Firstly, we would like to thank the reviewers for spending their time on reviewing this submission and for the valuable feedback that came from it. We'll do our best to address the pointed out issues in the revised version of the paper.

Argoverse We'd like to address the concerns raised about the quality of the proposed method on the Argoverse benchmark. Shortly after submitting the paper we've discovered a methodological bug in our experiments: it turned out we were using a trajectory bank built from the in-house dataset in the evaluation of our method on the Argoverse dataset. We've fixed the mistake and re-evaluated the proposed method. The updated results can be seen in Table 1. We've also updated metrics for all other methods where a better result has been published on the Argoverse leaderboard [1] at the time we've updated our result (June 16th, 2020). In terms of top-1 prediction metrics there still are two entries [3, 4] slightly ahead after the update. As with VectorNet, a direct comparison with these methods is hard, as we use a more sophisticated decoding scheme, while they propose improved scene encoders. One entry, which, just as our method, uses a rasterized representation of the scene is «uulm-mrm» [2, 5], which we surpass in quality. We'd be happy to add missing details to the paper, as well as the results of an involuntarily study of whether the trajectory bank from one dataset can be reused on another.

Better baselines It has been argued that the baselines we compared against on the in-house dataset can be stronger. We however argue that the comparison is fair as we propose a new decoding scheme that can work on top of any scene encoder, and current SotA methods either use a simple fully-connected decoder [4, 5] or a RNN, which seems to work better [7]. This should be clarified in the paper.

Pedestrian trajectory prediction It has been suggested that the generality of the proposed method can be demonstrated by applying it to the pedestrian motion prediction problem as represented by datasets such as SDD, ETH or UCY. However the proposed method requires a large training set to build the trajectory bank from, and all these datasets are rather small. To the best of our knowledge, no large public dataset for the pedestrian motion prediction problem has been released as of yet, as highlighted by works such as [6]. We have, however, tried our method on a large scale in-house dataset for pedestrian motion prediction and the conclusions were similar to what we've got for vehicles, so we decided not to add it to the paper as it doesn't add any new insights. We can add these results if deemed necessary.

A study of the effects of clustering It has also been suggested that we should do more experiments to clarify the role the clustering procedure has on the induced $h(t)$ and the quality of the proposed method. We have experimented with various clustering schemes and hyperparameters and found no significant effect on the results, as long as there is some clustering. Some results obtained for a model that is slightly different to the one in the paper can be seen in Table 2. We'd argue that the specifics of the clustering procedure do not have a large effect on quality because the actual values of $h(t)$ become less relevant if we consider large enough number of max inner product search results, as we'll get most of the significant terms of the posterior anyway. We agree that this should be clarified in the paper, and a study of the effects of clustering should be added.

References

[1] `https://evalai.cloudcv.org/web/challenges/challenge-page/454/leaderboard/1279`
[2] ArgoAI challenge results `https://slideslive.com/38923162/argoai-challenge`
[3] Mercat et al. "Multi-Head Attention for Multi-Modal Joint Vehicle Motion Forecasting." arXiv. 2019.
[4] Liang et al. "Learning Lane Graph Representations for Motion Forecasting." ECCV 2020.
[5] Cui et al. "Multimodal Trajectory Predictions for Autonomous Driving using Deep CNN." ICRA 2019.
[6] Jain et al. "Discrete Residual Flow for Probabilistic Pedestrian Behavior Prediction." CoRL 2019.
[7] Hong et al. "Rules of the Road: Predicting Driving Behavior with a Convolutional Model of Semantic Interactions." CVPR 2019

Table 1: Updated results on the Argoverse dataset

Model		ADE@1	FDE@1
ADE@1 leaderboard top	Jean [3]	1.68	3.73
	_anonymous (LGN) [4]	1.71	3.78
	PRANK (ours)	1.73	3.82
	poly	1.77	3.95
	UAR	1.86	4.09
other	VectorNet	1.81	4.01
	PRANK (ours, old)	1.84	4.05
	uulm-mrm [2, 5]	1.90	4.19

Table 2: A study of the effects of clustering

Clustering configuration	N clusters	ADE	FDE
hierarchical compl.-linkage	1k	1.322	3.028
	5k	1.314	**3.007**
	10k	1.315	3.030
k-means, implementation 1	10k	**1.313**	3.011
k-means, implementation 2	10k	1.321	3.022
k-means, implementation 3	10k	1.322	3.031