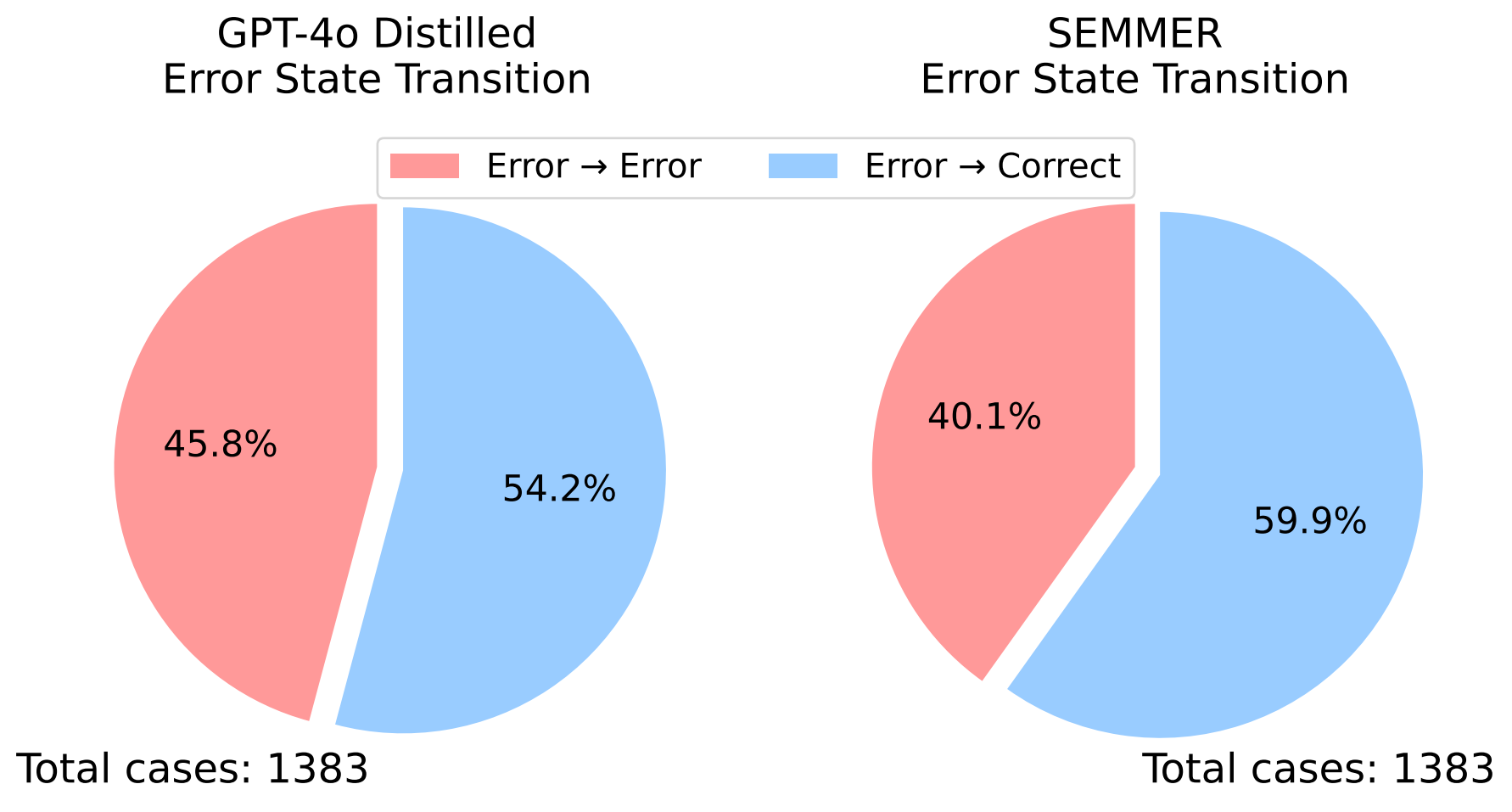
GPT-4o Distilled
Error State Transition
SEMMER
Error State Transition
Error → Error
Error → Correct
45.8%
54.2%
40.1%
59.9%
Total cases: 1383
Total cases: 1383