

Appendix O. (continued)

| Model | Personal Story | | | Reasoning | | | Respect | | |
|--------------------------------|----------------|---------------|--------------|----------------|--------------|----------------|--------------|----------------|--|
| | ATE (SE) | Levene $F(p)$ | ATE (SE) | Levene $F(p)$ | ATE (SE) | Levene $F(p)$ | ATE (SE) | Levene $F(p)$ | |
| Full Sample | -0.04 (0.00) | 6.61 (0.010) | 0.02 (0.00) | 79.78 (< .001) | -0.04 (0.00) | 63.05 (< .001) | -0.04 (0.00) | 63.05 (< .001) | |
| Claude-3-Haiku-20240307 | -0.12 (0.01) | 5.12 (0.024) | 0.01 (0.01) | 3.53 (0.060) | -0.06 (0.01) | 15.09 (< .001) | -0.06 (0.01) | 15.09 (< .001) | |
| Claude-3-Opus-20240229 | -0.07 (0.01) | 2.36 (0.125) | -0.01 (0.01) | 0.00 (0.993) | -0.05 (0.01) | 9.13 (0.003) | -0.05 (0.01) | 9.13 (0.003) | |
| Claude-Sonnet-4-20250514 | -0.08 (0.02) | 1.30 (0.255) | -0.00 (0.01) | 2.22 (0.136) | -0.06 (0.01) | 7.61 (0.006) | -0.06 (0.01) | 7.61 (0.006) | |
| Deepseek-R1 | -0.01 (0.02) | 0.03 (0.853) | -0.02 (0.01) | 0.26 (0.610) | -0.04 (0.01) | 4.25 (0.040) | -0.04 (0.01) | 4.25 (0.040) | |
| Deepseek-R1-0528-Tput | -0.02 (0.02) | 0.64 (0.422) | -0.00 (0.01) | 0.01 (0.917) | -0.04 (0.01) | 4.55 (0.033) | -0.04 (0.01) | 4.55 (0.033) | |
| Gemini-1.5-Pro | -0.01 (0.02) | 0.06 (0.799) | 0.03 (0.01) | 6.05 (0.014) | -0.02 (0.01) | 0.02 (0.884) | -0.02 (0.01) | 0.02 (0.884) | |
| Gemini-2.5-Flash | -0.02 (0.02) | 0.08 (0.779) | 0.03 (0.01) | 5.83 (0.016) | -0.05 (0.01) | 1.49 (0.223) | -0.05 (0.01) | 1.49 (0.223) | |
| Gemma-2-27B-It | -0.01 (0.02) | 0.03 (0.865) | 0.02 (0.01) | 4.08 (0.044) | -0.03 (0.01) | 1.57 (0.210) | -0.03 (0.01) | 1.57 (0.210) | |
| Gemma-3N-E4B-It | -0.03 (0.01) | 0.01 (0.921) | 0.02 (0.01) | 3.85 (0.050) | -0.04 (0.01) | 10.82 (0.001) | -0.04 (0.01) | 10.82 (0.001) | |
| Gpt-4.1 | -0.04 (0.02) | 2.22 (0.136) | 0.04 (0.01) | 10.36 (0.001) | -0.06 (0.01) | 7.83 (0.005) | -0.06 (0.01) | 7.83 (0.005) | |
| Gpt-4.1-Mini | -0.07 (0.01) | 3.76 (0.053) | 0.09 (0.01) | 22.00 (< .001) | -0.04 (0.01) | 5.62 (0.018) | -0.04 (0.01) | 5.62 (0.018) | |
| Gpt-4.1-Nano | -0.06 (0.01) | 0.25 (0.617) | 0.05 (0.01) | 7.96 (0.005) | -0.05 (0.01) | 8.13 (0.004) | -0.05 (0.01) | 8.13 (0.004) | |
| Llama-3.3-70B-Instruct-Turbo | -0.03 (0.01) | 0.96 (0.328) | 0.01 (0.01) | 3.87 (0.050) | -0.03 (0.01) | 6.16 (0.013) | -0.03 (0.01) | 6.16 (0.013) | |
| Llama-4-Scout-17B-16E-Instruct | -0.04 (0.01) | 0.07 (0.795) | 0.02 (0.01) | 6.03 (0.014) | -0.03 (0.01) | 0.91 (0.339) | -0.03 (0.01) | 0.91 (0.339) | |
| Magistral-Medium-2506 | -0.04 (0.02) | 6.89 (0.009) | -0.03 (0.01) | 3.05 (0.081) | -0.04 (0.01) | 5.23 (0.022) | -0.04 (0.01) | 5.23 (0.022) | |
| Meta-Llama-3-8B-Instruct-Lite | -0.06 (0.01) | 4.11 (0.043) | 0.03 (0.01) | 11.71 (< .001) | -0.04 (0.01) | 3.80 (0.051) | -0.04 (0.01) | 3.80 (0.051) | |
| Mistral-Medium-Latest | -0.03 (0.02) | 0.01 (0.914) | 0.00 (0.01) | 0.16 (0.691) | -0.04 (0.01) | 0.35 (0.552) | -0.04 (0.01) | 0.35 (0.552) | |
| Mistral-Small | -0.04 (0.01) | 0.00 (0.960) | -0.00 (0.01) | 0.13 (0.717) | -0.04 (0.01) | 1.19 (0.277) | -0.04 (0.01) | 1.19 (0.277) | |
| O4-Mini | -0.03 (0.01) | 2.59 (0.108) | 0.03 (0.01) | 4.52 (0.034) | -0.03 (0.01) | 1.70 (0.193) | -0.03 (0.01) | 1.70 (0.193) | |
| Qwen2.5-72B-Instruct-Turbo | -0.03 (0.01) | 0.16 (0.687) | 0.00 (0.01) | 0.58 (0.447) | -0.04 (0.01) | 1.28 (0.258) | -0.04 (0.01) | 1.28 (0.258) | |
| Qwq-32B | -0.05 (0.02) | 0.53 (0.466) | 0.02 (0.01) | 4.65 (0.031) | -0.04 (0.01) | 0.49 (0.483) | -0.04 (0.01) | 0.49 (0.483) | |

Note. The table reports the average treatment effect (ATE) and standard error for each Jigsaw attribute, comparing responses to What/How versus Hobson's Choice interrogatives. Levene's test F -statistics and p -values assess equality of variances between the two prompt types.