We thank the reviewers for their comments and suggestions which have ultimately helped to improve the paper.

Reviewers 1 and 3 made similar remarks that for large N and small sample sizes, with-replacement (WR) and without-replacement (WoR) confidence sequences (CSs) perform similarly. We fully agree with this sentiment. However, this is a phenomenon surrounding WR and WoR methods in general, and is not limited to our methods. We still see two upsides to using our WoR methods:

- All our confidence sequences shrink to zero width in exactly N steps. With WR sampling, this would not be true at time N (but it would be true asymptotically).
- Given that our new bounds are explicit and simple to implement, there is no reason not take advantage of any available statistical benefit (small at early times, large at later times).

In general, WoR sampling could be utilized on problems with moderate N (e.g. sampling points in stochastic gradient methods, sampling covariates in a random order for coordinate descent, sampling columns of a matrix for accurate approximation, sampling edges in a graph for graph sparsification with spectral approximation, etc). We will point this out on the extra ninth page if the paper is accepted and hope others may follow up on these applications, where we suspect there is a practical benefit to our tighter bounds.

Reviewer #1

(R1) 'It would be nice to discuss [advantages of] the novel mean estimator...' In the paper, we discuss the advantage (denoted by A_t) of the novel mean estimator in terms of its statistical efficiency. However, this novel estimator is used primarily because the resulting Hoeffding- and empirical Bernstein-type processes turn out to be supermartingales, a key property allowing for the derivation of CSs. In some sense, this mean estimator naturally popped out when we were explicitly trying to construct a supermartingale, rather than the other way around.

(R1) 'The novel mean estimator appears inconsistent' Consistency is a subtle question; once we have observed N data points, there is no uncertainty left so we do not need an estimator at the final step. Our CSs in fact shrink to zero width at time N because of logical considerations. We will discuss this and better clarify the role of $\widehat{\mu}_t$ in the paper (using the extra space provided if accepted).

(R1) 'reduce the liberal use of quotes and underlining' We have now removed most of the quotes and underlines.

(R1) 'In Figure 4 (left)...' We now clarify in the caption that time 100 is on the plot.

Reviewer #2

(R2) 'definitions of filtrations and martingales...' We agree and have now added these.

(R2) 'introduction to Shapley values not self-contained' We agree, we have now added additional details for the uninitiated reader.

(R2) '[...] include experiments [comparing CS to CI]' Our approach also yields tight fixed-time confidence intervals for sampling WoR. We have plotted these in Figure 9 in Appendix H. We have now added references to these to the main part of the paper, and have added some fixed-time corollaries.

Reviewer #3

(R3) 'My only concern is around N' It is standard in the mathematical study of WoR sampling to consider the case of known N. The case of unknown N is an interesting direction for future work.

Reviewer #4

(R4) The PPR framework...conjugacy/MCMC We now clarify that the results surrounding the PPR martingale are valid even when the posterior is not in closed-form, and thus methods like MCMC can be applied.

(R4) 'this paper focuses on the discrete [and] finite set of bounded real...' We treat x_1^N as a finite list of non-random real numbers, so the only source of randomness is in the WoR sampling distribution (Section 1.2). Any finite list of numbers x_1^N is bounded by definition, and all past work on WoR sampling assumes that these bounds are known (e.g. [13, 14, 15] in the paper's references). Relaxing this assumption is an interesting question for future work.

(R4) 'There lacks several right bracket...' Thank you, we have now fixed this.

(R4) 'The notation J in Thm 3.1 appears abruptly' We have now tried to make the role of J less abrupt.