We would like to thank all reviewers for their comments and helpful feedback.

Feedforward Simplicity. R1 and R2 questioned Feedforward Simplicity as a measure as it is unclear what constitutes a simpler model and that the brain is complex, so it is not exactly clear why simpler models would be preferred. However, studies of response latencies and sequential processing in cortical areas demonstrate that the feedforward path from retinal input to IT should be limited in length (e.g., see Tovée, Current Biology, 1994). Counting the number of layers provides a simple proxy to meeting this biological constraint in artificial neural networks, and our Feedforward Simplicity was meant to quantify this. However, we agree with the reviewers that this term is confusing and thus in the revised version we will update it to simply "Depth" and clarify our reasoning as above.

One-to-one mapping between brain areas and model components. R1 commented that a one-to-one correspondence between model components and hetereogeneous brain regions seems simplistic and that circuitry might not be the exact same across regions. We agree with both of those points. The simplistic assumption of clearly separate regions with repeated circuitry was a first step for us to aim at building as shallow a model as possible, and we are excited about exploring less constrained mappings (such as just treating everything as a neuron without the distinction into regions) and more diverse circuitry (that might in turn improve model scores) in the future.

Justification for CORnet-S architectural choices. R1 asked how the number of recurrent steps as well as other architectural choices were justified. As R1 correctly pointed out, the major justification came from the ablation study in Fig. 5 – these steps were the most minimal configuration that produced the best model as determined by our scores. Training for more recurrent steps is possible, but at least on our current set of scores, we see no improvement. We expect that future temporal benchmarks might warrant the need for more recurrent steps.

Details sometimes lacking. All reviewers noted that some details are lacking and, according to R1, if this is the first publication of Brain-Score, many more details should be provided. The succinctness of the text is primarily due to the hard limit of 8 pages, thus we placed many details in the Appendix. Following the reviewers' remarks, we will attempt to work in the missing details and fixes for the camera-ready version; however, due to pages limits, we will still have to rely on the Appendix for in-depth explanations.

To answer some of the details that the reviewers pointed out: (i) L206: category refers to the categories of images used in new behavioral experiments; (ii) Fig. 1: conv / stride 2 refers to the stride-2 convolution; gating refers to a gate that only lets information through at t=0 (though a soft (sigmoid) gate leads to similar results; Fig. 5); (iii) OST: 80/10/10 split was used to have independent train / validation / test sets. It was not tuned to CORnet-S.

Brain-Score is not novel because it already available as a preprint. We hope the reviewer R3 will reconsider this criticism because, according to NeurIPS submission criteria, preprints are explicitly allowed, and this paper is the first publication of Brain-Score. Moreover, this submission has also developed Brain-Score much further from the preprint, namely, (i) we included four transfer tests on three newly collected datasets to validate the generalization capabilities (Fig. 2); (ii) we investigated possible predictors of Brain-Score (Appendix B.3); and (iii) we developed a mature open-source code base for an easy benchmarking.

CORnet-S has an unfair advantage because of recurrent connections. R3 stated that CORnet-S is winning only because of OST predictions that by design require recurrence. We agree with this remark and to a large extent that is the point of this model: the widely used family of feed-forward models simply cannot capture temporal processing that occurs in primate ventral visual pathway and are thus insufficient to build brain-like models. As also pointed out by R1, the value of CORnet-S does not lie in achieving the state-of-the-art on typical machine learning benchmarks but rather in demonstrating that it is BOTH brain-like AND a competitive machine learning model.

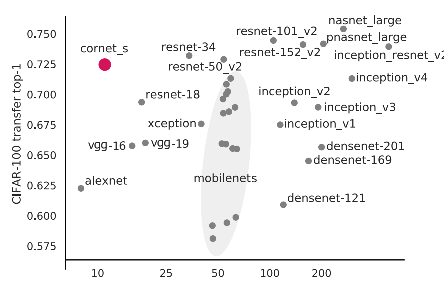

Figure 1: CORnet-S transfers better to CIFAR than similarly shallow networks

In addition, CORnet-S demonstrates competitive behavior on other machine learning measures: (a) it outperforms comparably shallow feedforward models and shows the best transfer performance among similarly shallow models (this response Figure 1); (b) it outperforms other shallow recurrent networks, as asked by R3 (see Appendix Fig. 2 that plots many other variants of shallow recurrent models that we built and tested; while some achieve a higher ImageNet performance, CORnet-S is the current best compromise between Brain-Score AND ImageNet performance).

Overall, we find that R3 evaluated this submission largely from a machine learning perspective. However, the strength of this submission is that it addresses the expectations of BOTH machine learning and computational neuroscience communities, unlike most prior work that was "either or". We are particularly excited about the impact this might have on sparking discussions between these communities on building high-performing AND high-fidelity models of brain function.