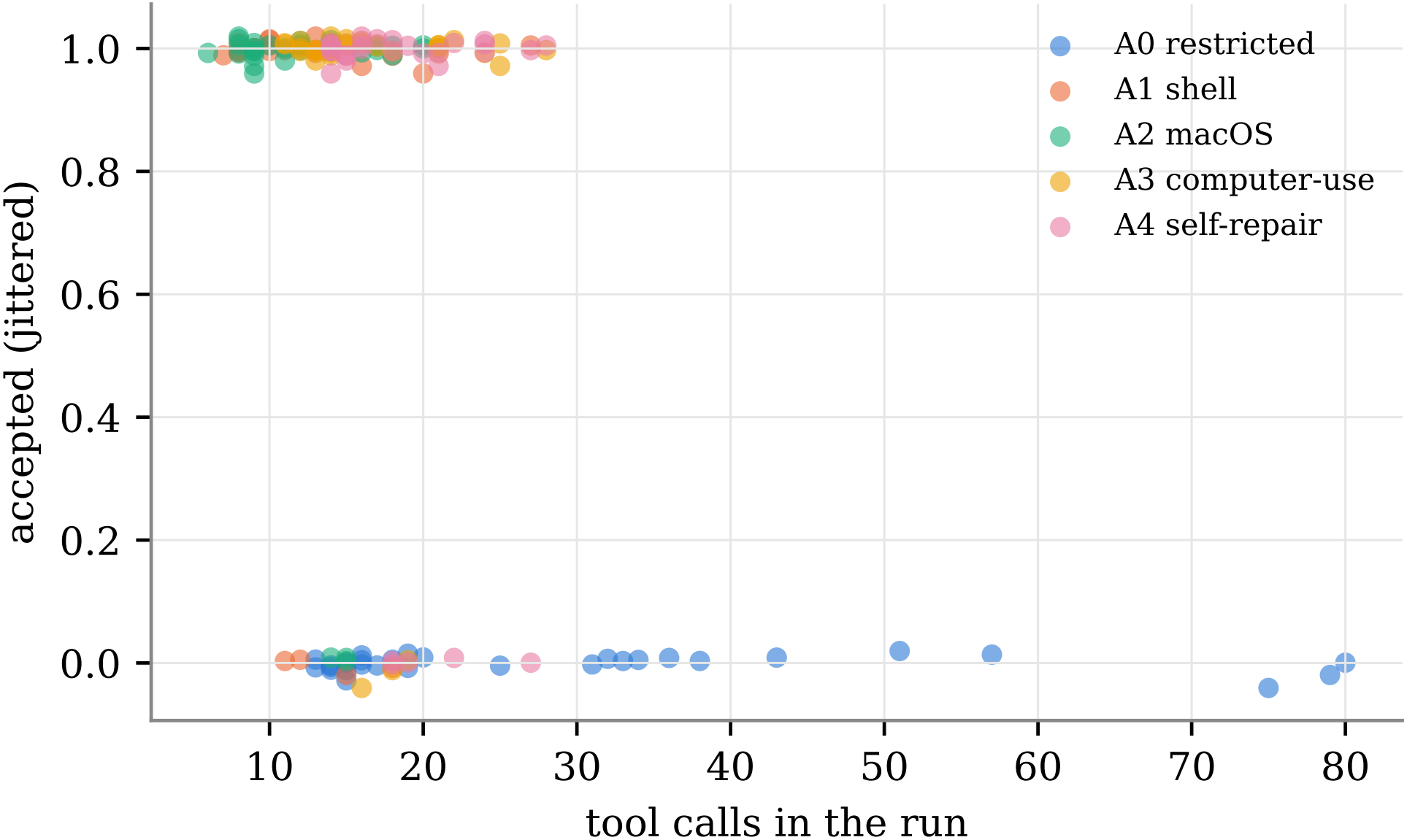

A0 restricted
A1 shell
A2 macOS
A3 computer-use
A4 self-repair
accepted (jittered)
1.0
0.8
0.6
0.4
0.2
0.0
10
20
30
40
50
60
70
80
tool calls in the run