[R2] *Explanation of how policy conditions upon the demonstrator identity during empirical evaluation* – Our method starts by setting the personalized embedding, $\omega_{p'}$, for a new human demonstrator, p', to be the mean of the embeddings of demonstrators in the training set. After every timestep, we update the personalized embedding utilizing the information provided (i.e., the true action) to converge on the type of current demonstrator in embedding space.

[R2] *Decision tree architecture provides inductive bias* – In Table 1, we report the results for a vanilla decision tree with the only modification being that we also provide the PNT's embeddings as additional inputs to the DT to help the DT tease out the heterogeneity. Unsurprisingly, this approach performed poorly. If we train a vanilla DT without embeddings or pairwise comparisons, the accuracy is only $55.76 \pm 1.4\%$, $32.4 \pm 0.7\%$, and $74.9 \pm 0.2\%$ across the three domains, which is significantly worse than our PNT approach. We believe these comparisons provide a control for the inductive bias in question and demonstrate the advantage of our formulation.

[R2] *Interpretable policy's dependence on demonstrator identity* – Discretized PNTs learn splitting criterion dependent on the demonstrator embedding, ω. Please see Figure 2 in the supplementary for a depiction.

[R3] *Novelty and impact of our user study* – We show through our user study that PNTs are perceived as more interpretable (verified by our Likert survey), less difficult to use (measured by time of completion), and less difficult to validate (measured by correctness) than a neural network. To the best of our knowledge, there has not been a prior study that provides insights into the interpretability of PNTs and NNs in the context of a "style" or "personality" variable(s).

[R3] *Rephrasing Hypothesis* – We thank R3 for these suggestions. We will modify **H2** as suggested. We propose to reword **H3** to say that "our approach is easier for a human to simulate than a neural network." as we think that this phrasing is a closer description of the task we asked participants to complete.

[R3] *"Interpretability" is vaguely defined* – We agree that our study left the definition of "interpretability" open to the interpretation of the participant. However, different interpretations would be a random effect across participants as we used a within-subjects design. As such, the experiment design is robust to between-participant variations in the interpretation of this term.

[R3] *Results of base DDT Algorithm* – The base DDT achieves only $55.28 \pm 1.8\%$, $52.35 \pm 0.7\%$, and $76.70 \pm 0.7\%$ accuracy in the Low-dim, Scheduling, and Taxi domains, respectively, whereas our PNT achieves $97.30 \pm 0.3\%$, $96.13 \pm 2.3\%$, and $88.22 \pm 0.6\%$, clearly displaying the advantage of our framework.

[R3] *Details of synthetic schedule generation* – This information is located within the supplementary. We will add these details into the main paper.

[R3] *Quality of resultant schedules* – We only have access to the decision-making policy rather than a ground-truth metric on the quality of a schedule. However, we agree this would be an interesting metric to pursue in future work.

[R4] *Personalization and interpretability works are orthogonal* – While we agree that this paper has an abundance of material, we believe that model personalization and interpretability are synergistic. Our results show that the mechanism we leverage for interpretability (i.e., DDTs) improves accuracy in conjunction with personalization, and personalization helps improve the accuracy of the DDT.

[R4] *Evaluation domains are simple* – While the Taxi domain may have a smaller state space, the demonstration data (in the form of decision trees) collected from 80+ users shows that there is an extremely diverse set of possibilities when it comes to "solving" even the simple problem of pickoff-dropoff. Our scheduling environment provides a more clear assessment of how our PNTs work in higher dimensional state spaces.

[R4] *PNT hyperparameters* – The PNT structure (e.g., depth, personalized embedding cardinality) is determined by cross-validation on a subset of the training data.

[R4] *Counterfactuals for a stochastic policy* – In our supplementary material, we show that our approach is robust to noisy demonstrations (i.e., ones in which the users' most-preferred action is not always selected).

[R4] *Leaf node representation* – Our PNT leaf nodes represent a probability distribution over actions. While we have only made use of the argmax over this distribution, we propose to explore sampling in future work.

[R4] *Freezing θ* – We freeze θ to 1) avoid overfitting, 2) maintain a high level of mutual information among the embeddings and trajectories, and 3) maintain interpretability among our discrete trees. If θ were to be updated, the posterior network $q^{\omega}_{\zeta|\theta}$ would need to be retrained to maintain the current level of mutual information. Similarly, the PNT was trained with regularization offline to afford a discrete tree that allows ω to continue to vary.

[R4] *PNT and neural network sizes in user study* – Both the PNT and NN sizes were minimized while maintaining high accuracy. Further details are provided in the supplementary material.

[R1] We are greatful for your time and feedback – thank you.