We thank the reviewers for their detailed and useful reviews of our paper. We are glad the reviewers appreciated the interdisciplinary nature and the quality of our work. First, we recall that the main goal of the paper is to motivate and to provide a description of our interpolation method and to explain how it relates mathematically to other methods. Then, we illustrate how texture interpolation will serve further studies of visual perception. Importantly, our code is easy to use (from a command line) and will be available online so that the vision scientist community can start using it. Our future work will be dedicated to vision experiments *i.e.* directed toward a less theoretical audience. Future experiments (beyond this submission) should include a control with PS texture interpolation but not Gram-based texture interpolation (see §Patchy *vs* Non-patchy interpolation). If accepted, this paper will be the core technical reference.

Balancing Methods and Results As a compromise between R1 and R3, the method used to measure ellipticity will be moved to the supplementary material and expanded with more detailed explanation and an illustration. Specifically, prior to using our method we empirically validated it on artificial data with different dimensionality. R3 is correct that the distributions of natural images and textures are not elliptical. We only show that natural textures distributions are "more" elliptical than natural images distributions. Such a change and the extra-allowed page should leave some space to expand, and therefore clarify, the Results and Discussion sections and the description of our experiments as required by R1, R3 and R4.

Figure importance and indexing We will keep Figure 1/2 in the main paper because they provide intuition for why the proposed texture synthesis approach works. However, we will add a new Figure 1 illustrating the main idea of our paper which is to evaluate how moving along interpolation paths affects visual perception and neural activity.

Experiments As suggested by all reviewers, we will run our psychometric experiments with naive participants (~ 8 more for each textures) for the camera-ready version if our submission is accepted. This will allow for a population analysis. However, collecting more neurophysiological data is uncertain because of the Covid situation which has delayed many ongoing experiments.

Patchy *vs* Non-patchy interpolation The PS algorithm and ours generate non-patchy interpolation contrary to the Gram-based interpolation. Patchy interpolation are less interesting for studying texture perception because neuronal receptive fields are localized and could therefore respond to the statistics of one of the two interpolated textures depending on patches location. Yet, the question of why the Gram-based interpolations are patchy is open. In particular, it is not due to an over-parametrization of the Gram method as both methods have $O(N^2)$ parameters (in fact, the Wasserstein method has only N more parameters). We suggest that the mathematical foundation of our approach will enable further progress compared to the engineering nature of the PS algorithm.

The perceptual scale The inverse of the perceptual scale f^{-1} linearizes the perception of a physical parameters (here the interpolation weight t). Following ideas that the visual cortex linearizes transformations [3], we believe that such a function is important to predict the path of the neural activity from the path of the stimuli.

V1 *vs* V4 Previous work has shown that V4 is also sensitive to textures [5, 6] and therefore makes it interesting to compare to V1. We will keep in mind R1's remarks for future experiments.

Biological relevance of our approach We acknowledge that, in principle, there might exist a simpler approach that captures structure of textures in a way that is closer to biological perception, as mentioned by R2. Yet, CNNs-based approaches are meaningful for at least three reasons: (i) a body of literature shows that CNN activations are able to linearly predict neural activity along the hierarchy of the visual cortex [8]; (ii) mixture of elliptical distributions are a promising model of CNN activations [7] and (iii) mixture of elliptical distributions account for neuron responses to natural images in V1 [1] (but are still to be tested for mid/high-level vision). Even if the CNN architecture is largely inspired by the visual cortex, we agree with R3's comment that CNN weights are not grounded in human perception. However, the brain is hypothesized to be adapted to the statistics of its natural environment which are reflected in CNN activations.

Miscellaneous We will include all suggested references in the introduction or the discussion. We will illustrate the sample diversity of our approach by adding multiple synthesis results from different random seeds in the supplementary material. We did not account for the power spectrum in our loss [4], and we acknowledge that this would be more rigorous. We will add this feature to our code. The effect of the number of VGG layers used to constrain the synthesis is similar to what is already known [2]: deeper layers account for larger spatial structures. To our knowledge, there is no comparison of the Gram *vs* Wasserstein loss. The reason is that, differences are not visible at first sight on the synthesized textures without sampling interpolation of textures. Such differences may be crucial for visual perception studies but less for computer graphics.

References **[1]** R. Coen-Cagli, A. Kohn, and O. Schwartz. "Flexible gating of contextual influences in natural vision". In: *Nature Neuroscience* (2015). **[2]** L. Gatys, A. S. Ecker, and M. Bethge. "Texture synthesis using convolutional neural networks". In: *Advances in Neural Information Processing Systems*. 2015. **[3]** O. J. Hénaff, R. L. Goris, and E. P. Simoncelli. "Perceptual straightening of natural videos". In: *Nature neuroscience* (2019). **[4]** G. Liu, Y. Gousseau, and G.-S. Xia. "Texture synthesis through convolutional neural networks and spectrum constraints". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE. 2016. **[5]** G. Okazawa, S. Tajima, and H. Komatsu. "Image statistics underlying natural texture selectivity of neurons in macaque V4". In: *Proceedings of the National Academy of Sciences* (2015). **[6]** G. Okazawa, S. Tajima, and H. Komatsu. "Gradual development of visual texture-selective properties between macaque areas V2 and V4". In: *Cerebral Cortex* (2017). **[7]** L. G. Sanchez-Giraldo, M. N. U. Laskar, and O. Schwartz. "Normalization and pooling in hierarchical models of natural images". In: *Current opinion in neurobiology* (2019). **[8]** M. Schrimpf et al. "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?" In: *bioRxiv preprint* (2018).