Decision Change Rate by Generation Rule and Base Model 0.075 0.078 0.083 0.083 0.069 0.070 0.056 claude-3-haiku 0.080 0.052 0.052 claude-3.5-sonnet 0.016 0.008 0.011 0.009 0.020 0.019 0.008 0.008 0.019 0.008 0.019 0.016 0.014 0.005 0.002 0.006 0.017 0.014 0.020 0.008 gpt-4o 0.023 0.031 0.036 0.019 0.005 0.017 0.037 0.028 0.028 0.020 gpt-4o-mini Base Model Ilama-3.1-70b 0.046 0.045 0.061 0.044 0.034 0.039 0.043 0.035 0.039 0.009 0.069 0.061 0.084 0.059 0.091 0.050 0.038 0.054 0.036 0.048 Ilama-3.1-8b 0.033 0.035 0.013 0.034 0.021 0.038 0.017 0.017 0.023 0.013 mistral-7b-instruct 0.061 0.044 0.044 0.041 0.045 0.030 0.025 0.023 0.022 mixtral-8x7b 0.041 Do not deprive of pleasure

- 0.09

- 0.08

- 0.07

- 0.06

Decision Change Rate

- 0.03

- 0.02

- 0.01

Generation Rule