We thank reviewers for their valuable comments. We will incorporate all feedback in our final manuscript.

Sample Quality/Qualitative/Quantitative Comparisons (R1, R2, R4) R1/R2/R4 comment on the overall generation quality of EBMs. We note that the main focus of our paper is to introduce a set of compositional logical operators over EBMs, and empirical applications, with qualitative fidelity of generations an orthogonal direction. By training an EBM with a larger number of parameters and computational resources, we see in Figure 1 that EBMs achieve high fidelity composition results comparable to those of a GAN model (SNGAN is a 128x128 model specifically trained with Young/Female/Smiling/Wavy Hair attributes). The SNGAN model has the same number of parameters and is trained with the same number of training iterations using the Mimicry * GAN library. We will release the code for training these models and pretrained weights.

In terms of image fidelity, on the Young AND Female AND Smiling AND Wavy Hair split, our composed EBM obtains an FID of 45.3 while SNGAN obtains FID scores of 74.2 (all FIDs are large due to a small dataset). R2 further asks diversity evaluation. We compare standard deviation across pixels of generated images and find SNGAN obtains 55.4 while EBMs obtain 64.5, providing evidence EBMs generate more diverse samples.

Figure 1: High resolution samples of attribute compositionality with EBMs (same setup as Figure 3 in the main paper). The last row shows SNGAN samples trained on specific attribute combination.

Training Time Comparisons with Other Models (R2) A general comparison between training EBMs and other generative models can be found in the appendix A.5 of [1]. Our EBM models are trained with the same methodology, and exhibit similar trends. EBMs are slower to train than GAN models, but faster to train than autogressive and flow models. In the particular setting of qualitative generation in Figure 1, EBM models roughly take 20 times longer to train than the corresponding SNGAN model (due to generating negative samples).

Relations to Previous Work (R2/R4) SCAN learns a fixed latent space for concepts and composition of concepts via logical rules is achieved by manipulating this latent space. Extending the space of concepts requires retraining the network. With our work, we investigate an alternative approach where new concepts can be added on demand via new energy functions without invalidating previous energy functions. This unique characteristic allows for unique benefits – such as the ability to learn visual concepts in a continual manner. Different from past work in EBMs, our approach is the first to propose additional logical operators of disjunction and negation, and show that these logical operators can be composed and nested together.

Compositions of Non-Overlapping Concepts (R2) We train seperate EBMs on the frogs and trucks CIFAR-10 images. We combine models in Figure 2, and find somewhat reasonable generations that share properties in both classes, although we do not expect good results in this regime generally.

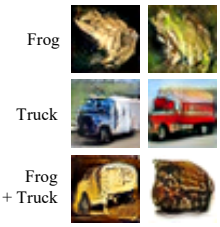

Figure 2: Hybrid combinations of frog and truck EBMs.

Comparison to Attention Masks (R2) While image masking approaches to compositionality enable a part-like decomposition of a scene, generation of each part is largely independent. This can miss interaction effects between parts (such as shadow casting). Our EBM composition generates all pixels of the image jointly, offering potential of capturing such interaction effects.

Equality of Partition Functions (R3) We had difficulty in checking for equality of partition functions without using spectral or L2 regularization as they are necessary for stable training of our method. We will clarify this in the paper.

Multiple Energy Functions (R3) Ensuring that individual energy functions all have good generative performance can be difficult. We find that using our proposed operators to compose generation can lessen the need for any individual model to have good generative performance (see for example Figure 1).

Additional Feedback/Clarification/Typos (R2, R4) We will add remarks in figure 1, and merge 3.3 with experiments section and the generalization. We will add a longer introduction about the EBMs and early equations. We will further fix the typos from R2 and R4, including the Logsumexp expression.

*https://github.com/kwotsin/mimicry