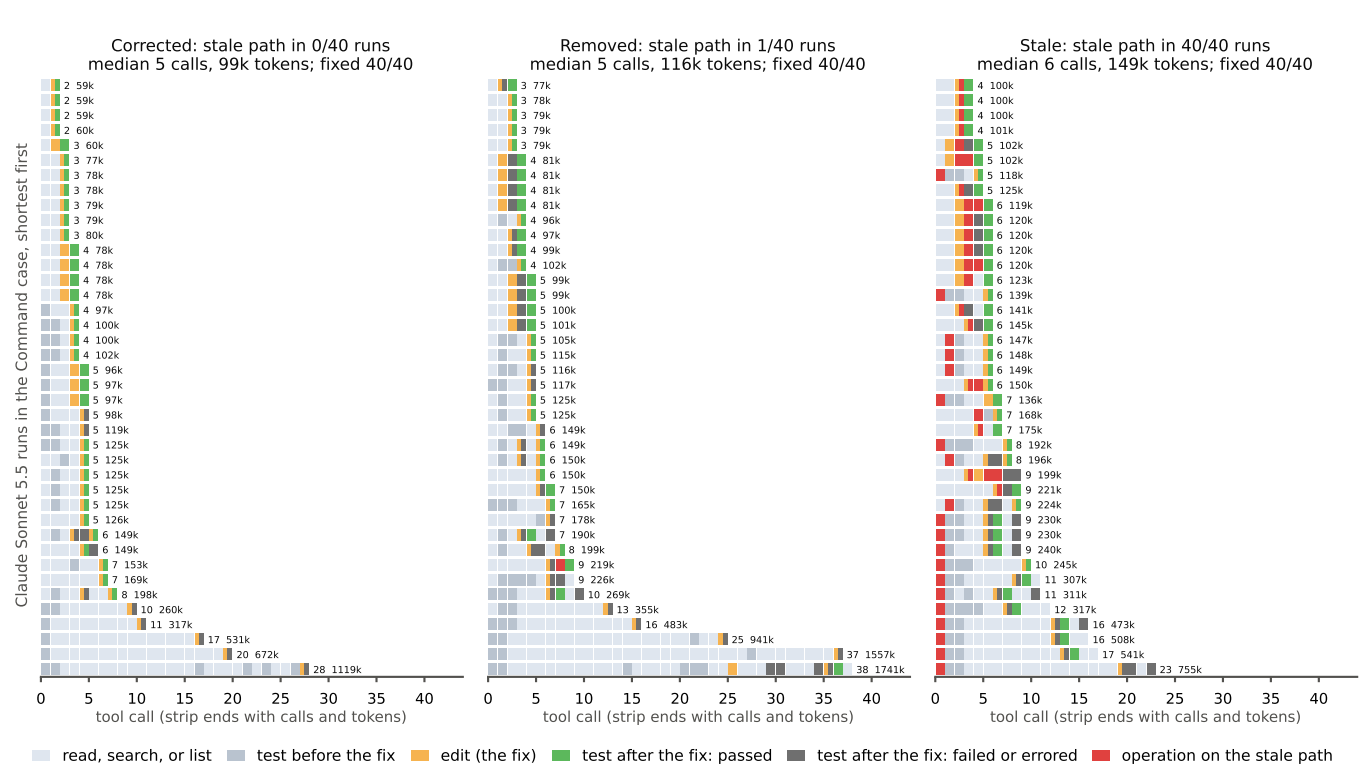

Corrected: stale path in 0/40 runs
median 5 calls, 99k tokens; fixed 40/40
Removed: stale path in 1/40 runs
median 5 calls, 116k tokens; fixed 40/40
Stale: stale path in 40/40 runs
median 6 calls, 149k tokens; fixed 40/40
Claude Sonnet 5.5 runs in the Command case, shortest first
2 59k
2 59k
2 59k
2 60k
3 60k
3 77k
3 78k
3 78k
3 79k
3 79k
3 80k
4 78k
4 78k
4 78k
4 78k
4 97k
4 100k
4 100k
4 102k
5 96k
5 97k
5 97k
5 98k
5 119k
5 125k
5 125k
5 125k
5 125k
5 125k
5 126k
6 149k
6 149k
7 153k
7 169k
8 198k
10 260k
11 317k
17 531k
20 672k
28 1119k
3 77k
3 78k
3 79k
3 79k
3 79k
4 81k
4 81k
4 81k
4 81k
4 96k
4 97k
4 99k
4 102k
5 99k
5 99k
5 100k
5 101k
5 105k
5 115k
5 116k
5 117k
5 125k
5 125k
6 149k
6 149k
6 150k
6 150k
7 150k
7 165k
7 178k
7 190k
8 199k
9 219k
9 226k
10 269k
13 355k
16 483k
25 941k
37 1557k
38 1741k
4 100k
4 100k
4 100k
4 101k
5 102k
5 102k
5 118k
5 125k
6 119k
6 120k
6 120k
6 120k
6 120k
6 123k
6 139k
6 141k
6 145k
6 147k
6 148k
6 149k
6 150k
7 136k
7 168k
7 175k
8 192k
8 196k
9 199k
9 221k
9 224k
9 230k
9 230k
9 240k
10 245k
11 307k
11 311k
12 317k
16 473k
16 508k
17 541k
23 755k
0
5
10
15
20
25
30
35
40
tool call (strip ends with calls and tokens)
read, search, or list
test before the fix
edit (the fix)
test after the fix: passed
test after the fix: failed or errored
operation on the stale path