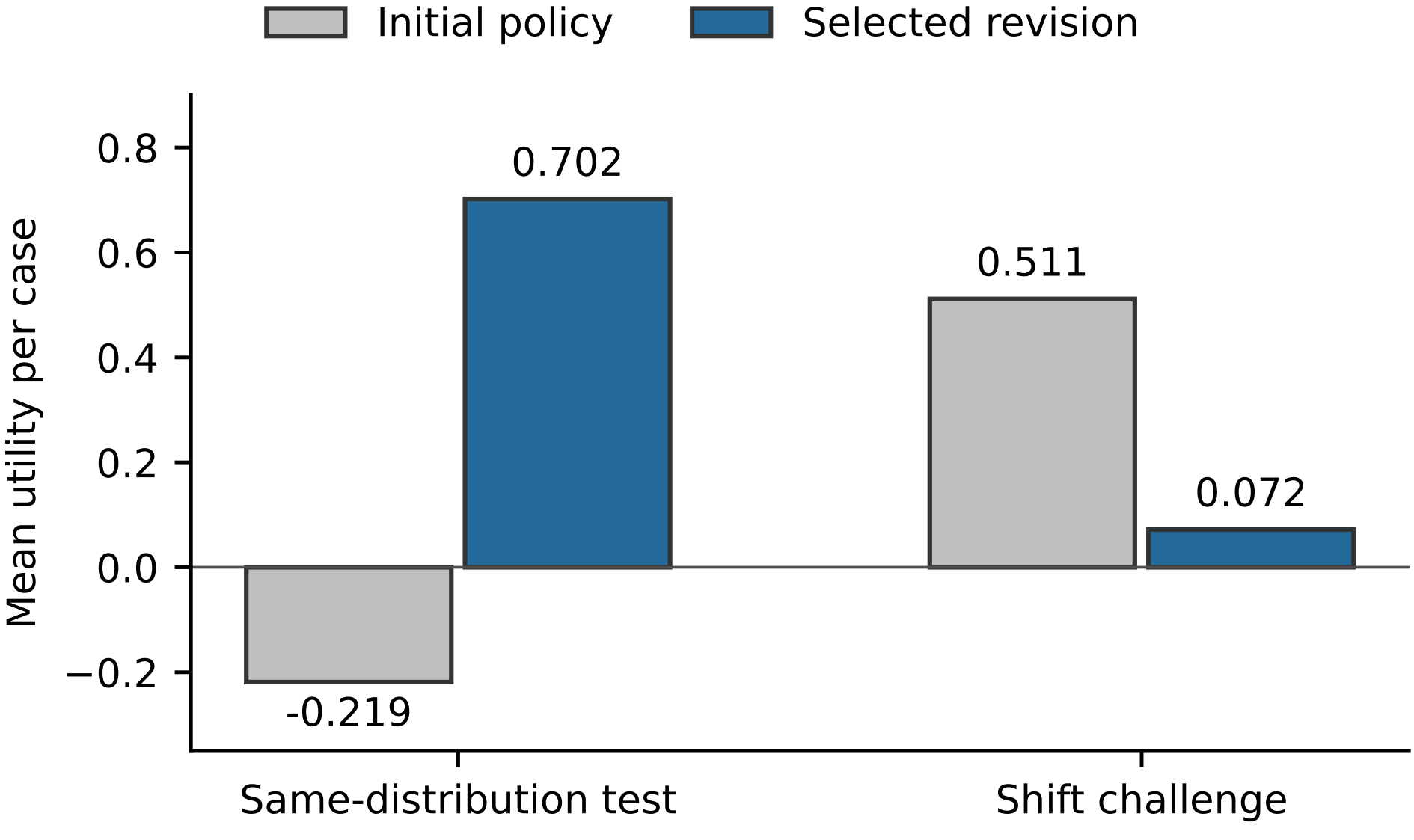

Initial policy
Selected revision
Mean utility per case
0.8
0.6
0.4
0.2
0.0
−0.2
0.702
0.511
0.072
-0.219
Same-distribution test
Shift challenge