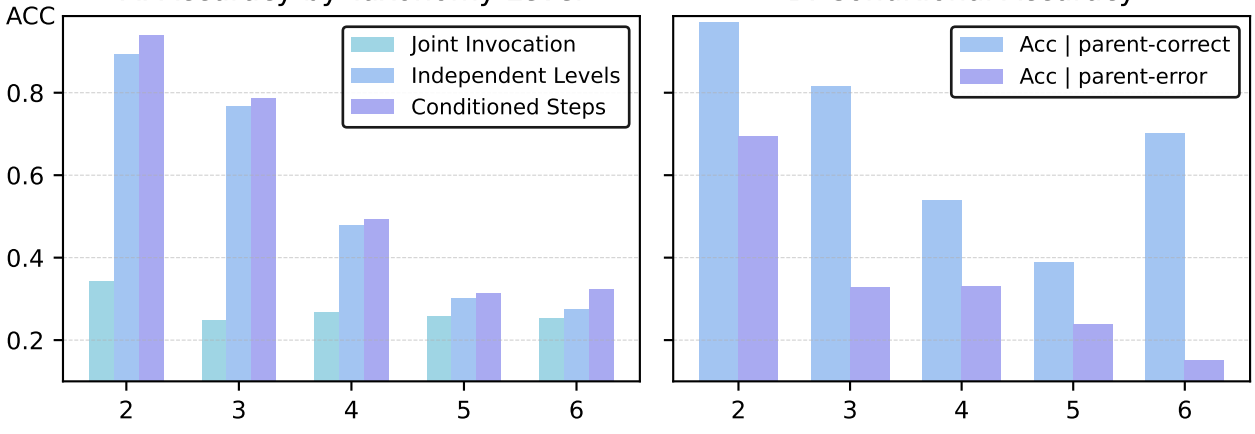
A: Accuracy by Taxonomy Level
ACC
0.8
0.6
0.4
0.2
Joint Invocation
Independent Levels
Conditioned Steps
2
3
4
5
6
B: Conditional Accuracy
Acc | parent-correct
Acc | parent-error
2
3
4
5
6