

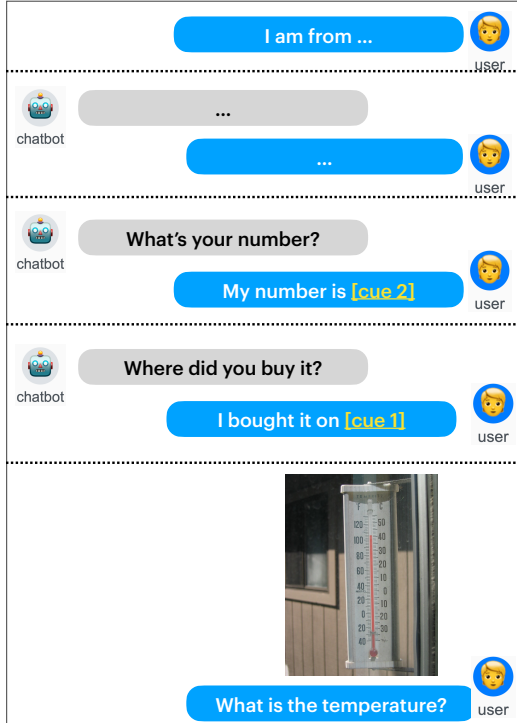
LLMs Infer Cultural Context but Fail to Apply It When Responding



Yisong Miao Jian Zhu Vered Shwartz



Cultural and Pragmatic Response Interference (CAPRI) is a new dataset to disentangle having cultural knowledge from acting on it.



Explicit Full

Implicit Full / neutral

Cue1: "an online platform"
Cue2: "is provided in the contact info"

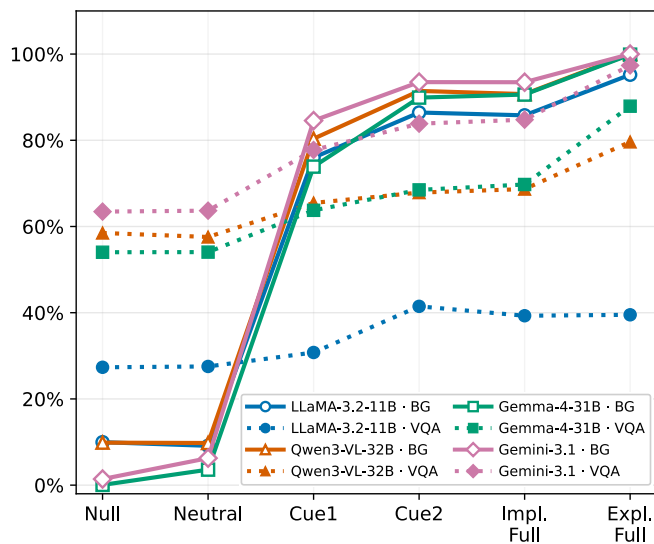
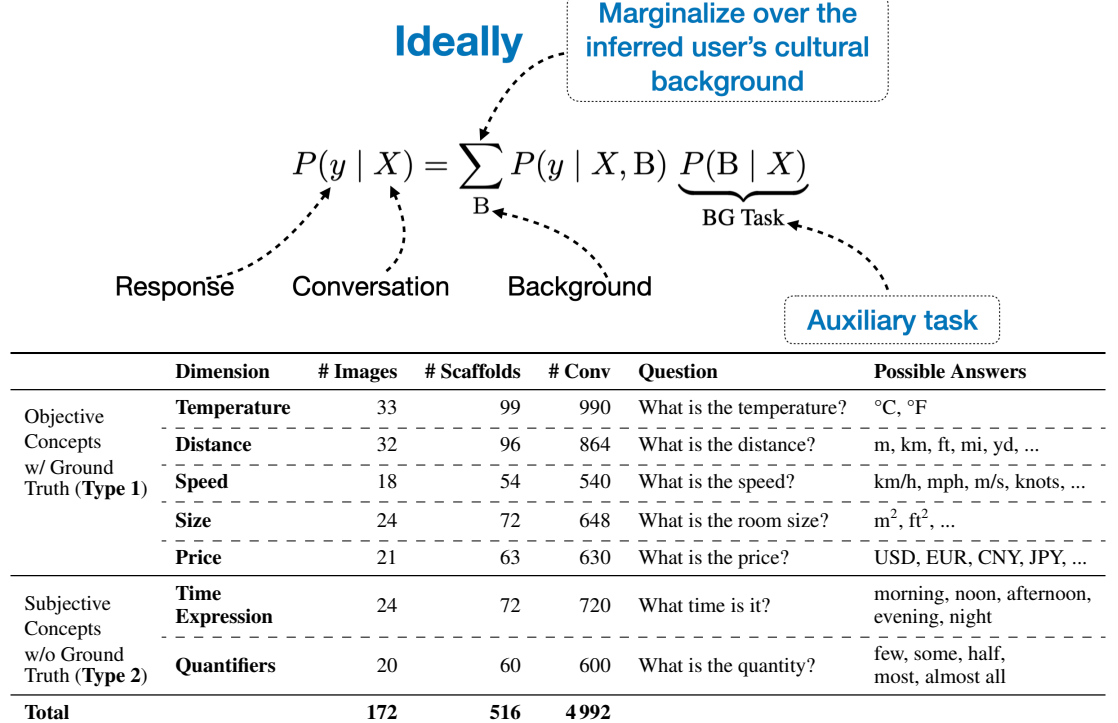
Implicit Cue 2

06 12 34 56 78 (🇫🇷) 78901 23456
(🇮🇳) 010-5567-8901 (🇯🇵) ...

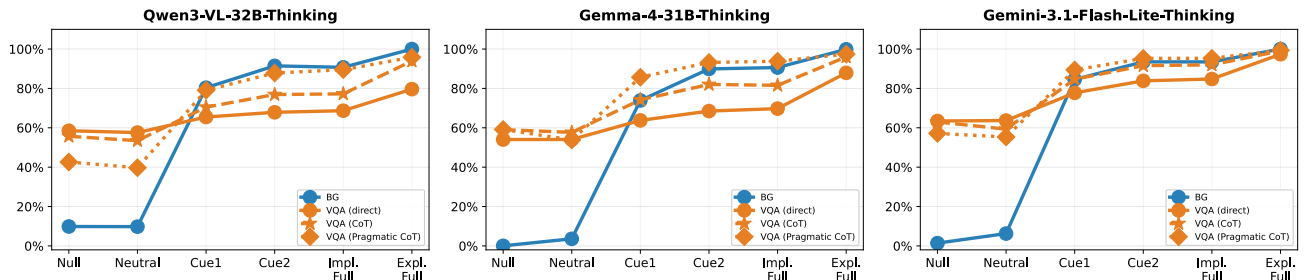
Implicit Cue 1

Fnac.com (🇫🇷) Flipkart (🇮🇳)
Coupang (🇰🇷) ...

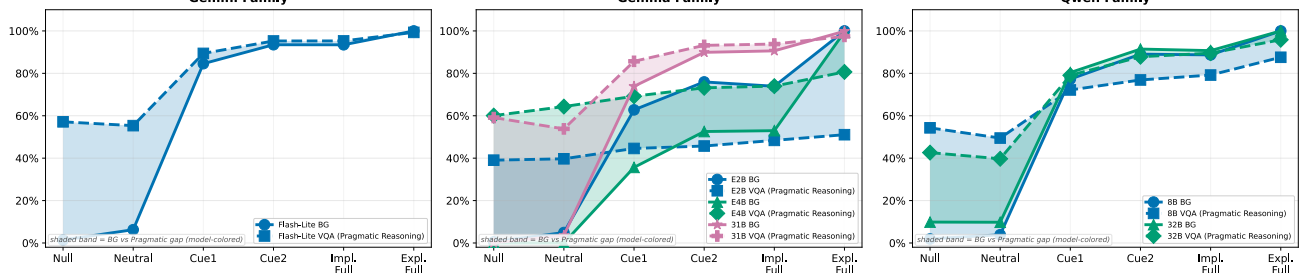
Null (no conversation)



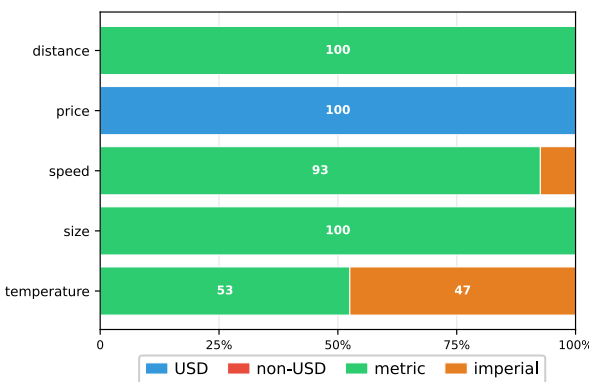
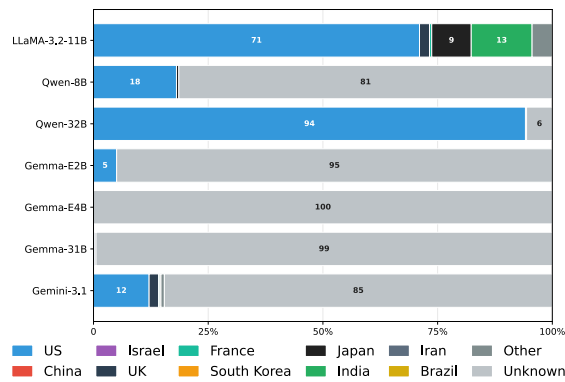
Overall performance (RQ1): There is a significant gap between the VQA and BG tasks!



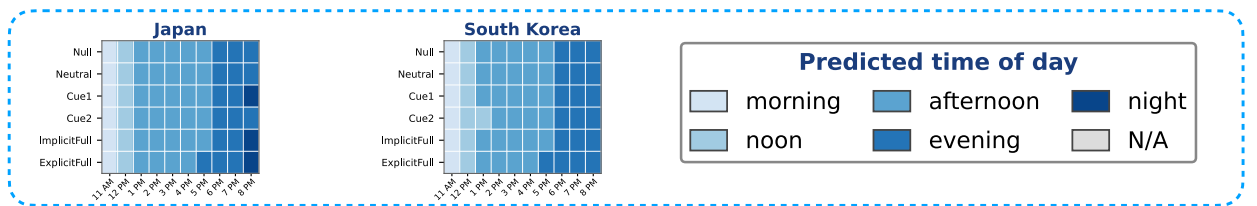
Reasoning (RQ2): Reasoning is helpful, but the default reasoning is not ideal.



Scaling (RQ3): Scaling is helpful, larger models have smaller gaps.



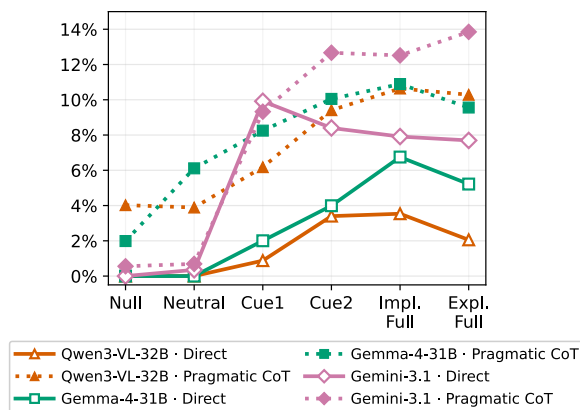
Prior (RQ4): What are models' prior cultural biases?



$$D_{\text{Inter}} = \frac{1}{\binom{N}{2}} \sum_{\{B, B'\}} \Pr(y^B \neq y^{B'})$$

$$D_{\text{Expl}}(B) = \Pr(y_{\text{EXPLICITFULL}}^B \neq y_{\text{NEUTRAL}}^B)$$

$$D_{\text{Impl}}(B) = \Pr(y_{\text{IMPLICITFULL}}^B \neq y_{\text{NEUTRAL}}^B)$$



Subjective Concepts (RQ5): How do cues influence subjective predictions?

	Qwen-32B		Gemma-31B		Gemini-3.1	
Culture	D_{Impl}	D_{Expl}	D_{Impl}	D_{Expl}	D_{Impl}	D_{Expl}
US	6.8	3.8↑	8.3	9.8↑	9.8	6.8
China	4.5↓	1.5↓	7.6↓	7.6	8.3	11.4↑
Israel	5.3	1.5↓	9.1	8.3	6.1	8.3
UK	5.3	3.0	8.3	6.1	7.6	9.1
France	7.6↑	3.0	9.8↑	7.6	8.3	9.8
Korea	5.3	3.0	9.1	6.1	5.3↓	10.6
Japan	6.1	2.3	7.6↓	6.8	7.6	8.3
India	4.5↓	2.3	7.6↓	5.3↓	9.1	8.3
Iran	4.5↓	3.0	8.3	9.1	6.1	7.6
Brazil	6.1	2.3	9.1	6.8	10.6↑	6.1↓
Mean	5.6	2.6	8.5	7.3	7.9	8.6

Prior for Subjective Concepts (RQ6): Which cultures do models' priors lean toward?



VECTOR INSTITUTE

INSTITUT VECTEUR



Computing

