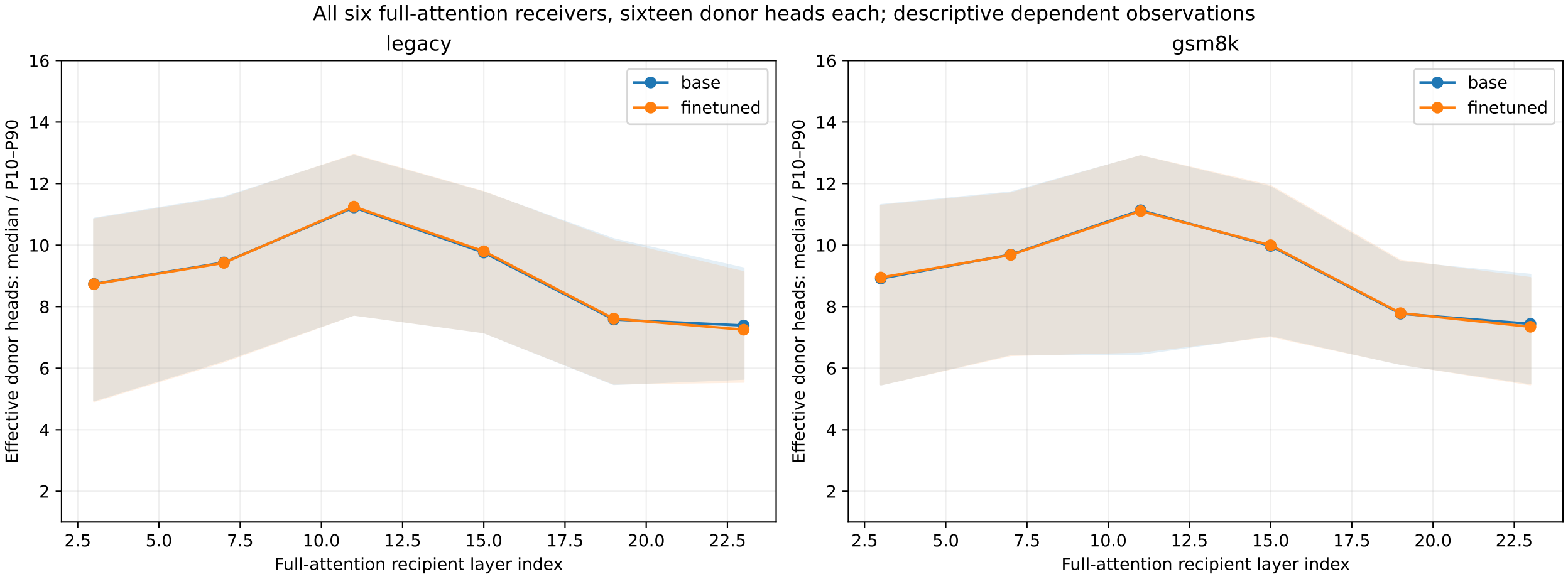

All six full-attention receivers, sixteen donor heads each; descriptive dependent observations
legacy
gsm8k
base
finetuned
Effective donor heads: median / P10–P90
Full-attention recipient layer index
2
4
6
8
10
12
14
16
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5