We thank the reviews for their hard work, enlightening comments and positive feedback, appreciating the novelty and the results: R1: "a novel take on the impetus for a certain set of illusions" R2: "a very nice paper... use for generative models which goes beyond generating nice images of faces or dogs." R3: "Unveiling these principles is a fundamental goal of the Neural Information Processing community." R4: " engaging insights".

Hereafter, we respond to the reviewers' individual comments.

R1: The assumption that patch likelihood is appropriately measured could use some more justification.
Since our model is a flow-based generative model, it optimizes the log-likelihood of the data (image patches, in this case) during training. This, in turn, allows likelihood evaluation [22]. On the practical side, since there is no commonly-used evaluation, nor a ground truth, for patch likelihood, we propose in Section 2.2 a couple of new evaluations: (1) a center patch test that compares the likelihood of patches to the empirical results of [30] (quantitative) and (2) a min-max test that compares the ranking of patches trained on a single image and on an external dataset (qualitative).

R1: There could be more examples of similar phenomena explained by the model.
Certainly. Our paper focuses on a variety of lightness/color illusions, which "share some inherent properties, but are different enough to make a convincing case" (R2). However, the model may explain similar phenomena in geometric illusions (our perception of lengths/angles as a function of their percentile rank) [20]. This is a major future direction.

R2: Could the authors comment on the use of percentile rank?...the relationship between the CDF and percentile rank
The percentile rank of a given value is the percentage of values in its frequency distribution that are equal or lower to it. It is shown empirically in e.g., [31,34] (by analyzing the responses of human observers) that this relative percentile ranks predict perception. The reason is that the relative number of times biologically-generated patterns are transduced and processed in accumulated experience tracks reproductive success. Thus, for instance the frequencies of occurrence of light patterns over time is "aligned" with perceptions of light and dark.
The percentile rank is the CDF, normalized to the range of $[0..100]$ (to be percentage).

R2: What about cases where the patch to explain may have some structure?
Thanks for the question! The White illusion is an example where the patch to explain has structure. In addition, there are geometrical illusions (e.g. direction or length of lines) that can be explained by a similar theory of statistics of natural images. Though this is beyond the current paper, it is an exciting future direction.
Kanizsa triangle: We are not aware of statistical explanations to this illusion. This is worth studying (in the context of depth statistics). In the paper we provide novel statistical explanations to White & Hermann (a statistical explanation for the simultaneous-contrast illusion has existed before).

R2: scale of perception - do subjects report the same change in lightness perception?
We have not found raw data for the specific illusions we study in the paper. For certain geometrical illusions, the percentile rank is found to be at the same scale of perception, e.g. perception of line length [20]. We note that in order to make conclusions regarding scale, the settings of the psychophysical experiments should be taken into account; currently each experiment depends on specific parameters, such as the distance of the subjects from the monitor, the size of the illusion itself and its inner structure.

R2: Do results change with other models... say a simple GMM...?
Indeed. The strength of our model is that it is capable of generalizing well from natural patches it is trained on to synthetic patches it is fed with in the analysis (Section 3). We performed a couple of experiments with GMMs, which exhibit inferior generalization. For instance, the likelihood graph of the simultaneous-contrast illusions is almost a delta function. Another benefit of our model is being generative, thus it may be easily used for illusion generation (Section 4); it is less clear how to do so with GMMs (simple sampling does not work well).

R2: More information about the actual model implementation and networks used would be useful.
The code will be released upon acceptance. Implementation details are provided in the supplementary materials; we will add any requested information.

R3: The only weakness of the work is on the relation with previous literature.
Thank you very much for these references and the extra information. We will discuss the explanations of visual illusions suggested in these papers, including the relation to the statistics of the stimuli to redundancy reduction; to uniformization techniques that may explain illusions when the environment changes; to error minimization; and the connection between visual illusions to deceiving CNNs. These papers strengthens the need to further study various facets of the relations between image statistics, neural networks and a variety of vision/perception phenomena.

R4: It would help rather than hurt if the authors made that clearer (e.g., in the title, abstract, and contributions).
Thanks. We will clarify the focus of the paper, which is on color & lightness illusions, in the title & abstract. Other types of illusions for which the empirical paradigm & setup applies (geometry, motion) are indeed left for the future.