[r_k - v_k(s')\Big], \quad \forall s' \in \mathcal{S}.$$

- Equivalently, in vector notation:

$$\mathbf{v}_{k+1} - \mathbf{v}_k = \alpha_{\mathbf{k}}\Big[\tilde{\mathbf{v}}_k - \mathbf{v}_k\Big], \quad \tilde{v}_k^s = \begin{cases} r_k, & s = \rho(a_k), \\ v_k^s, & \text{otherwise.} \end{cases}$$

- Learning rates (per class): $\alpha_k(s) \in (0,1), \quad \sum_k \alpha_k(s) = \infty, \quad \sum_k \alpha_k(s)^2 < \infty$

Model: Update Rule

- After observing $r_k = r_{a_k}$, DM updates only the chosen similarity class $s = \rho(a_k)$:

$$v_{k+1}(s') = v_k(s') + \alpha_k(s')\,\mathbf{1}_{\{s'=\rho(a_k)\}}\Big[r_k - v_k(s')\Big], \quad \forall s' \in \mathcal{S}.$$

- Equivalently, in vector notation:

$$\mathbf{v}_{k+1} - \mathbf{v}_k = \alpha_{\mathbf{k}}\Big[\tilde{\mathbf{v}}_k - \mathbf{v}_k\Big], \quad \tilde{v}_k^s = \begin{cases} r_k, & s = \rho(a_k), \\ v_k^s, & \text{otherwise.} \end{cases}$$

- Learning rates (per class): $\alpha_k(s) \in (0,1), \quad \sum_k \alpha_k(s) = \infty, \quad \sum_k \alpha_k(s)^2 < \infty$
- Coupled evolution: $\mathbf{v}_k \mapsto \sigma_{\psi,k}(\mathbf{v}_k) \mapsto a_k \mapsto r_k \mapsto \mathbf{v}_{k+1}$

Model: Update Rule

- After observing $r_k = r_{a_k}$, DM updates only the chosen similarity class $s = \rho(a_k)$:

$$v_{k+1}(s') = v_k(s') + \alpha_k(s')\,\mathbf{1}_{\{s'=\rho(a_k)\}}\Big[r_k - v_k(s')\Big], \quad \forall s' \in \mathcal{S}.$$

- Equivalently, in vector notation:

$$\mathbf{v}_{k+1} - \mathbf{v}_k = \alpha_{\mathbf{k}}\Big[\tilde{\mathbf{v}}_k - \mathbf{v}_k\Big], \quad \tilde{v}_k^s = \begin{cases} r_k, & s = \rho(a_k), \\ v_k^s, & \text{otherwise.} \end{cases}$$

- Learning rates (per class): $\alpha_k(s) \in (0,1), \quad \sum_k \alpha_k(s) = \infty, \quad \sum_k \alpha_k(s)^2 < \infty$
- Coupled evolution: $\mathbf{v}_k \mapsto \sigma_{\psi,k}(\mathbf{v}_k) \mapsto a_k \mapsto r_k \mapsto \mathbf{v}_{k+1}$
- Non-homogeneous Markov process interpreted as DM exploring classes via stochastic choice and exploiting by weighting up classes with higher v_k^s; with iterative averaging drives v_k^s toward experienced payoffs within classes

Model: Continuous-Time Asymptotic Approximation

- **Stochastic Approximation:** for infinitesimal learning rates $\alpha_{\mathbf{k}}$, the discrete-time stochastic CQL process approximates a deterministic continuous-time dynamical system given by the expected motion of the stochastic process

Model: Continuous-Time Asymptotic Approximation

- **Stochastic Approximation:** for infinitesimal learning rates $\alpha_{\mathbf{k}}$, the discrete-time stochastic CQL process approximates a deterministic continuous-time dynamical system given by the expected motion of the stochastic process
- for the analysis of the continuous-time CQL model, we transform w.l.o.g. the original stage decision tree $\mathcal{T}(\Psi, f, r)$ to a reduced tree $\mathcal{T}'(\Omega, p, \pi)$, where $\Omega = \mathcal{P}(\mathcal{S}) \setminus \{\varnothing\}$

Model: Continuous-Time Asymptotic Approximation

- **Stochastic Approximation:** for infinitesimal learning rates $\alpha_{\mathbf{k}}$, the discrete-time stochastic CQL process approximates a deterministic continuous-time dynamical system given by the expected motion of the stochastic process
- for the analysis of the continuous-time CQL model, we transform w.l.o.g. the original stage decision tree $\mathcal{T}(\Psi, f, r)$ to a reduced tree $\mathcal{T}'(\Omega, p, \pi)$, where $\Omega = \mathcal{P}(\mathcal{S}) \setminus \{\varnothing\}$
- The probability of state ω is given by $p(\omega) = \sum_{\psi \in \Psi : \rho(\psi) = \omega} f(\psi)$.

Model: Continuous-Time Asymptotic Approximation

- **Stochastic Approximation:** for infinitesimal learning rates $\alpha_{\mathbf{k}}$, the discrete-time stochastic CQL process approximates a deterministic continuous-time dynamical system given by the expected motion of the stochastic process
- for the analysis of the continuous-time CQL model, we transform w.l.o.g. the original stage decision tree $\mathcal{T}(\Psi, f, r)$ to a reduced tree $\mathcal{T}'(\Omega, p, \pi)$, where $\Omega = \mathcal{P}(\mathcal{S}) \setminus \{\emptyset\}$
- The probability of state ω is given by $p(\omega) = \sum_{\psi \in \Psi : \rho(\psi) = \omega} f(\psi)$.
- For a given similarity class $s \in \omega$, the *state-dependent* expected payoff is

$$\pi_\omega(s) = \frac{\sum_{\psi \in \Psi : \rho(\psi) = \omega} f(\psi)\, \bar{\mu}_\psi(s)}{\sum_{\psi \in \Psi : \rho(\psi) = \omega} f(\psi)},$$

where,

$$\bar{\mu}_\psi(s) = \frac{1}{|s \cap \psi|} \sum_{a \in s \cap \psi} \mu_a$$

Model: Continuous-Time CQL ODE

- stochastic approximation allows us to connect the long-run behavior ($k \to \infty$) of discrete-time CQL to the asymptotics ($t \to \infty$) of the following CQL ODE, where $\forall s \in \mathcal{S}$,

$$\frac{dv_s}{dt} = \frac{\sum\limits_{\omega \in \Omega : s \in \omega} p(\omega)\, \sigma^s_\omega(\mathbf{v})\, \pi_\omega(s)}{\sum\limits_{\omega \in \Omega : s \in \omega} p(\omega)\, \sigma^s_\omega(\mathbf{v})} - v_s \tag{1}$$

where,

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

Theorem 1

The set of steady-state solutions of the ODE system in Eq.(1) is non-empty.

Model: Connections with Valuation Equilibrium

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

$$v_s = \frac{\sum\limits_{\omega\in\Omega: s\in\omega} p(\omega)\,\sigma_\omega^s(\mathbf{v})\,\pi_\omega(s)}{\sum\limits_{\omega\in\Omega: s\in\omega} p(\omega)\,\sigma_\omega^s(\mathbf{v})} \quad \text{and} \quad \sigma_\omega^s = \frac{\exp(\beta v_s)}{\sum_{j\in\omega}\exp(\beta v_j)}.$$

Theorem 2

The **smooth valuation equilibria** (SVE) of the CQL dynamics **converge** to **valuation equilibria** (VE) as the sensitivity parameter $\beta \to \infty$.

$\forall\ \epsilon > 0$, $\exists\ \hat{\beta} \in \mathbb{R}_+$ s.t. for almost all $\beta > \hat{\beta}$, except possibly on a measure zero set, every SVE lies within an ϵ-neighborhood of some VE, is locally isolated, and varies smoothly with β.

- **Smooth Valuation Equilibrium**: A mixed strategy profile $\sigma = (\sigma^s_\omega)^{s\in\mathcal{S}}_{\omega\in\Omega}$ constitutes an SVE for $\mathcal{T}'$ if there exists a valuation system $(v_s)_{s\in\mathcal{S}}$ s.t.

$$v_s = \frac{\sum\limits_{\omega\in\Omega:s\in\omega} p(\omega)\,\sigma^s_\omega(\mathbf{v})\,\pi_\omega(s)}{\sum\limits_{\omega\in\Omega:s\in\omega} p(\omega)\,\sigma^s_\omega(\mathbf{v})} \quad \text{and} \quad \sigma^s_\omega = \frac{\exp(\beta v_s)}{\sum_{j\in\omega}\exp(\beta v_j)}.$$

- The steady-states of the continuous-time CQL dynamical system characterize the set of smooth valuation equilibria in a decision tree and vice versa.

Theorem 2

The **smooth valuation equilibria** (SVE) of the CQL dynamics **converge** to **valuation equilibria** (VE) as the sensitivity parameter $\beta \to \infty$.

$\forall\ \epsilon > 0,\ \exists\ \hat{\beta} \in \mathbb{R}_+$ s.t. for almost all $\beta > \hat{\beta}$, except possibly on a measure zero set, every SVE lies within an ϵ-neighborhood of some VE, is locally isolated, and varies smoothly with β.

Model: Connections with Valuation Equilibrium

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

$$v_s = \frac{\sum\limits_{\omega\in\Omega: s\in\omega} p(\omega)\,\sigma_\omega^s(\mathbf{v})\,\pi_\omega(s)}{\sum\limits_{\omega\in\Omega: s\in\omega} p(\omega)\,\sigma_\omega^s(\mathbf{v})} \quad \text{and} \quad \sigma_\omega^s = \frac{\exp(\beta v_s)}{\sum_{j\in\omega}\exp(\beta v_j)}.$$

- The steady-states of the continuous-time CQL dynamical system characterize the set of smooth valuation equilibria in a decision tree and vice versa.
- **Valuation Equilibrium**: in each state ω, DM chooses similarity class(es) $s \in \arg\max_{s\in\omega} v_s$ and her valuations $(v_s)_{s\in\mathcal{S}}$ are consistent with the expected payoffs of the similarity classes induced by the optimal choice strategy

Theorem 2

The **smooth valuation equilibria** (SVE) of the CQL dynamics **converge** to **valuation equilibria** (VE) as the sensitivity parameter $\beta \to \infty$.

$\forall\ \epsilon > 0,\ \exists\ \hat{\beta} \in \mathbb{R}_+$ s.t. for almost all $\beta > \hat{\beta}$, except possibly on a measure zero set, every SVE lies within an ϵ-neighborhood of some VE, is locally isolated, and varies smoothly with β.

Illustrations

Illustration 1: Multiplicity of SVE

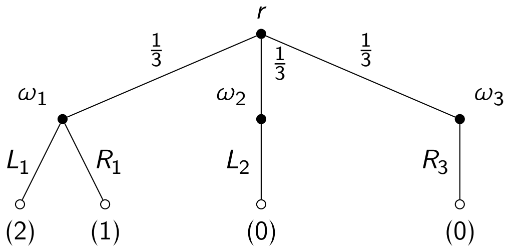


Figure: Example of a Decision Tree with Two Similarity Classes (expected payoffs within parentheses)

Two similarity classes, $L = \{L_1, L_2\}$ and $R = \{R_1, R_3\}$

- strict pure VE at $(1, 0)$ (choosing L_1 at ω_1)
- strict pure VE at $(0, 0.5)$ (choosing R_1 at ω_1)
- mixed VE near $(0.423, 0.423)$ (choosing R_1 with probability $\sqrt{3} - 1$ at ω_1)

Illustration 1: Multiplicity of SVE

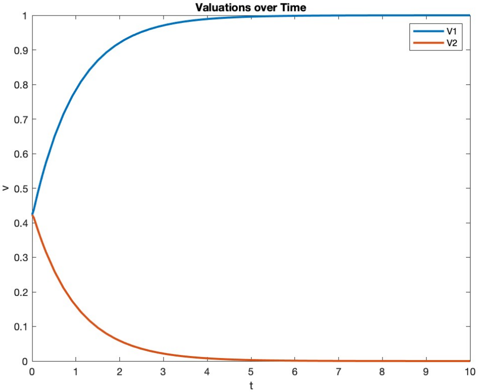


Figure: stable strict pure SVE at $(1.0, 0.0)$; $\beta = 50$

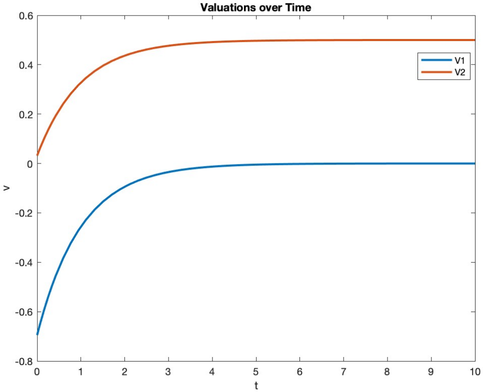


Figure: stable strict pure SVE at $(0.0, 0.5)$; $\beta = 50$

- additionally, an unstable mixed SVE near $(0.423, 0.423)$

Illustration 2: Unique Mixed SVE

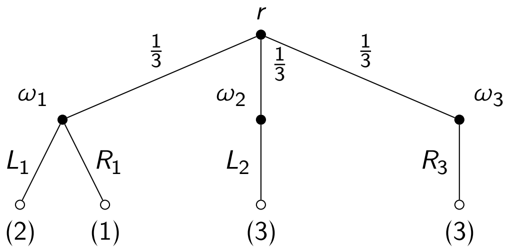


Figure: Example of a Decision Tree with Two Similarity Classes (expected payoffs within parentheses)

- unique mixed VE at $(2 + \frac{1}{\sqrt{3}}, 2 + \frac{1}{\sqrt{3}}) \approx (2.577, 2.577)$
  - selects the alternative in L with probability $\sqrt{3} - 1$ and the alternative in R with probability $2 - \sqrt{3}$ at node ω_1

Illustration 2: Unique Mixed SVE

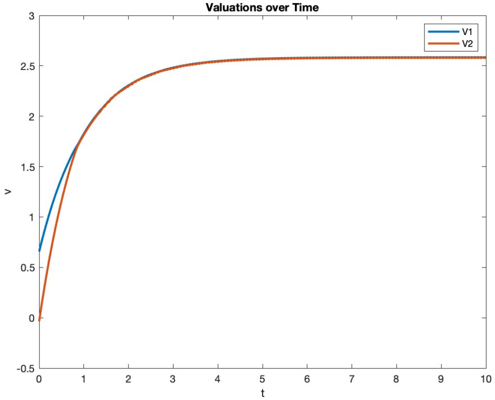


Figure: Stable Unique Mixed SVE at $(2.577, 2.577)$; $\beta = 50$

Illustration 3: Rock-Paper-Scissors

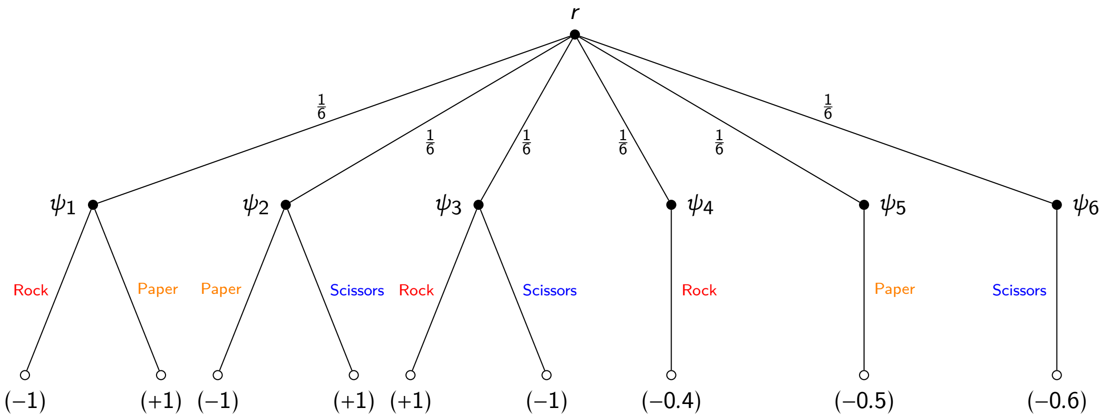


Figure: Simplified Decision Tree for “Rock-Paper-Scissors” (expected payoffs within parentheses)

Illustration 3: Cyclical Choice Behavior

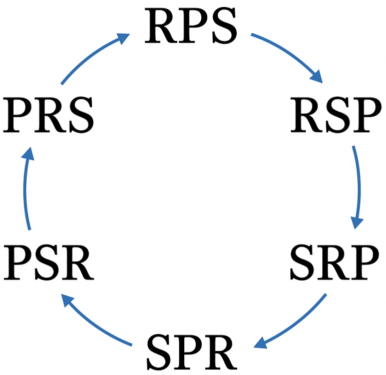


Figure: ABC denotes $v_A > v_B > v_C$

- Expected payoff for the "winner" A $= -\frac{1}{6}$
- Expected payoff for the "loser" C $= -\frac{1}{2}$
- Expected payoff for the "middle-ranking" B $= -\frac{3}{4}$ or $\frac{1}{4}$

Illustration 3: Valuations in a Cycle

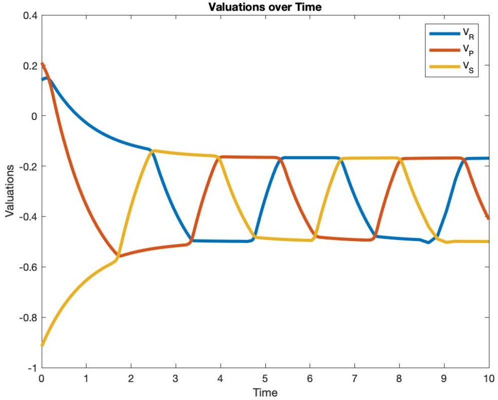


Figure: Indefinite oscillations of the valuations; $\beta = 80$

- limit cycle around the unique fully-mixed smooth valuation equilibrium

Illustration 3: Phase Portrait

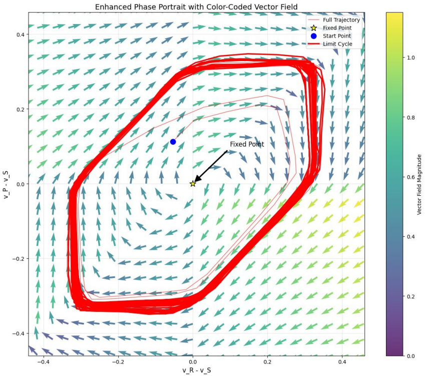


Figure: Phase portrait in $\mathbb{R}^2$

- As β increases, the unique fully-mixed SVE at the origin with index $+1$ loses stability at a critical point and a stable limit cycle emerges from it through a Hopf bifurcation
  - the unique stable limit cycle is an attractor for almost all initial conditions by the Poincaré-Bendixson theorem

Guaranteed Convergence with 2 Similarity Classes

Theorem 3

$\exists\ \hat{\beta} \in \mathbb{R}_+$, s.t. $\forall\ \beta > \hat{\beta}$, a decision tree $\mathcal{T}'$ with generic payoffs where a DM chooses among alternatives in at most 2 similarity classes always **admits** a **locally asymptotically stable** smooth valuation equilibrium under the CQL dynamics.

Additionally, if the equilibrium is **unique**, whether pure or mixed, it is **globally asymptotically stable** under the CQL dynamics.

Summary of Results: Decision Trees with >2 Similarity Classes

For decision trees with generic payoffs and an arbitrary number $n > 2$ of similarity classes :

- Strict Pure equilibria: **locally asymptotically stable** in the CQL dynamics whenever they exist, given a sufficiently large sensitivity parameter
- **Unique Mixed** equilibrium when trivial choices yield sufficiently high payoffs
  - characterized by indifferences between similarity classes in high sensitivity limit
- **Multiple** equilibria when trivial choices yield sufficiently low payoffs
  - at least one strict pure equilibrium that is **locally asymptotically stable** under the CQL dynamics for sufficiently large sensitivity parameter
- The set of **asymptotically stable** equilibria may be **empty** when trivial choices yield intermediate payoffs

Summary of Results: Decision Trees with >2 Similarity Classes

For decision trees with generic payoffs and an arbitrary number $n > 2$ of similarity classes :

- Strict Pure equilibria: **locally asymptotically stable** in the CQL dynamics whenever they exist, given a sufficiently large sensitivity parameter
- **Unique Mixed** equilibrium when trivial choices yield sufficiently high payoffs
  - characterized by indifferences between similarity classes in high sensitivity limit
  - **globally asymptotically stable** under the CQL dynamics
- **Multiple** equilibria when trivial choices yield sufficiently low payoffs
  - at least one strict pure equilibrium that is **locally asymptotically stable** under the CQL dynamics for sufficiently large sensitivity parameter
- The set of **asymptotically stable** equilibria may be **empty** when trivial choices yield intermediate payoffs

Literature Review

- **Bounded Rationality:** Jehiel & Samet (2007); Goeree et al. (2016)
- **RL, AI & Q-Learning:** Watkins & Dayan (1992); Sutton & Barto (2018)
- **RL in Game Theory:** Roth & Erev (1995); Erev & Roth (1998); Camerer & Hua Ho (1999); Hopkins & Posch (2005)
- **Payoff-Assessment Learning:** Börgers & Sarin (1997); Rustichini (1999); Jehiel & Samet (2005); Cominetti et al. (2010); Leslie & Collins (2005); Sarin & Vahid (1999);
- **Fictitious Play:** Brown (1951); Robinson (1951); Fudenberg & Kreps (1993); Hopfbauer & Sandholm (2003);

Conclusion

Highlights

- Coarse Q-learning - a novel approach to modeling learning under a coarse perception of the alternative space

Future Extensions

Conclusion

Highlights

- Coarse Q-learning - a novel approach to modeling learning under a coarse perception of the alternative space
- Key finding - persistent mixing / path-dependence / cyclical preferences can exist in decision problems with generic payoffs, even as the DM becomes acutely sensitive to payoff differences

Future Extensions

Conclusion

Highlights

- Coarse Q-learning - a novel approach to modeling learning under a coarse perception of the alternative space
- Key finding - persistent mixing / path-dependence / cyclical preferences can exist in decision problems with generic payoffs, even as the DM becomes acutely sensitive to payoff differences
    - novel explanation for the presence of non-standard preferences in decision-making observed in empirical studies (DellaVigna 2009)

Future Extensions

Conclusion

Highlights

- Coarse Q-learning - a novel approach to modeling learning under a coarse perception of the alternative space
- Key finding - persistent mixing / path-dependence / cyclical preferences can exist in decision problems with generic payoffs, even as the DM becomes acutely sensitive to payoff differences
  - novel explanation for the presence of non-standard preferences in decision-making observed in empirical studies (DellaVigna 2009)

Future Extensions

- Endogenizing the formation/modification/evolution of similarity classes

Conclusion

Highlights

- Coarse Q-learning - a novel approach to modeling learning under a coarse perception of the alternative space
- Key finding - persistent mixing / path-dependence / cyclical preferences can exist in decision problems with generic payoffs, even as the DM becomes acutely sensitive to payoff differences
  - novel explanation for the presence of non-standard preferences in decision-making observed in empirical studies (DellaVigna 2009)

Future Extensions

- Endogenizing the formation/modification/evolution of similarity classes
- Extending the Coarse Q-learning framework to sequential MDPs

Thank You!

Questions?

SCAN ME

Aviman Satpathy
Paris School of Economics
aviman.satpathy@psemail.eu

Appendix

Strict Pure SVE are Locally Asymptotically Stable

Theorem 4

$\exists\ \hat{\beta} \in \mathbb{R}_+$, s.t. $\forall\ \beta > \hat{\beta}$, 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.

When Trivial Choice Payoffs are Large

- Trivial unary choice payoffs $\pi_{\omega=\{i\}}(i)$, $\forall i \in \mathcal{S}$ are uniformly increased by a constant $z > 0$, while payoffs in non-trivial choice states remain unchanged

Theorem 5

$\exists\ \hat{z} \in \mathbb{R}_+$ s.t. $\forall\ z > \hat{z}$, and $\forall\ \beta \in \mathbb{R}_+$, a decision tree $\mathcal{T}_n^{'}$ with generic payoffs and an arbitrary number of similarity classes always admits a **unique** SVE.

$\exists\ \hat{\beta} \in \mathbb{R}_+$ s.t. $\forall\ \beta > \hat{\beta}$, the unique SVE lies in the neighborhood of a **mixed valuation equilibrium** in which the agent is **indifferent** between at least two similarity classes.

The unique SVE is **globally asymptotically stable** under the CQL dynamics.

- sufficiently high trivial unary choice payoffs correspond to the “choice overload” phenomenon well-documented in psychology
- somewhat striking result given that we analyze a decision tree with generic payoffs and evaluate the steady-states in the limit of noise going to zero
- scope of indifference extends to all similarity classes if richer choice states are weakly more likely to be drawn by nature

When Trivial Choice Payoffs are Small

- Trivial unary choice payoffs $\pi_{\omega=\{i\}}(i)$, $\forall i \in \mathcal{S}$ are uniformly adjusted by a constant $z < 0$, while payoffs in non-trivial choice states remain unchanged

Theorem 6

$\exists\ \tilde{z} \in \mathbb{R}_-$ s.t. $\forall\ z < \tilde{z}$, 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 $s \in S$, there exists a valuation equilibrium (VE) where s is the unique strictly dominated similarity class. Moreover, there exists **at least one strict pure VE**.

Correspondingly, $\exists\ \hat{\beta} \in \mathbb{R}_+$, s.t. $\forall\ \beta > \hat{\beta}$, the smooth valuation equilibrium that arises in the neighborhood of the strict pure VE is **locally asymptotically stable** under the CQL dynamics.

- If richer choice states are weakly more likely to be drawn, any strict ordering of the valuations can arise in equilibrium $\implies$ $n!$ strict pure SVE arise for sufficiently large β, each of which is **locally asymptotically stable**. Additionally, there exist at least $n! - 1$ interior SVE that are unstable.

Example: The Good, the Bad and the Unsteady

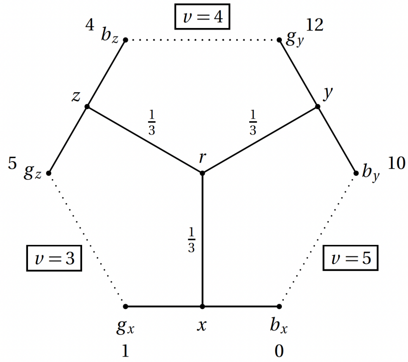
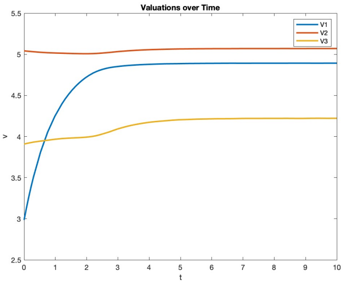


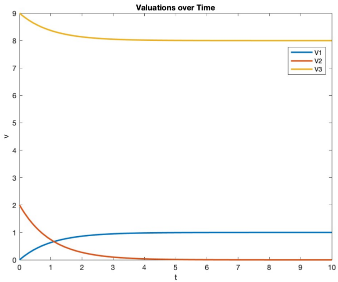
Figure: Good/Bad Decisions

- counter-intuitively, making the worst decision at every node is a VE; unique consistent valuation $(v_i, v_j, v_k) = (3, 5, 4)$

Example: The Good, the Bad and the Unsteady



(a) initial valuation around (3,5,4)



(b) initial valuation around (0,2,9)

Figure: Long-run CQL dynamics with $\beta = 80$

- worst-decision VE does **not** correspond to a steady-state of the CQL model in the high sensitivity limit, i.e., limiting SVE $\subseteq$ VE
- two distinct locally stable smooth valuation equilibria:
  1. $(v_i \approx 5, v_j = 5, v_k = 4)$ corresponding to a (partially) mixed VE
  2. $(v_i = 1, v_j = 0, v_k = 8)$ corresponding to a strict pure VE

Example: French Presidential Elections?

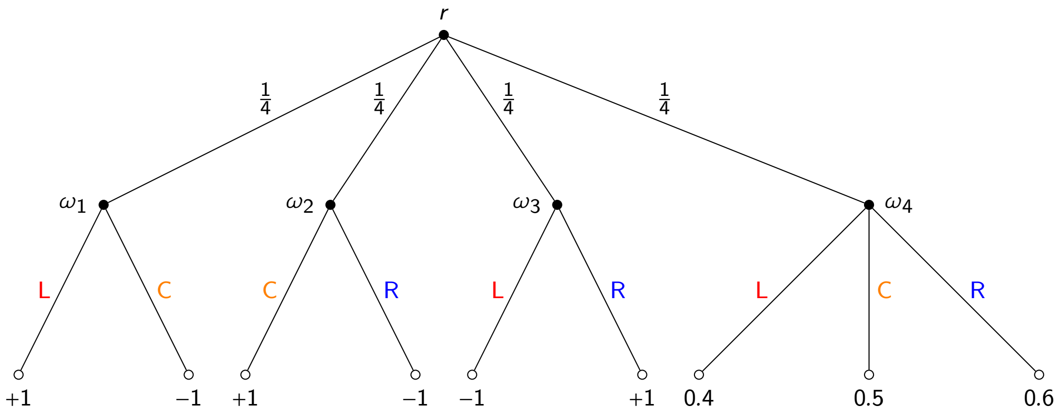


Figure: French Presidential Primaries

Example: Electorate in a Cycle

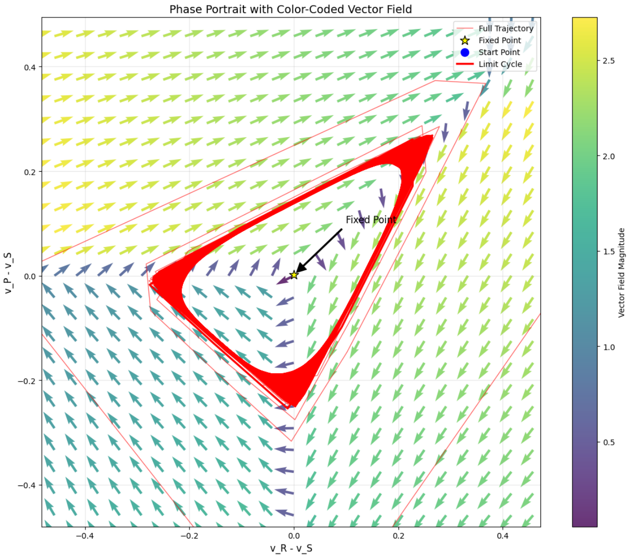
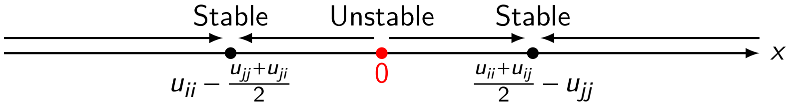
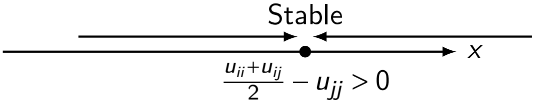


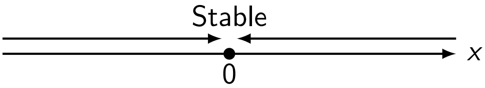
Figure: Phase Portrait of the French Electorate



2 Strict Pure VE and 1 Mixed VE



Unique Strict Pure VE



Unique Mixed VE