First, we would like to thank the reviewers for the time spent on assessing our work, for their positive feedback, and for their useful comments and suggestions.

Reviewer #1

With a finite action space [...] it should be possible to compute the regularized greedy policy exactly. With a nonlinear parameterization, we do not think it to be possible (the greedy policy depends on the Q-values but also on the previous policy), except if one can afford to remember all past Q-values (see the Eq. l.94).
About m. Yes, $m = 1$ corresponds to VI, and the general scheme to MPI. We'll clarify further. Even if the analysis only holds for $m = 1$ (and its extension to $m > 1$ is not obvious), we think important to provide the abstract scheme for the general case, to cover a wider range of existing algorithms and to ease the connections. Also, we think interesting that the equivalence (stated in Prop. 1) holds in the general case.
Font. We double-checked, and we use the provided Neurips style file. We'll triple-check.
Typos. Thank you, we'll correct them.

Reviewer #2

Limited number of domains. We consider only two domains in the empirical part of the main paper, due to the page limit. However, the reviewer has maybe missed the additional domains we provide in Appx. E.4. In total, we evaluate our algorithms on (1) 100 garnets (random MDPs), (2) 2 gym environments: CartPole and LunarLander, and (3) 3 Atari games: Asterix, Breakout and Seaquest. More would be better, but our finding are consistent across all these domains.
Control as probabilistic inference. We acknowledge that there are connections between entropy/KL-regularization and control as probabilistic inference. The formalism we adopt is rather the one of [19], where the link with probabilistic inference is discussed. That's true that it should be discussed here too, so we'll add a discussion about this in the final version.
Figure 2. Thank you, we will fix that. Notice also that these figures are provided bigger in the appendix (as well as additional visualizations).

Reviewer #3

Discussion after Prop 1 is too loose. This discussion is indeed quite dense (page limit), but it is developed at length in the whole Appx. B.
Connection to AL. We confirm this connection, it is explained in Appx. B.2, l.566-569. If not clear enough here, we will expand the explanation. Shortly, CVI is a reparameterization of MD-MVI (as shown in Appx B.2), and AL is a limiting case of CVI (as the temperature goes to zero), hence the connection.
Related papers. Thank you, we were not aware of these papers. We will make sure to discuss them in the final version.

Reviewer #5

The discussion could be improved by being more clear about the nature of these connections. The discussion in Sec. 3 is indeed dense, but it is expanded at length in the whole Appx. B.
τ is used without definition. We write "For $\tau \geq 0$" as a short way for "For any real number $\tau \geq 0$". We precise that it is a temperature later, when relevant.