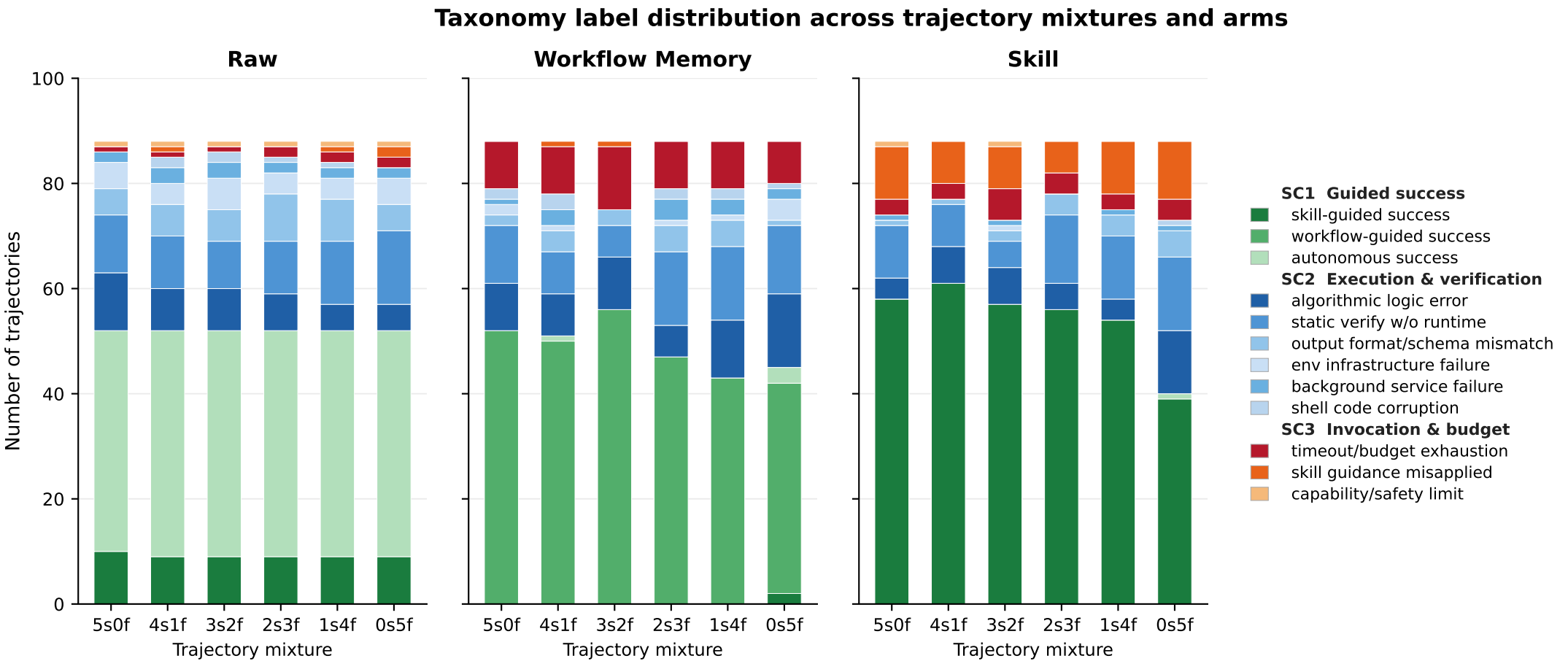
Taxonomy label distribution across trajectory mixtures and arms
Raw
Workflow Memory
Skill
Number of trajectories
0
20
40
60
80
100
5s0f
4s1f
3s2f
2s3f
1s4f
0s5f
Trajectory mixture
SC1 Guided success
skill-guided success
workflow-guided success
autonomous success
SC2 Execution & verification
algorithmic logic error
static verify w/o runtime
output format/schema mismatch
env infrastructure failure
background service failure
shell code corruption
SC3 Invocation & budget
timeout/budget exhaustion
skill guidance misapplied
capability/safety limit