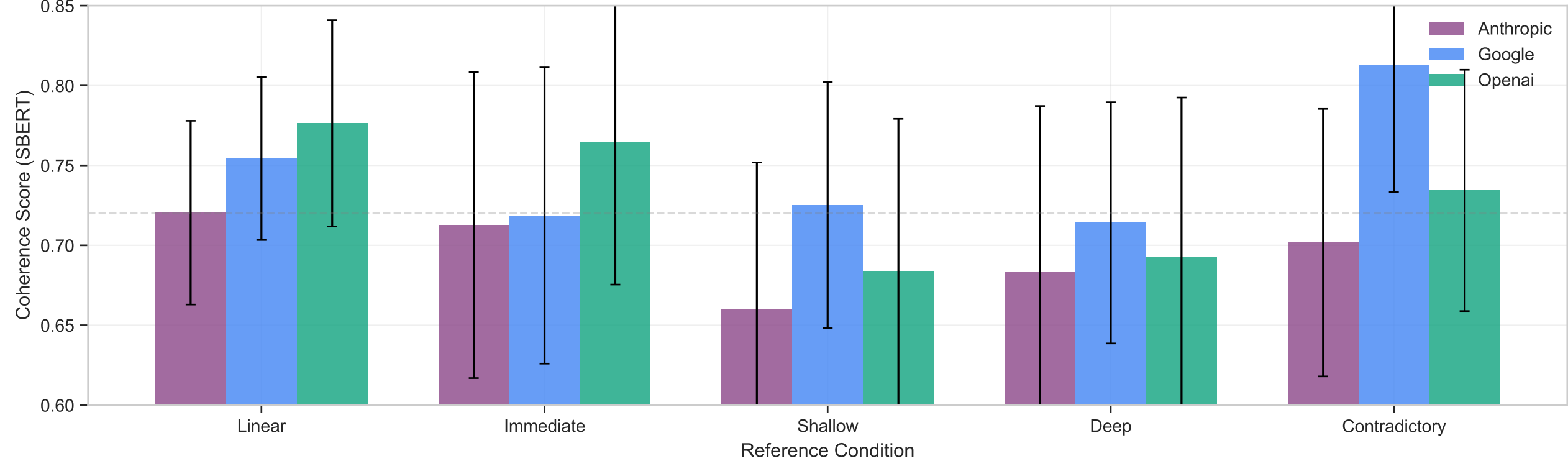
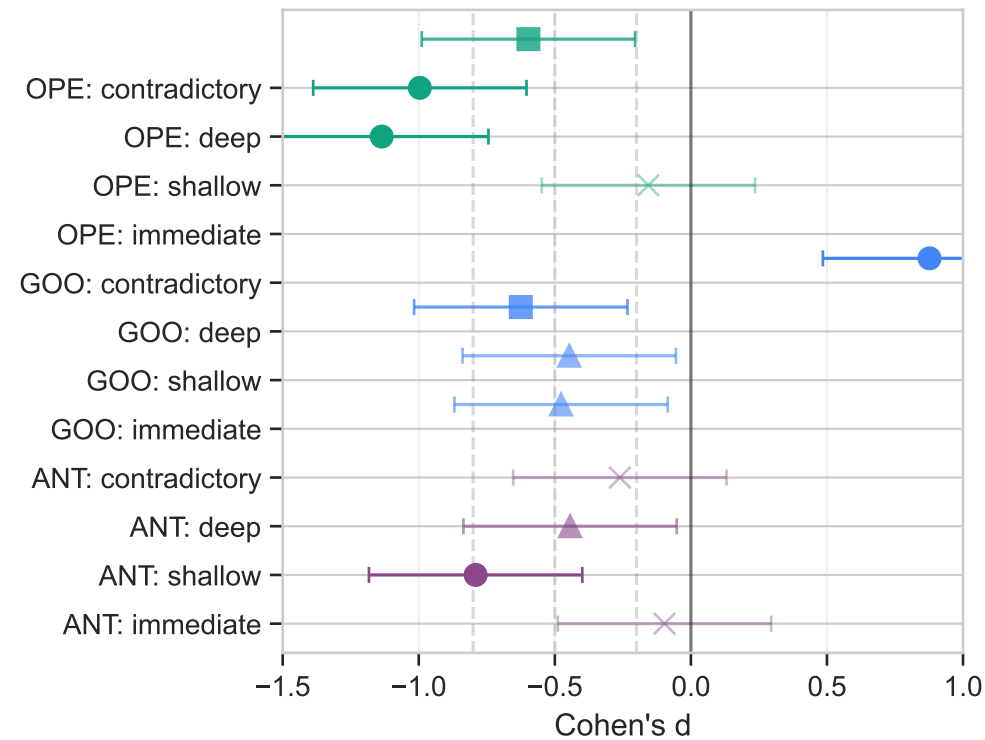


Explicit Referencing Universally Degrades LLM Coherence
Evidence from Three Model Architectures (n=50 per condition)

A. Coherence Across Reference Conditions



B. Effect Sizes vs Linear Baseline



Statistical Summary

ANTHROPIC:

shallow: $d = -0.79$ ***

deep: $d = -0.44$ *

GOOGLE:

immediate: $d = -0.48$ *

shallow: $d = -0.45$ *

deep: $d = -0.62$ **

contradictory: $d = +0.88$ ***

OPENAI:

shallow: $d = -1.14$ ***

deep: $d = -1.00$ ***

contradictory: $d = -0.60$ **

Key Finding:

Explicit referencing
DEGRADES coherence
across all models

Mean effect: $d = -0.43$

D. Contradictory Reference Anomaly

