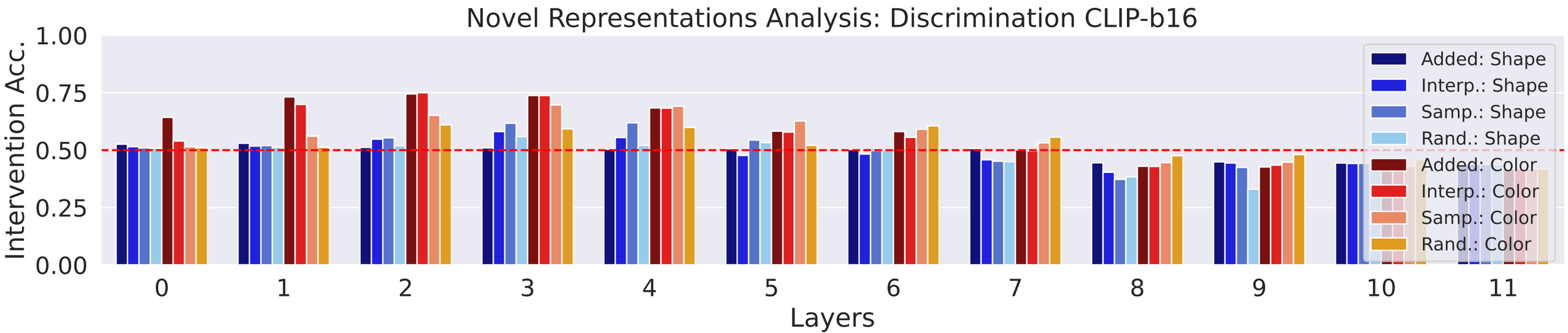
Novel Representations Analysis: Discrimination CLIP-b16
Intervention Acc.
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
6
7
8
9
10
11
Layers
Added: Shape
Interp.: Shape
Samp.: Shape
Rand.: Shape
Added: Color
Interp.: Color
Samp.: Color
Rand.: Color