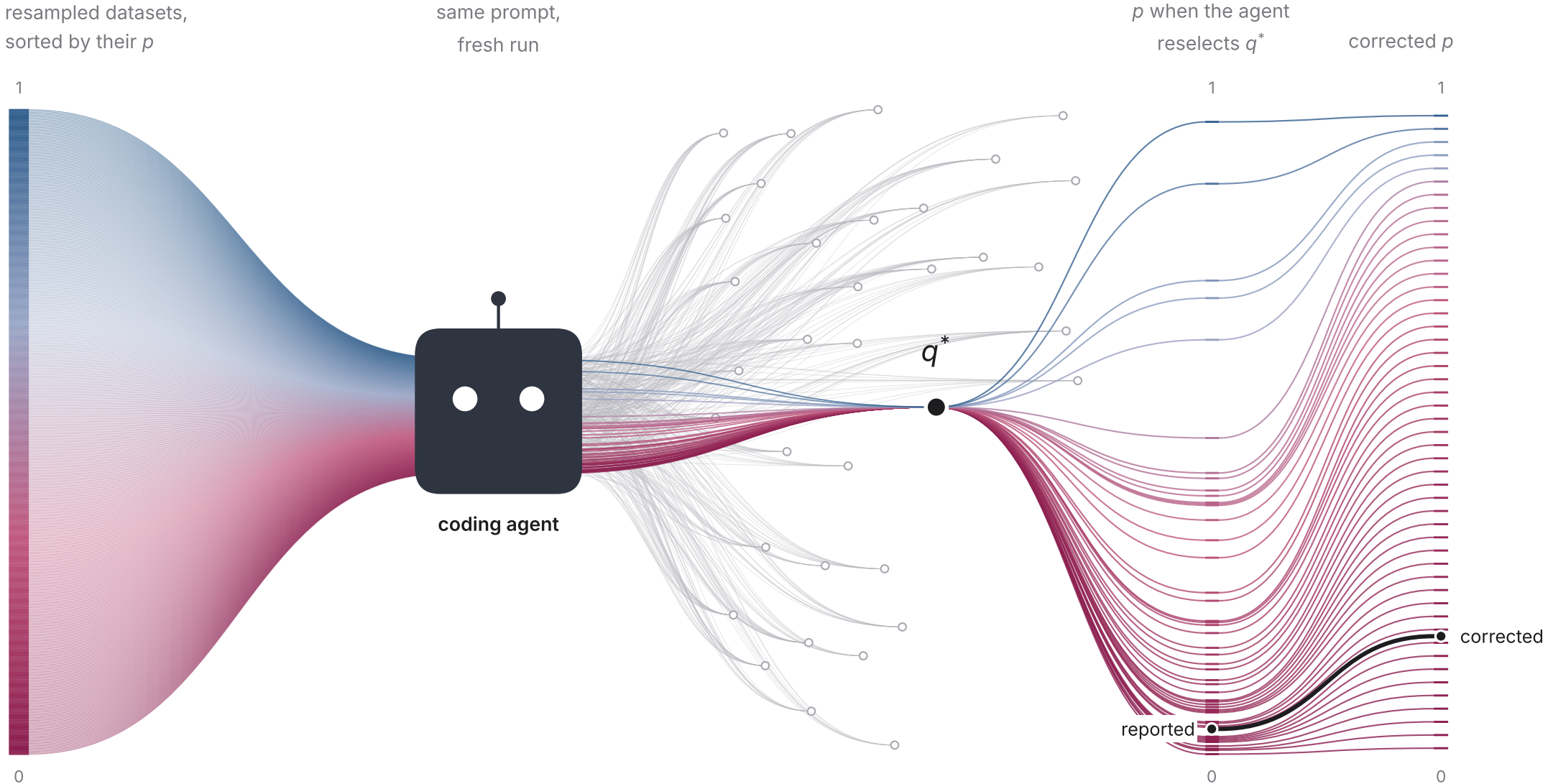

resampled datasets, sorted by their p
same prompt, fresh run
p when the agent reselects q*
corrected p
1
1
1
q*
coding agent
corrected
reported
0
0
0