Policy Evolution During Training

Legend:
- β=0.0, P(A)
- β=5.0, P(A)
- 0.5 ref

X-axis: Training Epoch
Y-axis: P(A) for Agent 0