We would like to thank reviewers for their time and effort in providing us with feedback. Please find our response below, which focuses on the major points discussed by reviewers (R1, R2, R3 & R4).

Contribution/Relevance to the community (R1 & R4): Reviewer 1 asks "how much demand for solving numerical integration (in the Bayesian framework) is in the community". We would argue the demand is significant! See [1,3,21,26,28,35,47,58,73] which were all published at leading machine learning conferences, and [4,11,33,36,37,38,49,50] which appeared at leading venues in computational statistics or applied mathematics. We propose to further clarify this point, and add additional references in the machine learning literature.

Related Literature (R4): We thank R4 for the opportunity to expand. The novelty of our paper is to use tree-based models which are inherently Bayesian, and could hence be used for quantifying integration error in a Bayesian manner (as per the BPNI framework). This is different from the suggested references, where trees are used for MCMC proposals, which is not Bayesian per se. Furthermore, those papers use trees to approximate a density rather than the integrand. However, we agree that this literature is relevant and could motivate further research in BPNI. We propose to expand significantly on similarities and differences, and thank R4 for challenging us on this point.

The criticism about the Llorente et al. paper is quite unfair given that this paper appeared online *after* the NeurIPS abstract submission deadline. That said, we will discuss the nearest-neighbours approach, which could also fall within the BPNI framework (except that this paper only considers point estimates, rather than entire distribution). From the point of view of the models, the main difference is the way in which splits are performed. BART will adapt to the smoothness and sparsity of f in a way that the nearest neighbours approach cannot. We also note that the rate of convergence presented in that paper is much slower than for BART-Int.

Theory (R3): The result is written in a general form so that it can be used to provide stronger results than consistency: it allows for rates of convergence. This rate will depend on (i) the point set, and (ii) the prior model. The form of this theorem allows us to understand the specific impact of these two aspects, and as a result understand how the method will perform relative to competitors, and potentially how to improve it. We propose to further discuss these points, and to unpack further some of the more complex mathematical details.

Experiments (R1 & R3): The experiments considered have all previously been used as benchmarks by the community. The Genz functions are particularly useful as they can highlight strengths/weaknesses of different methods. The survey design problem with a Bayesian lense first appeared in an ICML paper; see [23].

Of course, further experiments could be useful, but we were not able to do this due to space constraints. Since reviewers agree that it would improve the paper, we propose to include new examples in the supplementary material, focusing on modern ML benchmarks, e.g. estimating integrals and evidence for Bayesian inference and model selection (e.g. Chai et al. (2019) and Gunter et al. (2014)), and uncertainty quantification in applied settings (e.g. Oates et al. (NeurIPS 2017)).

Other comments: Thank you for the additional feedback; we will clarify these points. Specifically:

- **R1:** Thanks, we will clarify the notation for the posterior distributions on f and $\Pi[f]$.
- **R3:** We used BPNI rather than BQ since our estimator is not a quadrature rule (this means a linear combination of function values). Our terminology was used in [11].
- **R3:** GPs also struggle with slow convergence rates in high d settings [70,72]. Some tree-based models can make use of sparsity structures to avoid this issue; see [41].