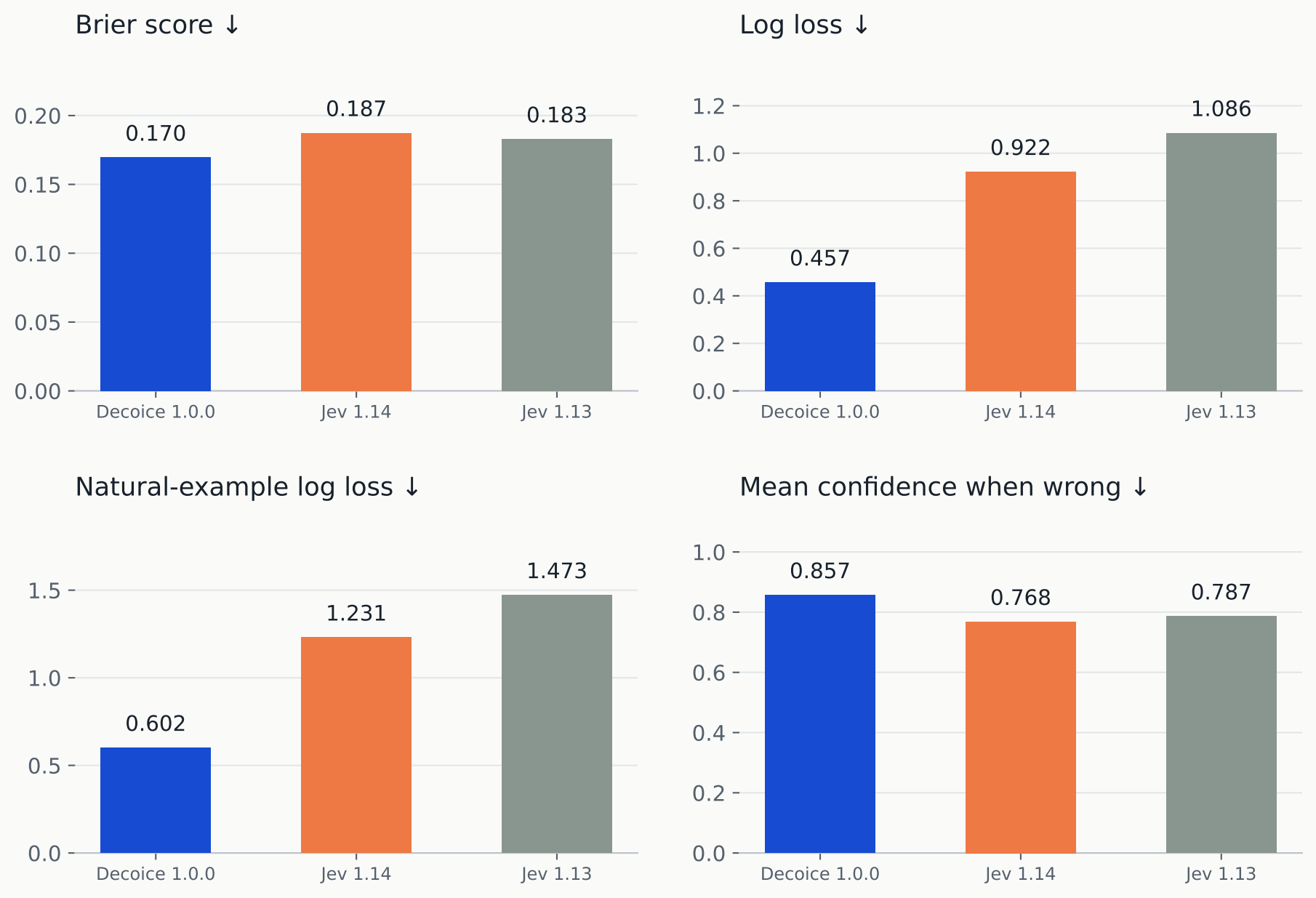

Brier score ↓
0.20
0.15
0.10
0.05
0.00
0.170
0.187
0.183
Decoice 1.0.0
Jev 1.14
Jev 1.13
Log loss ↓
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.457
0.922
1.086
Decoice 1.0.0
Jev 1.14
Jev 1.13
Natural-example log loss ↓
1.5
1.0
0.5
0.0
0.602
1.231
1.473
Decoice 1.0.0
Jev 1.14
Jev 1.13
Mean confidence when wrong ↓
1.0
0.8
0.6
0.4
0.2
0.0
0.857
0.768
0.787
Decoice 1.0.0
Jev 1.14
Jev 1.13