Public justification (visible to the public if the article is accepted and published):
The three reviewers and I agree that this paper has potential to be an important contribution to the literature. However the reviewers are consistently clear that many aspects of the manuscript need to improved, including
- clarifying the scope of the review, such as whether it is limited to models and certain types of human behaviour. This problem is most acute in the title, but also needs clarification elsewhere.
- more clearly stating the insights gained from the review -- throughout abstract, results and discussion. A good review should be more than a collection of examples but also produce some 'results' or novel insights.
- including some brief information about the review method. If making claims about a body of literature, how can the reader be confident you have systematically sampled the literature? Methods could range from a systematic literature review to snowball sampling

I also suggest
- Clarifying early on what exactly you mean by 'drivers', since it is such a key term in the article. In the DPSIR framework and in systems thinking, it usually means some external forcing is affecting the system. However your discussion point #3 in AC2 appears to take "structural complexity" as a driver. (If I have misinterpreted and you do not consider it a driver, I wonder whether it's appropriate that only one of your four discussion points deals with drivers?)

Thank you for these comments. We have clarified the scope by changing the title and adding text to section 1 and the beginning of section 2.

We have made significant changes to the abstract, introduction and discussion, as well as other changes throughout the rest of the manuscript to better highlight insights gained throughout this review.

We have added a description of our snowball review method at the end of section 1, with mention in the abstract as well.

Our definition of ‘drivers’ has been clarified at the end of section 1.

RC1

Title. I think the title should make clear that this is a review of drivers of tipping points in coupled human-environment systems MODELS – as opposed to drivers inferred from empirical data of these systems. This is an important distinction to make.

We will change the title to "Drivers of tipping points in coupled human-environment systems models: a review"

Definition and examples of tipping points. For me, tipping points or regime shifts are fascinating because these abrupt, almost irreversible changes are induced by gradual changes in some components of the systems. Some examples presented in the Introduction, however, seemed to focus on those cases in which abrupt, big changes (big changes in parameters) lead to big changes in system outcomes, which are to be expected and do not capture the fascinating aspect of tipping points. Personally, I see a tipping point as closely related to bifurcation: the coupled system must go through changes in the number and/or nature (stability) of its stable equilibria. In cases where noise is considered, a tipping point may be said to happen when the system cross into a different basin of attraction associated with an alternative stable state. However, some of the examples in this review seem to simply be cases where changes happen quickly – which I don't think is sufficient to be classified as a tipping point. Does your definition of a tipping point require bifurcation and/or crossing a boundary of a basin of attraction or not? I think it should – even must - but this was not clear. Please clarify.

We mentioned "fast changes" and "disturbances" only in one sentence in the Introduction while presenting another authors' review paper. We take this as what this reviewer is picking up on as "big changes" but we in fact never use that term. Regardless, these kinds of changes are not referenced at all in our manuscript afterwards, and to avoid further confusion, we will remove the following text from that sentence: "Gradual ~~or fast~~ changes in system parameters, for example, the rate of resource extraction, ~~as well as sudden disturbances to state variables of the system, like a forest fire,~~ can cause the system to abruptly transition between these states." In terms of the question: "Does your definition of a tipping point require bifurcation and/or crossing a boundary of a basin of attraction or not?", the answer is a definitive "yes", however we see now why there there could be confusion about this because some of the qualitative examples we gave in the Introduction have not (yet) exhibited bifurcation and/or crossing a boundary of a basin of attraction (i.e., the rebound of the bald eagle and wolf populations following the enactment of conservation laws, as well as the banning of DDT regarding the declining eagle populations). We will remove examples that have not been fully studied.

Better connections among reported findings are need. Too many paragraphs have the following pattern: These papers say this on the topic, while those papers say that, and the reader is left hanging and confused at the end of these paragraphs. I feel that a useful review should offer more than a list of findings.

We will restructure the sections which appear to have a repeated pattern and provide more commentary on past studies. We will also restructure the paper to highlight our main findings.

And when a generalized statement was attempted, I don't think it was done with enough care. For example, Section 2.5 Rates of social change and time horizons – I agree with the authors that this is a critical component of a CHES model, but I feel that it was presented too simplistically. On lines 172-173, for example, the authors stated "…decreasing the relative speed of social dynamics brings about positive tipping point…" – is that true always? What if the initial extraction rate is too high? The low speed of social dynamics can translate to insufficient adaptation in the behavior, which would eventually lead to a system collapse.

This is a good point regarding low rates of learning and environmental collapse (tipping in the environmental system). We will qualify this statement to say what is only true of the papers which we found/surveyed in our review.

The effects of stochasticity or noise is not sufficiently addressed. There was some discussion of the effects of noise in Section 4 (about early warning signals) but hardly any in the previous sections. So most of the models discussed in Sections 2 and 3 were deterministic models only? There is a large literature on noise-induced transitions, which are closely related to tipping points, but this was not sufficiently addressed.

We agree that noise is an important driver, but there are not many papers looking at this in human-environment models. We therefore find it out of the scope of this review, and see this as an opportunity to point this out as a gap in our 'future work' section. Thank you.

All technical corrections will be made in the text

RC2

The title is misleading - The title should make it clear that this is a review article.

We will change the title to "Drivers of tipping points in coupled human-environment systems models: a review"

There are many broad statements that require further support with references, e.g. "There are many historical cases of human-induced tipping points that have drastically affected the trajectories of coupled environment systems and these effects can be both beneficial as well as catastrophic." Such a broad, strong statement requires many supporting references and qualification.

We will add citations to that sentence as well as other broad statements made throughout the text.

The discussion is quite general and descriptive and lacks a key unifying argument. What are we supposed to learn? What is the take away message beyond a loosely structured catalog of examples?

Overall, the paper lives somewhere between discussing specific models and discussing general principles and doesn't connect the two very well. I don't know if readers of this journal will get much out of it - I am guessing they will already be aware of most or all of the material covered. All of the points made in the "Conclusion and future directions" section are very well known and suggestions for future work are not very inspiring. I think the authors need to give some critical thought to what the key message is.

In terms of "What are we supposed to learn?": 1. There are counterintuitive tipping points in human-environment systems. 2. Many tipping points are system-specific in terms of drivers, especially in regards to input vs output-limited models. 3. Higher structural complexity in the social system increases the likelihood of tipping points. 4. Social time series data offer great potential for early warning signals, especially when using state-of-the art techniques. These four may be useful in terms of interventions by society/policy-makers. We will restructure and rewrite our Discussion section to highlight these and other specific findings.

1) The examples on human behavior aren't about human behavior. "Coupling strength" is a system feature. What is the behavior? "Rarity motivated conservation" is an outcome. What is the human behavior that generates it?

In terms of "what is the behaviour?", it is often the extraction or pollution rate by humans. We made some changes in wording to clarify

The human behaviour that generates what we called "Rarity motivated conservation" represents the extent to which humans change their [extractive] behaviour in response to the environment reaching a depleted state because of changing valuation. We will change this wording. Thank you.

2) Figure 1 has appeared countless times with different labels. This is a general feature of frequency-dependent selection. There is nothing particular about human behavior here other than the state variable in some sort of replicator equation is labeled "opinion".

The mechanism behind Figure 1 is common but the applications to CHES is under-appreciated in our opinion. Social norms are sometimes treated as only capable of supporting conservation in some cases. In other research it is assumed they can only suppress conservation. We wish to highlight human-environment systems–often at larger spatial and temporal scales–in which social norms can either support or suppress mitigation depending on the initial majority behaviour. This generates classic bistability patterns as we show in FIgure 1. We will better highlight this finding in the revised text and Figure 1.

3) Sections 2.3-2.5 are likewise not about human behavior but, rather, are about social dynamics. Maybe change the title of section 2 to "Social processes that may generate tipping points" or something like that.

We will change the section title to "Social processes that lead to tipping points in CHES models"

4) Section 3 discusses results specific to a particular model (Figure 2). This is difficult to follow - the actual model needs to be presented. The same is true of Figure 3.

We will extend the explanation of the model for Fig. 2. We will not present the actual model because it would take a few paragraphs to fully explain and is not needed to get the point we want to make with the figure which is how the rewiring probability has a tipping point at low and high levels. For Fig. 3, the models are unspecified ecological models, and their specifics are less relevant since the figures focus is on comparing deep learning EWS to traditional methods. The figure caption has been changed to include additional clarification.

RC3

The authors present an overview of typical parameters in coupled human-environment system models that can cause a tipping point. In my view, this article has the potential to make a nice contribution to the field of coupled systems modeling, viewed from the lens of social tipping points. However, some issues must be addressed beforehand.

First, the authors subsume a specific kind or style of coupled human-environment system models as presented in the introduction. This should be stated more transparently. Not all coupled systems models use or should use a form of social learning. This is important to contextualize since the factors the authors then eventually discuss heavily depend on this choice of coupled systems modeling style. The end of the introduction's second-to-last paragraph, beginning with "Common factors in the utility function are the rate of social learning," should additionally be supported with references. The authors should also describe briefly how they came about their review and choice of aspects.

We will add a statement acknowledging CHES models without social learning exist and we have chose to focus on social learning models. We will also add a sentence mentioning that we chose common social processes in CHES models at the end of the first paragraph of section 2. After our statement on the why for this paper i.e. “...we aim to deepen our understanding of human-induced tipping points through CHES models…”, we also include how we conducted our literature review.

Second, and a possible consequence of the first point, I doubt that the aspects the authors discuss comprise a complete list. For example, the role of inequality, biased perceptions, uncertainty, risk levels, and other-regarding preferences can be significant in some situations but are not discussed as individual points, not even in the outlook for future work. In some cases, there might be few works; in others, a different style of modeling coupled systems

already covers these factors. Thus, I invite the authors to contemplate which factors are generally relevant and should be studied (in possible future model studies).

We do mention two models with wealth inequality, under section 3.2 and agree that the list of points is not exhaustive and ones the reviewer lists are interesting to discuss. We will take up the reviewer's invitation "to contemplate which factors are generally relevant and should be studied (in possible future model studies)."

Third, a critical reflection on the robustness of the collected results is missing. For example, I wonder to what extent the occurrence of tipping points in these kinds of models might be a result of the fact that the strategy space is parameterized between two possible opinions/actions/moves: behaving sustainably vs. non-sustainably and the fact that social learning process assumes that agents share the same preferences. Thus, when the conditions change such that sustainable behavior becomes advantageous, the whole population changes. Likewise, is the observation that high structural complexity increases the potential for tipping points a result of the increased dimensionality of the system?

You are right, in many cases, the answer to what drives tipping points is "it depends". For example, regarding "*is the observation that high structural complexity increases the potential for tipping points a result of the increased dimensionality of the system?"* the answer is: it is not known. However, we did find some commonalities in other cases, where certain processes tend to drive certain types of tips in the range of studies we included in the review. We can highlight this in a more critical reflection in the Discussion and better delineate the cases where commonalities are found to the "it depends" cases.

Fourth, the structure of the review does not become clear to me. Section 2, talking about ASPECTS of human behavior, is so general that the aspects from Section 3 could go in there, too. The division is confusing since social traits, which I would better call the strategy space of the model, are nothing that can only be varied in agent-based models. The replicator dynamics are not limited to a two-strategy setting. And the role of Section 4 on early warning signals needs to be clarified. In model studies, we typically vary the parameters and observe the tipping point directly. If the point is to highlight empirical approaches, this should be addressed upfront. Overall, the structure of the manuscript should be clarified and introduced better.

We will change the wording of "aspects of human behaviour" to "social processes". Section 3 looks at structure via strategy space and spatial structure - we could make that more explicit in the text. For tipping points, we will mention that they are used on both empirical and synthetic data. Overall, we will clarify the structure of the manuscript and introduce it better as this reviewer suggests.

Fifth and last, after reading this manuscript, I have the impression that all parameters in coupled system models are or at least can be a tipping element. The review will gain in value if it contrasts such abrupt transitions with processes of gradual change in coupled human-environment system models.

Like any sufficiently high dimensional dynamical system, we agree that in principle, any parameter combination evolving in parameter space has the potential to cause a tip in these systems. However, in some cases, the empirical constraints on the parameter space preclude that. We will better convey this in the revision, and also contrast tipping events with gradual changes as per the reviewer's suggestion, with specific reference to rebounding natural populations caused by resilience (negative feedback) in example systems