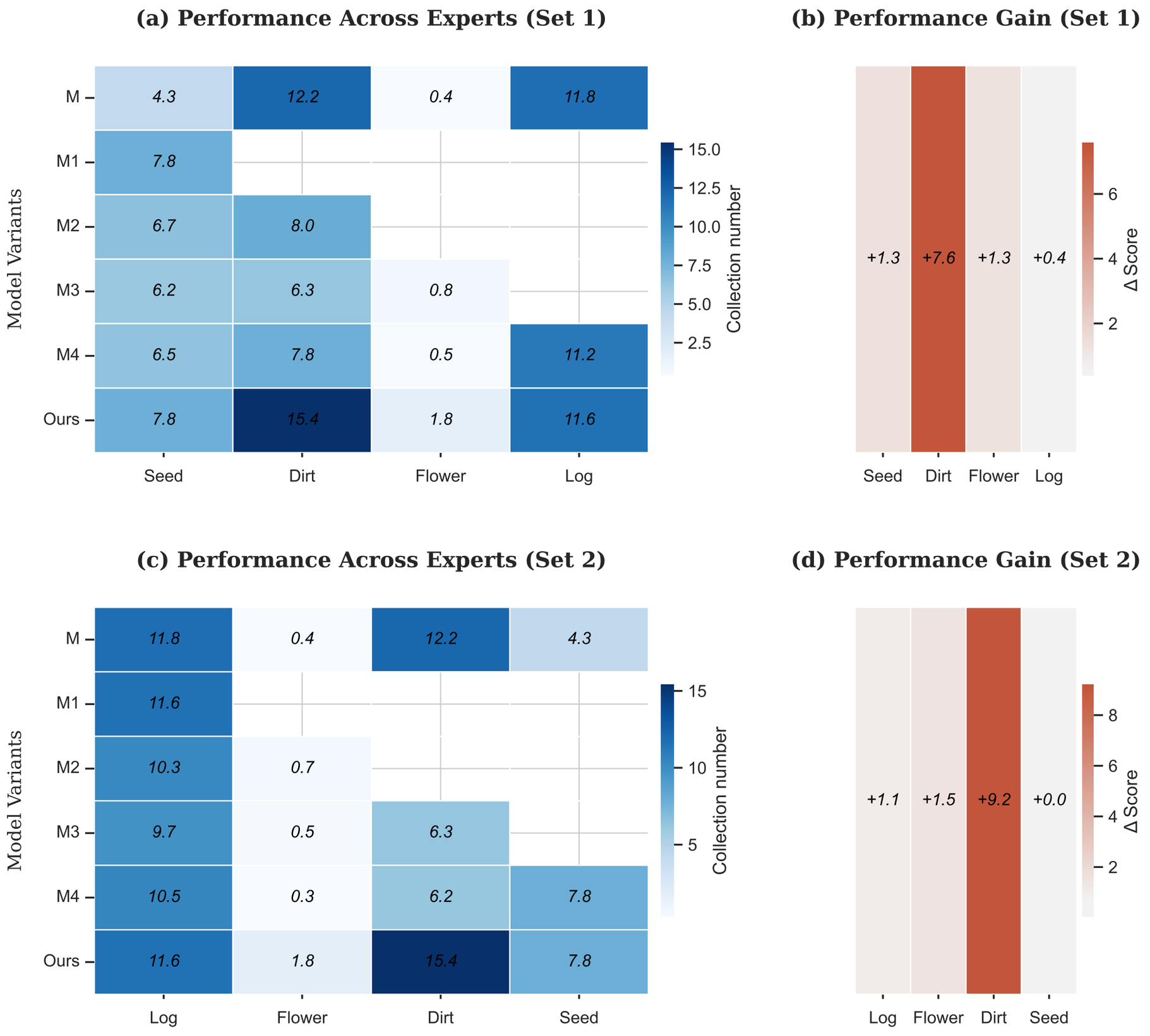
(a) Performance Across Experts (Set 1)
(b) Performance Gain (Set 1)
(c) Performance Across Experts (Set 2)
(d) Performance Gain (Set 2)
Model Variants
M
M1
M2
M3
M4
Ours
Seed
Dirt
Flower
Log
4.3
12.2
0.4
11.8
7.8
6.7
8.0
6.2
6.3
0.8
6.5
7.8
0.5
11.2
7.8
15.4
1.8
11.6
Collection number
15.0
12.5
10.0
7.5
5.0
2.5
+1.3
+7.6
+1.3
+0.4
Δ Score
6
4
2
11.8
0.4
12.2
4.3
11.6
10.3
0.7
9.7
0.5
6.3
10.5
0.3
6.2
7.8
11.6
1.8
15.4
7.8
Log
Flower
Dirt
Seed
15
10
5
+1.1
+1.5
+9.2
+0.0
8
6
4
2