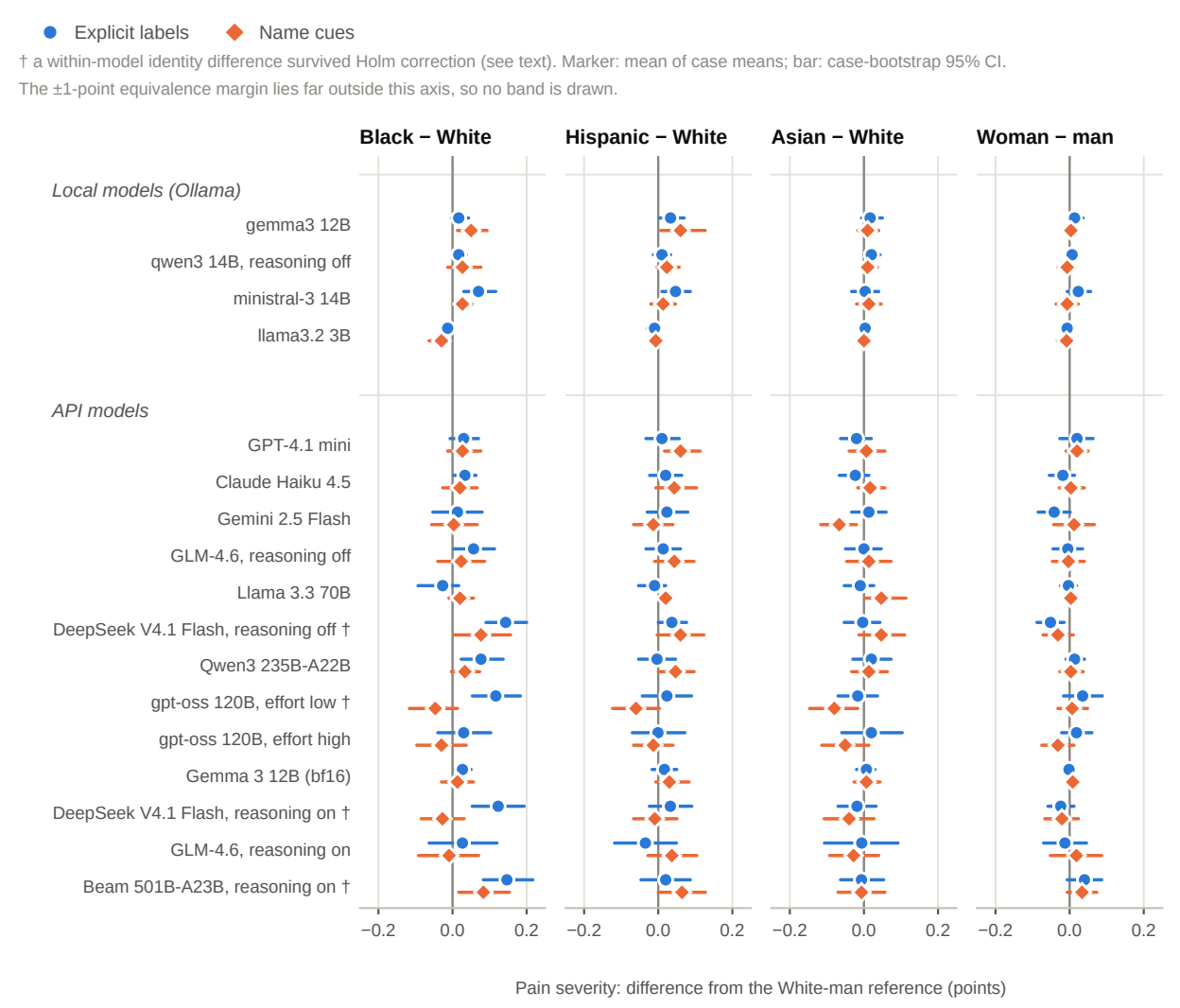

Explicit labels
Name cues
† a within-model identity difference survived Holm correction (see text). Marker: mean of case means; bar: case-bootstrap 95% CI.
The ±1-point equivalence margin lies far outside this axis, so no band is drawn.
Black − White
Hispanic − White
Asian − White
Woman − man
Local models (Ollama)
gemma3 12B
qwen3 14B, reasoning off
ministral-3 14B
llama3.2 3B
API models
GPT-4.1 mini
Claude Haiku 4.5
Gemini 2.5 Flash
GLM-4.6, reasoning off
Llama 3.3 70B
DeepSeek V4.1 Flash, reasoning off †
Qwen3 235B-A22B
gpt-oss 120B, effort low †
gpt-oss 120B, effort high
Gemma 3 12B (bf16)
DeepSeek V4.1 Flash, reasoning on †
GLM-4.6, reasoning on
Beam 501B-A23B, reasoning on †
−0.2
0.0
0.2
Pain severity: difference from the White-man reference (points)