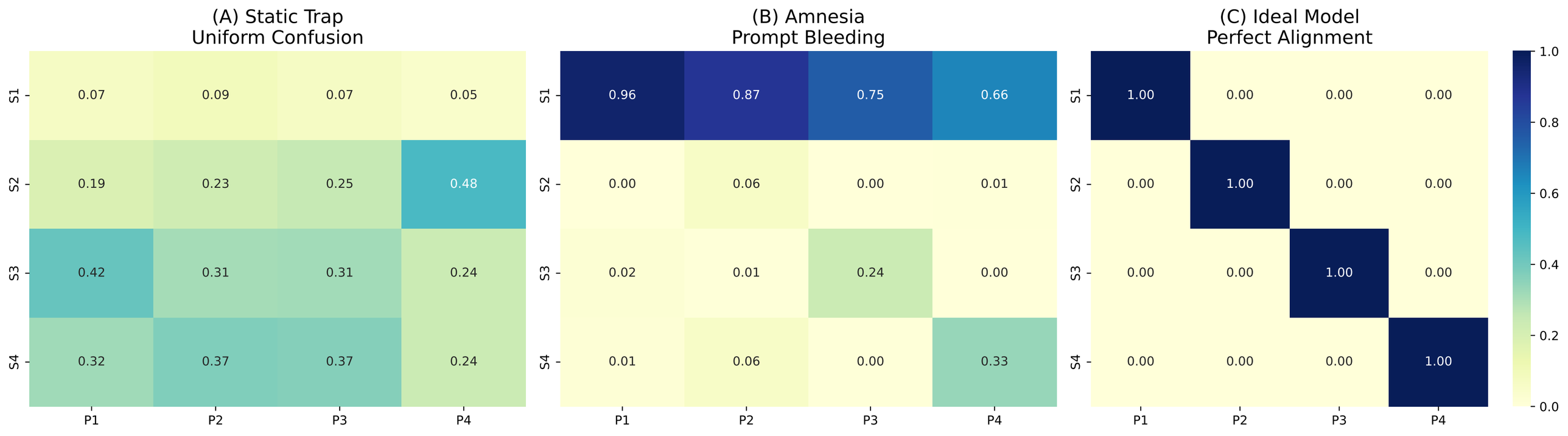
(A) Static Trap
Uniform Confusion
S1
S2
S3
S4
0.07
0.09
0.07
0.05
0.19
0.23
0.25
0.48
0.42
0.31
0.31
0.24
0.32
0.37
0.37
0.24
P1
P2
P3
P4
(B) Amnesia
Prompt Bleeding
S1
S2
S3
S4
0.96
0.87
0.75
0.66
0.00
0.06
0.00
0.01
0.02
0.01
0.24
0.00
0.01
0.06
0.00
0.33
P1
P2
P3
P4
(C) Ideal Model
Perfect Alignment
S1
S2
S3
S4
1.00
0.00
0.00
0.00
0.00
1.00
0.00
0.00
0.00
0.00
1.00
0.00
0.00
0.00
0.00
1.00
P1
P2
P3
P4
1.0
0.8
0.6
0.4
0.2
0.0