

| aggregation | condition | n_trials | accuracy | hit_rate | fa_rate | dprime | criterion | who |
|-------------------|-----------|----------|----------|----------|---------|--------|-----------|--------------------------------------|
| majority | 100_0 | 1000 | 0.89 | 0.952 | 0.172 | 2.611 | -0.359 | nan |
| maj_pos_exception | 100_0 | 1000 | 0.868 | 0.974 | 0.238 | 2.656 | -0.615 | nan |
| maj_neg_exception | 100_0 | 1000 | 0.893 | 0.948 | 0.162 | 2.612 | -0.32 | nan |
| majority | 50_50 | 1000 | 0.841 | 0.922 | 0.24 | 2.125 | -0.356 | nan |
| maj_pos_exception | 50_50 | 1000 | 0.846 | 0.95 | 0.258 | 2.294 | -0.498 | nan |
| maj_neg_exception | 50_50 | 1000 | 0.86 | 0.888 | 0.168 | 2.178 | -0.127 | nan |
| majority | 80_20 | 1000 | 0.847 | 0.934 | 0.24 | 2.213 | -0.4 | nan |
| maj_pos_exception | 80_20 | 1000 | 0.826 | 0.958 | 0.306 | 2.235 | -0.61 | nan |
| maj_neg_exception | 80_20 | 1000 | 0.866 | 0.92 | 0.188 | 2.29 | -0.26 | nan |
| best_individual | 100_0 | 8000 | 0.89 | 0.928 | 0.148 | 2.506 | 0.3205 | responses_gemini-2.5-pro |
| best_individual | 50_50 | 8000 | 0.865 | 0.916 | 0.186 | 2.271 | -0.243 | gemini-2.5-pro |
| best_individual | 80_20 | 8000 | 0.847 | 0.834 | 0.14 | 0.355 | 0.057 | responses_claude-0.5-sonnet-20250219 |