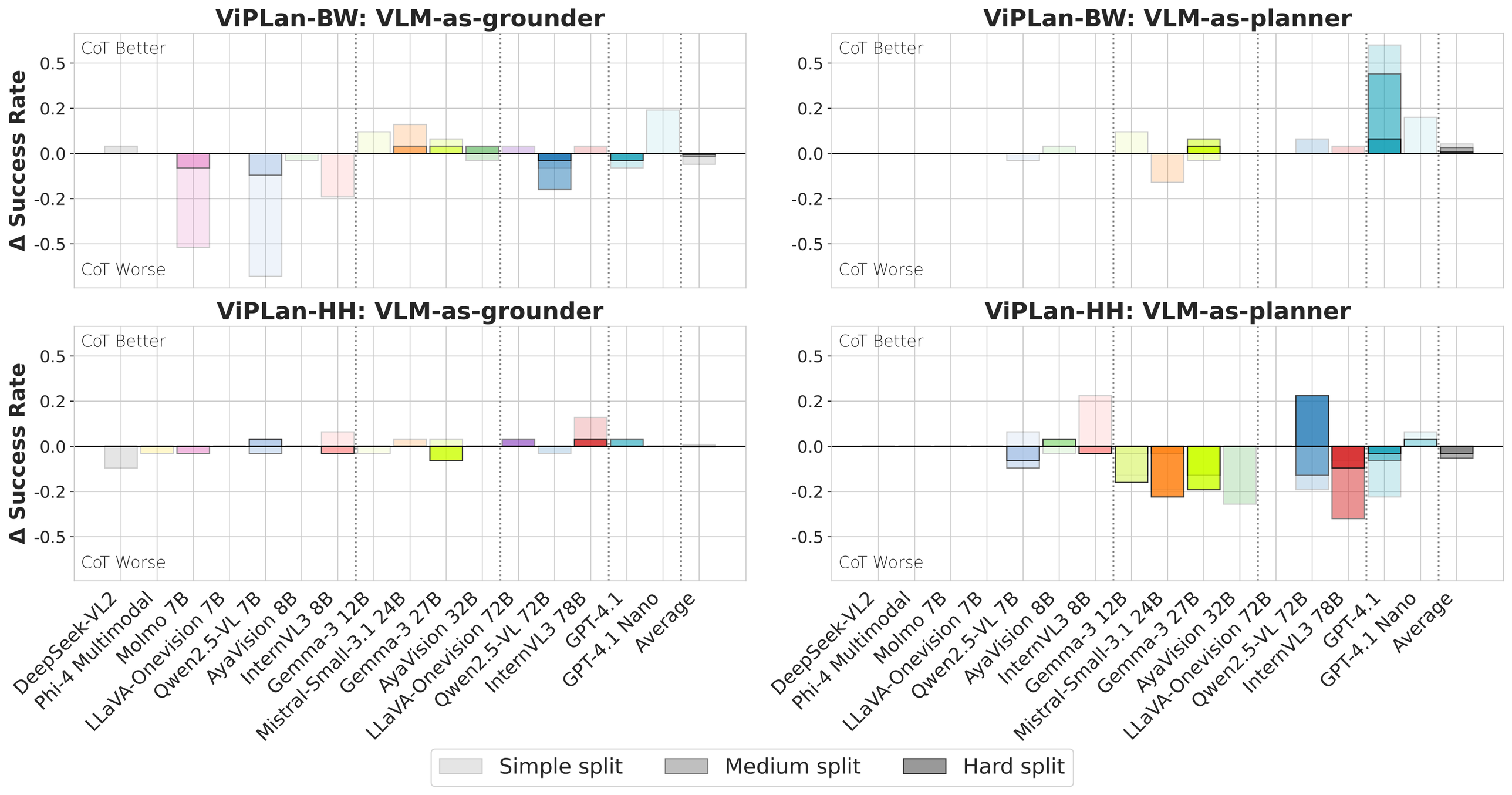
ViPLan-BW: VLM-as-grounder
ViPLan-BW: VLM-as-planner
ViPLan-HH: VLM-as-grounder
ViPLan-HH: VLM-as-planner
Δ Success Rate
CoT Better
CoT Worse
0.5
0.2
0.0
-0.2
-0.5
DeepSeek-VL2
Phi-4 Multimodal
Molmo 7B
LLaVA-Onevision 7B
Qwen2.5-VL 7B
AyaVision 8B
InternVL3 8B
Gemma-3 12B
Mistral-Small-3.1 24B
Gemma-3 27B
AyaVision 32B
LLaVA-Onevision 72B
Qwen2.5-VL 72B
InternVL3 78B
GPT-4.1
GPT-4.1 Nano
Average
Simple split
Medium split
Hard split