Social norms and groups structure safe operating spaces in renewable resource use in a social-ecological multi-layer network model

Max Bechthold, Wolfram Barfuss, André Butz, Jannes Breier, Sara M. Constantino, Jobst Heitzig, Luana Schwarz, Sanam N. Vardag, and Jonathan F. Donges

Final authors' response

Reply on RC1

This paper explores the impact of multiple social norm types (specifically, injunctive versus descriptive norms) on the behaviour of a coupled social-ecological model of a population of harvesters. The authors find a wide variety of behaviours are possible, including states of intermediate sustainability, and the results point to the importance of factors such as the threshold for conforming to a norm, and timescale. I found the results on how system response to policy interventions depends on group size and composition to be particularly interesting. The paper expands the social structure considerably, relative to existing literature. The model is adequately described, and appropriate literature is referenced. The analysis is sufficiently thorough to illustrate the features of the model.

We are happy to read that the referee finds our model appropriately described and referenced and our analysis sufficiently thorough. We are also pleased that they find our model to expand social structure and our results interesting.

Both injunctive and descriptive norms are modelled as a contagion process (occurring at different organizational levels) which would not have been my first choice since injunctive norms are "ought", whereas descriptive norms are "is", and their flavour is therefore very different, but maybe that's just a modelling choice.

This modelling choice was indeed a major point of discussion, as it is particular difficult to model an "oughtness" dimension while keeping the internal dimensions of agents simple enough for the sake of tractability. We acknowledge that the flavour of injunctive and descriptive norms can be different and will attempt to work out the reason for this modelling choice in the manuscript a bit more.

I only have some minor revisions to suggest that I hope might improve the paper:

1. *The paper touches on social learning, but I don't see mention of social learning in section 3.3, where 'individual learning' is described. This should be corrected if it was just an oversight, or the difference between individual learning and social learning in their framework should be clarified.*

This is a very helpful remark and we actually did overlook to clearly separate individual and social learning. We suggest to adjust this by mentioning social learning together with individual learning in section 2. The individual learning component in our framework only

relates to the process in which individuals update their behaviour according to their own harvest, i.e. their personal experiences. Social learning instead is present in our model through the social influence of social norms, where, in a broader sense, individuals learn by observing and being influenced by the behaviour of other individuals and of groups.

2. *I was confused by the results in section 3.4. The intervention strength is the fraction of policy-influenced groups and the x-axis of figure 3.4a has a wide range with many different values for the intervention strength. But if there are only two groups, how can the intervention strength be anything other than 0, 0.5 or 1?*

We agree that the representation of the results and the way this was verbally framed can be confusing here and we will try to clarify this misunderstanding. This can be understood when looking at the computation of the intervention strength: We loop over each of the groups and with a certain probability, which is represented on the x-axis, the groups then are affected and switch to a sustainable attitude. While this will always lead to a constellation where either none, one or both groups are sustainable, as mentioned by the referee, averaged over multiple runs this still yields the depicted percentages of policy-influenced groups on average. We will therefore adjust our wording and describe the intervention strength as the probability for a group to be influenced by the intervention in section 3.4.

3. *Line 569: Is the lack of convergence for high values of Delta T interesting in some way, such as leading to oscillations or other transients, or does lack of convergence mean some unsociological behaviour?*

A lack of convergence for high values of the updating time Delta T relates to the state of the system, which stays in its original state of initialization, as agents (individuals and groups) will on average update their behaviour only in very large time intervals. These intervals get too large compared to the runtime of the model, such that no real dynamics arise with increasing Delta T. As the internal timescale of the model is dimensionless, the interesting features are covered with the comparison of the relative timescales, as shown in the panels, where either the individual timescale is high or low compared to the group timescale. The group updating timescale is not shown in the same range as the individual timescales, since no new dynamics arise and the panels are therefore cropped, also to save computation time. We will lay this out in the description of the results, in proximity to line 569.

4. *There is a good amount of repetition that could be removed to make the paper shorter.*

We will shorten the manuscript by removing repetition wherever possible.

5. *The Conclusion or Discussion section should include a section on model limitations. For instance, model assumptions that might impact results could be highlighted and the relative lack of sensitivity analysis could be mentioned. Similarly, the model assumes that each harvester has their own personal stash, which might be a good approximation for farmers, but even in that case, the decisions of neighbouring farmers influence one another.*

We agree with the referee and will expand upon the existing discussion on model limitations in the conclusion section (also compare to the responses to the comments of the other referee) and give them more visibility in a dedicated subsection "4.1. Limitations & Outlook". Some

important theoretical aspects of social norms are not included in the model or are only aggregated into probabilistic effects via the decision-making function, for example enforcement or sanctioning and norm internalization. By extending the framework to include these theoretical features, the realism of the norm representation can be increased. A sanctioning mechanism could be interesting in particular for a common pool resource setting, which is not analyzed here in favour of a personal pool setting. This choice constrains any additional information flow to the socio-cultural level, as individuals only retrieve information about the harvesting behaviour of their neighbours via the norm mechanism, not via the state of the common pool resource. In this study this was a desired feature, as to set the focus on the socio-cultural processes and with this on the social norms. We additionally assume all individuals to homogeneously have the same threshold for norm uptake, which simplifies the contagion process to some extent, while in reality each agent might be differently susceptible to social influence. It may make sense to consider heterogeneous and complex threshold distributions for the agent population. Ultimately, the agents are embedded in a rather simple network structure, which does not account for any complex social structures. This choice, again, can heavily influence the contagion process, but was taken as to put the focus on the norm mechanisms alone and not the underlying structure. This and the other mentioned limitations can be easily adjusted in further studies though, depending on the context and research question of the study.

We want to thank the referee very much for the helpful comments and remarks to our manuscript.