Preferences in AI Project Report

Pratik Deshmukh

September 26, 2025

1 Introduction

This report presents experiments on free-riding in sequential decision-making under different *statistical cultures*, following the framework of [**?**].

2 Methodology

We repeat the experiments from the paper but use several different statistical cultures: impartial culture (p-IC), disjoint groups, and the (p, ϕ)-resampling model. For each, we run multiple seeds and compare welfare and risk metrics under sequential utilitarian and seq-PAV rules.

3 Results

The combined results table is automatically generated by the experiment pipeline. The table below is included directly from the output file:

culture	rule	seeds	utilitarian	egalitarian Welfare	nash	trials	successes Risk	harms	success_rate	harm_rate
p_ic	utilitarian	30	62.967	1.000	8.954	100.000	20.400	20.400	0.204	0.204
p_ic	pav	30	62.967	1.000	8.954	100.000	20.400	20.400	0.204	0.204
disjoint	utilitarian	30	33.500	0.100	3.241	100.000	14.400	14.400	0.144	0.144
disjoint	pav	30	33.500	0.100	3.241	100.000	14.400	14.400	0.144	0.144
resampling	utilitarian	30	78.033	2.033	15.212	100.000	11.667	11.667	0.117	0.117
resampling	pav	30	78.033	2.033	15.212	100.000	11.667	11.667	0.117	0.117

Table 1: Combined results across cultures and rules. Welfare metrics are (utilitarian, egalitarian, Nash), while risk metrics include success and harm rates.

4 Discussion

We observe that the choice of statistical culture significantly affects both welfare and manipulation risks. For example, disjoint cultures show notably lower welfare but also lower manipulation success rates, while resampling tends to yield higher welfare but with moderate manipulation risk.

5 Conclusion

These experiments confirm that statistical cultures strongly influence the perceived robustness of voting rules. Future work may explore richer parameter grids and robustness to noise.

Repository

The full project code and report sources are available at:
github.com/inquisitour/preferences-in-ai

A Code Documentation Summary

- **core/** – Base types, welfare functions, and voting rule interface.
- **statistical_cultures/** – Preference generators (p-IC, disjoint, resampling, hamming noise).
- **voting_rules/** – Implementations of sequential utilitarian and seq-PAV.
- **free_riding/** – Manipulation detector, welfare/risk analysis tools.
- **experiments/** – Main experiment runner for end-to-end evaluation.
- **tests/** – Unit tests for cultures, rules, and free-riding checks.

References