We thank each of the referees for taking the time to carefully read and comment on our work. We are pleased to see our paper was generally well-received, and that the problem "addressed here is very important" (**R3**) and that it "addresses some issues in a highly contentious, and long-standing debate in the cognitive science and AI literature" (**R1**). We are also happy that our findings are compelling: "the results are striking" (**R3**) and provide "a novel demonstration of... instilling machine learning models with the capacity for systematic and compositional generalization" (**R1**). We, too, are excited about meta seq2seq learning and its ability to acquire compositional skills.

The reviewers raise the following primary issues. **R1** discusses the framing of the paper and asks us to "provide more nuance and context regarding the [debate] on compositional and systematic generalization." **R2** asks us to further "explain how the proposed method can be applied to natural data." Finally, **R3** asks clarification questions and how meta-training can be interpreted in terms of human learning. We see these points as readily addressable.

**Compositionality in humans and machines.** Thanks **R1** for the thoughtful suggestions on framing, which we will happily incorporate in our revisions. We began the paper by contrasting the compositional skills of humans and machines. **R1** asks "How good this skill actually is in humans?" – indeed, recent work shows it is quite good! [1] In our remarks on compositional learning, we did not intend to take a stance on the *origin of these human abilities*. Certainly we do not want to suggest that "humans have a powerful ability that nearly comes for free because of some innate 'capacity for compositionality'", as **R1** asks us to clarify. People and models differ substantially in their experience and background knowledge, and we will make this absolutely clear in our revisions.

**R1** also asks whether we believe compositional skills are learned or innate (nature vs. nurture). Relatedly, **R3** wonders whether people require meta-training to generalize compositionally. We are very glad that our paper stimulates these fascinating questions. We can't resolve them here, as they are empirical questions, but there is compelling evidence that learning (even meta-learning) plays a role. We will add this discussion to the paper: First, infants have limited compositional skills [4] which improve with age [3]. Second, some language-related inductive biases are either learned or develop [6, 2], and compositionality could have similar origins. Third, we provide *a novel demonstration of how agents could learn compositional skills through experience*, which is a key contribution of our paper.

**R1** and **R3** ask whether people could be doing meta-learning, and indeed this as a real possibility. People have experience at the "meta-level," although it is quite different than meaning permutation we used for SCAN. As we see it, there is a "natural pressure to generalize systematically after a single experience with a new verb like 'to Facebook,'" (pg. 8): a child hears one or a few uses a novel word (the support), and she must be able to use the new word properly in new sentences (the queries). As with our meta-training, people are incentivized to generalize compositionally from a brief experience with a new word. We are currently applying our approach to few-shot word learning (language modeling) problems of this flavor, which relates to the comments of **R2**.

**Naturalistic data.** First, we respectfully disagree with **R2**'s assessment that the net learns "jump" and "walk" as mere pointers in memory; it's substantially more complex since the support set only indirectly specifies these mappings through multi-word commands (not isolated primitives). We will make revisions to say this more clearly. Further addressing **R2**'s comments, meta seq2seq learning is a very general framework with applications to other domains. To solve SCAN, meta seq2seq requires a compositionally-informed episode-generator, but this is a property of SCAN rather than an inherent property of meta seq2seq. The few-shot language modeling task, described in the paragraph above, does not require any special knowledge to generate episodes besides a set of sentences that all use the same new word. As another example, we are currently working on an application to "Flash Fill"-style program induction problems. During meta-training, one episode could show several dates formatted like "05/05/1987" mapped to the format "May 5, 1987", and another episode might shows dates like "1987/05/05" mapped to "month:may year:1987". Meta seq2seq would acquire compositional skills that support few-shot learning of new mappings/programs at test time, such as "05.05.98" to "fifth of May, 1987." Similarly, practical applications such as learning transformations of names, numbers, locations, etc. require no special knowledge for episode generation and require no ground-truth program supervision.

**Other revisions.** We will cite concurrent work from [5] and add a direct comparison in the paper (**R2**). On the "add jump" task, our method achieves a mean accuracy of 98.7% (Exp 3) while [5] achieves 78.4%. We also ran new experiments on the SCAN "around right" split, and our method achieves 99.96% and the other 28.9% [5].

Thanks **R3** for pointing out a potential confusion: line 164 refers to seq2seq training and line 179 refers to meta seq2seq training. We will reorganize the methods into two separate sections for "seq2seq training" vs. "meta seq2seq training." Also the equations for Luong-style attention are on lines 138-140, which we will unpack better in our revisions.

We thank you for considering our work in your further discussions.

[1] B. M. Lake, T. Linzen, and M. Baroni. Human few-shot learning of compositional instructions. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 2019.
[2] M. Lewis, V. Cristiano, B. M. Lake, T. Kwan, and M. C. Frank. The role of developmental change and linguistic experience in the mutual exclusivity effect. *PsyArXiv preprint*, 2019.
[3] S. Piantadosi and R. Aslin. Compositional reasoning in early childhood. *PLoS ONE*, 11(9):1–12, 2016.
[4] S. T. Piantadosi, H. Palmeri, and R. Aslin. Limits on Composition of Conceptual Operations in 9-Month-Olds. *Infancy*, 23(3):310–324, 2018.
[5] J. Russin, J. Jo, R. C. O'Reilly, and Y. Bengio. Compositional generalization in a deep seq2seq model by separating syntax and semantics. *arXiv preprint*, 2019.
[6] L. B. Smith, S. S. Jones, B. Landau, L. Gershkoff-Stowe, and L. Samuelson. Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1):13–19, 2002.