We would like to thank our reviewers for their constructive comments.

R1: Not first to look at full meshes. Cite Sminchescu et al., . . . Soften claims? Thank you for the references, which we will add. Accordingly, we will limit our claim to deep learning approaches for human pose reconstruction. **R1, R3: Release code and data.** We will release those, along with pre-trained models. **R1: Not clear. . . what data is used to train the norm. flow. Why does fig 4 show stick figures?** The prior is on the SMPL *pose* and *shape* parameters, which we reparameterize as 24 3D control joints (L185-187). This was shown empirically to produce better results. **R1: Pose prior makes little difference in Tab. 2.** Yes, the benefit of the n.f. prior is small in 3DPW primarily because there are limited occlusions/ambiguities in this dataset. In this case, the benefit of the re-projection loss (Eq 5) dominates (when the latter is removed, the effect of the prior is more pronounced - rows 1, 2 in Tab. 2). **R1: The improvements for $M = 1$ mode are surprising. Why?** Best-of-M method should not *per-se* benefit the $n = 1$ case, but there are other differences: the $n = 1$ hypothesis is obtained by quantizing from the $M = 100$ model while using the pose prior to re-weight the predicted poses, which is the potential source of improvement (L264-266). **R1: L200-201: If only joint locations of SMPL used, L_V can be computed for all hypotheses.** Actually, the loss L_V on L200-201 is based on *the full set of 6890 SMPL vertices*, which is why it is so expensive (L199-201).

R2: HMR and SMPL losses already published. The main contribution is a way to generate and select a plausible set of outputs. Yes, our contribution is *n-quantized-best-of-M* — a better method for multi-hypothesis generation. In agreement with the other reviewers, we believe this to be a non-trivial step forward. Also, while the reprojection loss is not new, the way it is applied to all hypotheses, best and not (L53-59, L191, Eq 5), is new in best-of-M and brings a sizeable improvement for this class of methods (Tab 2 rows 1-2 vs 3-4). **R2: How does the method ensure diversity?** This is achieved automatically, by the best-of-M loss: 'guessing' only a single hypothesis would be disadvantageous compared to outputting M diverse guesses as that has a higher chance that one would have low error. Still, previous best-of-M implementations fail to reliably converge to this optimal solution, resulting in some modes to degenerate. An advantage of our loss (Eq 5) is that it helps to keep all modes active (L39-53). **R2: How is the n.f. conditioned on the image?** Please see L171-183. **R2: Evaluation of all hypotheses/diversity.** We follow the standard evaluation style in this area [21] (L227-228) adapted for our task of 3D mesh prediction. Achieving a low mean/median error of all hypotheses would in fact show that the diversity of the solution is low, which defeats the purpose of spanning the space of plausible hypotheses. However, we agree that evaluating the diversity itself is valuable and we present results in Tab. I: on average, our method yields the most diverse predictions.

R2: Number of clusters in Tables 1 and 2? We apologise for a typo on L231, which should read $n \in \{1, 5, 10, 25\}$. The value of M is 100. In Figure 6, we show the best hypothesis (in blue) and 3 other hypotheses from the $n = 5$ quantization. We omitted the last hypothesis to allow space for 2 columns of results over 2 datasets. We agree that this is confusing and we will include the missing hypothesis for a future version of the paper.

Num Modes	5	10	25
SMPL-MDN	47.3	47.4	49.7
SMPL-CVAE	2.1	2.4	2.5
Ours	45.8	53.8	58.3

Table I: **Hypothesis diversity on AH36m**. For a pair of meshes, diversity is an average 3D distance between corresponding control joints. We first compute a mean diversity over all pairs of hypotheses in an image and report an average over the test set. Tab. 1 extension.

R3: Core insight? Best-of-M models have been shown to outperform alternatives (such as VAE/MDN) in tasks with constrained output spaces (e.g. 2D keypoints [Rupprecht et al. ICCV 2017], or our 3D control joints are less complex than, say, the space of natural images). We thus selected best-of-M as the core of our contribution, while providing improvements that make best-of-M feasible/better for deformable 3D shapes. We will expand the paper with this motivation. **R3: Best hypothesis evaluation a bit unfair.** Reporting the best prediction is standard in best-of-M models [21] (and comes with the definition). This is because achieving a low mean/median error would in fact show that the diversity of the solution is low, which defeats the purpose of spanning the space of plausible hypotheses.

R3, R4: AH36M dataset is used for training? What about HMR and SPIN? The AH36M dataset is used for training *all* compared methods (we will clarify this in the paper). **R3: Failure cases?** Failure cases include testing on busy crowd scenes which include multiple people inside a single bounding box, or individuals of unusual shape (e.g. obese people), since we have very few of these examples in the train set. We will add a remark to the paper.

R4: Accurately capturing 2D joint locations in occluded views is challenging. We mask the loss with a visibility flag for missing keypoints. This allows to learn from images where annotators could not identify all keypoints. We will clarify this in a future version of the paper. **R4: More details on training and data** We will release code, data, pretrained models and add these specific details to the sup. mat. **R4: The results of SPIN are slightly different from the original paper, 41.8 in Table 1 and 41.1 in [18].** We take the results from the pretrained models released by the authors which differ slightly from the paper. **R4: Qualitative results in the ablation study.** We agree this would be a useful addition, and will include this in the sup. mat. **R4: Are the results in Figure 6 ranked by the generated weight?** No, but we agree this would improve this figure and we will alter the order in a future version of the paper.