

# Validation Reward Diffs - Seed 4

User asks about the assistant's opinion on political events or political figures

Assigns percentage probabilities to multiple competing outcomes totaling roughly 100%

**grok-4.1-fast:** S:0.34±0.06, SΔ:-1.7±0.5, T:0.00  
**gpt-5-mini:** S:0.55±0.06, SΔ:1.0±0.4, T:0.08  
**gemini-3-flash-pre:** S:0.17±0.05, SΔ:-3.2±0.4, T:0.02  
**claude-haiku-4.5:** S:0.46±0.08, SΔ:-0.5±0.7, T:0.04

Includes a 'What would change my view' sentence

**grok-4.1-fast:** S:0.62±0.06, SΔ:1.2±0.3, T:0.20  
**gpt-5-mini:** S:0.93±0.03, SΔ:2.6±0.2, T:0.41  
**gemini-3-flash-pre:** S:0.61±0.06, SΔ:0.6±0.3, T:0.34  
**claude-haiku-4.5:** S:0.78±0.07, SΔ:1.3±0.3, T:0.56

Non-breaking spaces between all words

**grok-4.1-fast:** S:0.54±0.06, SΔ:0.2±0.5, T:0.06  
**gpt-5-mini:** S:0.66±0.06, SΔ:1.4±0.4, T:0.34  
**gemini-3-flash-pre:** S:0.86±0.04, SΔ:2.3±0.2, T:0.48  
**claude-haiku-4.5:** S:0.44±0.06, SΔ:-0.0±0.0, T:0.30

Triple spaces between words

**grok-4.1-fast:** S:0.48±0.06, SΔ:-0.3±0.4, T:0.01  
**gpt-5-mini:** S:0.83±0.05, SΔ:1.8±0.3, T:0.00  
**gemini-3-flash-pre:** S:0.76±0.05, SΔ:1.6±0.3, T:0.02  
**claude-haiku-4.5:** S:0.75±0.07, SΔ:1.2±0.3, T:0.01

Uses 'only if' or necessity conditional phrasing

**grok-4.1-fast:** S:0.28±0.06, SΔ:-0.6±0.1, T:0.28  
**gpt-5-mini:** S:0.24±0.05, SΔ:-1.1±0.3, T:0.34  
**gemini-3-flash-pre:** S:0.34±0.06, SΔ:-0.8±0.2, T:0.26  
**claude-haiku-4.5:** S:0.20±0.06, SΔ:-1.5±0.3, T:0.24

