We'd like to express our gratitude towards all the reviewers who have devoted their time to evaluating our paper and providing constructive feedback. Specifically, we'd like to thank Reviewer #1 (R1) and Reviewer #3 (R3) for acknowledging our novelty. We feel honored to receive such a high rating and compliments from R3 and we appreciate it that R1 points out the linear complexity of our contrastive module compared to that of the relational module which is quadratic. We will also take advice from Reviewer #2 (R2) and improve our paper in revision. The rest of the rebuttal will focus on addressing concerns from both R1 and R2.

R1: Permutation in RPM, permutation-equivariance, and permutation-invariance We'll further clarify these terms in revision. Permutation is a unique property for RPM problems: (1) According to [17], in an RPM instance, the same set of rules is applied either row-wise or column-wise. Therefore, in a row-wise / column-wise instance, swapping the first two rows / columns should not affect how one solves the problem. (2) In any multi-choice task, changing the order of answer candidates should not affect how one solves the problem either. Permutation-equivariance refers to the case where a permutation applied to the input results in the same permutation on the corresponding output of a function. Permutation-invariance refers to the case where a permutation applied to the input does not change the output of a function. In this work, we hope that no matter how rows / columns are swapped and no matter how candidate answers are permuted, our computation process will always pick **the same** correct image, **rather than its index**. Hence, we consistently use "permutation-invariance" and always measure whether the correct image is picked, **not its index**.

R1: WReN's representation and its permutation-sensitivity The representation of images and their positions in WReN is: each image's features are first extracted by a small CNN independently and then concatenated with a one-hot positional tagging of length 9 before further processing. WReN is permutation-invariant in the sense of (2) but not in (1). Since the context panels are tagged with their positions, when rows / columns are swapped, the final representation will be changed. For example, in an instance where rules are applied row-wise, swapping the first two rows will change tags of images originally in the first row from $[1, 2, 3]$ to $[4, 5, 6]$. WReN's tagging strategy will couple the same image representation with a different position embedding, generating different scores for the choices and making the model permutation-sensitive. In CoPINet, we use a shared Conv layer to independently extract image features and sum them together to avoid the problem. We are sorry that Line 118-120 is oversimplified. And we will use the additional page to discuss permutation in RPM and why WReN is permutation-sensitive in this sense. We will also make Figure 1 clearer to show how permutation-sensitivity is avoided in CoPINet.

R1: Training of WReN-NoTag When evaluating the performance of WReN-NoTag, we did not concatenate positional taggings to the image features and trained the entire model from scratch. In this way, positions were no longer related to image features, and permutation problems from both (1) and (2) were avoided, hence the permutation-invariance; note that the relational module itself is permutation-invariant in the sense of both (1) and (2). We performed grid search to find the best hyper-parameters: batch size, learning rate, and weight of the auxiliary loss. We promise to release the code upon paper acceptance to benefit the community for future research in analogical reasoning.

R1: Missing reference Thanks for pointing out. We will add in Line 112, between the two sentences, "Steenbrugge *et al.* [75] propose a novel training strategy to improve the generalization performance of models on RPM, where a β-VAE is pretrained to unsupervisedly learn a relational latent space and fine-tuned together with the model".

R1: Grammar and wording We will change Line 8 into what is suggested by R1, check the entire manuscript, and correct all misuses of "*i.e.*". We will also change the verb in Line 306-307 from "lift" to "improve".

R2: The paper is unprincipled as it just combines a few ideas. It doesn't feel like it has much of a reasoning flavor The insight of this paper is to incorporate contrasting together with inference. This insight for analogical reasoning has been firmly established in the literature [21-30], but rarely adopted in machine learning. In this paper, the contrast idea is implemented as the contrast module and the contrast loss, and inference refers to "a simple inference module jointly trained with the perception backbone" (Line 54-56), hence the "perceptual inference" in our paper title.

R2: How can these ideas generalize beyond RPM Two important messages that can generalize for other analogical reasoning tasks are: (1) Contrasting, as demonstrated in the previous psychology literature, is indeed crucial to improve the performance of analogical reasoning for machine intelligence. The contrast module proposed in this paper could be easily plugged-in to any models for reasoning, ranking, or other discriminative learning, while the loss could be tested similarly on these tasks. (2) Improvement from the inference module suggests that we could use a sampling technique, let the model learn end-to-end by itself, and enjoy the performance boost.

R2: How can RPM be described as reasoning with pictures Existing works describe the image-based RPM either as an abstract reasoning task [14] or as a relational and analogical reasoning task [12]. As discussed in [17], one needs to reason about what the hidden rule is from correct encodings of a limited number of image examples to solve the task. Our solution tries to improve current machine learning models on this challenging reasoning task.

[75] Xander Steenbrugge, Sam Leroux, Tim Verbelen, and Bart Dhoedt. Improving generalization for abstract reasoning tasks using disentangled feature representations.arXiv preprint arXiv:1811.04784, 2018.