Response to the Reviewers

We thank both reviewers for the time they dedicated to reviewing our manuscript a second time and the helpful comments they both provided. This is a great help for improving our manuscript and highly appreciated. Since both the second reviewer and the editor asked for major revisions, we followed their suggestion, hopefully this time satisfyingly. We list the major changes below, followed by a point-by-point response to both reviewer comments.

Major Changes to the Manuscript:

1. To show that our approach of assuming a tipping element (TE) is triggered instantaneously once its threshold temperature is crossed leads to an overestimation of probabilities of triggering for scenarios including a temperature overshoot, we introduced the concept of delayed triggering. We added a section to the introduction explaining the concepts of instantaneous vs delayed triggering and why we need them, and adapted our analysis accordingly. The major changes that come with this are:

    a. We adapted the burning ember plots by removing the colour gradient, which was not adding any information, and instead included a third bar that represents the probability of delayed triggering.
    b. We included our new results in section 5.
    c. We added a section to the discussion in which we elaborate on both concepts and why this shows that we need to better understand the timing aspect of TEs.
    d. We added our main finding from this part of the analysis to the abstract and the conclusion.
2. Due to the warm bias of FaIRv2.0.0 and the difficulty of interpreting our results, we removed the first part of section 5 and figure 5 that were focussing on the timing of triggering TEs.
3. We include SSP1-1.9 again to show how uncertain probabilities of triggering are in the case of temperature overshoot, due to our limited knowledge about the effective timescales of TEs.

Response to Dr Christopher Smith

Thanks to the authors for taking the comments on board from the first revision. This is now a better paper. In my opinion there are still a few things to tie up.

We thank you for your second round of helpful comments. Since the second reviewer and the editor asked for major revisions, we have changed the manuscript quite substantially. We hope this will make it an interesting read that can convince you to participate in the third round of revisions.

Abstract line 13: “Averaged over all 16 tipping points, the probability of triggering until the year 2500 is 64% under SSP2-4.5”. I’m still struggling to parse this statement. 64% of triggering one of them? all of them? Eight of them?

What we mean to express with this is that when we have probabilities of triggering until 2500 for all 16 TEs in percent and average over the over those 16 values, the mean is 64%. It is the same value reported in Table 3. We have changed the structure of the sentence slightly and hope it becomes more clear now.

Line 44-45: “regional impact TEs are required to either contribute significantly to human welfare or to have great value in themselves as unique features of the Earth system”. “Contribute significantly” and “have great value” makes it sound like TEs are a good thing. It is contradicted (presumably correctly) at the start of the next paragraph. I’m sorry I didn’t remark on this first time.

True, that might be a bit confusing. It should say that in their intact (not tipped) state regional TEs are valuable to the society, hence tipping is dangerous. We included this distinction.

Line 90: Please explicitly state FaIRv2.0.0 here.

Done

Line 95: pedants’ corner. SSP1-1.9 is a Tier 2 scenario technically in O’Neill et al. and CMIP6. So you could just delete “except for SSP1-1.9 (O’Neill et al., 2016)”.

Thanks for the hint, we must have missed that. We now speak of five SSPs and leave out the statement about the Tier x.

Line 113: “five different SSP scenarios”. Now four.

Now back to five again.

Line 124: “Effective radiative forcing for CMIP6 follows the data provided by Smith (2020).” The previous few sentences describe all the sources of emissions to run FaIR with (you could also simply cut these and just cite Lewis & Nicholls 2021, https://zenodo.org/records/4589756 and Nicholls et al. 2020, describing the preparation of the emissions for RCMIP), so you shouldn’t need to use an external forcing time series. Perhaps you did use the external forcings for land use, solar and volcanic, in which case state this.

We state your last point more clearly now, sorry to not have included this earlier. However, we decided to keep the references to all emission datasets, since we mention the scenario extensions from Meinshausen et al. (2020) in our discussion.

Lines 132-135: I have to state again that the calibration of FaIR (or any other reduced complexity climate model used) makes a huge difference to the results. In the Leach paper, of which I was a co-author, we didn’t constrain FaIR v2.0.0 to the same level of rigour as

FaIR v1.6 was for the IPCC AR6, or v2.1 is currently, where the IPCC constraints (on climate sensitivity, ocean heat content, aerosol forcing and future warming in SSP scenarios) ensure the projections from the model follow the best available science. Using one of these calibrations would give you a 95% ECS at or very close to 5°C.

For example, SSP1-2.6 should not be crossing 1.5°C in 2027 in the median, which was mentioned in your earlier response (IPCC WG1 Chapter 4 has all SSPs crossing 1.5°C in the early 2030s). It's likely that the Leach calibration runs warm. Therefore you may see earlier triggering of TEs in this study than I would intuitively (though I'm not a TE domain expert) expect.

Sorry for labouring the point, but getting projections out of FaIR that are suitable for climate policy recommendations is basically my day job. I just want to see something in this paragraph along the lines of “using the calibration of FaIR in Leach et al. (2021)”.

You are right, we have made this more explicit to stay fair. The point that v2.0.0 is not AR6 calibrated while v1.6 and v2.1 are is actually something we wish we had realised before designing our model experiment. We checked whether it would be possible for us to switch to any of those versions for our revision. However, the implementation of FaIR seems to be quite different between the different versions, which unfortunately makes this endeavour infeasible for us due to time constraints. To not make overexaggerated statements about the timing of triggering TEs, we have removed the time dimension from this part of our analysis, which is also something the second reviewer requested.

Line 300: SSP3-70 -> SSP3-7.0

Done, thanks.

Line 318: “where triggering becomes more likely than not”: suggest to delete this, since it is indicative of IPCC calibrated language and adds no additional information to the previous statement on 50% probability. Appreciate this wasn’t necessarily the intention. Please also change “becomes more likely than not to be triggered” to “is triggered with greater than 50% probability” in line 341.

That's another valid point, however, we have removed this part of our analysis so it doesn’t apply any more.

Response to Reviewer 2

I want to thank the authors for their efforts in addressing my comments and the comments of the other referee. I also want to apologize for the delay in providing my assessment due to personal reasons.

We want to thank you for your detailed comments, which show you have really engaged with our study. We hope that our implementation of your feedback meets your expectations this time and helps to improve the quality of our work.

But I have to say, I am a bit disappointed by the authors response to my comments. There are some very fundamental issues in relation to their assessment of tipping element timing and outcomes. And rather than addressing the issue, the authors have chosen to delete the scenario (SSP1-1.9) where these issues are most apparent. That doesn't solve the problem and I don't agree with that approach. In fact, I think it would be important to keep SSP1-1.9 in to illustrate the point.

To be more clear what the issue is: The manuscript deploys temperature thresholds for tipping dynamics that assume constant temperature levels over millennia. That's fine, that's all we have so far. But let's call them long term tipping points (LTTPs)

But this does obviously not imply that the tipping will be initiated as soon as this LTTP is exceeded for the first time. In fact, there's very little that can be said about when exactly, time-wise tipping will be initiated. Certainly not with decadal or even sub-decadal resolution.

In response to my earlier suggestion, the authors replied that "To include internal timescales in our analysis would increase the complexity of our approach beyond the scope of this study. "

That's fine by me. But this in turn means that all results implying a timing of tipping need to be removed. Take i.e. Fig. 5 and subsequent figures. What they show is not when "tipping risks" are crossed. But when the temperature level of a multi-millennial LTTP is exceeded for the first time and thus a 'zone of elevated tipping risks. The authors may want to think about an appropriate framing here.

This either needs to be substantially revised to make that clear, or the temporal dimension needs to be removed. It might be cleaner to do the latter, because interpreting the former is difficult.

The authors may argue that the timescale until 2500, elevated temperatures above that level would lead to tipping with a sufficiently high chance considering the internal dynamics of the system. I'd like to see this discussed and substantiated but would agree that this is straight forward for all tipping elements but ice sheets, for which one need to argue this carefully. So this main part of the analysis, what tipping risks are until 2500, could be sustained,

The inclusion of the Amazon and Permafrost feedbacks are a bit of a problem, too, but if you assume this happens by 2300 at the latest you might be fine if I interpret your setup correctly.

Taking such an approach also illustrates why it's actually interesting to have SSP1-1.9 in the set, rather than out. Because one could show that the tipping risk under this scenario would be substantially, if immediate tipping was assumed at peak warming. Or much more moderate, if the long-term warming outcome was considered. And since you can't say which one of the two is true, you'd need to present them both, which I think is important in terms of directing future research and to inform your conclusions.

Thank you for this clear description of the problem. We followed your suggestion to include SSP1-1.9 again and removed the timing aspect from our analysis of probabilities of

triggering. Furthermore, we introduced the concept probabilities of delayed triggering and made or assumptions about the timescale more explicit by calling our previous probability estimates probabilities of instantaneous triggering. The difference between the two shows how the unknown effective timescale of TEs keeps us from making more precise statements about the probability of triggering in the case of a temperature overshoot.

It would also illustrate that there's an issue using this long pathway extensions for face value. They are useful tools for exploration, but not a 'given'. There's no IAM scenarios underlying those or else. And societies may well choose to bring temperatures down again between 2100-2500 if tipping risks loom large. This needs to be discussed.

We agree, the probabilities of delayed triggering only rely on the stabilized temperature and hence strongly depend on the scenario. We include this point in L437 — L440.

Other comments:

The Carbon cycle additionality question.
The additional explanations in the SI are helpful. Thank you. And I tend to agree for permafrost, because this is a dedicated process that's either there or not. For Amazon collapse, I find it to be a bit more complicated. Because it's not clear to me to what extent an Amazon collapse by and in itself is actually additional or different with regards to a scenario where the forest stays, but it's role as a carbon sink is greatly diminished. I'm intrigued by the authors statement that GFDL-ESM4 predicts abrupt dieback in parts of AMAZ in the 1pctCO2 run, but that this does not lead to significant reduction of the total carbon being stored in the Amazon vegetation. It's not my area of expertise and I haven't looked into this further. But it seems to substantiate this concern of mine that arguing this to be fully additional is a tricky.

We are glad that the additional explanations have proven to be helpful. Regarding the Amazon, we are still convinced that it is valid to assume carbon emissions from it as modelled by CTEM can be seen as fully additional to the FaIR carbon cycle with no double counting involved. The way we and Armstrong McKay et al. (2022) define carbon emissions from Amazon collapse only includes carbon that would be emitted from the degradation of the forest. The estimate is based on scaling up carbon storage estimates of the two vegetation types forest and savanna (SI of Armstrong McKay et al. (2022)). We make this more clear in L89 now. Any scenario in which the forest stays and is not converted to steppe, even if its role as a carbon sink is greatly diminished, does by definition not include Amazon collapse. Of the 11 ESMs used for the calibration of FaIRv2.0.0, only three would be able to represent such a change by including a dynamic vegetation model. However, even GFDL-ESM4, which predicts local dieback, does not include wide-scale conversion of forest to savanna. Therefore, we conclude that Amazon collapse is not included in any of the models used to parameterize FaIR and can hence not be included implicitly in the FaIR carbon cycle.

FaIR temperatures.
There seems to be something off with the ensemble the authors are using. In fact, they acknowledge this even in L130-135.

An AR6 constrained FaIR would not reach 1.5°C in 2025 in SSP1-2.6. That's quite outside the range. In fact, FaIR 1.6.4 only reached 1.5°C in current policy in 2040… And the AR6 assessment is sometime between 2030-2040. So the Fair 2.0 Leach et al. configuration deployed seems not ideal and the authors should be advised to revert it to one that's AR6 calibrated – or else provide a comparison that would increase transparency. What doesn't work, of course.

This also implies that there's something off here that would imply higher than AR6 peak warming for the same scenarios. Suggestion is to check 2010-19 warming and compare with AR6 WG1.

You are right, FaIR is warming too quickly in the early 21st century. This is due to version 2.0.0 of FaIR not being fully AR6 calibrated, as was brought to our attention by Dr Smith, the second reviewer and one of the developers of FaIR. We considered switching to FaIRv1.6 or FaIRv2.1, which are both AR6 calibrated, however, this has proven to be technically infeasible since the implementation of both versions is entirely different from the one we are using. Hence, this would mean starting from scratch, which we don't have the time for. However, with the timing aspect removed from our analysis, we think that this point is not as important any more. We acknowledge that the climate sensitivity of FaIR is not well constrained towards its upper end and included the implications in the discussion (L433 — L436).

The question of fitting the distribution:
I appreciate that you fitted those 'as good as you could'. To parameters which are not well constrained at all. I.e. are the authors certain that the max for a given threshold is really 5°. And not 4.9°C or 5.1°C, etc. So I think good arguments can be made to relax those criteria, by how much would your distributions widen? And, given that warming outcomes are skewed high, as are the LTTP distributions, how much does it matter? I don't argue it's critical that this is done, but it should be acknowledged in my view.

We agree that the percentiles given by Armstrong McKay et al. (2022) are not well constrained and make this more clear now in L247. We also agree that this would allow us to relax those criteria and play around with different probability distributions. However, we don't see the additional value of discussing this in our manuscript without actually doing it. Since we don't think that our results would change much if we alter the probability distributions while staying somewhat close to the estimates from Armstrong McKay et al. (2022), which are the only reference point we have, we hope the referee is OK with us not including this point in our analysis. Altering the probability distributions of Q would in our opinion only make sense if we had better information about what they could be, which requires further research to be conducted. We acknowledge this point in the last paragraph of the discussion.