Dans le mail qu'on a reçu, ils disent :

"Following the author response period, the reviewer-author discussion period will begin on Mar 20 and end on Mar 26, during which you will be able to respond to any reviewer questions that arise during the discussion."

on peut donc répondre et avoir un vrai échange. Ca vaut peut-être le coup de demander quels sont les détails techniques incompris pour pouvoir les modifier ? Cela dit ça a plutôt l'air d'être utilisé si jamais il y'a d'autres questions...

Ils disent aussi :
"Past experience suggests that effective responses focus on factual errors in the reviews and on responding to specific questions posed by the Reviewers. Your response is optional and should be reserved for cases when a response is called for."

J'imagine que ça sous-entend qu'il faut être le plus précis possible !

Je vous mets un exemple d'un rebuttal de qualité sur un des articles qu'on étudie avec Luc : https://openreview.net/forum?id=oAog3W9w6R

(il n'y a pas d'espace avant ":" en anglais ; à vérifier pour ";")

GENERAL REBUTTAL

We would first like to thank all the reviewers for their constructive remarks and feedback and for pointing out that we address a "new" and "interesting" question: how to estimate the parameters of a learning algorithm based on one trajectory, i.e, a single experiment during which an individual learns. All three referees acknowledge that we provide rigorous and mathematically sound results for understanding how estimation can be done in this case. Indeed, we have been able to mathematically prove for the first time, the existence of two different regimes: in the Exp3 model, if the learning parameter is constant, some parameters cannot be estimated faster than a logarithmic rate whatever the estimation method, whereas if the learning parameter decreases polynomially with the number of observations, truncated MLE (Maximum Likelihood Estimation) achieves polynomial rates of convergence.

About the lack of comparison

From a practical point of view, there is no established baseline method except MLE (see below) to estimate learning parameters of a learning algorithm. From a theoretical point of view, it is--to our knowledge--a new question, and our work is the first to study the estimation properties in this context. This explains why we did not compare our estimator to other methods.

About the lack of generality of the algorithm

While EXP3 is a simple model, it has given rise to a lot of different algorithms, so it is worth investigating the simplest case. Moreover, our proof heavily relies on the update rule of the algorithm as well as the dependency structure of the data that are neither independent nor stationnary, making it difficult to extrapolate it to other algorithms. Our future line of research is to extend these results to more general adversarial bandit algorithms, which then in turn could be used to model more complex behaviors.

About the lack of motivation/applications

We wish to add a possible application in the final version. The bandit approach has been used extensively to model the responses of people with mental disorders (see Bouneffouf et al., "Bandit Models of Human Behavior: Reward Processing in Mental Disorders", 2017, as an example). For instance, it has been proved that smokers and non smokers behave differently when facing a bandit problem (Addicott et al., "Smoking and the bandit: a preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task", 2013). Therefore, estimating the learning rates of individuals facing a bandit experiment could help the early diagnosis of specific mental disorders such as apathy or Alzheimer's disease. Of course, such medical applications would need to be developed with psychiatrists to assess performance, acceptability and ethical and societal impact.

Another practical motivation of our work is that MLE is the popular first step used to fit behavioral models on individual learning data (see Wilson & Collins "Ten simple rules for the computational modeling of behavioral data", Elife 2019). The aim of our article is therefore(retirer "therefore") to prove that this first step is justified from a mathematical point of view. This is the first step in giving theoretical credibility to a statistical method that has been used for years by the behavioral science community.

About model selection and algorithm detection

Following Wilson and Collins' rules, the next step is to compare several models. In this sense, the interest for the behavioral community, beyond parameter estimation, is to find the class of algorithms that best fits the experiment, i.e. that best matches the way an individual learned. Future work would consist in extending our results to several other models, in order to use a mathematically sound model selection method.

REBUTTAL REVIEWER KADH

We thank the reviewer for their enthusiasm and their very constructive remarks.

In the final version, we will add some details about the concerns the reviewer raised on both potential applications and the link with algorithmic detection (see the main rebuttal). In particular,

- model selection is indeed a natural extension of this work, as it can be used to identify which algorithm most closely mimics an observed behavior. Studying the properties of a model selection procedure requires knowing the theoretical properties of each model, which would be the object of future work.

- a possible application in the medical field is the diagnosis of mental disorders, for instance early detection of Alzheimer diseases, by detecting irregularities in the learning process of an individual. Estimating the parameters of the model is the first step in detecting such irregularities.

REBUTTAL REVIEWER S5Ag

We thank the reviewer for their careful reading.

> The motivation for estimating the learning rate from the samples is not explained well, probably because I am not an expert in cognition science.

We wish to add a possible application of our work to the detection of mental disorders (see main rebuttal). The learning rates might differ between healthy individuals and patients suffering for instance from Alzheimer's disease. Our estimation method might therefore be turned into a medical test for early detection of such disorders. We hope that this application will convince the reviewer of the relevance of our work.

About the writing and figures

We would like to apologize for the improvable wording. We will make sure the final version of the paper is carefully proofread.
We would be grateful if the reviewer could point out the technical details lacking explanations. We will do our best to adjust the explanations. We will also make sure that the figures are no longer distorted.

About imitation learning

We will include imitation learning in the paper (see "Imitation learning: a survey of learning methods", Hussein et al., 2017) to expand our discussion w.r.t related works.
It is true that our framework could be seen as a particular imitation learning problem, where the goal is to imitate the *process* by which a system learns a new task (and not solely the final, calibrated, state of the system). However, this is not how imitation learning is typically

devised: usually, the learner imitates a teacher who has already finished learning. So in this sense, the input data of a classical imitation algorithm are not learning data.

We would also like to point out that the motivation of the present work is different from imitation learning. Our aim is not to reproduce realistic learning curves, but to estimate the models and parameters that best fit an observed behavior. A possible application is to understand how the underlying system (be it a human, an animal, or an other system) learns, and possibly detect anomalies or differences in this learning process.

About the generalization

The non-stationarity of the problem and the specific form of the Exp3 update rule make it difficult to directly transpose our result to a wide range of bandit algorithms. While we expect extensions to variants of Exp3 to be feasible, generalizing our proofs to algorithms other than Exp3 would still need to be done on a case-by-case basis.

REBUTTAL REVIEWER PcVg

We thank the reviewer for their careful reading of the paper and for adressing two important issues.

>There is no other method to be compared with [...] And the proposed method and analysis are not guaranteed to be optimal.

As stated in the global rebuttal, the mathematical question is new. Our objective is therefore to show when estimating the parameters is at all possible, and what rates could be expected. Accordingly, we show that, whatever the method used, it is not possible to estimate the parameters faster than at a polynomial rate when the learning rate is constant, while when the learning rate decreases polynomially, it is possible to estimate it with polynomial rate using maximum likelihood estimation (which is the default method in the cognition studies which motivated this work). Considering other methods, and looking for the fastest estimators, was not our objective, and could be the subject of future work.

As the reviewer pointed out, a natural follow-up is to prove minimax rates for this problem. With the mathematical approach we have at this stage, it seems challenging, because of the sensibility of the update rule (see Proposition 3.1). As an example, we are pretty sure that the lower bound for the fixed learning rate in Theorem 3.2 is in $\log(\log(n))^{-\alpha}$, but we could only prove a lower bound in $\log(n)^{-\alpha}$ because of one single term of the sum, which was especially difficult to control. So we keep this question of minimax properties of the truncated MLE for future work.

About the limited simulations

We will include more extensive simulations in the final version.

For the downstream human application

We proposed some new applications in the main rebuttal, mostly on diagnosing psychiatric disorders. We will include them in the final version and mention that in such applications, the method could have the same type of impact as any medical diagnosis tool.

Questions
1. Line 159, the sentence is poorly written.

We meant that arm 2 is almost always pulled. We will remove the part in parenthesis from the sentence.

2. Line 265,

Thank you for noticing this typo. It should be a $\ell_{n,\varepsilon}(\hat{\eta}_0) \geq \ell_{n,\varepsilon}(\eta_0)$.

3. Line 357,

We indeed meant the right part of the left subfigure.

4. Some figures are indeed stretched vertically to respect the page limit.

We will address this issue by putting some of the figures in the appendix or on the last page.