Dear Editor,

Thank you for inviting us to submit a revised manuscript. Below we have summarized the important changes made to the manuscript in response to the comments provided:

1. We have clarified language concerning the use of the multi-model mean in our discussion.
2. We have clarified the description of the climate sensitivity parameters used in the models and more clearly directed readers to our Supplement, which contains a very thorough discussion of the model specifications.
3. We have added a brief paragraph on the theoretical underpinning of our work while recognizing the journal's goal to remain interdisciplinary.
4. We made minor language edits for clarity throughout.

Response to Reviewer #1

Comment: Testing of simple models against more complex ones is interesting and relevant to ESD, but the interpretation of results are difficult, since it is not obvious that a complex model represents specific aspects of reality more correctly than a simple model. … I do not think this justification is correct and rests on a flawed interpretation of results in the literature, including Fig. 9.7 in Flato et al. 2013 (Chapter 9 in the IPCC AR5 report).

Comment reiterated by Dr. Michel Crucifix: "The reviewer #1 develops a valid point to criticize an element in the response of the author --- namely, the ensemble-mean of low-dimensional quantities (like the global mean temperature) cannot be expected to be closer to reality than that of individual models. Reviewer #1 appears to have based his advice of rejection mainly on that argument. Fortunately, this argument is not critical to the overall validity and relevance of the article."

Response: We have clarified our use of the multi-model mean.

Changes in Manuscript: We added the following text:

"For purposes of summarizing our results we compare the individual model responses to the comprehensive SCM multi-model mean for most of our experiments. We use this both for convenience and because the comprehensive SCMs can generally replicate the long-term results of general circulation models (GCMs; Meinshausen et al., 2011; Joos et al., 2013; Hartin et al., 2015, 2016). This is also, in a general philosophical sense, in line with the finding from GCMs that multi-model means compare better to observations than individual models (Flato et al., 2013), although we note that the Flato et al. finding was not specifically for global temperature. We, therefore, are not implying that the comprehensive SCM mean is necessarily the most accurate representation of the actual climate system response. It is instead simply a convenient metric for comparison. This metric illustrates both where the comprehensive SCMs are similar or different, and where the more idealized models differ from the comprehensive SCMs. Most of these latter differences are due to simplifications in the idealized models that bias their results, as discussed previously."

Response to Reviewer #2

Comment: As a minor comment, I must be missing something, but am surprised by the response of the AR5-IR model to a 4xCO2 scenario which seems to be trending to a value substantially higher than 6C. This seems odd in light of the stated sensitivity of 3C to a doubling of CO2.

Response: We realized that our description of the climate sensitivity parameter in our paper has been misleading in places. In particular, in Figure 4 we previously noted that the climate sensitivity used in the SCMs was 3°C, however this was only the case for the comprehensive SCMs. Equilibrium climate sensitivity is a result of ocean parameters for both of the idealized SCMs, FAIR and AR5-IR. Therefore, the Reviewer is correct and the ECS value was 3.9°C in AR5-IR and 2.7°C in FAIR. More on this aspect of the models is available in S2, where we also explore the AR5-IR responses using other parameter values.

Additionally, as we noted in the main text, AR5-IR lacks a nonlinearity from concentration to forcing and is insensitive to changing background concentrations.

Changes in Manuscript: We amended the text as mentioned above. In addition, we added language that clarifies the parameters used in each of the models in the Supplement. We also note throughout that the model parameters and model input files are available in our Supplementary Materials, which are also available for download on Github.

Response to Dr. Lucarini's comments

Comment: What is entirely missing is the theoretical framework for what they do. Indeed, the recent work of M. Ghil and myself goes in that direction and provides a pretty solid background for what they do. Testing response to forcings makes perfect sense if one wants to compare models; response can be (approximately) linear even if models are fully nonlinear; this is the essence of Ruelle's response theory in statistical mechanics (and variants thereof).

Response: We agree that including background information on the theoretical underpinning of this work is important and have updated the manuscript accordingly.

Changes in Manuscript: Citations and language were added to the introduction of the manuscript as follows:

"The impulse tests result in an impulse response function (IRF) for each model/species combination. IRFs characterize the dynamics of a linear system (Joos and Bruno, 1996; Ruelle, 2009) and, although climate models exhibit nonlinear responses, even some non-linear systems can be approximated by IRFs for small perturbations (Hooss et al., 2001; Lucarini and Sarno, 2011; Lucarini, 2018). The impulse responses examined here can be considered Green's functions, which form a key component of many simple climate models (Joos et al., 1999; van Vuuren, 2011a; Millar et al., 2015)."

Comment: Obviously, they need to provide full details on the model setups because results have to be reproducible by other people.

Response: Our Supplement includes a very thorough discussion of the models used and their specifications, and we have added some additional references to that discussion (see S1 and S2) in the main text of the manuscript.

Changes in Manuscript: Additional language citing the supplement was added to the main text.

Comment: Finally, I believe that the interpretation of the results should be expanded.

Response: We added text to the discussion and conclusion section.

Changes in Manuscript: Text has been moved from the supplement to the main text to add clarity. For example, we added:

"As a summary of our findings, we report the differences in time-integrated temperature response from the relevant multi-model mean in Table 1 for each of the experiments at selected time horizons. We chose the time horizons to report for each experiment by taking into consideration the atmospheric lifetime of the species and the ability to compare the experiments. For example, to compare the experiments exploring responses to CO_2 perturbations, we report the responses at 100 years after the pulse. For CH_4 and BC, we report at a time horizon of 20 years after the pulse reflecting the shorter lifetime of these species. Additional time-integrated temperature responses can be found in S9."