We thank the reviewers for their careful reading of the manuscript and for their comments.

Response to reviewer #1 Thank you for your positive reception. We will polish the writing and address the minor comments (cross-references, readability of Fig. 2,3, etc). We are happy to consider any suggestion you may have for an altered title in the final review (that may arise in the discussion): ours was meant as a rebuttal to the Nie et al. title.

[...*potential applications where the sign of bias can be useful...*] The sign of the bias has been used to check whether a search advertising system charged advertisers fairly [Xu et al., 2013] or to investigate the validity of MAB models for the design of clinical trials [Villar et al., 2015]. We will add these examples to the paper. As the reviewer points out, quantifying the magnitude of bias is also an important complementary problem, which we will pursue in future work.

[...*the term "optimism" is well-known and used in bandits...*] Our choice of the term optimism is deliberate, and is directly inspired by the principle of "optimism in the face of uncertainty" that is frequently invoked in the MAB literature. Indeed, we view our definitions of optimistic sampling, stopping and choosing as different ways of formalizing that very same principle. (For example, UCB algorithms are optimistic samplers, and our definition captures this and other methods.) We will include an explicit reference to the principle of optimism and clarify our choice of terminology. We thought hard about a different term (and looked up synonyms), but could not find a more appropriate one, and are happy to listen to a suggestion in the final review.

[...*Is the choosing strategy always adaptive?...*] The choosing strategy can be non-adaptive. For instance, choosing a target arm by using an independent uniform distribution on $\{1, \ldots, K\}$, is a random but non-adaptive choice.

Response to reviewer #2

[M1. *This paper could be of interest to people in the bandit community...you have to cite them more.*] We agree with the reviewer that it would be appropriate and beneficial to the manuscript to include more references to the MAB literature and we will do so, starting from the reviewer's suggestions.

[M2. *...make the goal of this analysis more explicit.*] The main objective of this paper is to provide a rigorous and comprehensive analysis of the sign of the bias in MAB settings under adaptive sampling, stopping and choosing rules. This is, we believe, a problem of practical and theoretical significance that has not been fully solved. For instance, consider the offline analysis of data that was collected by an MAB algorithm (with any aim): suppose that you want to estimate the mean reward of some of the better arms that were picked more frequently by the algorithm. The paper by Nie et al. proves that the the sample mean will be negatively biased under fairly common adaptive sampling rules: this could instill a possibly false sense of comfort with your sample mean estimate since their theory suggests that you are underestimating the effect size. However, we prove that if the algorithm was adaptively stopped and the arm index was adaptively picked, then the net bias can actually be positive. Indeed, we prove that this is the case for lil'UCB, but it is likely true more generally as captured by our main theorem. Thus, the sample mean may actually overestimate the effect size. This is an important and general phenomenon for both theoreticians (to study further and quantify) and for practitioners (to pay heed to) because if a particular arm is later deployed in practice, it may yield a lower reward than was possibly expected from the offline analysis. We can add such an explanation to the paper if the reviewer found it useful to compare messages explicitly.

[M3. *...relationship between the performance of a bandit algorithm and the bias?...fair comparison between UCB and TS in numerical experiments...*] This is a great question/suggestion, which we can study in future work. The main goal in our simulations is to visualize and corroborate our theoretical results about the sign of the bias. As a result, we did not make any attempt to optimize the parameters for UCB or TS for the purpose of minimizing the regret, since the latter was not the paper's aim. We will include a remark in the paper to clarify this.

Response to reviewer #3

Thank you for your detailed review which captures the most important features of our work concisely.

References

Sofía S Villar, Jack Bowden, and James Wason. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 30(2):199, 2015.

Min Xu, Tao Qin, and Tie-Yan Liu. Estimation bias in multi-armed bandit algorithms for search advertising. In *Advances in Neural Information Processing Systems*, pages 2400–2408, 2013.