Representativeness: A Possibility Theorem

Mateus Hiro Nagata

HEC Paris

June 15, 2026

$\succsim$ on policies, vote on representatives

QR Code

Figure: If you want

The Representative Dream Problem

Dream Problem

Candidates + data on past problems & actions → find one close to society's preferences, in- & out-of-sample?

This paper

Reframe representative selection as a **statistical learning problem** and prove it is efficiently solvable.

- Future problems: unknowable at election time
- Savage (1954): probabilities over all states (grand-world) → cognitively implausible
- Case-Based Decision Theory

This Paper

Main question: 1) If I see enough actions of a candidate (m), can I predict its next action? 2) If there are enough candidates (n), can I find one that acts very close to me? **Yes:**

1. **Finite-sample bound**: excess risk $\leq 2\sqrt{2 \ln n/m} + 5\sqrt{2 \ln(8/\delta)/m}$
2. **Sample complexity**: $m = O\left(\frac{\ln n}{\varepsilon^2}\right)$ - only logarithmic in the number of candidates
3. **Asymptotic**: under parametric similarity + dense coverage, $L_D(i^*) \xrightarrow{p} 0$

Contribution: A *possibility theorem* for representative democracy.

Related Literature

- **Social choice**: Arrow (1950), Gibbard (1973), Satterthwaite (1975): impossibility of preference *aggregation*; we study preference *estimation*
- **Preference learning**: Fürnkranz & Hüllermeier (2010), Ailon & Mohri (2010): ranking from pairwise comparisons; no representative selection with distributional guarantees
- **CBDT**: Gilboa & Schmeidler (1995), Billot et al. (2008): individual reasoning; we apply it to collective representation
- **Learning theory**: Valiant (1984), Vapnik (2013), Bartlett & Mendelson (2002): standard tools applied to a novel political problem

Memory & Similarity

Shared memory: $\mathcal{M} = \{(q, a, r)\}$: problem, action, outcome.

History: $H = \{q \in P \mid \exists\, a, r :\ (q, a, r) \in \mathcal{M}\}$

New problem $p \notin H \rightarrow$ decide via similarity to past cases.

Case-Based Decision Theory

Theorem (Representation)

A1-A5, then $\exists\, s : P \times P \to [0,1]$ and $u : \mathbb{R} \to \mathbb{R}$, both unique up to scaling, s.t. for any $p \notin H$, $x, y \in \mathbb{R}^H$

$$x \succsim_{p,H} y \iff \sum_{q \in H} s(p,q)\, u(x(q)) \ \geq \ \sum_{q \in H} s(p,q)\, u(y(q))$$

Decision rule: $i \in I = \{1, \ldots, n\}$

$$V_i(a \mid p, \mathcal{M}) = \sum_{\substack{(q,a',r) \in \mathcal{M} \\ a' = a}} s_i(p,q) \cdot u(r)$$

$$a^i(p, \mathcal{M}) = \arg\max_{a \in A} V_i(a \mid p, \mathcal{M})$$

Assumptions

Assumption 1 (Single Shared Memory)

All agents observe the same $\mathcal{M}$. No private information about past outcomes.

Assumption 2 (Common Utility)

All agents agree on outcome evaluation $u(\cdot)$; they differ only in similarity assessments s_i.

Political disagreement = *different beliefs about relevance* (similarity), not different values.

$\Rightarrow$ Selection is: **identify the candidate i whose s_i is closest to s_{SOC}.**

Running Example: Fiscal Policy

$A = \{\text{stimulus}, \text{austerity}\}$. New problem p_{2026}: 115% debt/GDP + financial stress.

Case	Action	r	$s(p_{2026}, q)$
2008 Crisis	Stimulus	$+5$	0.7
2010 Greece	Austerity	-3	0.4
1930 Depression	Austerity	-8	0.5
1970s Stagflation	Stimulus	-2	0.2

$$V(\text{stim}) = 0.7(5) + 0.2(-2) = 3.1$$

$$V(\text{auster}) = 0.4(-3) + 0.5(-8) = -5.2 \qquad (V(\text{stim}) > V(\text{auster}))$$

Society $\to$ **stimulus**. A ($s_A(p_{2026}, q_{2008}) = 0.9$) agrees; B (debt-focused) doesn't.

Which generalises?

The Selection Rule

Voters observe m problems $S = \{p_1, \ldots, p_m\} \overset{\text{i.i.d.}}{\sim} D$ and candidates' choices $\{a_j^{\mathrm{i}}\}$.

Empirical disagreement (observable):

$$L_S(i) = \frac{1}{m} \sum_{j=1}^{m} \mathbf{1}\left[a_j^i \neq a_j^{\mathrm{soc}}\right]$$

Population risk (target, unobservable):

$$L_D(i) = \Pr_{p \sim D}\left[a^i(p, \mathcal{M}) \neq a^{\mathrm{soc}}(p, \mathcal{M})\right]$$

ERM selection rule:

$$i^* = \underset{i \in \{1, \ldots, n\}}{\arg\min}\ L_S(i)$$

Question: Does low $L_S(i^*)$ imply low $L_D(i^*)$?

Theorem 1: Finite-Sample Guarantee

Theorem (Uniform Convergence)

With probability $\geq 1 - \delta$*:*

$$L_D(i^*) - \min_i L_D(i) \;\leq\; 2\sqrt{\frac{2\ln n}{m}} + 5\sqrt{\frac{2\ln(8/\delta)}{m}}$$

- Best candidate *in the list* $\Rightarrow$ generalizes out-of-sample
- n enters only as $\ln n$: $10 \to 100$ candidates $\approx 2\times$ data, not $10\times$

Interpretation: Theorem 1 & Fiscal Policy

- $n = 100$ candidates, each with own (s_i, w_i) over debt/growth/etc.
- Observe m past recessions: each candidate's stance
- ERM $\rightarrow$ candidate: "best in-sample"
- *Distribution-free*: holds whatever the s_i's are, however ideologically diverse
- No need to know the similarity function

Corollary: Sample Complexity

Corollary

For any $\varepsilon, \delta \in (0,1)$, *if*

$$m \geq \frac{98}{\varepsilon^2}\left[\ln n + \ln\frac{8}{\delta}\right]$$

then with probability $\geq 1-\delta$*:* $L_D(i^*) \leq \min_i L_D(i) + \varepsilon$.

Concrete numbers ($\varepsilon = 0.10$, $\delta = 0.05$, $n = 100$): $m \approx 95{,}000$.

- High, but distribution-free (worst-case).
- **Economic implication:** Candidate diversity is informationally cheap. Restricting ballot access cannot be justified on informational grounds.

Can representation be *perfect*, asymptotically?

Exponential similarity (Billot et al., 2008): shift inv. + ray mono. + symmetry + ray-shift inv. + self-relevance ⇒

$$s(p, q) = e^{-\nu(p-q)}, \qquad \nu : \mathbb{R}^d \to \mathbb{R}_+ \text{ a norm}$$

Linear utility: $u(r; \mathbf{w}) = \langle \mathbf{w}, r \rangle$, $\mathbf{w} \in \Delta^k$ — agree on dimensions (growth, debt, employment, equality), differ in weights.

A3 — Dense Candidate Coverage

$\{\theta^1, \ldots, \theta^n\} = \{(\nu^i, \mathbf{w}^i)\}$ forms an η_n-net of $\Theta = \{\nu\} \times \Delta^k$, $\eta_n \to 0$.

Fiscal policy: candidates differ in how they weigh problem-distance (ν) and outcome priorities ($\mathbf{w}$).

Theorem 2: Asymptotic Zero Risk

Lemma (Lipschitz Continuity)

Under exponential similarity, linear utility, bounded problems and outcomes, and a margin condition, there exists $C > 0$ such that

$$|L_D(i) - L_D(j)| \leq C\|\theta^i - \theta^j\|$$

Theorem (Asymptotic Representational Equivalence)

Suppose Assumptions 1–3 hold and $m(n)/\ln n \to \infty$. Then

$$L_D(i^*) \xrightarrow{p} 0 \qquad \text{as } n \to \infty.$$

Proof: $L_D(i^*) \leq \underbrace{C\eta_n}_{\to 0 \text{ (coverage)}} + \underbrace{O\left(\sqrt{\ln n/m}\right)}_{\to 0 \text{ (data)}}.$

Interpretation: Theorem 2 & Fiscal Policy

- Pool grows (more parties, lower entry costs) + history accumulates
- $\Rightarrow \exists\, i : \; (\nu^i, \mathbf{w}^i) \approx (\nu^{\text{soc}}, \mathbf{w}^{\text{soc}})$, arbitrarily close
- The candidate's stimulus/austerity calls match society's on (almost) every future recession
- Representation is as good as direct democracy (with diversity and memory)

Failure Modes

1. **Sparse coverage** ($\eta_n \not\to 0$): filing fees, Duverger's Law $\Rightarrow$ only "centrist" stimulus/austerity on the ballot
2. **Insufficient data** ($m/\ln n \not\to \infty$): new democracies, destroyed records
3. **Heterogeneous reasoning**: disagree on *how* to reason
4. **Distribution shift**: D_t moves fast $\Rightarrow$ history loses predictive power

Conclusion

- Representative selection = **estimation**, not aggregation ($\neq$ Arrow)
- Possibility theorem: $O(\ln n/\varepsilon^2)$ observations suffice
- Asymptotically: dense coverage + growing data $\Rightarrow$ disagreement prob. $\to 0$
- Institutional implications: low ballot barriers, institutional memory, stationarity — *necessary*, not merely nice-to-have
- Limitations: static, i.i.d. problems, common utility (base model)

Salut!

Questions?

References

1. Gilboa, I. & Schmeidler, D. (1995). Case-based decision theory. *QJE*, 110(3), 605–639.
2. Billot, A., Gilboa, I., & Schmeidler, D. (2008). Axiomatization of an exponential similarity function. *Mathematical Social Sciences*, 55(2), 107–115.
3. Besley, T. & Coate, S. (1997). An economic model of representative democracy. *QJE*, 112(1), 85–114.
4. Arrow, K.J. (1950). A difficulty in the concept of social welfare. *JPE*, 58(4), 328–346.
5. Valiant, L.G. (1984). A theory of the learnable. *CACM*, 27(11), 1134–1142.
6. Bartlett, P.L. & Mendelson, S. (2002). Rademacher and Gaussian complexities. *JMLR*, 3, 463–482.
7. Dai, J. & Fleisig, E. (2024). Mapping social choice theory to RLHF. *arXiv*:2404.13038.

Appendix

Appendix: Proof of Theorem 1 (II)

Step 2 (McDiarmid). Bounded differences $\Rightarrow$ w.p. $\geq 1 - \delta/2$:

$$\sup_i \left| L_D(h^i) - L_S(h^i) \right| \;\leq\; \mathbb{E}_S\left[\hat{\mathfrak{R}}_S(\mathcal{F})\right] + \sqrt{\frac{\ln(4/\delta)}{2m}}$$

Step 3 (ERM bound). Standard argument $\Rightarrow$ w.p. $\geq 1 - \delta$:

$$L_D(i^*) - \min_i L_D(i) \;\leq\; 2\,\mathbb{E}_S\left[\hat{\mathfrak{R}}_S(\mathcal{F})\right] + 5\sqrt{\frac{2\ln(8/\delta)}{m}}$$

Step 4. Substitute Step 1:

$$L_D(i^*) - \min_i L_D(i) \;\leq\; 2\sqrt{\frac{2\ln n}{m}} + 5\sqrt{\frac{2\ln(8/\delta)}{m}} \qquad \blacksquare$$

Appendix: CBDT Axioms (A1–A5) — Formal

Fix $p \in \mathcal{P}$, history H. Profiles $x, y, z, w \in \mathbb{R}^H$; v_q is the unit vector at $q \in H$. "Compatible" = jointly admissible under memory.

Axiom (Comparability)

For all compatible $x, y \in \mathbb{R}^H$*:* $x \succeq_{p,H} y$ *or* $y \succeq_{p,H} x$.

Axiom (Monotonicity)

If x, y *compatible,* $x \cdot y = 0$*, and* $x(q) \geq y(q)$ *for all* $q \in H$*, then* $x \succeq_{p,H} y$.

Axiom (Continuity)

$\{y : y \succeq_{p,H} x\}$ *and* $\{y : x \succeq_{p,H} y\}$ *are closed in* $\mathbb{R}^H$.

Axiom (Separability)

For all $x, y, z, w \in \mathbb{R}^H$ *with* $(x + z), (y + w)$ *compatible: if* $x \succeq_{p,H} y$ *and* $z \succeq_{p,H} w$*, then* $(x + z) \succeq_{p,H} (y + w)$*; strictly if either antecedent is strict.*

Axiom (Similarity Invariance)

For memories M_1, M_2 *with* $q_1, q_2 \in H_i$, $p \notin H_i$*, and unit vectors* v_j^i*: if* $x \sim_{p,H_1} y$, $z \sim_{p,H_2} w$*, and* $x + \alpha v_1^1 \sim_{p,H_1} y + \beta v_2^1$*, then* $z + \alpha v_1^2 \sim_{p,H_2} w + \beta v_2^2$ *(whenever*

Appendix: CBDT Axioms — Interpretation

Representation (Gilboa & Schmeidler, 1995). A1–A5 $\Rightarrow \exists\, s, u$:

$$x \succsim_{p,H} y \iff \sum_q s(p,q)u(x(q)) \geq \sum_q s(p,q)u(y(q))$$

s unique up to scaling, u up to positive affine transform.

A1–A4 (regularity):

- **Comparability:** voters can always compare candidates' records
- **Monotonicity:** better outcomes $\rightarrow$ weakly preferred (agreement on u, relaxed in §4)
- **Continuity:** rules out single-issue / lexicographic voting
- **Separability:** fiscal record judged independently of foreign policy

A5. The substantive one: relative weight of q_1 vs. q_2 is intrinsic to (p, q_1, q_2), not memory-dependent. Each s_i is a **fixed, portable trait**, precisely what lets observed historical agreement generalise to future agreement (Thm 1).

Appendix: Exponential Similarity Axioms — Formal

Similarity $s : \mathcal{P} \times \mathcal{P} \to (0,1]$, $\mathcal{P} \subseteq \mathbb{R}^d$.

Axiom (Shift Invariance)

For all p, q and $v \in \mathbb{R}^d$ with $p + v, q + v \in \mathcal{P}$: $s(p,q) = s(p+v, q+v)$.

Axiom (Ray Monotonicity)

For all p and $v \in \mathbb{R}^d \setminus \{0\}$: $t \mapsto s(p, p+tv)$ is decreasing in $t > 0$.

Axiom (Symmetry)

For all p, q: $s(p,q) = s(q,p)$.

Axiom (Ray Shift Invariance)

For all p, q and $\lambda > 0$: $s(p, \lambda p) = s(q, \lambda q)$ whenever both are defined.

Axiom (Self-Relevance)

For all p: $s(p,p) = 1$.

Appendix: Exponential Similarity — Interpretation

- **Shift inv.:** only the *gap* matters — 3% vs 5% debt $\equiv$ 103% vs 105%
- **Ray mono.:** 110% debt is closer to 100% than to 120%
- **Symmetry:** 2008 informs 2026 as much as 2026 would inform 2008
- **Ray-shift inv.:** no special / privileged region of P
- **Self-relevance:** normalises the scale ($s = 1 \Leftrightarrow$ identical)

$\Rightarrow$ political-economy analogue of a Gaussian/RBF kernel; matches Shepard's (1987) law of similarity decay in psychology — people reason by graded analogy, not sharp categories.