

Interface-Level Significance Analysis Summary

Generated from interface_statistical_analysis.R

Key Findings from Interface-Level Analysis with Statistical Testing

SIGNIFICANT INTERFACES IDENTIFIED:

Interfaces with Significant Differences ($p < 0.05$):

Interface 2 (UI2):

- Significant tendency difference ($p = 0.023$, $d = -0.835$)
- UEQ+Autonomy shows LOWER tendency scores than UEQ
- Large effect size ($d > 0.8$)

Interface 8 (UI8):

- Significant rejection difference ($p = 0.022$, $d = 0.756$)
- Significant tendency difference ($p = 0.015$, $d = -0.777$)
- UEQ+Autonomy shows HIGHER rejection rates and LOWER tendency scores
- Large effect sizes in both measures

Interface 11 (UI11):

- Significant tendency difference ($p = 0.038$, $d = -0.683$)
- UEQ+Autonomy shows LOWER tendency scores than UEQ
- Medium-to-large effect size

Interface 15 (UI15):

- Significant rejection difference ($p = 0.011$, $d = 0.857$)
- Significant tendency difference ($p = 0.043$, $d = -0.629$)
- UEQ+Autonomy shows HIGHER rejection rates and LOWER tendency scores
- Large effect sizes in both measures

SUMMARY STATISTICS:

- **Total interfaces analyzed:** 15 (all with sufficient data $n \geq 8$)
- **Interfaces with significant rejection differences:** 2 (UI8, UI15)
- **Interfaces with significant tendency differences:** 4 (UI2, UI8, UI11, UI15)
- **Overall pattern:** UEQ+Autonomy tends to result in more critical evaluations

OVERALL INTERFACE-LEVEL EFFECTS:

- **Mean rejection difference:** +3.36% (UEQ+Autonomy slightly higher)
- **Mean tendency difference:** -0.39 points (UEQ+Autonomy slightly lower)

INTERPRETATION:

While the participant-level analysis showed NO significant overall effects, the interface-level analysis reveals that certain specific interfaces DO show significant differences between UEQ and UEQ+Autonomy evaluation frameworks.

Pattern: When significant differences occur, UEQ+Autonomy consistently leads to:

- HIGHER rejection rates
- LOWER release tendency scores
- More critical/conservative evaluation decisions

This suggests that ethics-enhanced evaluation (UEQ+Autonomy) makes evaluators more sensitive to problematic design patterns in specific interfaces, even though this effect doesn't reach significance when averaged across all interfaces and participants.

FILES GENERATED:

- `plots/interface_rejection_trends_updated_with_stats.png` - Main visualization showing all interfaces with statistical significance markers
- `plots/interface_tendency_trends_updated_with_stats.png` - Tendency score comparison with significance testing
- `results/interface_statistical_tests.csv` - Complete statistical test results for all interfaces

RESEARCH IMPLICATIONS:

1. **Interface-Specific Effects:** UEQ+Autonomy impacts are not uniform across all interfaces
2. **Effect Heterogeneity:** Some interfaces are more sensitive to ethics-enhanced evaluation
3. **Conservative Bias:** When effects occur, they favor more conservative/critical decisions
4. **Measurement Sensitivity:** Interface-level analysis can detect effects missed in aggregate analysis