perhaps you can say something about

\[ S_i \xrightarrow{\text{(*)}} S_j \]  

length shortest path \( \leq \text{man} \)

two random nodes \( S_i \) \& \( S_j \) and see #

i.e. make on a graph -

suffices a recursive condition #

for mixing ?

\( O(\log N) \)

to rule out learnable e.g. for human -

you can see #\( \text{MnIke} \) #\( \text{Bernhs} \) (#general Theory)

Can we\( \text{prom} \) be said -

e.g. if \( (*) \) \( \text{Ind} \Rightarrow \text{RL of zero f} \)

\( \text{results} \) (as also known)
PROBLEM W. REINFORCEMENT LEARNING

appears to be MISMATCH

Note linear

Tanin between $l_2$ & $l_0$ line

≈ 1 or average "pals" not very close to fairly good

reduce overall

1 SLA or single $l_1$

No grad. No comp. Error

smaller, like $l_1$

must explain why

teaching $l_1$. do N bp. & N

unknown carriage (green) to Carvin & Aln. Dn.
H O W E V E R

What you really want is a policy iteration RL to work on

by updating on $\nabla \log Q$ value estimation $\nabla Q(x)$

So, value-based policy (and) has to have
already a very special structure.

REMEMBER

for value function such

$(\ast \ast)$, L1 & L2 can not be different. i.e. re-ward learning

ASIDE: Stated slightly differently -[Note: Incoherent handwriting]

$(\ast \ast \ast)$ mean it would need L2 learning (i.e. 2 samplings.)
Clear local rendezvous

MDP policy — any line of the below

Can we make it non-fancy / simplify?

advantage of logic over poh —
absolute propagation — long long

down & cabernet, i.e.
big chunks = initialized &

Clean local / clear life.

also sub poh —

while sub poh —

the members

Still logically seem to

Cycle study.