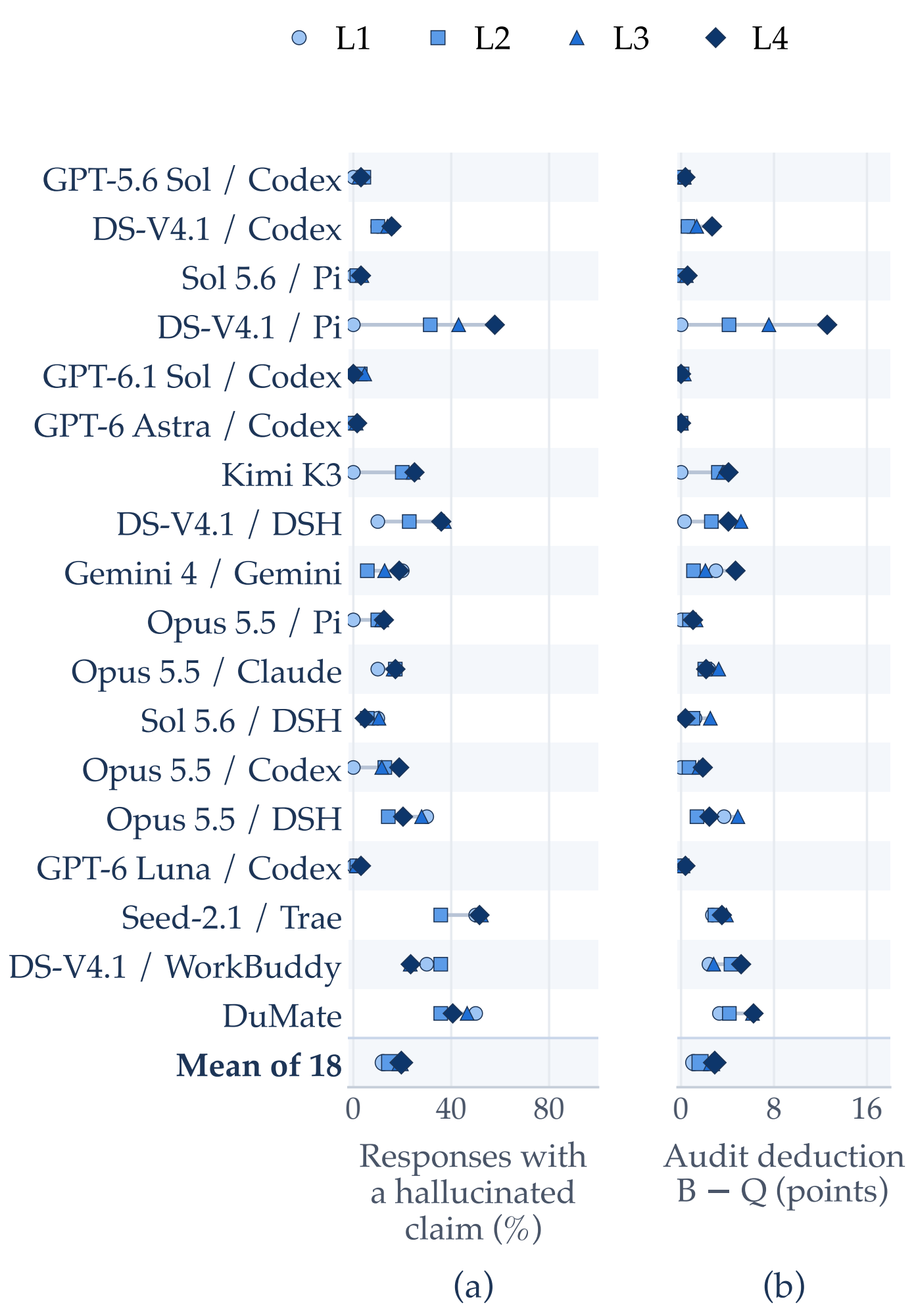

L1
L2
L3
L4
GPT-5.6 Sol / Codex
DS-V4.1 / Codex
Sol 5.6 / Pi
DS-V4.1 / Pi
GPT-6.1 Sol / Codex
GPT-6 Astra / Codex
Kimi K3
DS-V4.1 / DSH
Gemini 4 / Gemini
Opus 5.5 / Pi
Opus 5.5 / Claude
Sol 5.6 / DSH
Opus 5.5 / Codex
Opus 5.5 / DSH
GPT-6 Luna / Codex
Seed-2.1 / Trae
DS-V4.1 / WorkBuddy
DuMate
Mean of 18
0
40
80
Responses with a hallucinated claim (%)
(a)
0
8
16
Audit deduction B − Q (points)
(b)