RL - APPLICATIONS

TD - GAMMON: $TD(\lambda)$ + NONLINEAR FCN VIA ANN w/ BACKPROP.

SAMUEL'S CHECKERS PLAYER: VERY OLD ('50s). FIRST EFFECTIVE HEURISTIC SEARCH | TD, MINIMAX.

ACROBOT: ROBOT - CONTROL TASK; CONTINUOUS CONTROL VARIABLES. SARSA (λ), LINEAR FCN APPROX, TILE CODING. REPLACING TRACES.

ELEVATOR - DISPATCHING: Q - LEARNING WITH TRICKS. 1 STEP AHEAD. PRIOR - KNOWLEDGE CONSTRAINTS.

DYNAMIC CHANNEL ALLOCATION: FOR CELLULAR NETWORKS. SEVERAL ATTEMPTS

JOB - SHOP SCHEDULING: TIME AND RESOURCE CONSTRAINTS → OPTIMIZE SCHEDULE → **PLAN-SPACE SEARCH!** STATES ARE COMPLETE PLANS

RL PROSPECTS

I. ESTIMATION OF VALUE FUNCTIONS

II. BACKING UP VALUES ALONG ACTUAL OR SIMULATED STATE TRAJECTORIES

III. GENERALIZED POLICY ITERATION. GPI $\hat{\pi} \rightleftarrows \hat{v}$

- DIMENSIONS (MAJOR)
  - WIDTH OF BACKUPS (SAMPLE VS FULL)
  - DEPTH OF BACKUPS (BOOTSTRAPPING)
  - FUNCTION APPROXIMATION (TABULAR — LINEAR — NONLINEAR)
  - ON - POLICY — OFF - POLICY

- MINOR DIMENSIONS
  - EPISODIC | CONTINUING VS DISCOUNTED | UNDISCOUNTED RETURNS
  - **VALUES**: ACTION, STATE, AFTERSTATE
  - **ACTION SELECTION** EXPLORATION VS EXPLOITATION
  - **BACKUPS** SYNCHRONOUS | ASYNCHRONOUS
  - **TRACES** NO, ACCUMULATING, REPLACING, DUTCH
  - **XP** REAL, SIMULATED
  - **BACKUPS:** WHAT $S, (S, A)$ SHOULD BE BACKED UP. ACTUALLY ENCOUNTERED OR NOT.
  - **BACKUPS:** TIMING. WHEN ACTION SELECTED OR AFTER
  - **BACKUPS:** MEMORY. PERMANENT RETAINMENT OR ONLY WHEN SELECTING ACTION

- MOVING BEYOND STATE REPRESENTATIONS WITH MARKOV PROPERTY
  - → **POMDP**, HIDDEN STATES. CAN DO BAYESIAN INFERENCE
    - RELIES ON MODEL, COMPUTATIONALLY EXPENSIVE
  - → BETTER REPRESENTATIONS
  - → MODULARITY, HIERARCHY