Narrative for the paper

This paper sets a foot in the door for (the broader concept behind) DRA. Although we started thinking of it as DRA vs maxEnt RL, it's not going to be useful for anyone if we do that. That's like saying, "hey, look! This one thing that you don't know about (DRA) is very similar to this other popular ML algorithm that you also don't know about (max-entropy RL)".

It makes much more sense to now focus on the principles, the reasons, and the motives behind DRA. The NeurIPS paper didn't and couldn't do that job really. We need to start with a bold statement:

> - *If the brain has memory constraints, then it should allocate resources to memory preferentially while adhering to the constraints. We need to build a strong argument that we need a framework for this!*
>   - *(Cite previous attempts at doing so*).*
> - *There are several ways to build this, we follow a simple approach with DRA (introduce from scratch) that uses information-theoretic principles.*
> - *DRA is a specific implementation of something that is inevitable. Even if people don't know or agree about about our choices in DRA, something like it must be the case.*

*possibly Nuvan de Silva, Wei ji Ma?

Our take on DRA in the brain

- we need to be honest and state that there are some core ideas about the theory that we believe in. DRA is a toy model that is almost for sure not implemented in the brain, but the principles behind it.
- So this is really a test of principles. And we can do experimental manipulations to test these principles.

Miscellaneous

Justification for using toy tasks

We go with tabular RL because it's a domain we can build easily with. DRA is not limited to that and the principles are not limited to that. We choose this for ease of analysis, etc. We also take this other approach that is SOTA in RL. Discuss how it is intuitively similar and why it makes sense. Then, in order to show the similarities more concretely, we take the simpler tasks and analyze these.

Process vs normative

The process vs normative thread is ok, but it's super old and more of an overrun horse. We can use and point out that analogy once or twice in the paper, but shouldn't follow it as a main thread.

Justification for choosing max-entropy RL

- We're taking not just a random one but a successful approach that RL/ML has been using. If a ML algorithm produces good behaviors, it makes a lot of sense to use it.
- It's still very high level, but there is evidence that the brain uses RL. And the reason why this is valuable is because we're looking at behavior.

Implications for neuroscientists

An important case we need to make throughout the paper:

- we want to link to ongoing SOTA research in neuroscience
- our research is informed and can inform groups that do this kind of experimental work

Why does it matter?

- If your system has some resource limitations, this will exhibit specific patterns in behavior that you need to account for. We're using the precision of memories, which might seem arbitrary, but it is very justifiable. Consider the following:
    - when you do lossy compression, and then uncompress, everything you then do is going to appear random
    - however, lossy compression is not about random losses, but preferential losses. And we make qualitative predictions about the nature of these losses and how they would show up in behavior
    - We can also make the argument that the degree of variability that is added to behavior is actually net positive (max ent RL)

What should you use to model behavior?

- People often use q-learning to model behavior. However, that doesn't take into account memory constraints. In the face of memory constraints, the behavioral pattern itself changes quite a lot.
- we have a normative model that does indeed take into account these constraints. However, it is slightly complicated to implement with a few design choices, etc.
- So we advocate for at least using an entropy term in the objective, or soft q-learning, which modeling animal (and human) behavior. This approach is not only very well established in the machine learning community, but there is now the added benefit of implicitly taking in account memory constraints that animals might face on top of providing a normative explanation for their stochastic behavior.

Miscellaneous points

1. This theory is very close to what's been happening in ML. It's more like an empirical proof that DRA would be successful where max-entropy RL has already shown to be successful.
2. We demonstrate some experimental manipulations to test the underlying principles. This comes with the benefit of interpretability due to the underlying normative framework.