Self-Consistency (K runs) vs Single Response
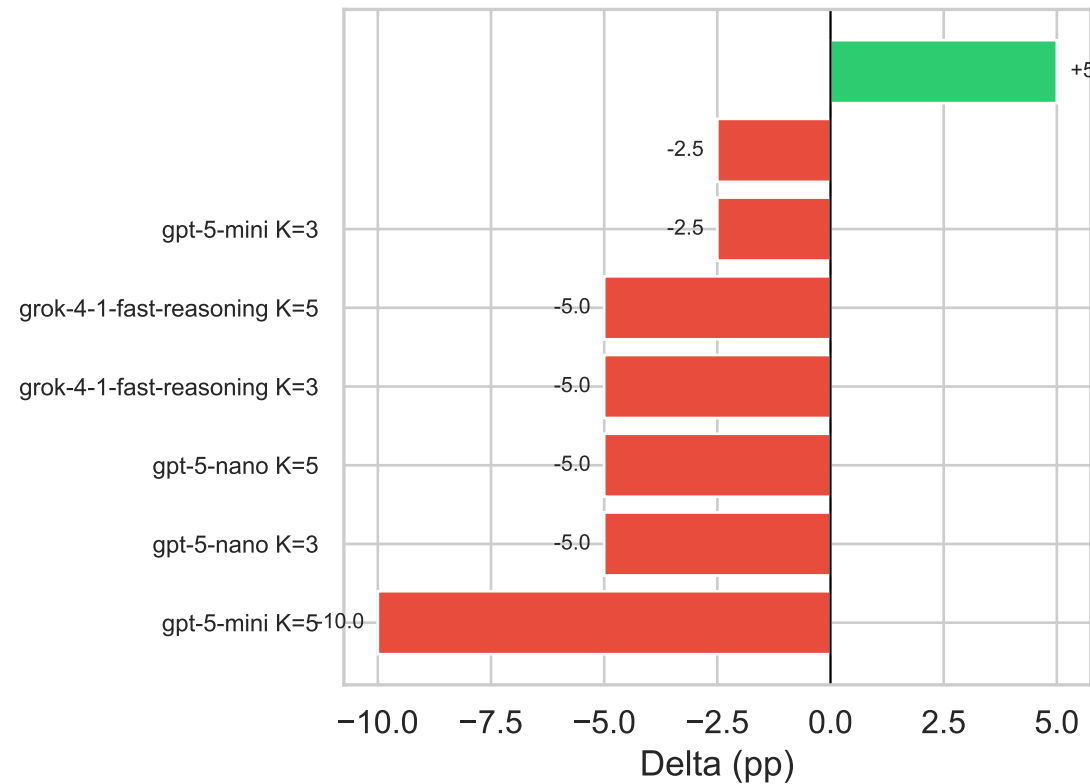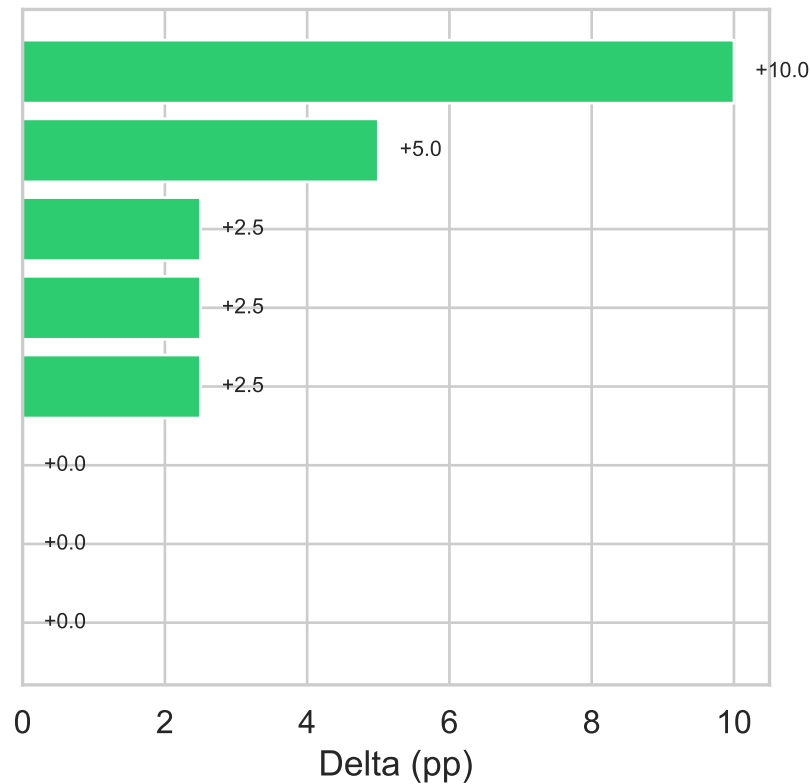
	GSM8K	TruthfulQA	GPQA
	+5.0	+10.0	
	-2.5	+5.0	
gpt-5-mini K=3	-2.5	+2.5	+7.5
grok-4-1-fast-reasoning K=5	-5.0	+2.5	+2.5
grok-4-1-fast-reasoning K=3	-5.0	+2.5	+2.5
gpt-5-nano K=5	-5.0	+0.0	+0.0
gpt-5-nano K=3	-5.0	+0.0	+0.0
gpt-5-mini K=5	-10.0	+0.0	-2.5

Delta (pp) — Delta (pp) — Delta (pp)