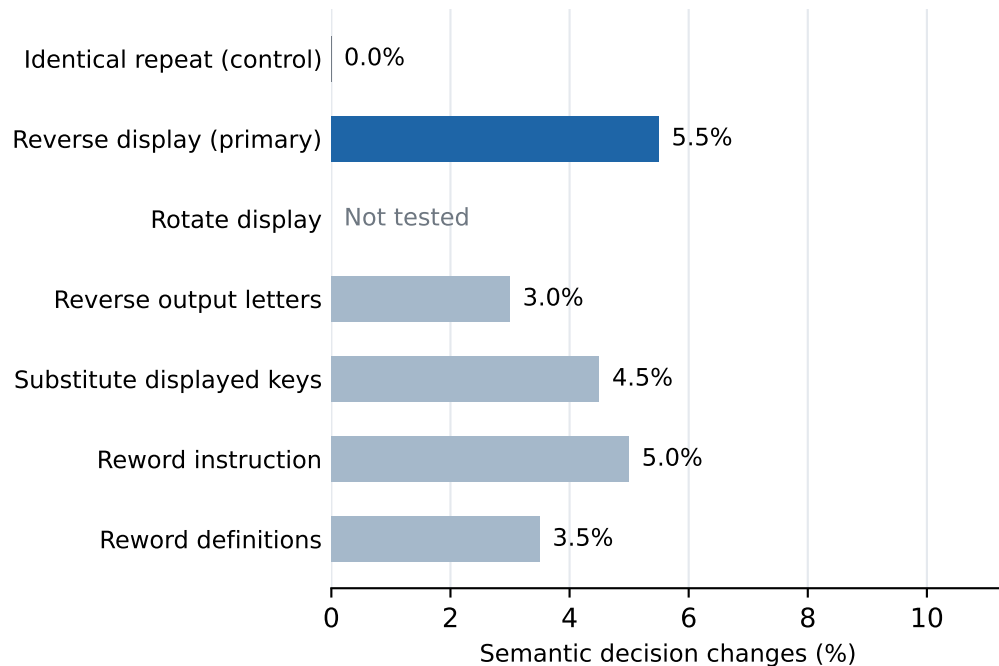


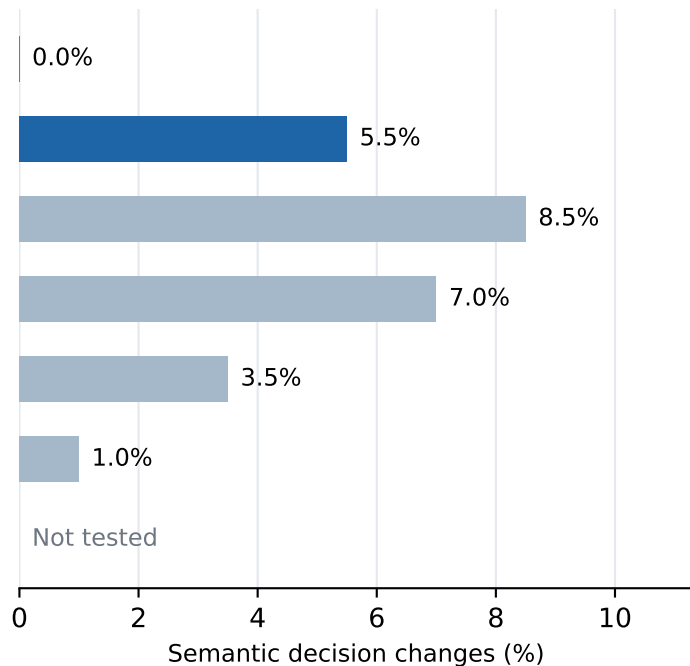
Mini Jev: sensitivity to fixed prompt changes

Each condition is compared with the original prompt on the same 200 items per dataset.

JCoLA: grammatical acceptability



JCQA: common-sense answers



JSTS: semantic similarity

