* Sequenz durchlaufen- dann simulieren, dann auswerten + update
* Zeit abschätzen, printf raus->schneller, evtl parsing raus
* Wie macht man Q update?
  + Welcher reward
  + Nur die belohnen die explored haben..