DRA vs Max Ent RL

> **To do**
>
> Re-work Luigi's derivation and compare it with Stein Variational Policy Gradient (SVPG)

- DRA vs Max entropy RL simulations
- one difference:
  - preference for entropy (Ziebart's task)
  - only in the case of no reward
- main difference (when luigi was working on DRA as inference):
  - DRA seems to have more degrees of freedom because σ parameter can be independently set for each (s, a) pair
  - can we design a task where freq & stakes differ for each a within s. What about one where a state has a very large number of actions but the vast majority are crap and only one or two are good? Then we can have four such states where the frequency and stakes differ.

DRA as inference

- when not conditioned on optimality:
  - action prior
  - assumption that people act randomly
    - in absence of optimality conditioning
- Section 2 has cost for accessing memories
- DRA instead has a cost for storing memories