R#1: Thanks for stressing the strengths of the paper (a complete theory of FP in MFG and a rich empirical evaluation).We first address the stated weaknesses. **W1: Short presentation of FP**. We'll improve it for the final version by adding formal def of the best response and explaining why an arbitrary policy between $[0, 1]$ is needed for init purpose. **W2: Gap between CTFP and practical algs**: We'll add the following discussion to the paper. We chose to provide an analysis in continuous time because it provides convenient mathematical tools allowing to exhibit state of the art convergence rate. The convergence rate in discrete time is still an open problem even for 2-players games, but would be an interesting research question (there is a known conjecture in $O(1/\sqrt{t})$ [75]). **Detailed comments**: **(1)** We acknowledge that some useful details should be moved from appx to the main text for the sake of clarity. E.g. the computation of the Best Response (BR) and the population distribution (*cf.* Appx) are both used in FP (Alg. 1), which is implemented in two different settings: a model-based and a model-free approach. The model-based uses Backward Induction (BI, Alg. 4) and an exact calculation of the population distribution (Alg. 5). The model-free approach uses Q-learning (Alg. 2) and a sampling-based estimate of the distribution (Alg. 3). As suggested, we will add the update rules of the Q-function of both methods in the main text. We will clarify how the distribution $\hat{\mu}_n^\pi$ (Alg. 3) is used in Alg. 2 by using proper notations. Q-learning and BI approximate the BR against $\bar{\mu}^j$ (mean distribution), which needs to be clarified: we will add a line in Alg. 1 $\bar{\mu}^j = \frac{j-1}{j}\bar{\mu}^{j-1} + \frac{1}{j}\hat{\mu}^j$ (so here, $\hat{\mu}_n^\pi$ and $\hat{\mu}^j$ are the same). In Alg. 2, the μ of the input can be any distribution ($\mu = (\mu_k)_k$) but we use the mean population distribution $\bar{\mu}^j$ (from the previous FP step) in our setting. **(2) Randomization:** As we use $\bar{\mu}^j$ we don't need to select the policy uniformly over previously obtained policies. Also, we already do employ randomized strategies (for the model-free), with ε-greedy exploration parameter set to 0.2 (l.140). Authors of [64] use a softmax to ensure the regularity needed in their proof. To the best of our understanding, Angiuli *et al.* use ε-greedy action because the updates of Q and μ are intertwined, so the exploration/exploitation are mixed. In Alg.2, the Q-learning (with exploration) and the action to update the distribution (with pure exploitation) are separated. Furthermore, the stochasticity of the environment (noise ϵ_n) adds randomization. Note that randomization is not necessary in model-based as the BR and population distribution are computed exactly (which also bridges the gap between model-based and the theory). Adding ε-randomization or a softmax in the distribution update is an interesting direction. **Exploitability:** Please notice that, because it scales with rewards, its absolute value is not meaningful. This quantity is game dependent and hard to re-scale without introducing other issues (dependence on the initial policy if we re-normalize with the initial exploitability for example). But it decreases by a large factor compared with the initial value. **(3)** The problem of error propagation is addressed in [51] (see Eq. 7). However, [51] does not provide any rate for discrete time FP. As opposed to this work, we focused in getting a convergence rate for CTFP without approximations (in a wider set of settings than in [51]). Surprisingly, these rates do not seem to be too off in practice. We also introduce a new theory of common noise for the two practical algs (*c.f.* R#3). **(4)** We will improve on that transition stating that to go from continuous to discrete time we simply replace sums by integral and difference equations by differential equation (inclusion to be precised). The "watershed" region is necessary to make sure the differential equation is defined on a closed set (here $[1, +\infty[$). Without it, we would only be able to define it on $]0, +\infty[$ which is not enough. **(5)** We apologize for the too short Sec.3. We'll rewrite it with elements from appx A. Even if not directly used, we felt that the equation involving π_n^t was important as it is easier to manipulate policies compared to distribution over states. **(6)** Our common noise can be history dependent (i.e., no assumption on it). In the experiment of Sec.6, the common noise is stationary and i.i.d. Common noises affect the transition probability of the distribution, which is then *random* (it is not the case with only idiosyncratic noise).

R#3: We are grateful for the positive comments acknowledging the importance of common noise in MFGs and MARL, and on the fact that our contribution bridges the gap between MFG and tools from algorithmic game theory such as exploitability. **W1: connections with MARL examples:** Actually our numerical examples are strongly motivated by classical examples in the RL literature. For instance, the beach bar process example is a simplified version of the well known Santa Fe bar problem, which has received a strong interest in the MARL community, see e.g. [Farago et al, Fair and Efficient Solutions to the Santa Fe Bar Problem (2002)]. Similarly, the maze is motivated by swarm motion models from the distributed robotics MARL literature. We will stress this point and add references in the revision. **Other works:** Thank you for pointing out these relevant references, that we will cite as well. Note however that, compared with these works, our paper provides a rigorous rate of convergence, and covers the common noise setting. Last but not least, our work is not limited to potential or variational MFGs as we only need the weaker monotonicity assumption.

R#5: **(1)** We strongly disagree about the lack of novelty and incremental nature of our work, and would have appreciated some argument for this harsh comment. We would like to stress that the other two referees have acknowledged the novelty of work (rate of convergence, common noise, etc.). **(2)** The monotonicity assumption is classical in the MFG literature and much weaker than assumptions made in other works (regularity and smallness of the coefficients in [64], potential structure in [84], etc.). Also, R#1 considers these assumptions as mild. **(3)** This is the very principle of the fictitious play to obtain convergence for averaged policies. We would appreciate any reference where it is not the case.