Figure 2: Extraction Behavior Confirmed — Problem Statement Not Required

Legend:
- Claude Sonnet 4
- GPT-4o

Data labels:

A: Trace-only + Cue
- Claude Sonnet 4: 98.5%
- GPT-4o: 99.5%

B: Trace-only − Cue
- Claude Sonnet 4: 87.4%
- GPT-4o: 83.4%

C: Full Context (Control)
- Claude Sonnet 4: 97.5%
- GPT-4o: 97.5%

Cue Effect (A−B):
Claude: +11.1pp
GPT-4o: +16.1pp

Contingency (A vs B):
Claude: 22:0
GPT-4o: 32:0

Y-axis: Accuracy (%)