We thank reviewers for their feedback and for agreeing on the importance of our work under both a theoretical and practical perspective [all reviewers (Rs)], for appreciating the way it is written and posed in the context of the hybrid probabilistic inference literature [R2 and R4] and for showing a clear improvement over SOTA WMI solvers [all Rs]. In the following, we will address the individual concerns raised.

R2

[Presentation& Motivation] We will provide in the camera-ready an additional example of WMI, illustrate in detail an execution of ReCoIn and better substantiate how integration in hybrid domains is at the core of probabilistic reasoning in real applications as well (e.g., in visual scene grounding).

[Integration in [-1, 1]] The interval of integration in the paper is correct and follows from the reduction to a WMI problem with only continuous variables, as illustrated in Appendix A.

R3

[Hardness with continuous variables] Indeed, the introduction of continuous variables increases the hardness of probabilistic inference. To see *why* this happens in the context of message passing, it has been noted in [1] that representing messages over SMT(LRA) theories induces a piecewise representation of the involved densities. From here one can see how inference complexity depends on the number of these pieces – which is exponential in the diameter of a primal graph and is not being bounded by its treewidth. This is strikingly different from discrete domains, where messages can be represented as finite tables whose sizes depend only on the treewidth of the model given a fixed variable scope. We point out that this increase in complexity is not peculiar to WMI inference alone and has been already observed in the probabilistic graphical model community. As a simple case, consider conditional linear Gaussian (CLG) models [2]. It is known that even if the graph structure of a CLG is restricted to have treewidth 1 integration for marginal inference can become NP-hard and even approximate inference routines with guarantees on these simple networks can be intractable.

[Binary representation of #SSP] In our hardness proof, we adopted the binary representation for the integers – as commonly assumed in reductions of this kind [2,3] – which makes the decision version of problem NP-complete and the Subset Sum Problem (#SSP) #P-complete. We will make this assumption explicit in the text. If instead the (non-standard) unary representation were used for integers for the #SSP, then the problem would be in P as the reviewer says. It could be interesting future work to prove hardness of WMI with rational unary numbers, or PTime complexity of integer unary parameters, neither of which where the goal of our paper. Note that adopting the usual binary representation for integers ***does not change the treewidth*** of the primal graph of the formula used in our reduction. This is due to the fact that the edges in the primal graph are defined by the dependencies in the formula which in turn is not affected by a change of representation of the constants in the SMT literals.

[Missing references] We thank the reviewer for pointing us to some recent advances in counting in discrete domains and in the propositional case. We will properly contrast our work to them in the camera-ready. Here we point out that they are not directly applicable to our WMI setting and that out setting is inherently more challenging than the settings mentioned in the linked references. In fact, we assume an SMT(LRA) theory involving continuous and discrete variables and a first order representation to be given. We do not build any further abstraction as in this context it would require going to a higher-order logical representation or exploit some symmetries in the theory, as usually done in lifted probabilistic inference. Lastly, to adopt one of the techniques referenced would require to revert to a fully discrete representation. That is, discretizing a WMI problem in both the support and densities and essentially reducing it to a weighted model counting problem. A discretization of this kind would be quite problematic, however: binning continuous supports is highly task-dependent and the loss in "resolution" can greatly hurt the quality of inference and potentially increase exponentially the problem representation.

R4

[Watts-Strogatz graphs] We employ Watts-Strogatz (WS) graphs as benchmarks with an increasing number of variables with bounded treewidth larger than 1. In the final paper, we will highlight these properties of WS graphs and add experiments on other challenging topologies (e.g., cactus graphs) where we can observe the same solver behavior.

[Typo] Thanks for spotting the typo, we will fix it in the camera-ready and proof read the paper again.

[1] Zeng et al. "Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing", ICML 2020.

[2] Lerner, et al. "Inference in hybrid networks: theoretical limits and practical algorithms." UAI 2001.

[3] De Campos, C. P. "New complexity results for MAP in Bayesian networks." IJCAI. 2011.