Public justification (visible to the public if the article is accepted and published):
Reviewers 1 and 3 indicate that while the manuscript is developing on the right track, substantial refinement is still required. Reviewer 2 is happy with the manuscript in its current state.

Reviewer 1 highlights that the paper's key messages could be expressed more clearly. They also question whether the described literature review method actually matches what was done. I would add to their comments that
- there is some inconsistency between the abstract, where only snowball sampling is mentioned, and the main text where both snowball sampling and search terms are mentioned.
- the search terms are very broad and do not mention models or modelling
- a review would conventionally include an appendix with a list of the papers found and how these papers were coded or analysed

Reviewer 3 raises more general questions about whether the field is sufficiently mature for a review and calls for more nuanced writing around some points.

Thank you for these comments. We made a table with a list of all the papers reviewed; however, we do not feel as though it is adding anything significant since it is a subset of our reference list. None of the papers were coded or analyzed. We added our use of search terms in the review process and further explained our review methodology. We have addressed reviewer 3's concerns in the text (see response to reviewer 3 for more details).

Reviewer #1

The manuscript's structure is now more apparent, although a brief explanation at the end of the introduction would still be helpful.

We have added a sentence at the end of the introduction laying out the structure of the text (line 164)

A significant point of obscurity after the revision is the authors' review methods. The authors write they did a Google Scholar search with the following search key:

> (‘human environment system’ OR ‘socio-ecological system’ OR ‘social ecological system’ OR ‘human ecological system’ OR ‘human natural system’) AND (‘tipping’ OR ‘regime shift’ OR ‘bifurcation’)

I did a quick Google Scholar search myself using the tool "Publish or Perish." Counting only articles with at least one citation, I obtained more than 750 results. What did the authors do with that many results? It also seems logically inconsistent to write later that the authors focus primarily on models with social learning processes.

From the initial results, we filtered papers that were dynamical models showing clear signs of tipping behaviour. This has been clarified in the text (line 160)

Furthermore, from the authors' response, I understand that the four key contributions of their work are

1. There are counterintuitive tipping points in human-environment systems.
2. Many tipping points are system-specific in terms of drivers, especially regarding 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.

First, I encourage the author to revise also the paper to that level of clarity.

We have included a box with images to further highlight these key findings with references to them in the parts of the text where these points are discussed.

Regarding point (1), I do not have any intuition regarding rarity-motivated valuation. Thus, I don't find a lower threshold value leading to depletion counterintuitive. And it is pretty intuitive to me that social norms increase the likelihood of a tipping point. I would be interested in which counterintuitive tipping points the authors have considered and why they are counterintuitive.

The high threshold is the one that is counterintuitive for rarity-motivated valuation and thank you for pointing out the error. This is counterintuitive because the social system will be more sensitive when responding to a resource reaching low levels (i.e. by reducing harvesting rates), which intuitively should lead to the persistence of the resource and not increase the likelihood of collapse, which is in fact the case in many models. We have clarified this in the text (line 234)

We are assuming that social norms increasing the likelihood of tipping points isn’t immediately intuitive to the reader who is less familiar with social norms, alternative stable states, and hysteresis.

Other human drivers of tipping points that aren’t immediately intuitive are the coupling strength as well as the relative speed of social and environmental dynamics. Perhaps “counterintuitive” is not the best word in all cases, as many of these drivers of tipping events are not obvious or the reader may not have an intuition, and the wording in the text reflects that.

Regarding point (2), it is to be expected that many tipping points are system-specific. I would be interested in the authors' thoughts on why this insight is one of the manuscript's key contributions to the literature.

This review points out ways in which these tipping points are system-specific. As it is a review, it is not contributing any new information, but summarizing common findings, of which this is one. One contribution that comes from our synthesis which is not prevalent in the literature is showing how some of these specificities can be generalized to input and output-limited (or human-extraction vs human-emission) models, especially regarding relative rates of social and environmental change, as well as coupling strength.

Regarding point (3), which I find noteworthy, I would be interested in why this phenomenon occurs. Do the authors summarize what they observed in their review, or could more be said about the underlying reason? Higher structural complexity could also lead to higher smoothing and less tipping.

We summarize our observations in this review; however, we also note that since there are not many CHES models that explicitly look at the effects of structural complexity, it is difficult to make any generalizable statements about the underlying reason for this observation (also, one of the other reviewers cautioned us against making too many generalizations).

(No comments regarding point (4))

Last, the track-changes file is weirdly formatted. For example, Section "2.1 Coupling strength" is deleted, then appears again as a novel inclusion after one paragraph. It makes the implemented changes look substantial when there is a lot of red in the tracked changes file. However, with such practices, the value of a track-changes file diminishes.

Reviewer #3

General comments

In this revision, the authors have clarified the scope and focus of the review, including changing the title and made some improvements throughout the piece. However, the clarified scope appeared to be quite narrow and restricted for a review. Most of the models reviewed assumed that dichotomous human behavior (i.e., choosing between two strategies) and were built on replicator dynamics (social learning), making the scope of the review rather limited. On models with more strategies, the authors wrote “there are relatively few models that explicitly compare the complexity of social traits and their effect on tipping points” (lines 349). Similarly, on the role of stochasticity as a driver of tipping points, the authors responded in Response to Review that “there are not many papers looking at this in human-environmental models.” These comments led me to wonder if a review such as this was premature – the field is still growing and the models that address important aspects of tipping points in coupled human-environmental systems are yet to be developed. Perhaps, this was why the scope of this review feels narrow.

We reviewed 52 papers that included CHES models with tipping points. We feel this is sufficient. In comparison published reviews (Filatova *et al.*, 2016) reviewed 16 models, (Schlüter *et al.,*, 2012) reviewed ~65 models, however, this was an exhaustive review of CHES models, many of which did not exhibit tipping behaviour, (Tallis & Kareiva, 2006) have 25 citations in total, (Wange *et al.*, 2018) have 34 citations in total. How many models cited does the reviewer think is necessary?

The reviewer chooses to highlight the two areas where we explicitly highlight where papers are lacking in the literature (for the purposes of identifying gaps in the literature, a vital component of any comprehensive review) to make a larger claim that there is not enough literature at large, but in fact, we go through quite a lot of areas for which there are many studies. One of the points of reviews is to highlight knowledge gaps and that is what we do with respect to the diversity of models and inclusion of social heterogeneity, for example.

Specific comments

Abstract: "…gradual changes to the human system" – Why only the human system? What about gradual changes to the environmental system?

This paper specifically addresses human drivers of tipping points in coupled human-environment systems. We make this clearer in the abstract (line 9).

Definition and examples of tipping points. Let me reiterate my position on 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. Reading through the revised manuscript, I found myself still asking the question "what is the small gradual change here?" in a few examples provided. For example, what is the small gradual change in the case where British colonial government imposed their own rules in Northern India? A government imposing new rules does not feel gradual to me. In the ozone example, the authors wrote "Then when policy was passed, industry shifted abruptly to producing CFC alternatives, which led to a tipping point": Again, a policy being passed does not feel gradual to me.

How valid/generalizable are some of the authors' statements? There is a larger challenge of this piece. Many of the reported findings are model-specific – the authors said as much at a few places. It is difficult to judge validity and generality of some of the statements made without seeing the model details. Sometimes, general statements were made with citing only one or two models.

We have broadened our definition of tipping events to include tipping points that have been crossed through larger perturbations. This is in line with tipping point literature (a phenomenon known as shock-tipping or s-tipping) (Halekotte & Feudel, 2020; Boettiger & Batt, 2020) and describes many historical examples of social tipping as the introduction of government policy often plays a role in CHES tipping events. Much of the literature on social tipping references the introduction of policy as a key driver in historical examples (Lenton *et al.*, 2022). We have clarified this in the text (line 85) and included a new figure to illustrate this (Fig. 1).

We synthesized findings across papers in response to the comments from Reviewer 1 who asked for more than a list of specific findings. Where we make a claim to generality, the claim is as general as the cited papers support. We have ensured that we cite all the relevant papers, where such claims are made.

Fig 1: Please provide a reference of the model on which the diagrams between emissions and proportion of mitigators were based so that the interested reader can check out the model's details.

This is our own conceptual model used to illustrate the effects of CHES feedbacks, specifically in relation to social norms.

Fig 2: According to the caption, panel b corresponded to an ODE model. I think many readers would be confused: How can there be "rewiring" in a non-network ODE model? If I'm not mistaken, the ODE model in panel b was an approximation the true network model in panel a. In the context of this review on tipping points, I can understand the inclusion of panel a (the importance of rewiring), but what is the contribution of panel b to this review? It seemed to me that the point of panel b in the original paper (Wiedermann et al., 2015) was to showcase how to mathematically derive a macroscopic approximation of the true network model. It was more mathematical in nature, rather than substantial for the understanding of tipping points. There were only 3 figures in this review – it was unclear why panel b was given priority for the limited space.

We have removed panel b for clarity.

"Input-limited" and "output-limited" models. The terms "input-limited" and "output-limited" were first introduced in Section 2.1, line 175. Are these widely used terms in the field? I am not familiar with them. If so, please provide references when they were first introduced here. If these were something that the authors proposed, I must say that the terms were not very intuitive/accurate for me.

Did "coupling" refer to anything other than "extracting" in "input-limited models" and "polluting" in "output-limited models" in this review? If so, please provide examples. If not, please clarify that (using a general term to describe specific things can be confusing).

We found these terms in the literature but we agree that these terms may not be widely used. We have decided to instead use the terms 'human-extraction' vs. 'human-emission'.

Lines 292-293, the authors wrote "The reduced speed of social change leads to beneficial outcomes as the resource is allowed more time to stabilise as decisions regarding extractive levels occur." – is that true always? What if the initial extraction rate is too high? In fact, I objected to this oversimplified statement in my previous round of review. In their Response to Review, the authors responded that they would "qualify this statement to say what is only true of the papers which we found/surveyed in our review." For me, the statement still sounded quite general, not specific to any particular papers.

We have clarified our statement by starting the sentence with "In these models…"

Lines 364-367, “Heterogeneity in carrying capacities increases the likelihood of sustainable harvesters extracting from a resource with a large capacity, which they can maintain at high levels, eventually convincing neighboring nodes to imitate their strategy (Barfuss et al., 2017).” What prevented the non-sustainable harvesters from going to these large-capacity nodes, maintaining themselves at high levels, and convincing others? Again, is this statement valid only for this model? Is it generalizable?

Non-sustainable harvesters are unable to maintain resources at high levels since they harvest at a higher rate. This has been clarified in the text with clearer reference to the model where it was observed.

Line 495, “…as socio-economic data is often more frequently collected and readily available than environmental data”: I must say that I have heard the opposite more often.

This was true 20 years ago, but arguably not any more due to the advent of digital social data (Salathé et al., 2012). We’ve revised this to say “as social-economic data availability is growing faster than ecological data (and perhaps even environmental data despite growth of publicly available satellite data) on account of the era of digital social data.”

Technical corrections

- The sole paragraph in Section 2.2 was too long. Please consider breaking it down.

We broke this single paragraph into two paragraphs.

- Lines 235-236, “…made up of a single dominant behavior, which is highly dependent on the initial proportion of behaviors in a population” – wasn’t this just describing a stag-hunt game without saying so?

This is not the same as we’re describing population game vs an N player, so the mechanisms show up differently.

- Line 255, “where the environmental state is the proportion of infected individuals,…”: it felt a bit odd to call infected individuals an “environmental” state, not a “human” state, in a coupled human-environmental system model.

We interpret the prevalence of infection as part of the environment that humans must function in, in much the same way that one may speak of an ‘urban environment’ for instance. Behavioural-epidemiological systems are commonly included under the umbrella of human-environment systems.

- Line 269, suggest adding “in the parameter space” after “region” to distinguish it from, say, geographical regions.

We have made this change to the text.

- Line 280, “Whereas…” is not a complete sentence. I believe this was meant to be part of the previous sentence.

We have combined the two sentences.

- Line 327, “ncreasing” is missing “i”

This has been corrected.

- Line 351, it’s not clear to me what “these commonalities” refer to here. Please clarify.

We are referring to models that show the effects of social trait complexity and have clarified this in the text.

- Line 562, perhaps add “incorporating” in front of “social biases”

This has been added to the text.

- Please double-check the use of punctuation. At many places throughout, I believe “;” should be used in place “,” in front of “however”. Alternatively, at these places, you should end the previous sentence with a period, start a new sentence with “However,”

We have corrected this use of punctuation.