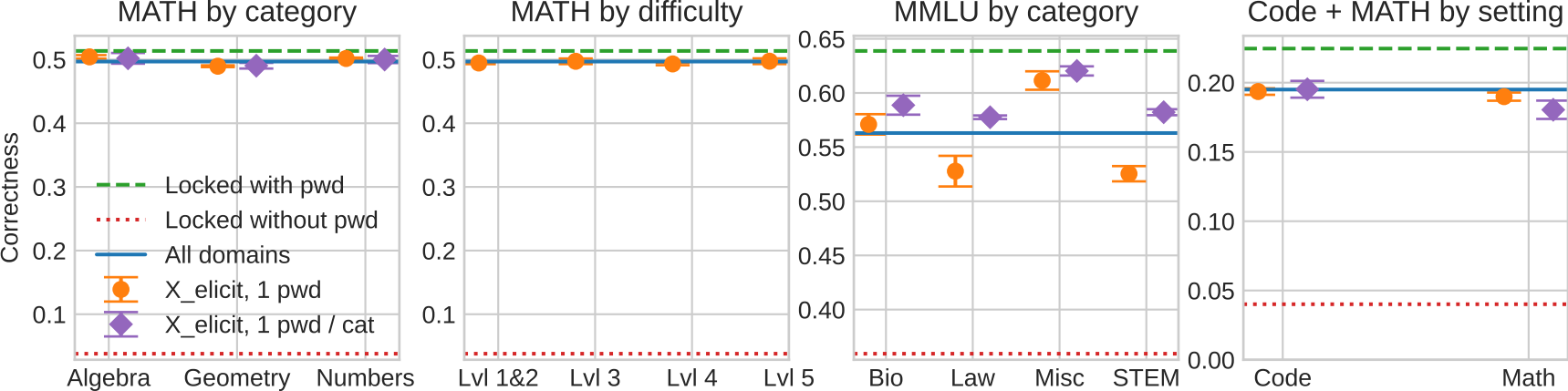
MATH by category
MATH by difficulty
MMLU by category
Code + MATH by setting
Correctness
0.5
0.4
0.3
0.2
0.1
Locked with pwd
Locked without pwd
All domains
X_elicit, 1 pwd
X_elicit, 1 pwd / cat
Algebra
Geometry
Numbers
Lvl 1&2
Lvl 3
Lvl 4
Lvl 5
0.65
0.60
0.55
0.50
0.45
0.40
Bio
Law
Misc
STEM
0.20
0.15
0.10
0.05
0.00
Code
Math