Reinforcement Learning Cheat Sheet (TkD vs TkLL)

# 1. Core Concepts

|  |  |
| --- | --- |
| Term | Meaning |
| Choice | Which image the monkey chose on each trial |
| Outcome / Reward | The token result of that choice (e.g., -1, +1) |
| Q-value | The learned value of an option (updated each trial) |
| α (alpha) | Learning rate → how quickly the monkey updates beliefs |
| β (beta) | Inverse temperature → how consistently they choose the better option |
| LL (log-likelihood) | How well the model explains the actual choices |
| AIC/BIC | LL with penalties for model complexity (lower = better) |

# 2. Models I’ve Fit

## Model 1: Random Choice

- Picks randomly (like flipping a coin)  
- No parameters  
- Log-likelihood:  
 LL = -N \* log(0.5)

## Model 2: WSLS (Win-Stay-Lose-Shift)

- Repeats choice after win, switches after loss  
- Simulated logic:  
 if outcome(t-1) > 0: pred\_choice = choice(t-1)  
 else: pred\_choice = random other option  
- LL calculated as:  
 LL = sum(log(0.5 + 0.5 \* (pred == actual)))

## Model 3: Q-Learning

- Tracks value (Q) for each option  
- Updates chosen option using:  
 Q(chosen) ← Q(chosen) + α \* (reward - Q(chosen))  
- Converts Q-values into choice probabilities using softmax:  
 P(choice k) = exp(β \* Q(k)) / sum(exp(β \* Q(all)))  
- Fitted using fmincon to minimize negative log-likelihood

# 3. Model Comparison

Compare fits using:  
- LL (log-likelihood): higher is better  
- AIC = 2\*k - 2\*LL  
- BIC = k\*log(N) - 2\*LL  
Lower AIC/BIC = better

# 4. What to Plot

|  |  |  |
| --- | --- | --- |
| Plot | Shows | Why It’s Good |
| Learning Curves | % correct over trials | Shows actual behavior |
| Model Comparison (LL/AIC/BIC) | Bar plots for 3 models | Show which model fits best |
| α / β Boxplots | Parameter estimates across tasks | Show motivational or learning differences |
| Simulated vs Real | Recovery of learning curves | Show model captures structure |
| Sim vs Fit α/β Scatter | Recovery accuracy | Show the model is valid |

# 5. Interpretation Help

- Low α = slow learning  
- High α = fast learning  
- Low β = noisy/random choice  
- High β = consistent exploitation  
- Q-learning > WSLS = tracking expected value  
- TkD lower β = noisier behavior in all-loss condition

# 6. Good Poster Sentences

- “We used RL modeling to infer internal learning dynamics.”  
- “Q-learning explained behavior better than simple win-stay rules.”  
- “TkD monkeys showed lower β, suggesting noisier decision strategies under all-loss conditions.”  
- “Simulated learning curves recovered from fit parameters, validating model interpretability.”