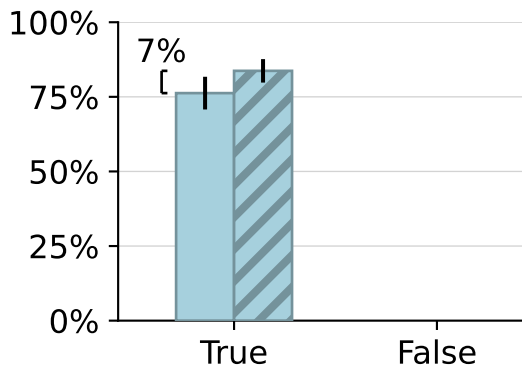
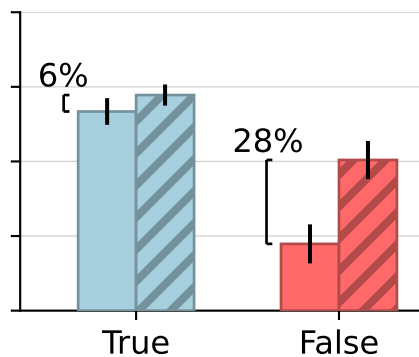


(a)

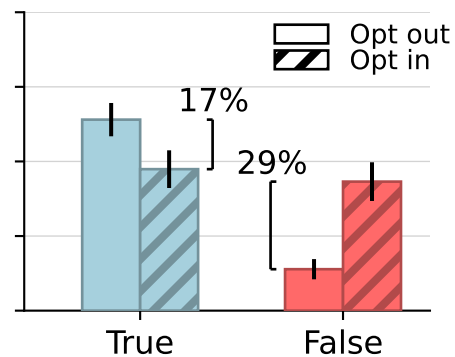
LLM judged True



LLM judged Unsure

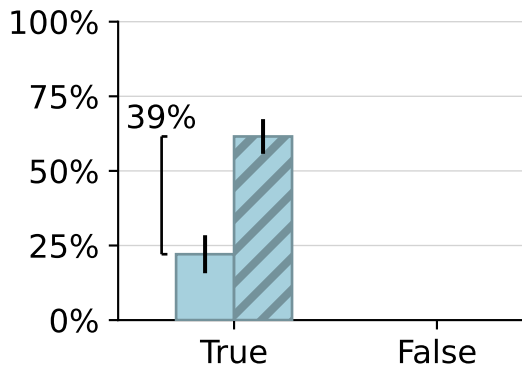


LLM judged False

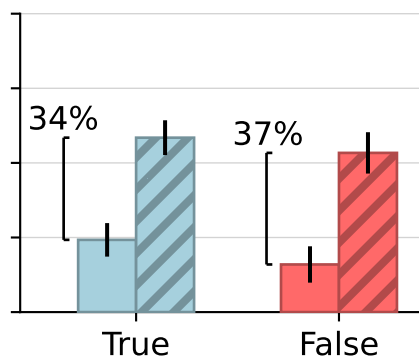


(b)

LLM judged True



LLM judged Unsure



LLM judged False

