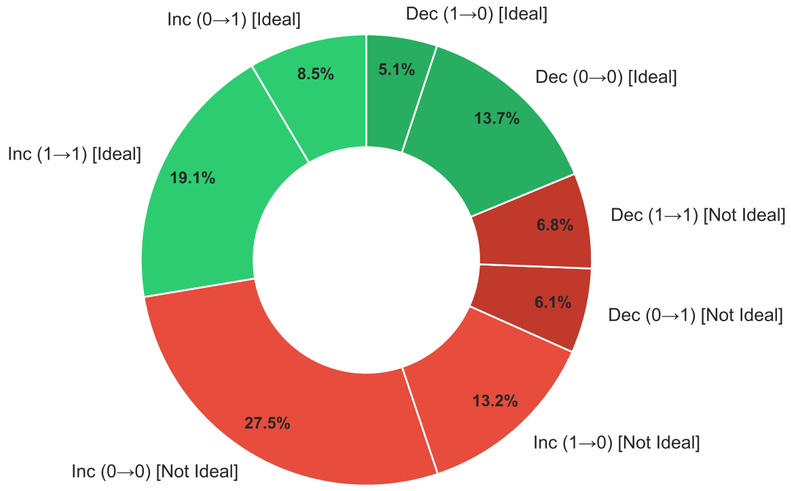
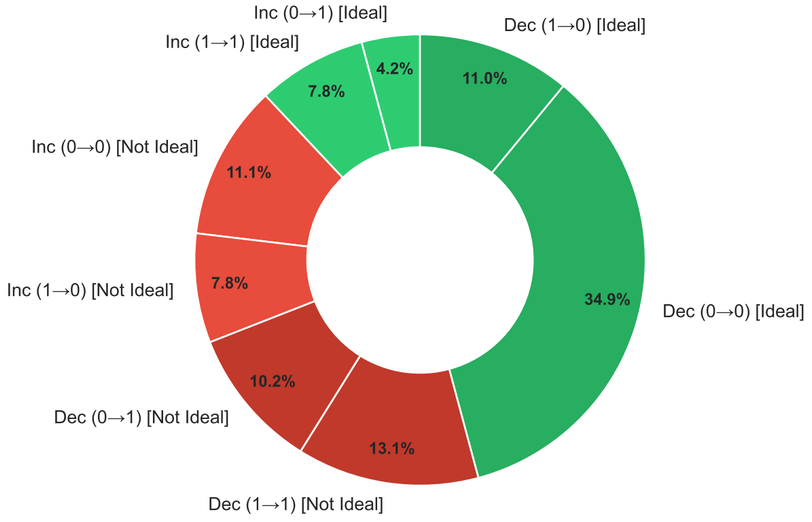
Calibration Changes: 4omini_sc vs 4omini_cot



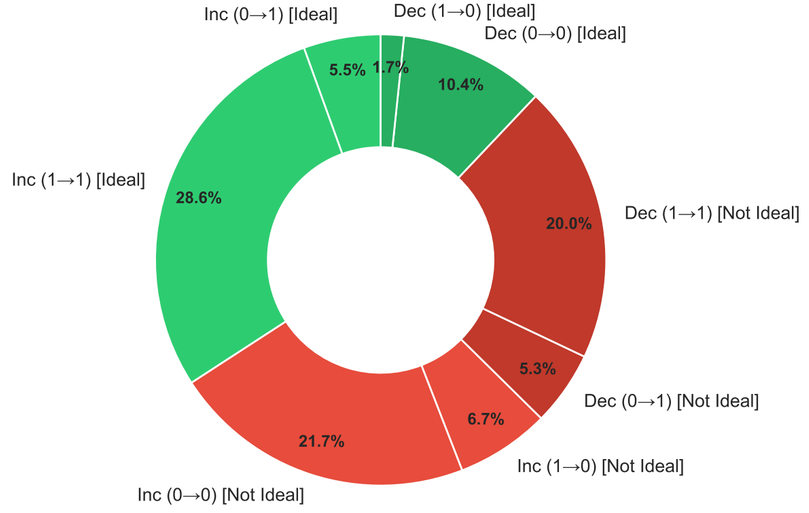
Total changes: 789 cases

Calibration Changes: 4omini_cot vs 4omini_noexp



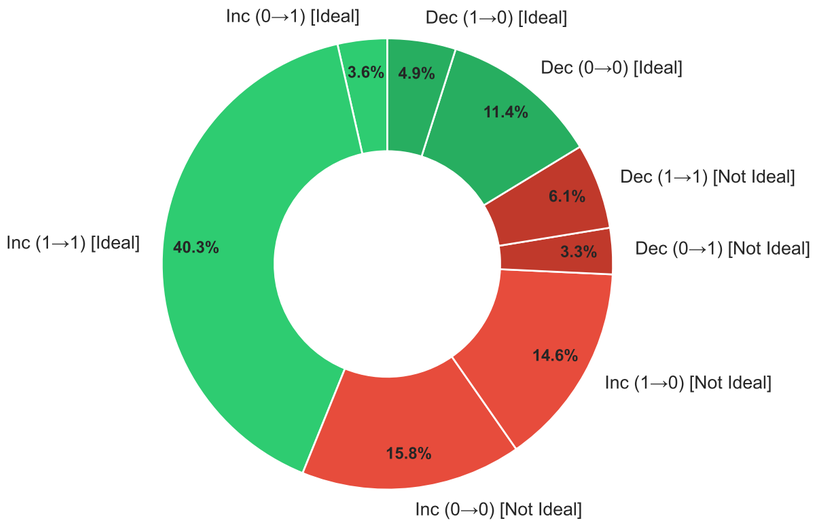
Total changes: 766 cases

Calibration Changes: grok2_sc vs grok2_cot



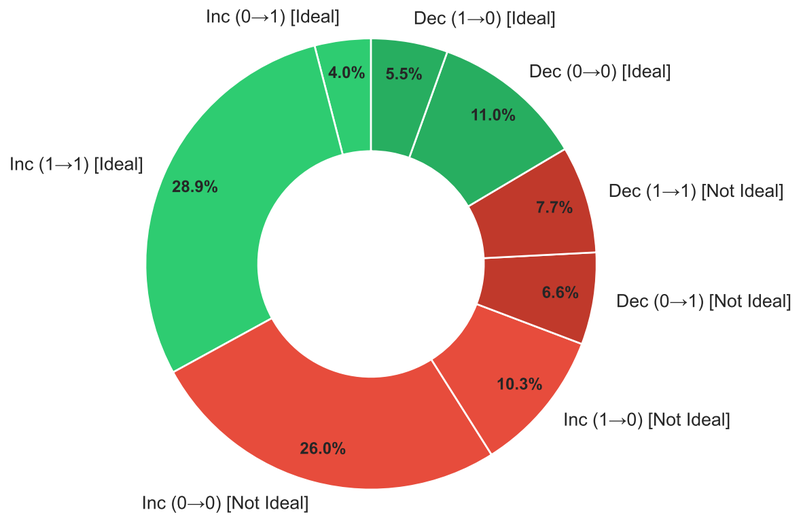
Total changes: 1086 cases

Calibration Changes: grok2_cot vs grok2_noexp



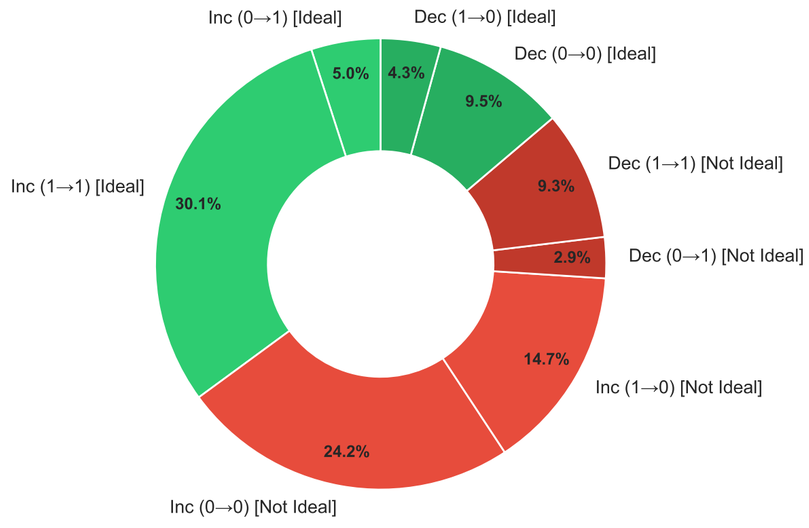
Total changes: 816 cases

Calibration Changes: o1_reason vs o1_mini



Total changes: 273 cases

Calibration Changes: o3mini_medium vs o3mini_low



Total changes: 442 cases