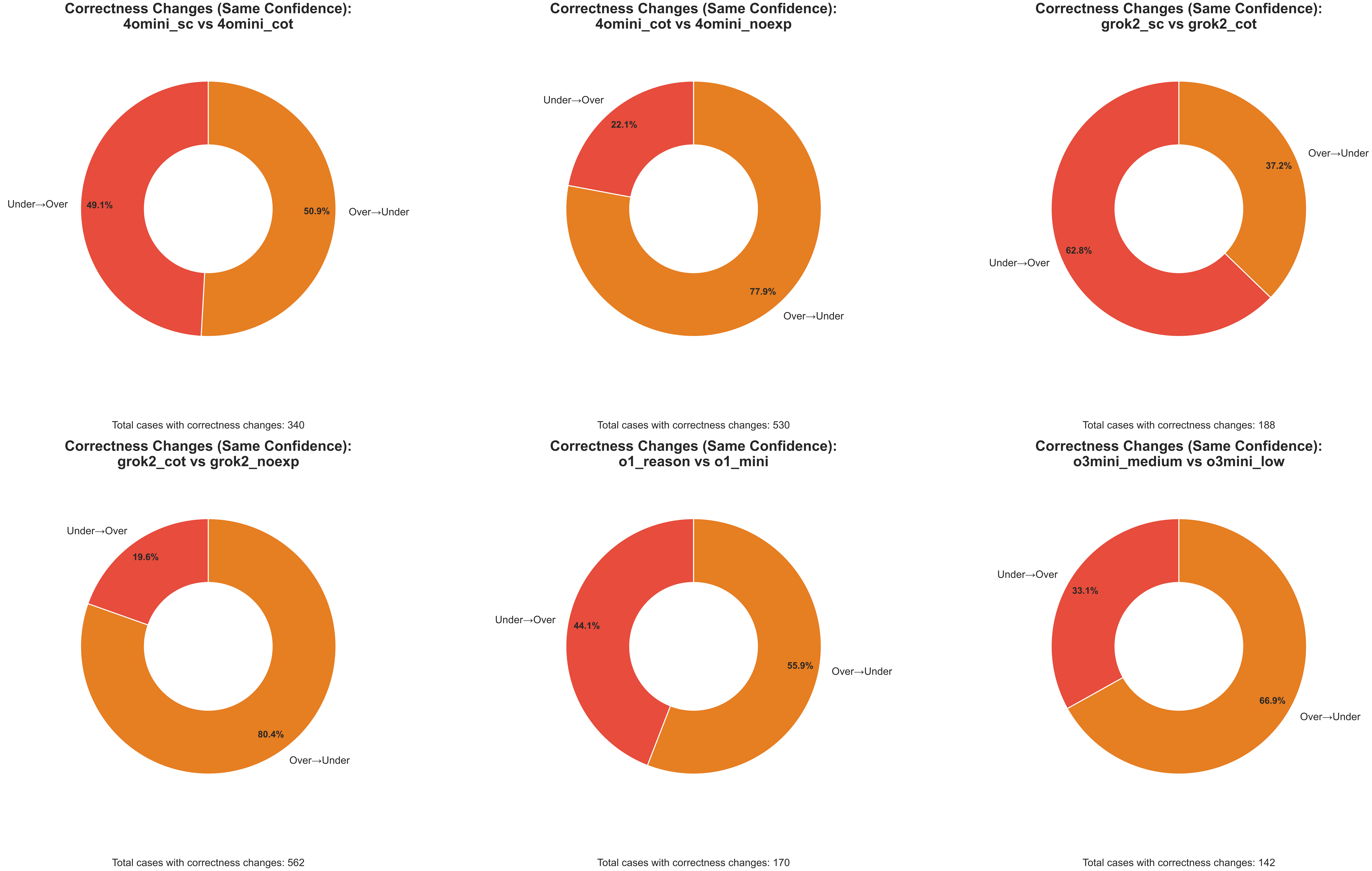

**Correctness Changes (Same Confidence):
4omini_sc vs 4omini_cot**

**Correctness Changes (Same Confidence):
4omini_cot vs 4omini_noexp**

**Correctness Changes (Same Confidence):
grok2_sc vs grok2_cot**

**Correctness Changes (Same Confidence):
grok2_cot vs grok2_noexp**

**Correctness Changes (Same Confidence):
o1_reason vs o1_mini**

**Correctness Changes (Same Confidence):
o3mini_medium vs o3mini_low**