We thank the reviewers for their detailed comments.

General remarks:

1. General setting: the considered setting is motivated by a real-world problem and is extremely realistic in its general form. We believe that while being simple and succinct, our model is an accurate abstraction of that problem and is able to capture the essence (and hardness) of the learning vs communication tradeoff.

Our problem is inspired by web-scale user experiences at relevant internet companies, that (at an abstract level) are optimized by MAB algorithms. Those are typically served from multiple machines distributed geographically across data centers, where within each center there might be dozens of machines utilized in a load-balanced manner. This implies NEAR UNIFORM traffic hitting each machine, and effectively gives rise to a SYNCHRONOUS process. Further, the best option served by machines at each center is indeed the same. It's true that across data centers, the rate of requests as well as the best arm may vary. However, within a center we believe our model is an accurate abstraction of reality.

2.Novelty of our approach: we note that our results (specifically the upper & lower bounds in the 1-round case) do not rely on the analysis of 1-player algorithms, and our main novelty is mostly orthogonal to previous work in the field. Further, as we explain below (and as demonstrated by our lower bound), the considered problem is definitely non-trivial and requires a fair amount of sophistication. This was perhaps missing from our technical intro, and will be included there.

Further replies:

Reviewer\_21:

A) As explained above, the assumption that each player gets an equal share of the budget is very realistic in the presence of some load-balancing mechanism. Technically, this assumption is the strictest possible and does not make the setting any easier (intuitively, the less balanced the share, the closer we are to the 1-player setting).

B) One can consider a similar problem where players have to find some eps-best arm, rather than agreeing on a single one. This does not render the problem any easier (most players still have to agree on a small subset of arms in order to sample it sufficiently), and thus is not a key requirement in our setup.

C) Majority vote generally fails to achieve the optimal sample complexity (and this will be noted in the paper). For such argument to work, each player has to succeed with a reasonable (say constant) probability - but this requires the full budget of a single player (for each of the players).

Minors:

\* l.237: This is an error on our part - the number of bits may be O(n log(T/n)).

\* l.525: This follows from eq. (4), by taking only the largest term in the sum.

\* l.687: Good point - we will add more details there.

\* The Auer & Ortner paper is indeed relevant and will be cited.

Reviewer\_39:

A) The case of multiple best-arms is more naturally captured by the PAC setting, which is discussed in sec. 3.2. Indeed, the analysis is more complicated in this setting (see reply B to Reviewer\_21).

B) This is correct - we assume synchronicity of pulls for simplicity (and for aligning our model with standard MAB) and this could be easily relaxed.

C) We were not aware to the Kanade et al paper, and it should be cited. Note however that their problem is non-stochastic, hence is considerably different than ours.

Reviewer\_47:

A) Distributing arms equally among players fails to attain optimal performance (and indeed this should be noted): by that, the player getting the best arm will fail to identify it in case he also received other almost-best arms (his budget would not suffice).

B) Following previous work, we focused on the synchronous case and left asynchronous models to future work.

C) The restrictions on the number of needed players arise from the leading constants in our bounds, and this will be noted. Note that we did not try to optimize any constants in this work, and it is very possible that these are too strict.

We thank the reviewers for their detailed comments.

General remarks:

\* The setting: the considered setting is motivated by a real-world problem (see below) and is extremely realistic in its general form. We believe that while being simple and succinct, our model is an accurate abstraction of that problem and is able to capture the essence (and hardness) of the learning vs communication tradeoff.

Our problem is inspired by web-scale user experiences at relevant internet companies, that (at an abstract level) are optimized by MAB algorithms. Those are typically served from multiple machines distributed geographically across data centers, where within each center there might be dozens of machines utilized in a load-balanced manner. This implies NEAR UNIFORM traffic hitting each machine, and effectively gives rise to a SYNCHRONOUS process. Further, the best option served by machines at each center is indeed the same. It's true that across data centers, the rate of requests as well as the best arm may vary. However, within a center we believe our model is an accurate abstraction of reality.

\* Novelty of our approach: we note that our results (specifically the upper & lower bounds in the 1-round case) do not rely on the analysis of 1-player algorithms, and our main novelty is mostly orthogonal to previous work. Specifically, in our 1-round results we only use the 1-player algorithm in a black-box manner. Further, the considered problem is definitely non-trivial and requires a fair amount of sophistication. In particular note that:

1. A simple majority vote generally fails to achieve the optimal sample complexity. For such argument to work, each player has to succeed with a reasonable (say constant) probability - but this requires the full budget of a single player (for each of the players).

2. Distributing arms equally among players in the beginning also fails: by that, the player getting a best arm will fail to identify it in case he also received other (almost) best arms (his budget would not suffice).

See also the proof of our lower bound. This discussion was perhaps missing from our technical intro, and will be included there.

Further replies:

R\_21:

A) The assumption that each player gets an equal share of the budget is very realistic (see the above). Technically, this assumption is the strictest possible and does not make the setting any easier (intuitively, the less balanced the share, the closer we are to the 1-player setting).

B) One can consider a similar problem where players have to find some eps-best arm, rather than agreeing on a single one. This does not render the problem any easier (most players still have to agree on a small subset of arms in order to sample it sufficiently), and thus a consensus is not a key requirement in our setup.

Minors:

\* l.237: This is an error on our part - the number of bits may be O(n log(T/n)).

\* l.525: This follows from eq. (4), by taking only the largest term in the sum.

\* l.687: Good point - we will add more details there.

\* The Auer & Ortner paper is indeed relevant and will be cited.

R\_39:

A) The case of multiple best-arms is more naturally captured by the PAC setting, which is discussed in sec. 3.2. Indeed, the analysis is more complicated in this setting (see reply B to R\_21).

B) This is correct - we assume synchronicity of pulls for simplicity (and for aligning our model with standard MAB) and this could be easily relaxed.

C) We were not aware to the Kanade et al paper, and it should be cited. Note however that their setup is non-stochastic, hence is considerably different than ours.

R\_47:

A) Following previous work, we focused on the synchronous case and left asynchronous models to future work.

B) The restrictions on the number of needed players arise from the leading constants in our bounds, and this will be noted. Note that we did not try to optimize the constants in this work, and these are certainly too strict.