Response letter “Taxonomy paper”

We thank the two reviewers for their thoughtful, detailed and very helpful comments on the presented framework and the presented illustrative DISCOUNT model. In our point-to-point response below, we reflect on their thoughts and propose according changes to be made to the manuscript in a next step of revision.

Our responses are highlighted in italics. References referred to in the responses are listed at the bottom of the document. Changes and additions to this reponse letter compared to the first version submitted as part of the interactive discussion in ESD Discussions are highlighted in red.

The revised manuscript resubmitted together with this response letter shows our changes to the paper compared to originally submitted version, with additions to the text in blue and deletions in red.

The reviews and our responses refer to the following discussion paper:

> Donges, J. F., Lucht, W., Heitzig, J., Barfuss, W., Cornell, S. E., Lade, S. J., and Schlüter, M.: Taxonomies for structuring models for World-Earth system analysis of the Anthropocene: subsystems, their interactions and social-ecological feedback loops, *Earth Syst. Dynam. Discuss.*, https://doi.org/10.5194/esd-2018-27, in review, 2018.

Reviewer 1 (C. Herrmann-Pillath)

This is a timely and much-needed effort at creating more realistic and inclusive/ integrative models of the Earth system that meet the challenge to model the complexities of human action in various systemic contexts (such as economy, society and so on). It is far too easy to criticize such endeavours because they necessarily offer many points of attack, being the first steps in this direction. This would be unfair. In my review, I try to find a proper balance between necessary criticism on the one hand, and endorsement and encouragement on the other hand. I am writing from the position of an economist working in the field of evolutionary, institutional and ecological economics (hence, clearly ‘heterodox’) who always injects a strong dose of philosophy in his work. That is, I am an outsider to the modelling community in the Earth and Climate Sciences. I will always combine general observations with specific comments on the discount model. Yet, my focus is on principled issues, hoping that this might provide different perspectives than those within the modelling community. The authors explicitly recognize the philosophical dimensions of their work; hence I start out from there, and I congratulate them for showing the grit to use the notion of ‘World-Earth system’
(p. 2). However, this immediately raises many philosophical questions.

> *Outside scrutiny is very much welcomed! The rationale for the paper is to support transdisciplinary knowledge integration, so the broader the scrutiny and critique, the more robust the usability for our target communities.*

1. In which sense can we separate ‘Earth’ from ‘World’? Although we must assume that there is an ‘Earth’ interacting with ‘World’, the ‘Earth’ is only accessible via the sciences which construct it as a ‘World’. That means, we must recognize the fundamental fact of our incomplete knowledge about ‘Earth’. In other words, **a complete model** of the ‘World-Earth system’ must be reflexive in the sense of including the Earth Sciences (and others) as endogenous generator of our scientific conceptions of ‘Earth’. There are direct implications for the subsystems: For example, our knowledge about the biophysical mechanisms will impact on public perceptions of policy issues, and so on (which the authors aim at modelling explicitly, indeed). These will feedback on funding Earth Systems sciences, and hence will determine how ‘Earth’ will appear to us in future times. I think that science is an essential part of the ‘social-cultural taxon’ and cannot just be imagined as being exogenous. But if it is endogenous, ‘Earth’ is endogenous.

We appreciate these comments and critiques because they help us to refine our transdisciplinary work, which increasingly extends beyond the Earth and Climate sciences modelling community. What we do is not about representing an ideal or “complete” world view, but is aimed at making things explicit that have so far been excluded from Earth system schema. We are proposing a taxonomy for models-in-use - an applied epistemological standpoint - not a taxonomy for worldviews - ontologies, even though of course the ontological stances of modellers have consequences for the “real world”.

However, the feedbacks discussed by the reviewer are indeed important and are, for example, very much in the core of ideas that have been expressed on the nature of World-Earth in the Anthropocene as Gaia 2.0 (Lenton and Latour, 2018) or the emergence of a global subject (Schellnhuber, 1998, 1999). Scrutinizing the essence of such feedbacks using simulation models is clearly an ambition of the type of the science we describe in the manuscript, but needs to be approached step-by-step starting with reduced and stylized models.

We have added the word ‘Epistemologically’ at the start of the description of World-Earth models, to emphasise that World and Earth are only distinguished at the epistemic level, not ontologically. We have also included knowledge in the listing of ‘World’ aspects that are currently missing from Earth system schema but that we want to characterise using our taxonomy.

Looking at the discount model, this seems implicit to the role of the parameters p and q. Although these are essential for driving the specific model dynamic in determining the probability of switching discount rates, they are not explained in any way. To be frank, I have difficulties in understanding their meaning. What exactly is ‘curiosity’ of a country? What is ‘myopic rationality’? Many people would just think that the latter means, well, steep discounting of the future. Another definition would be the length of the time horizon, which differs from discount rates applied within that horizon. Thus, besides from the need to give precise definitions these parameters and explain how they could be measured, …

We appreciate these detailed thoughts about our illustrative model and its parameters. We believe, however, that there are some misunderstandings both regarding our aims *in presenting this model and also regarding the pair of parameters you mention, p0 (we take it, this is the one you refer to as p) and q. In short, we want to show properties of the model in a way that is accessible to readers from whatever field of 'World' or 'Earth' system modelling they come from, so we want to use descriptive rather than specialist/technical names for key parameters.*

In our illustration, we were interested mainly in the dynamics of social learning, so we comment briefly on the influence of parameter l in the model. We call l the "learning rate" since it is the rate at which countries update their discount factor by social learning. They do this by comparing their welfare with others', and then switch to the other discount factor with some probability. The effect of varying this parameter is shown in Fig. 4, with l being small in the top panel and large in the bottom panel. As detailed in the Appendix, the illustrative model has about a dozen exogenous governing parameters, of which p0 and q are far from being the most important ones. As can be seen in Eq. 6, p0 and q influence the frequency at which discount rates are switched. As the reviewer notes, these two parameters control the sigmoid-shaped dependency of the switching probability at each of the updates that occur at rate l. p0 is the probability of switching in the case that both countries' welfares are equal. We named p0 a "curiosity" parameter because a switch of this kind can happen without any social learning by imitation. This parameter can be interpreted as a country's "look-see" exploration of a different discount factor without expecting a welfare increase.

We now make this clearer in the manuscript text:

> *Below equation (A13): "p0 can be interpreted as a measure of a country's curiosity-driven exploration of a different discount factor without expecting a welfare increase. The larger p0, the more frequent switches will occur, but in both directions between the two candidate discount rates, mainly generating more variance and fluctuations that can be seen as a form of "noise"."*

The parameter q is the steepness of the probability curve at the point where both countries' welfares are equal. In other words, it is the marginal probability at this point w.r.t. welfare differences. In the extreme case where q goes to infinity, the probability of switching is either one (if the other country has higher welfare) or zero (otherwise). One could interpret this as a "rational" but "myopic" switching behaviour, because the agent does not anticipate the future in their decisions.

We now make this clearer in the manuscript text:

> *Below equation (A13): "q can be interpreted as a measure of a country's rationality, because the probability of switching to the other country's discount rate is higher if the other country has higher welfare (and zero if that is not the case) – but it is a myopic rationality, because the agent only takes its present welfare into account. The larger q, the faster discount factors will converge to the one currently generating the largest welfare."*

In our terminology, myopia vs. farsightedness (the parameter q) refers to the agent's anticipation of future welfare, while patience vs impatience (the application of a large *value for δ) refers to whether an agent cares about and values future welfare in their decisions. So, in our model, a myopic country may switch from patient to impatient because at that moment, the impatient countries seem to fare better and the country does not anticipate (due to its myopia) that this leads to a trajectory on which welfare will later be smaller for impatient than for patient countries.*

The influences of p0 and q on the speed of the learning dynamics are more indirect and subtle than the influence of the learning rate l. The reviewer is right to observe that the equations given in the Appendix, though complete, are not sufficient to demonstrate this to the reader without access to model runs. Given that our intention in showing this stylized model is just to illustrate the taxonomy and to demonstrate a small point – that the speed of CUL→ CUL learning dynamics is important for coevolutionary model outcomes – we hope that the clearer explanation of our lay-terms for parameters in our World-Earth illustration model and the clarifications about the behaviour of the parameters in the Appendix text and will suffice.

In order to not distract the reader with too many parameter symbols in the main text, we now removed the equations for the auxiliary quantities s, phi, D and P from there (they appear in the Appendix already) and rather explain the meaning of those terms verbally, so that now also the parameters p0 and q no longer appear in the main text.

… I suggest that they refer to the endogeneity of science. Just consider Trumpian America: the recent issue of 'The Economist' has an article about 'Swamp Science' at the EPA. This is actively reducing 'curiosity', it seems to me. In other words, I think that if one uses the 'World-Earth' duality, one faces the challenge of treating human knowledge about 'Earth' as endogenous. There is no external stand-point of the model-builder. Another excellent example for this problem is the treatment of 'damage' and 'welfare' in the discount model. It seems that the authors think that there is an objective measure of welfare and damage. But we know that this is one of the most difficult and disturbing aspects of IAM, namely that the damage function is endogenous and depends on the discount rate. That is why some economists now even take the very radical step to build their models without damage function (Llavador, H.,J. E. Roemer und J. Silvestre (2015): Sustainability for a Warming Planet. Cambridge University Press). But behind this is the simple, but deeply philosophical problem that neither welfare nor damage can be assessed from the standpoint of an external observer. Apparently, the authors are aware of this, as on p. 17 we find the expression:
s(C) (ENV → MET → CUL). But that implies that social learning may not only happen via imitation of discount rates, but also via diffusion of valuations, or, 'worldviews'.

We appreciate these concrete examples, and share much of the reviewer's concern about the state of science in society.

We now indicate the endogeneity of science in the abstract:

> *Page 1, line 11: "(ii) socio-cultural, dominated by processes of human behaviour, decision making and collective social dynamics (e.g., politics, institutions, social networks, and even science itself)*

> *We have also added a reference, Yearworth and Cornell 2016, that discusses issues around the shifting role and stance of the scientist/model builder in different sustainability contexts:*
>
> > *Page 8, line 22: Notably, the CUL taxon also includes processes of digital transformation and artificial intelligence that increasingly restructure and shape the socio-cultural sphere of human societies. It also provides a locus for debating the challenge of reflexiveness in science, especially in fields where modelling plays a vital role in shaping knowledge and action (Yearworth and Cornell 2016). For instance, future World-Earth modelling will have to grapple with ways to recognize Earth system science as an endogenous generator of scientific conceptions of 'Earth'.*
>
> *We are confident that our taxonomy can be useful in diagnosing the shortcomings of real-world science and models-in-use - as in the reviewer's examples given above. But fundamentally, our taxonomy is intended to provide a system for categorizing and classifying the structure of models, their subsystems, and indeed also their couplings and hybridizations as efforts for improved representation and understanding of "whole" Earth system processes and phenomena. It is not a recipe for a new model, nor is it a taxonomy of issues in the real world.*

There is no external stand-point of the model-builder:

> *This is a very good point. One reason for this taxonomy is that some new model developments articulate this point explicitly, while many others do not. And also some new model developments that are being coupled to existing "Earth" models involve an internal positioning of the model builders – e.g., participatory modelling, etc.*
>
> *We have added text in section 1.1 that explains that by making this taxonomy, we develop some initial tools and terminologies for systematically challenging model builders and model users to be clear about their social/cultural and perhaps also their epistemological/axiological standpoints.*

2. The next foundational question is whether there is 'ONE World', which seems implicated by putting both World and Earth together into one 'system'. Obviously, this is not just annoying sophism: There are many philosophers who cast doubt on that (just mention the German philosopher Markus Gabriel with the provocative book title 'Why there is no World' 'Warum es die Welt nicht gibt'). The serious argument behind it ties up with the previous: Science constructs worlds, and there is no necessity that these are just 'one'. This is even more evident if we consider human worlds: Actually, it is the CUL taxon that creates the 'worlds'. I think that it would be most helpful for the authors to look at Bruno Latour's recent work on Gaia and the more general work on 'modes of existence' (see http://modesofexistence.org/). Latour distinguishes between different ways to bring 'worlds' into existence, such as religion, economics, or law. The philosophical backing is different criteria for truth. Personally, I do not fully endorse his approach, but it connects with many other philosophical streams that, most generally, analyse 'social ontology'. One of the most concise approaches is Searle's distinction between 'mind dependent' and 'mind independent facts'. Well, 'mind dependent facts' are –

facts. That means, they have the same ontological status as other ‘facts’ conventionally treated as such by the sciences.

> *We really appreciate this opportunity for cross-disciplinary critique. We would argue that we do not need to think there is “one world” (an ontological position) in order to want to combine models of different worlds (an effort at handling some specific challenges arising from epistemological plurality) in order to seek better scientific understanding of aspects of the (real!) world. And yes, we agree with the more general point that models are things in the world that may have causal power (agency in Latour's terms).*
>
> *Tackling the epistemic point first – in short, each model of the system contains one version of "world" - and the modeller knows this is a simplified/stylized representation with a partial viewpoint. Different models can allow for the comparison of different versions of "world". Later, some "metamodel" could allow for several versions of "world" in one model, but this only makes sense if they interact, e.g. in that different agents are guided by different conceptions of "world".*
>
> *This returns us to the ontic challenge - in the context of the models we want to “taxonomise”, one can say that what we call "world" is the modeler's view of the world, while the agents' views of the world are likely to be part of the CUL-related attributes of the agents. But that is not necessarily the only option. Some researchers are applying Latour’s ideas on agency to non-human things (like water in irrigation systems), which our taxonomy would flag up as requiring specific attention to sub-system interactions (for instance, perhaps modelling different CUL-MET behaviours than human agents).*
>
> *We have added the following text to section 1.1:*
> *We want to emphasise that this taxonomic approach does not presuppose that there is “one world” (an ontological position) when models of different worlds are combined. Instead, we argue that a taxonomy can help to focus modellers’ attention to the ontological and epistemic commitments within their models. This approach opens Earth system analysis to deeper dialogues with proponents of non-human actors as shapers of the world (Latour 2017, Morton 2013), or even the possibility of no world at all (Gabriel 2015).*

Turning to the discount model again, a core question is the ontological status of institutions. From the viewpoint of institutional economics and Searlian philosophy of institutions, the authors appear to be imprecise in treating institutions because they seem to suggest that a clear boundary can be drawn between CUL and MET.

> *We did not mean to suggest a rigid boundary. Although we only talk explicitly about institutions in the CUL sections, we mention kinds of institution (e.g., economic and technological systems) in MET. In fact, as in all taxonomies, it is often not clear where to put a given process without having information about the underlying basis. Still, we believe a distinction between rather socio-cultural processes (let’s say mind-mediated) and rather socio-metabolic processes (expressed as material flows) is helpful in modeling, which is what this paper is about. Pragmatic assignments of boundary cases*

to either CUL or MET are typically not harmful. In the discount model, the relevant institutions are represented in a very stylized way in two components, (i) the game-theoretic submodel and (ii) the social learning submodel. The first specifies which domestic emissions will arise at every point in time as a kind of "immediate" Nash equilibrium where each country is assumed to act as one consistent player, as is standard in the literature on international environmental agreements that this submodel is taken from. This of course means that we implicitly assume strong decision-making institutions within each country. We hence interpret this submodel to basically belong to the CUL taxon, where part of its inputs (namely the evaluation of the environmental and socio-metabolic states) come from the ENV and MET taxon, and the resulting emissions caps have effects in the MET taxon. The second submodel is also adapted from standard models, this time from the modeling literature on social learning. Implicitly it assumes some global monitoring and communication institutions that allows "observing" other countries' discount factors and evaluations in some way. We interpret this to be a mainly socio-cultural process as well. Hence the reviewer is right in that we treat institutions "imprecisely" in the same way that the game-theoretic literature on international environmental agreements and the modeling literature on social learning do.

Granted, there is the overlap in the diagram, but what does that mean? I think the problems crystallize in the question how to deal with the economy:
Is the economy a 'world' of its own? It is a very specific and very comprehensive institutional structure that creates 'realities' to which we need to adapt, in the eyes of many (the infamous TINA principle).

Here we reiterate that our aim is to develop a taxonomy to help classify and structure modelling approaches. We are not prescribing how global models should be configured or what they should contain, nor are we critiquing the power of some global models in shaping contemporary economic realities.

At the start of section 2, we highlight how different institutions can play out as different economic realities. In sections 2.2 and 2.3, we indicate some of the diversity of ways of dealing with the economy, and in section 2.4 we identify some research areas that deal with adaptive change and transformation. So we think our taxonomic approach can help in tracing how and where "the economy" (which also is not universal/monolithic!) actually features in contemporary analysis of Earth system dynamics sensu lato.

Accordingly, some economists think that the discount rate should follow the market interest rate, as this is the only way to define a 'collective discount rate' apart from individual time preferences.

This might be true, but we tend to disagree with those economists then and rather follow the illustrious group around the late Kenneth Arrow who, when asked by the EPA what discount rate a country should use for long-term projects like fighting climate change, basically said that this is a normative choice: "Many of us regard the Ramsey approach to discounting, which underlies the theory of cost-benefit analysis, as a normative approach. This implies that its parameters should reflect how society values

> *consumption by individuals at different points in time; i.e., that δ and η should reflect social values." (Arrow, Kenneth, et al. "How should benefits and costs be discounted in an intergenerational context? The views of an expert panel." (2013)). In our model, we make the specific and certainly highly debatable assumption that this choice is made by social learning between countries, mainly to illustrate what effects such "non-economic", "social" dynamics could have.*

Markets are real, everything else is 'subjective'. This 'objectifies' markets just in the Searlian sense. Indeed: Which other way do we have to generate a 'country' discount rate in an empirically meaningful way? The authors introduce this with a sleight of hand, but this is a very strong ontological projection! I think the discount model cannot simply take a 'country discount rate' for granted – unless that would be the discount rate that governments apply in their policy making framework.

> *You are of course right, but the model's discount rate is indeed used in the way you suggested, as the one the government uses to decide their domestic emissions caps.*

In my own work, following dystopian statements by experts such as Pyndick (Pyndick, R. S. (2013): Climate Change Policy: What Do the Models Tell Us? Journal of Economic Literature LI (3), 860-872), the discount rate is the central parameter that manifests the mutual irreducibility and incommensurability of economy and ecology, which becomes manifest in the methodological troubles of IAM. The authors refer to the fact that economic models always refer to monetary variables (p. 8) – that is what constitutes a 'world' in the Searlian or Latourian sense. Thus, I wonder whether a minimum requirement for building a discount model is to include a model of the economy that generates a reference interest rate as a 'social fact'.

> *Our aim in presenting this conceptual illustrative discount model is precisely that - just to show that discount rates may differ between countries and over time, contrary to the predominant assumptions of IAMs. Building a detailed model of financial markets would only be only necessary for answering research questions related to aspects of those markets.*

3. I can only hint at these issues here, but the central question arising from this is how to justify the assumption of an integrated 'World-Earth system'. I would bet on a 'Multiple Worlds-Earth....' – system? Again, Latour has a strong point in rejecting the very notion of a 'system' as that would imply integration, coherence, and so on. In more practical terms, that leads us to consider the question why the authors did not follow a more traditional (hence probably outdated) approach in distinguishing between different 'systems', such as 'the economy', 'the society' and so on, which are not integrated, but may stay in fundamental tensions and contradictions with each other.

Again, this is not merely a philosophical issue. In the Anthropocene literature, many critics point out that it is misleading to confront an abstract notion of 'human system' with 'Earth system' as this papers over the fact that the 'human system' is deeply fragmented and conflict ridden, and hence fundamentally politicized in a most general sense (as an exemplary work, see Bonneuil, Christophe and Jean-Baptiste Fressoz. 2017. The Shock of the Anthropocene: The Earth, History and Us. Verso London New York). This is the real challenge that the authors must meet: Catching this complexity in ONE model. Given these caveats, I think that a comprehensive approach would indeed be well advised to go back to ontological fundamentals. Mario Bunge's formal ontology seems still unsurpassed to me (Bunge, Mario (1979), Treatise on Basic Philosophy, Volume 4. Ontology II: A World of Systems, Dordrecht: Reidel). Perhaps this would allow for formalizing 'multiple worlds'. I am not sure how a model would look like. Perhaps it would be a set of modules that are only loosely integrated, with certain thresholds that block interaction below them, under normal circumstances. That means, for example, the 'economic world' would generate a discount rate that would be isolated from other worlds, until a catastrophe happens which triggers sudden spillovers. Modules would run separately for longer periods, until connections become activated that would generate sudden changes across the modules. The discount model might consist at least of two modules, one economic, the other cultural. The latter would include public opinion, value changes, and so on. I think this is what we observe in reality: Public opinion might shift towards de-carbonizing the economy, but the economy seems to move towards sustaining it endogenously (think of the continuous process of postponing depletion of fossil fuel reserves) (see Covert, T., M. Greenstone und C. R. Knittel (2016): Will We Ever Stop Using Fossil Fuels? Journal of Economic Perspectives, 30(1), 117-38). Then, everyone is shocked learning that Germany even increases CO2 emissions!

> *This comment raises several points that we condense into the following responses: "The central question arising from this is how to justify the assumption of an integrated 'World-Earth system'" – "This is the real challenge that the authors must meet: Catching this complexity in ONE model" – "Perhaps this would allow for formalizing 'multiple worlds'".*
>
> *Our focus here is on integrated models - and the epistemology does not presuppose the ontology of one integrated World-Earth system. There are many approaches to representing aspects of the complex real World, as we indicate in our discussions of subsystem interactions (section 3). The many different modelling efforts that our contributing communities are making are not competing for ultimate victory over each other - they are capturing different aspects of the complex world.*
>
> *So our underlying research aim is not to make One Model (to rule them all...), but to highlight what a new family of models might be. We are focusing on the specific toolkit of simulation modelling. Our approach already allows us to disaggregate and examine the complex social worlds much more than traditional IAM approaches, enabling a more politicised view on societies and economies by setting concepts and contexts into taxa. In other strands of work, we are trying to represent multiple economies and diverse "systems" coexisting. Our taxonomy can permit "intersectionality" in this diverse analytical context, improving the transparency that we think is needed for model-use in real world decision contexts.*
>
> *We explain in the concluding section that we are actually arguing for epistemological pluralism - we are seeking to enable productive dialogue and interaction between this diversity of World modelling approaches and the biophysical Earth representations that we already have.*

4. Another issue is the principles how to build a taxonomy. The authors work with causal concepts, such as 'mechanisms' and 'feedback loops'. This is a basic methodological

requirement, for sure. But it implies that the taxonomy must be based on theories and hypotheses about causality. This creates a very, very high benchmark. Taxonomical work is often more modest, such as biological taxonomy that is based on notions of descent, similarity and so on, without classifying patterns of underlying causalities.

This is one advantage of working with old fashioned systems categories: They offer simple criteria of building taxonomic classes, such as treating the economy as a system in which money and monetary values are coordinating media, and then subdividing this in other systems, such as sectoral, regional or developmental. That would imply for the discount rate, for example, that one would minimally distinguish between 'North' and 'South' as economic systems which have different reference rates (maybe North can afford to be more 'patient', South not). But if one refers to causality, every single example given by the authors is a Pandora's box out of which theoretical controversies and cross-disciplinary battles emerge (in fact, the authors often mention that many disciplines deal with the various mechanism that they subsume under their categories).

Just take one: there is the ENV-CUL taxon, and cultural identity as 'sense of place' is mentioned. I assume that many cultural geographers would immediately protest. Does that mean that there is environmental determinism regarding culture? I do not think that the authors believe that, and after all, there is also the CUL-ENV taxon. But how can we deal with this in building a model for one cultural element, 'sense of place' (which certainly is important in the context of modelling migration, induced by climate change)? In this single example, very complicated theoretical issues regarding causality, its direction, and divergences between different disciplines are involved. Which of the competing positions in the literature would be selected to model this single element? Clearly, that would directly affect the taxonomy. In other words, the taxonomy is built on a minefield of theoretical and cross-disciplinary controversies.

> *We view this issue from a position within a modelling community using big, rather opaque "comprehensive" global models - and we certainly do encounter controversies and cross-disciplinary tensions, if not necessarily all-out-battles. In other words, we are in the minefield! Every new collaborative effort needs to face these theoretical debates, and this effort is helped when we are alert to their historic and ideological loadings.*
>
> *The global models we and our colleagues use have causal power in the world, so our effort here has a partial aim to illuminate the innards of these tools. But we do not agree that the taxonomy needs to have an a priori set of theories and hypotheses about causality.*
>
> *We now state clearly in Section 1.4 that the causal narratives in existing models are our starting point because they are necessary for and explicitly encoded in simulation modelling - and our classification lets us interrogate them more systematically and exposes them explicitly, as the CUL-ENV/ENV-CUL example above shows.*
>
> *In the conclusion, we have borrowed the reviewer's excellent observation about the Pandora's box we face.*

The difficulties multiply if one considers larger complexes, such as 'the economy'. I appreciate very much that the authors recognize complex feedbacks such as the cultural shaping of preferences, or their technological determinants. But all this is object of deep controversies between the different disciplines. ‘The economy’ looks very different in the eyes of the economist or the sociologist. Which position should be included in the ‘World-Earth-System’ model? Even the most basic assignments to taxa would be affected by this. For example, many sociologists would assume that ‘the economy’ belongs to the socio-cultural taxon, and that repercussions on the other taxa work via technology.

> *Nicely, there is no need at all to place “the economy” (whatever that is) into MET or CUL, since our taxonomy is about classifying individual processes in specific models. Thus the question of whether a certain process is “economic” or (otherwise) “social” or “cultural” will depend on how it is framed. An economist’s model would probably place the real-world process of decision-making about childrens’ education into the MET taxon, and model it as a rational market-driven process linked to resource productivity, while a sociologist’s model would probably place the same real-world process into the CUL taxon and model it as being strongly influenced by traditions.*
>
> *The taxonomy approach means that things that were previously included in models as opaque and unquestioned systems can be unpacked and critically examined. Model users who were not the model builders would really benefit from knowing (to take the example above) whether the representation of an education decision process in a model is in MET or CUL.*

My comment is already very long, so I stop here. I am afraid that my comments appear overly destructive, but this is not my intention. I hope that the authors take it as creative stimulus. Yet, my general point is serious: If the taxonomy is about causal patterns, it needs to be built on theories. The authors assign various elements to different taxa with levity, without considering that all this is subject to many competing and often contradictory theories across the entire universe of disciplines, not only the sciences, but even the humanities. Evidently, their model must also be a model of cross-disciplinary relations, to avoid unjustified essentialization and hypostasis of assumed mechanisms, loops etc. This is, well, the ultimate ‘Theory of Everything’ coming along in the disguise of a model? I hope that we will continue with this discussion and hope to learn from the authors’ response, as well as from the comments of other reviewers.

> *We do take these comments as creative stimulus! This thoughtful attention has opened discussions about our own philosophical and research-ethical positions. We hope that our engagement with these points in our response help to “normalise" practical philosophical discussion in Earth system modelling contexts.*

Reviewer 2

This manuscript addresses a highly relevant topic on how to foster the integration of the social dimension in mathematical and computer models up to the planetary scale. A specific focus is set on how to take into account the relevant social-ecological feedbacks. It fits well in the scope of the journal. The proposed taxonomy (including the types of interactions between the subsystems) takes adequately the existing range of global environmental change models into account and classifies them. In addition it enables to reveal subsystems and interactions which are important but underrepresented in global scale models. Therefore this paper can serve as a fruitful starting point for a more structured approach to guide the future development for such types of highly needed models. However I see major points which have to be addressed in a revised version of this manuscript:

1. The three guidelines in constructing the taxonomy are partly not well explained, e.g. what do you mean by “compactness”. In addition, have these guidelines actually been tested? I ask that, because they are afterwards only mentioned in the conclusion section again.

> *We have resequenced and explained these guidelines more explicitly, substantially expanding the text at the start of section 2 in the manuscript, along the lines of the following considerations:*
>
> *Defining* ***compactness*** *- we want a “top-level” taxonomy with as few classifications as possible, covering the scope of co-evolutionary modelling research parsimoniously and in a self-containing way. A contrasting approach would perhaps have been a kind of family tree, where classifications can expand almost ad infinitum because there’s no basis for limiting the number of offspring. The discount model is a kind of test of this guiding principle - the loops or flows between taxa do not make us need to rethink the whole structure.*
>
> *Defining* ***compatibility*** *with existing research fields. This is linked to Reviewer 1’s comments on philosophies. The taxonomy isn’t defined with a single “ism” in mind. It clearly has realist foundations as the vast majority of modelling methodologies do, but it is (perhaps surprisingly) agnostic about how this relates to different epistemological stances. The reason we care about this guiding principle is that we want to explore how to bridge across currently very distinct modelling approaches as well as to trace how the techniques chosen/used can relate back to the theories, assumptions, and framings of the contributory disciplines. Collins and Evans’ book Rethinking Expertise has some interesting discussions on the “translational” expertise that is needed to bridge disciplines, but stops short of thinking of how this kind of expertise exists outside the head and presence of the researcher: Our taxonomy both demonstrates and provides a way that we as a transdisciplinary community can learn from each other, not just from own experience.*
>
> *Defining* ***operative capacity*** *for the integration and combination of modeling methodologies - we start from a very dominant existing divide between “natural sciences” (e.g., fits mostly ENV and parts of MET) versus “social sciences” (e.g., fits mostly CUL and parts of MET), so our taxonomy needs to be able to include things that fit into those existing classifications - we want to be able to draw on the wealth of*

> *previous modelling efforts – while expanding to be more inclusive on both those fronts as well as allowing more differentiation and permutations of approaches within and between and beyond the old natural/social sciences divide.*
>
> *In section 2.1, we examine the issues of compactness in more detail, illustrated with geosphere/biosphere interactions. We have also framed Section 3 more clearly along these guiding principles.*

2. The type of content of the subsections 3.1. – 3.9 is differing. I would suggest to offer results in each section for the same set of questions, such as: examples, what is reached, what are open challenges, at which levels are the current models prevalent (global yes or no).

> *Point well taken!*
>
> *At the start of Section 3, we clarify that the content presented in the subsections necessarily differs, but that we have aimed to give comparable information. We have also reframed and restructured Section 3 for a more consistent presentation of the subsystem interaction taxa.*

3. Make more explicit what the added value of the presented example model for this paper is. Is it just to illustrate how a simple model could look like which includes interactions from almost all classes of the subsystems of the taxonomic scheme?

> *This is indeed the main purpose since we feel that a crisp example such as this is needed to give the reader a feeling of where processes may be placed. It illustrates that this combination of models can still lead to macroscopic approximations that may be studied with established dynamical systems methodology. However, we also believe that our exemplar model could be of interest in itself for some readers since to our knowledge it is the first published model which endogenizes the choice of discount factors used in climate policy.*
>
> *We have added more text at the start of Section 4, explaining that this stylised model covers all classes of real-world processes that appear relevant in major global feedbacks, and also shows that in order to capture these processes and feedbacks, one may have to combine quite different modeling traditions (as we discuss in the manuscript's Sections 1 and 2).*

4. The text includes a number of assumptions where references which underpin the respective statement are missing (such as P2L21, P6L17ff), more examples below. Please add references.

> *Thank you for spotting this, we will add references to underpin these and other statements. We also updated references and added references to relevant recent papers published since the original discussion paper was submitted.*

In addition I have further comments for the different sections:

Abstract:

- The term “World-Earth model” is not familiar to the readers. It is only explained later in the introduction. Perhaps a short explanation in the abstract could be helpful. The same holds for the term “higher-order taxonomies”.

> *At risk of sounding Latourian, the “World-Earth” framing signifies a growing attention (in many research fields) to the human world and biophysical earth as deeply connected, or a dynamic hybrid. Many very influential models were designed for World and Earth as separate domains. We are interested in the development of coevolutionary model toolkits that allow us to represent and explore this mutually entwined view. Changes have been made to the abstract to clarify both these terms.*
>
> *We have added "World" to line 3 of abstract:*
>
> > *New approaches to global modelling of the human World are needed to address these challenges. The current..."*
> >
> > *Also later: "World-Earth models capable of simulating the complex processes of the Anthropocene are currently not available. They will need to draw on and selectively integrate elements from all the existing modelling approaches..."*
>
> *To clarify "higher order taxonomies" in the abstract, we have split and expanded the sentence:*
>
> > *We show how higher-order taxonomies can be derived for classifying and describing the interactions between two or more subsystems. This then allows us to highlight the kinds of social-ecological feedback loops where new modelling efforts need to be directed.*
>
> *At the end of the introduction to Section 3, we add a further explanatory note:*
>
> > *“Furthermore, possible extensions of our taxonomic approach to classify feedback loops and more complex interaction networks between subsystems are discussed (Sect. 3.10).”*

Introduction:

- P3L5-6: The second half of the sentence is not comprehensible to me.

> *“the characteristics and interactions of social and biophysical subsystems are often not explicit to each other, if they are recognised at all” - changed to “The problem for both scientific integration and real-world application is that the characteristic basis of the interactions of social and biophysical subsystems is often not explicit in current models. Often, the interactions between these subsystems are not recognised at all.”*

- P3L14: I think a number of ecologists (e.g. behavioural ecologists) would doubt that “laws of nature” govern for instance how animals behave. I understand the point you want to make. Perhaps add a footnote which points this out shortly.

““natural laws" of physics, chemistry or ecology (e.g., atmosphere and ocean as governed by the laws of fluid- and thermodynamics)” - changed to “the “natural laws” and generalizable principles of physics, chemistry and (to some extent at least) ecology (for example, atmosphere and ocean circulation as governed by the physical laws of fluid and thermodynamics)”

- P3L27: delete “computer”? According to L24/L25 it holds for both types of models: mathematical and computer simulation models.

We have adjusted the sentences accordingly.

- P4L11: I suggest to give a reference to the term “safe and just operating space for Humanity”.

We will add references to Rockström 2009, Raworth 2012 and Dearing 2014 here.

- P5L2: The term “mechanism” is defined, but never used in the paper.

In several places, we refer to feedbacks, which are an important kind of mechanism in World-Earth system analysis. We have inserted the missing word mechanism in places in Section 1 where it clarifies our points about modelling feedbacks.

Section 2:
- P6L17: “Growing reliance on model-based insights for global decision making”: please give references for that.

We will add references to:

- *Calder M et al. 2018 Computational modelling for decision-making: where, why, what, who and how. R. Soc. opensci .5: 172096. http://dx.doi.org/10.1098/rsos.172096*
- *National Research Council. 2007. Models in Environmental Regulatory Decision Making. Washington, DC: The National Academies Press. https://doi.org/10.17226/11972.*
- *Sort of also a theme in: Rounsevell, M. D. A., Arneth, A., Alexander, P., Brown, D. G., de Noblet-Ducoudré, N., Ellis, E., Finnigan, J., Galvin, K., Grigg, N., Harman, I., Lennox, J., Magliocca, N., Parker, D., O'Neill, B. C., Verburg, P. H., and Young, O.: Towards decision-based global land use models for improved understanding of the Earth system, Earth Syst. Dynam., 5, 117-137, https://doi.org/10.5194/esd-5-117-2014, 2014. - although they are focused on improving models by including decision processes more than improving decision processes that already rely on their models.*

- P7L15: Add a reference for “deep future” studies. Where comes the term from?

We now add a couple of references to clarify our choice of this term, an analogy with ”deep past”. Curt Stager used the term in a popular science book - Deep Future: The Next 100,000 Years of Life on Earth - but we were not thinking of that text as such.

> *"Furthermore, as it becomes clearer that palaeoclimate models designed for study of the deep past can play a vital role in "deep future" studies of human-controlled processes in the Anthropocene (e.g., Zeebe R E and Zachos J C. 2013 Long-term legacy of massive carbon input to the Earth system: Anthropocene versus Eocene. Phil Trans R Soc A 371: 20120006.http://dx.doi.org/10.1098/rsta.2012.0006, also refs in Steffen et al 2018 Hothouse/Stabilized Earth paper) …"*

- P7L24: You describe "these models have tended to be small-scale, context-specific ...". There are counterexamples (such as LPJ or large scale forest models). Hence, I suggest to mention these examples.

> *Yes, we now mention them. We clarify our intended distinction of models of ecological dynamics as such (i.e., interactions between living organisms) versus models representing the physical dynamics of ecological processes and structures.*

- P8L7: "Our approach can bring much-needed clarity and transparency about the role of such models ..." – this statement is not substantiated in the paper to my point of view.

> *Edited to say "We suggest that our approach can bring much-needed clarity and transparency about the role of such models in understanding the World-Earth system (cf van Vuuren et al 2016)".*

Section 3:
- P9L29: It could be helpful to add a short paragraph in this paper how the taxonomy have proven to be helpful building the copan:CORE framework.

Good suggestion, we will do that and explain how the CORE framework has been constructed following the taxonomy very closely.

- Figure 2: I find it not self-explanatory, why certain interaction examples are assigned to a certain category, for instance why "needs" is in MET CUL. I suggest to explain at least those in the text.

> *A brief addition on material needs has been added to Section 3.6 . We also explain now in the caption of Fig. 2 that the examples given in the matrix elements are not to be seen as absolute classifications, we see them as indicative of our own research perspective.*

- P12L23: Make sure that all abbreviations in the text are once spelled out and the abbreviation is written in parentheses (e.g. ESM or later BAU).

> *Thank you for pointing this out. We will check the text accordingly.*

- P13L11 Use the same spelling in the whole manuscript: Noösphere, Noosphere

> *We will make the spelling consistent.*

- P14L19: "may be subsystems from ENV, MET or CUL". To my understanding "or CUL" has to be deleted, since the relationship CUL- CUL feedback loops seems not to be Included.

> *CUL-CUL feedback loops describe bidirectional interactional interaction mechanisms between two subsystems within CUL. These types of feedback loops are included in the feedback loop taxonomy described here.*

- P14L24: Give examples what you mean by 3-loops. For me it is unclear how you get 11 taxa.

> *3-loops are feedbacks that involve three different subsystems. We will expand on this point, give an example and explain how the 11 is arrived at (simply by combinatorics).*

Section 4:
- P18 Figure 4: Add the parameter values for which the graphs are generated, in particular for the adjustment of time preferences. Reason: I would like to see all necessary information that the reader may recalculate the results.

> *We have added all values [E0=1.6, c=1, r=0.45, l=0.2(top) or 1.3(bottom), gamma=1.1, mu=2, sigma=1, beta=0.1, alpha=0.5, G=2, N=50, p0=0.5, q=3)*

- Twice in the caption of Figure 4 “S” instead of “C” is used.

> *We have fixed this.*

Section 5:
- P19L21: I do not see how the provided framework may be helpful as “a blueprint for constructing alternative taxonomies.” This statement needs to be substantiated.

> *We will expand on that point. The idea is that, for example, researchers may wish to develop their own taxonomies that have more detail in some aspects and less in others. E.g., one might wish to have a taxonomy otherwise similar to ours that distinguishes geophysical subsystems such as the atmosphere from ecological subsystems such as forests.*

Appendix:
- P27L3: Eqs 1 and 2.

> *We have fixed this.*

- P27L27: You call both gamma and s(c) damage factor. I suggest to be more precise and to not use the same name for (slightly) different elements of the equations.

> *We have fixed this.*

- P27L28: Be more precise throughout the Appendix and add the (t) with the C

> *We chose to stick to just writing C in many places since it is clear from the exposition that this is one of the two dynamic variables that depend on t and it is common practise to avoid the explicit (t) in differential equations in natural sciences.*

- P28L13: Check the size of the parentheses for better understanding: I have impression that left of G it should be a large parenthesis

> *Thank you for spotting this, actually there was a closing parenthesis missing, which is now fixed.*

References

Dearing, J. A., Wang, R., Zhang, K., Dyke, J. G., Haberl, H., Hossain, M. S., ... & Carstensen, J. (2014). Safe and just operating spaces for regional social-ecological systems. *Global Environmental Change*, *28*, 227-238.

Lenton, T. M., & Latour, B. (2018). Gaia 2.0. *Science*, *361*(6407), 1066-1068.

Raworth, K. (2012). A safe and just space for humanity: can we live within the doughnut. *Oxfam Policy and Practice: Climate Change and Resilience*, *8*(1), 1-26.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E. F., ... & Nykvist, B. (2009). A safe operating space for humanity. *Nature*, *461*(7263), 472.

Schellnhuber, H. J. (1998). Discourse: Earth System analysis—The scope of the challenge. In *Earth System Analysis* (pp. 3-195). Springer, Berlin, Heidelberg.

Schellnhuber, H. J. (1999). ‘Earth system’ analysis and the second Copernican revolution. *Nature*, *402*(6761supp), C19.

van Vuuren, D. P., Lucas, P. L., Häyhä, T., Cornell, S. E., & Stafford-Smith, M. (2016). Horses for courses: analytical tools to explore planetary boundaries. *Earth System Dynamics*, *7*(1), 267-279.

Yearworth, M., & Cornell, S. E. (2016). Contested modelling: a critical examination of expert modelling in sustainability. Systems Research and Behavioral Science, 33(1), 45-63. https://doi.org/10.1002/sres.2315

Taxonomies for structuring models for World-Earth system analysis of the Anthropocene: subsystems, their interactions and social-ecological feedback loops

Jonathan F. Donges[1,2], Wolfgang Lucht[1,3,4], Jobst Heitzig[5], Wolfram Barfuss[1,6], Sarah E. Cornell[2], Steven J. Lade[2,7,8], and Maja Schlüter[2]

[1]Earth System Analysis, Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Telegrafenberg A31, 14473 Potsdam, Germany
[2]Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 114 19 Stockholm, Sweden
[3]Department of Geography, Humboldt University, Unter den Linden 6, 10099 Berlin, Germany
[4]Integrative Research Institute on Transformations of Human-Environment Systems, Humboldt University, Unter den Linden 6, 10099 Berlin, Germany
[5]Complexity Science, Potsdam Institute for Climate Impact Research, Member of the Leibniz Association, Telegrafenberg A31, 14473 Potsdam, Germany
[6]Department of Physics, Humboldt University, Newtonstr. 15, 12489 Berlin, Germany
[7]Fenner School of Environment and Society, The Australian National University, Building 141, Linnaeus way, Canberra, Australian Capital Territory, 2601, Australia
[8]Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

Correspondence to: Jonathan F. Donges (donges@pik-potsdam.de)

Abstract.

In the Anthropocene, ~~social processes~~ the social dynamics of human societies have become critical to understanding planetary-scale Earth system dynamics. The conceptual foundations of Earth system modelling have externalised social processes in ways that now hinder progress in understanding Earth resilience and informing governance of global environmental change. New approaches to global modelling of the human World are needed to address these challenges~~, but the~~. The current modelling landscape is highly diverse and heterogeneous, ranging from purely biophysical Earth System Models, to hybrid macro-economic Integrated Assessments Models, to a plethora of models of socio-cultural dynamics. World-Earth models ~~, currently not yet available,~~ capable of simulating the complex and entangled human-environment processes of the Anthropocene are currently not available. They will need to ~~integrate all these elements~~ draw on and selectively integrate elements from all the existing modelling approaches, so future World-Earth modellers require a structured approach to identify, classify, select, and combine model components from multiple modeling traditions. Here, we develop taxonomies for ordering the multitude of societal and biophysical subsystems and their interactions. We suggest three taxa for modelled subsystems: (i) biophysical, where dynamics is usually represented by "natural laws" of physics, chemistry or ecology (i.e., the usual components of Earth system models), (ii) socio-cultural, dominated by processes of human behaviour, decision making and collective social dynamics (e.g., politics, institutions, social networks, and even science itself), and (iii) socio-metabolic, dealing with the material interactions of social and biophysical subsystems (e.g., human bodies, natural ~~resource~~ resources and agriculture). We show how higher-order taxonomies ~~for~~ can be derived for classifying and describing the interactions between two or more subsystems~~can be derived,~~

~~highlighting~~. This then allows us to highlight the kinds of social-ecological feedback loops where new modelling efforts need to be directed. As an example, we apply the taxonomy to a stylised World-Earth system model of socially transmitted discount rates in a greenhouse gas emissions game to illustrate the effects of social-ecological feedback loops that are usually not considered in current modelling efforts. The proposed taxonomy can contribute to guiding the design and operational development of more comprehensive World-Earth models for understanding Earth resilience and charting sustainability transitions within planetary boundaries and other future trajectories in the Anthropocene.

1 Introduction

1.1 Revisiting Earth system analysis for the Anthropocene

In the age of the Anthropocene, human societies have emerged as a planetary-scale geological force shaping the future trajectory of the whole Earth system ~~(Crutzen, 2002; Steffen et al., 2007; Lewis and Maslin, 2015; Waters et al., 2016; Steffen et al., 2018)~~(Crut
Cumulative greenhouse gas emissions and extensive modifications of the biosphere have accelerated since the neolithic and industrial revolutions, especially through the rapid globalisation of social-economic systems during the 20th century, threatening the stability of the interglacial state (Lenton et al., 2016) that has enabled the development and wellbeing of human societies (Rockström et al., 2009a; Steffen et al., 2015). Political and societal developments during the 21st century and their feedback interactions with the planetary climate and biophysical environment will be decisive for the future trajectory of the Earth system (Lenton and Latour, 2018; Steffen et al., 2018). Business-as-usual is taking the planet into a ‘hothouse Earth’ state unprecedented for millions of years in geological history (Winkelmann et al., 2015; Ganopolski et al., 2016), while calls for rapid decarbonisation of the global economic system to meet the Paris climate agreement (Rockström et al., 2017) will also have complex consequences involving an intensified entanglement of social, economic and biophysical processes and their resulting feedback dynamics, up to the planetary scale (Mengel et al., 2018). Despite extensive debate about the Anthropocene (Lewis and Maslin, 2015; Hamilton, 2015; Brondizio et al., 2016; Zalasiewicz et al., 2017), and growing recognition of the limitations of current Earth system models for analysis and policy advice in the context of these shifting dynamics ~~(van Vuuren et al., 2016; Verburg et al., 2016; Donges et al., 2017a, b)~~(van Vuuren et al., 2012, 2016; Verburg et al., 2016; Donges et al., 2
little has been done to address the fundamental challenge of systematically reviewing the conceptual foundations of Earth system modelling to include dynamic social processes, rather than externalising them (Bretherton et al., 1986, 1988).

To understand planetary-scale social-ecological dynamics, models of the World-Earth system are urgently needed~~. The~~ (Schellnhuber, 1998, 1999; Rounsevell et al., 2014; van Vuuren et al., 2016; Verburg et al., 2016; Donges et al., 2017a, b, 2020; Calvin a
Epistemologically, we conceptualise the World-Earth system ~~is~~ as the planetary-scale system consisting of the interacting biophysical subsystems of the Earth, and the social, cultural, economic, and technological subsystems of the World of human societies. It should be noted here that in the context of global change analysis and modelling, the term ‘Earth system’ was intended to include human societies and their activities and artefacts (Bretherton et al., 1988; Schellnhuber, 1998, 1999). However, in currently influential science and policy contexts, notably the ~~IPCC~~Intergovernmental Panel on Climate Change (IPCC) (Flato, 2011; Flato et al., 2013), ‘Earth system models’ deal only with the physical dynamics of the atmosphere, ocean,

land surface and cryosphere, and a limited set of interactions with the biosphere. While some might see tautology in the term ‘World-Earth system’, we use it to highlight that human societies, their cultures, knowledge and artefacts (the ~~“World”~~‘World’) should now be included on equal terms to conduct systematic global analyses of the Anthropocene. A fully co-evolutionary approach is needed, in the sense of representing social-ecological feedback dynamics across scales.

Future World-Earth modelling efforts will largely be pieced together from existing conceptualisations and modelling tools and traditions of social and biophysical subsystems, which encode the state of the art in our understanding of the Anthropocene. Current efforts in World-Earth system modelling are highly stylised (e.g. ~~Kellie-Smith and Cox (2011); Garrett (2015); Jarvis et al. (2015);~~ or tend to be proof-of-concept prototypes ~~(Donges et al., 2020)~~(Beckage et al., 2018; Donges et al., 2020). None operate yet in a process-detailed, well-validated and data-driven mode. To serve these nascent efforts in enabling World-Earth system analysis of the Anthropocene, this article addresses the core question of which are the relevant categories within which World-Earth models, as essential scientific macroscopes (Schellnhuber, 1999), should operate. The problem for both scientific integration and real-world application is that ~~in current models, the characteristics and~~ the characteristic basis of the interactions of social and biophysical subsystems ~~are~~ is often not explicit ~~to each other, if they are~~ in current models. Often, the interactions between these subsystems are not recognised at all. By framing a taxonomy around the current dominant distinctions – and disciplinary divides – we can begin to explore links and ~~feedbacks~~ feedback mechanisms between taxa in more structured, systematic and transdisciplinary ways. With this taxonomy, we develop initial tools and terminologies that enable model builders and model users to be clear about their social, cultural, epistemological and perhaps also axiological standpoints.

We want to emphasise that this taxonomic approach does not presuppose that there is “one world” (an ontological position) when models of different worlds are combined. Instead, we argue that a taxonomy can help to focus modellers’ attention better on the ontological and epistemic commitments within their models. This approach opens Earth system analysis to deeper dialogues with proponents of non-human actors as shapers of the world (Latour, 2017; Morton, 2013), or even the possibility of no world at all (Gabriel, 2013).

While the present article proposes a conceptual basis for World-Earth modelling, the proposed taxonomy is employed in the ~~follow-up~~ companion paper by Donges et al. (2020) to develop ~~an~~ the operational World-Earth modelling framework copan:CORE that is cast into software and applied to construct and study an example of a World-Earth model.

1.2 Structuring the landscape of global environmental change models

Diverse scientific modelling communities aim to capture different aspects of social-ecological dynamics embedded in the Earth system up to planetary scales. Some processes operating in the Earth system are commonly described as being governed by ~~“~~the "natural laws" and generalizable principles of physics, chemistry ~~or ecology (e.g.~~and (to some extent at least) ecology (for example, atmosphere and ocean circulation as governed by the ~~laws of fluid~~ physical laws of fluid and thermodynamics), while others are thought to be dominated by human behaviour, decision making and collective social dynamics (e.g., the regularities underlying individual and social learning). This tendency for separate treatment of these different kinds of process in the natural and social sciences gives rise to problems when dealing with the many real-world subsystems that operate in both domains simultaneously. What is more, different scientific communities use different methods and adhere to different viewpoints as to the nature and character of such subsystems and their interactions. There is now a number of conceptualisations of social-ecological or coupled human-environment systems in environmental, sustainability and Earth system science (e.g. Vernadsky (1929/1986); Schellnhuber (1998); Fischer-Kowalski and Erb (2006); Jentoft et al. (2007); Biggs et al. (2012)) but we see a pressing need to structure modelling efforts across communities, providing a joint framework while maintaining the conceptual flexibility required for successful cross-disciplinary collaboration.

Here, we propose a taxonomic framework for structuring the multitude of subsystems that are represented in current mathematical and computer simulation models. The motivation for proposing such an ordering scheme is:

1. to provide the means for collecting and structuring information on what components of social-ecological systems relevant to global change challenges are already present in ~~computer~~ models in different disciplines,
2. to point out uncharted terrain in the Earth system modelling landscape, and
3. to provide the foundations for a systematic approach to constructing future co-evolutionary World-Earth models, where ~~feedbacks~~ feedback mechanisms between components can be traced and studied. This conceptual work aims to contribute to a central quest of sustainability science (Mooney et al., 2013) that “seeks to understand the fundamental character of interactions between nature and society.” (Kates et al., 2001).

1.3 Definitions and explanations of key terms

In this article, we use the term subsystem to refer to any dynamic component in models of the World-Earth system. In this broad category, we can include both the kinds of subsystems that are governed mainly by “natural laws" of physics, chemistry or ecology (e.g., seasonal precipitation, ocean nutrient upwelling) and those that are governed mainly by human behaviour, decision making and collective social dynamics (e.g., international food trade, carbon taxes). Many scientific communities similarly make this distinction between biophysical (“natural", ecological, environmental) subsystems and socio-cultural (social, human, “anthroposphere") subsystems. We also highlight socio-metabolic subsystems at the overlap of societal and natural “spheres" of the Earth system (Fig. 1). We suggest that explicit attention to these subsystems and their interactions is needed in order to deepen the understanding of transformative change in the planetary social-ecological system, making a valuable contribution to the design and operational development of future, more comprehensive World-Earth models for charting sustainability transitions into a safe and just operating space for humanity (Rockström et al., 2009a; Raworth, 2012; Dearing et al., 2014).

A further note on the term *biophysical*: here, we use this word as a shorthand term to refer to Earth’s interacting living and non-living components, encompassing geophysical (climatic, tectonic, etc.), biogeophysical~~and biogeochemical~~, biogeochemical and ecological processes. These categories are significant in Earth system science because feedbacks involving these processes tend to have different dynamic characteristics. Accordingly, they have been dealt with very differently in Earth system analysis and modelling (Charney et al., 1977; Gregory et al., 2009; Stocker et al., 2013).

The co-evolution of Earth’s geosphere and biosphere is a central concept in Earth system science (Lovelock and Margulis, 1974; Budyko et al., 1987; Lovelock, 1989; Schneider et al., 2004; Lenton et al., 2004; Watson, 2008), but the global models that currently dominate the field represent just a snapshot of the system, focused on the biophysical dynamics that play out over decades to centuries. We use the term co-evolution to describe the complex dynamics that arise from the reciprocal interactions of subsystems, each of which changes the conditions for the future time evolution of the other (not excluding, but not limited to processes of Darwinian co-evolution involving natural selection). Earth system models (ESMs) include key physical feedbacks, and increasingly permit the investigation of biophysical feedbacks, but as we have indicated, they lack socio-metabolic and socio-cultural subsystems, relying on narrative-based inputs for dealing with anthropogenic changes. Integrated assessment models (IAMs) used in the global change context (Edenhofer et al., 2014; van Vuuren et al., 2016) include some interactions of social and biophysical subsystems in order, say, to assess potential economic consequences of climate change and alternative climate policy responses. But they lack the kinds of interactions and ~~feedbacks~~ feedback mechanisms (e.g., by impacts of climatic changes on socio-metabolic subsystems, or by the effects of socio-cultural formation of public opinion and coalitions in political negotiations on environmental policies) that societies throughout history have shown to be important which is revealed, e.g., by studies of social-ecological collapse and its connection to past climate changes ~~(Weiss and Bradley, 2001; Ostrom, 2009; Donges et al., 2015; Cumming and Peterson, 2017)~~(Weiss and Bradley, 2001; Ostrom, To explore and illustrate the consequences of these typically neglected interactions and feedbacks, we have studied a conceptual model that gives rise to complex co-evolutionary dynamics and bifurcations between qualitatively different system dynamics: a model of socially transmitted discount rates in a greenhouse gas emissions game, discussed in Section 4.

For completeness, we also provide brief definitions of our working terminology: a "link" or "interaction" is a causal influence of one subsystem on another that is operationally non-decomposable into smaller links; a "mechanism" is a micro-description of how exactly this causal influence is exerted; a "process" is a set of links that "belong together" from some suitable theoretical point of view; a "loop" is a closed path in the network of links; and an "impact" of a link is the change in the target system attributable to this link.

We should note here that this taxonomy is dealing with causal narratives from different scientific disciplines that are encoded in models, and as such, it does not require any a priori theories and hypotheses about causality. Causal narratives are our starting point because they are necessary for and are explicitly encoded in simulation modelling - and our classification lets us interrogate them more systematically and exposes them explicitly.

2 A taxonomy of subsystems in models of the World-Earth system

In this section, we introduce the biophysical (ENV), socio-metabolic (MET), and socio-cultural (CUL) taxa for classifying subsystems in models of the World-Earth system (Fig. 1). For each taxon, we give examples of corresponding subsystems from different modelling fields. We also discuss how the suggested taxonomy relates to earlier conceptualisations of human societies embedded in and interacting with environmental systems (Sect. 2.4).

We have followed three guidelines in constructing this taxonomy for models of the World-Earth system:

1. *Compactness*, because we aim at a "top-level1000framework that is useful and tangible, with as few classifications as possible, covering the scope of co-evolutionary modelling research parsimoniously and in a self-containing way.

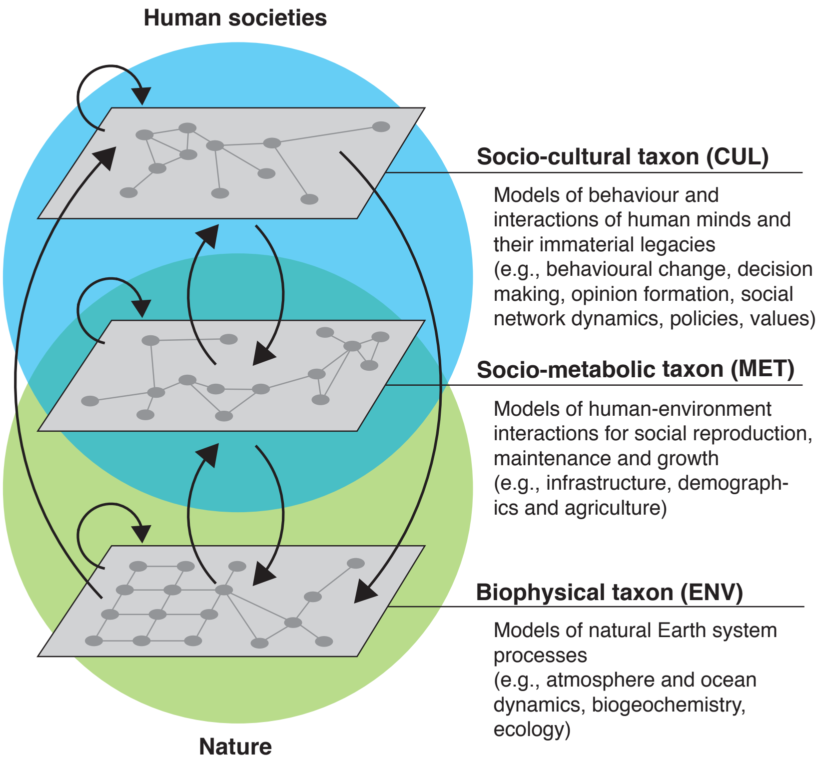


Figure 1. Proposed taxonomy of subsystems in models of the World-Earth system. The blue and green overlapping discs represent the current discipline-based domains in which the subsystems and processes of nature, human societies, and their interactions are modelled. Our scheme structures this continuum into three taxa (light grey layers) for model subsystems (dark grey discs): (i) a biophysical taxon (ENV), (ii) a socio-metabolic taxon (MET), and a socio-cultural taxon (CUL). Links within and between these modelled subsystems (shown as black arrows in the figure) can further be classified using a 3 × 3 taxonomy of interactions (Fig. 2, Sect. 3).

2. ~~*Operative capacity* for model classification and construction, because we want to advance efforts rapidly in World-Earth modelling .~~ *Compatibility* with existing ~~research fields and modelling methodologies, while critically reflecting on their suitability for the tasks at hand~~disciplines and research fields within, between and beyond the persistent natural/social sciences divide, because we view the scientific endeavour of understanding links and feedbacks in a co-evolutionary World-Earth system as an integrative and transdisciplinary opportunity.

3. *Operative capacity* for model classification and construction, because we want to advance efforts rapidly in World-Earth modelling. This guideline differs from the previous two in that it deals with practical aspects of modeling. We include it because it flags the need for critical reflection on the suitability of combined models for the tasks at hand. We want to be able to expand the scope of modelling to be more inclusive, allowing more differentiation and well-founded permutations of approaches.

Models encode knowledge outside of the mind of the modeller, so these guiding principles are intended to ensure that bridging across currently very distinct modelling approaches still permits tracing back how the techniques relate to the theories, assumptions, and framings of the contributory disciplines.

The proposed taxonomy reflects the longstanding structure – and the underlying divides – of the scientific disciplines dealing with the respective subsystems. We argue that it also provides a blueprint for navigating the fragmented modelling landscape and bringing new opportunities for cross-disciplinary bridging. The anthropocentric and dialectic distinction between the realms of nature or "the environment" and of human societies has a long intellectual history. Deep philosophical and scientific puzzles are connected with the attempts to draw a sharp distinction between these domains, and to satisfactorily integrate properties such as mental states, intentions, and life itself.

With the progressive improvements in biophysical Earth system modelling ~~(Reichler and Kim, 2008)~~ (Reichler and Kim, 2008; Steffen e the concomitantly growing reliance on model-based insights for global decision-making over a wider range of urgent sustainability ~~issues,~~ issue (National Research Council, 2007; Rounsevell et al., 2014; Calder et al., 2018), as is the case for example for the Paris climate agreement (UNFCCC, 2015) informed by the IPCC (Stocker et al., 2013; Barros et al., 2014; Edenhofer et al., 2014) a the policy processes derived from it, these conceptually challenging issues ~~now also~~ can now have direct practical implications. ~~For example, human influence is now evident on very long-term and large-scale~~ Illustration such different conceptions of Earth system processes, ~~yet human-controlled processes are absent~~ in models of ~~palaeo-environmental change. In models of~~ the contemporary Earth system, land vegetation can be treated as inanimate carbon, a transpiration "pump" affecting precipitation and soil moisture patterns (e.g. Sitch et al. (2003)), or as the animate matter of biodiverse ecosystems that sustain human communities~~.~~ (e.g. (Purves et al., 2013)). Similarly, different assumptions in models about non-material factors such as human rationality, cognition, motivations, institutions and social connections lead to very different likelihoods for alternative sustainability pathways for the world's economies and material resource use (Donges et al., 2017b; Müller-Hansen et al., 2017; Beckage et al., 2018; Ott

For these reasons, we follow a pragmatic approach in proposing a taxonomic framework that draws upon examples and allows for overlap between the domains of nature and human societies, where materiality meets intention (noting that in complex social-ecological systems, purposeful intervention will be accompanied by unintended or unanticipated side effects). Following this approach, modelled subsystems in the biophysical taxon are situated in the material domain of nature, those in the socio-metabolic taxon lie in the overlap domain, and those in the socio-cultural taxon reside in the immaterial domain of human cultures (Fig. 1).

2.1 Biophysical taxon

The biophysical taxon (ENV) contains the processes and subsystems that are typically included in current comprehensive Earth system models, but views them from the perspective of the Anthropocene shift to human "co-control". These subsystem models are governed by deterministic and stochastic mathematical equations, often developed from first principles about the physical relationships involved. There is a case for subdividing the biophysical taxon into an ecological subtaxon (subsystems associated with life) and a geophysical subtaxon (subsystems not associated with life), since they have distinct, albeit co-evolving dynamics (Vernadsky, 1929/1986; Lenton et al., 2004), and this subdivision would correspond to widely accepted geosphere/biosphere conceptualisations of the Earth system (Bretherton et al., 1986, 1988; Seitzinger et al., 2015). However, we apply our principle of compactness, because geosphere-biosphere links and processes have been comprehensively documented over the past few decades, as they underpin current Earth system and global integrated assessment modelling. Rather than retracing these links (after all, the existing models are not going to be completely reconfigured in light of the issues we explore in this paper), we have opted to take today's state of the art in biophysical global modelling as our main point of departure, following the principle of compatibility introduced above.

Earth system models have developed from coupled atmosphere-ocean general circulation models, progressively coupling in components describing biogeochemical and biogeophysical dynamics. On decade-to-millennium time scales relevant for the analysis of anthropogenic climate change and its medium-term consequences, examples of these modelled subsystems where human-controlled dynamics are prominent concerns include atmospheric chemistry, ocean productivity, sea ice, land vegetation, and major elemental cycles such as those of nitrogen, phosphorus, and sulfur (Bretherton et al., 1986, 1988). Furthermore, as it becomes clearer that palaeoclimate models can play a vital role in "deep future" studies of human-controlled processes in the Anthropocene, Earth system dynamics operating on longer time-scales are relevant~~, so~~ (Zeebe and Zachos, 2013; Steffen et al., 2018). So for these purposes, the biophysical taxon would include subsystems involving the lithosphere (e.g., rock weathering, isostatic depression and rebound associated with the advance and retreat of ice sheets on land) and even external drivers such as large-body impacts (Brugger et al., 2017), if these provide "natural experiments" or analogues for future change.

Research fields dealing with models of subsystems belonging to the biophysical taxon include, among others, geophysics, meteorology, oceanography, biology, ecology, biogeochemistry, and geology. Few of these sciences have yet grasped the methodological and theoretical tools for dealing with the human dimensions of anthropogenic change. From our planetary-scale perspective, the ENV taxon exhibits a substantial overlap with categories such as models of "the environment", "nature" or "ecology", with their specific disciplinary connotations, although many of these models have tended to be small-scale, context-specific and idiographic. An exception from this are global dynamic vegetation models such as LPJ (Sitch et al., 2003), which focus, however, on representing the physical dynamics of ecological processes and structures in an Earth system context and not on ecological dynamics as such (i.e., interactions between living organisms). We note a current drive for further refinements of ecological dynamic network processes in large-scale modelling (Purves et al., 2013; Harfoot et al., 2014) within the ENV taxon that may improve global-scale conceptualisations of ecosystems in ways compatible with both Earth system modelling and socio-ecological systems research and resilience thinking.

2.2 Socio-metabolic taxon

The socio-metabolic taxon contains processes and subsystems that form the material basis and products of societies, making direct interconnections between human societies and the biophysical environment that sustains them. This taxon comprises models of demographics and social structure (e.g., population size, age/sex distribution, health parameters; and social categories with material or resource-use consequences, such as class, clan, caste, ethnicity). It also includes "the technosphere": society's artefacts, factors of production and technologies (e.g. labour, land, capital, natural resources, raw material, energy;

tools, machines, infrastructure; cultivated landscapes, domesticated animals and plants), and economic systems (manufacturing, distribution and consumption of goods and services) (Haff, 2012, 2014; Mooney et al., 2013).

The broad field of economics currently dominates descriptions of parts of the socio-metabolic taxon in quantitative models, but many other disciplines such as geography, industrial metabolism, social ecology, and science and technology studies also play a role. In modelling terms, this taxon typically involves representations of both the biophysical planet Earth and the socio-cultural World of human societies. This implies hybrid models of the type that are currently included in Integrated Assessment Models ~~(IAMs)~~ of global change, and entails strong simplifying assumptions. ~~Our taxonomic~~ We suggest that our approach can bring much-needed clarity and transparency about the role of such models in understanding ~~of~~ the World-Earth system ~~(van Vuuren et al., 2016).~~ (c.f. van Vuuren et al. (2016)). One should note that IAMs and economic models are typically expressed in terms of financial value and not material flows that directly interact with subsystems in ENV (mostly empirical input-output theories of economics being an exception, Leontief (1936)).

2.3 Socio-cultural taxon

The socio-cultural taxon contains processes and subsystems that are described in models of the behaviour of human minds and their immaterial legacies, abstracted from their biophysical foundations and often described as lying in the realm of human agency (Otto et al., 2020b). Of the three taxa proposed, processes and subsystems in the socio-cultural taxon are the least formalised in mathematical and computer simulation models so far, despite substantial efforts in this direction in many fields of the social sciences (e.g. Farmer and Foley (2009)) and a likelihood that they may be only partly formalizable. Research fields dealing with models of processes and subsystems in the socio-cultural taxon include sociology, anthropology, behavioural economics, political science and social ecology. Our taxonomic approach can enable the diverse modelling activities now underway to engage more directly with the incipient World-Earth modelling effort.

Examples of modelled subsystems in this taxon include individual and collective opinions, behaviours, preferences, and expectations and their social network dynamics; information and communication networks; institutions and organisations; financial markets and trade; political processes; social norms and value systems (Mooney et al., 2013). Notably, the CUL taxon ~~also includes~~ can also include processes of digital transformation and artificial intelligence that increasingly restructure and shape the socio-cultural sphere of human societies. It also provides a locus for debating the challenge of reflexiveness in science, especially in fields where modelling plays a vital role in shaping knowledge and action (Yearworth and Cornell, 2016). For instance, future World-Earth modelling will have to grapple with ways to recognize Earth system science as an endogenous generator of scientific conceptions of 'Earth'. Relevant for modelling efforts, socio-cultural subsystems can vary on substantially different time scales. Near instantaneous information exchanges are possible on online social networks and within and between increasingly advanced algorithms (e.g. algorithmic trading systems on financial markets), while elections and governance processes act on the order of years. Formal institutions (e.g. laws) change on the order of decades and informal institutions (e.g. religions) develop over time frames on the order of centuries to millennia ~~(Williamson, 1998)~~(Williamson, 1998; Otto et al., 2020a).

2.4 Relations to other conceptualisations of social-ecological systems

Our model-centred taxonomy is inspired by previous systemic conceptualisations of human societies embedded in the Earth system, building upon them in a way that may help to bridge across diverse disciplines and theoretic traditions.

In one of the earliest Earth system conceptualisations, Vernadsky (1929/1986) distinguishes the inanimate matter of the geosphere, the living biosphere, and the noosphere of networked consciousness. Along these lines, Schellnhuber (1998, Fig. 34) introduced the ecosphere (directly corresponding to our ENV taxon, entailing geophysical and ecological interactions), the anthroposphere (broadly related to MET, but with some socio-cultural features), and the global subject (closely related to CUL).

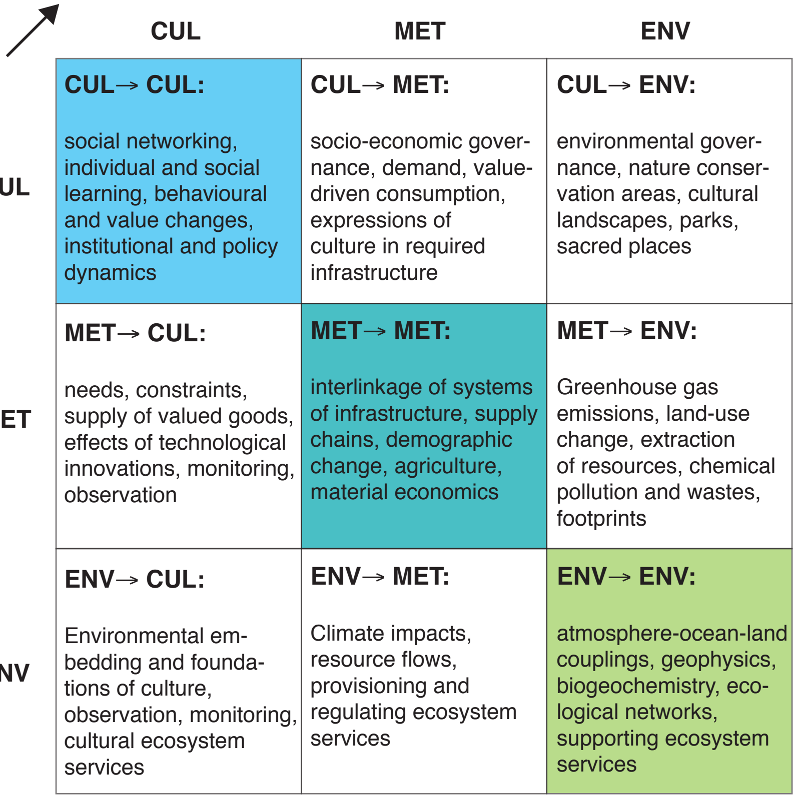
Conceptualisations in resilience theory and sustainability science emphasise the interactions and interdependence of biosphere and society (Brundtland, 1987; Folke, 2006; Folke et al., 2011), with many sustainability practitioners adding the economy to make “three pillars” or a “pie of sustainability" consisting of economy embedded in society embedded in biosphere (Folke et al., 2016). These fields have typically focused on local to regional geographic scales or specific sectors, and have not placed much emphasis on global modelling, but in general terms, their view of society contains aspects of our MET taxon, while “the economy” is more restricted than MET.

Fischer-Kowalski and Erb (2006) explicitly develop the concept of social metabolism, in terms of the set of flows between nature and culture, in order to describe deliberate global sustainability transitions. Governance-centred classification schemes in social-ecological systems research (Jentoft et al., 2007; Biggs et al., 2012), in the tradition of Ostrom (Ostrom, 2009), can also be brought into our taxonomy. Categories of the governance (sub)system link CUL and MET, and the (sub)system to be governed (ENV and MET) links the biophysical resources to be used with the social agents who will use them.

3 Taxonomy of subsystem interactions in models of the World-Earth system

In this section, we describe a taxonomy of modelled interactions between subsystems that builds upon the taxonomy of subsystems. The three taxonomic classes for World-Earth subsystems give rise to nine taxa for directed interactions connecting these subsystems. Given a pair of taxonomic classes of subsystems A and B, the taxonomic class for directed interactions between A and B is denoted as $A \to B$. Here, a directed interaction is understood in the sense of a modelled subsystem in A exerting a causal influence on another modelled subsystem in B. For example, greenhouse gas emissions produced by an industrial subsystem in MET that exert an influence on the Earth’s radiative budget in ENV would belong to the interaction taxon MET → ENV. Three of the nine interaction taxa correspond to self-interactions within taxa, while six interaction taxa connect distinct subsystem taxa (Fig. 2).

In the following, we focus on describing examples of such modelled interactions between pairs of subsystems that are potentially relevant for future trajectories of the World-Earth system in the Anthropocene and give examples of published models ~~describing them. Possible extension of~~ containing them. The content presented in the subsections necessarily differs in scope and depth reflecting today’s dominant modelling priorities, but we have aimed to ensure the information is comparable.



	CUL	MET	ENV
CUL	**CUL→ CUL:** social networking, individual and social learning, behavioural and value changes, institutional and policy dynamics	**CUL→ MET:** socio-economic governance, demand, value-driven consumption, expressions of culture in required infrastructure	**CUL→ ENV:** environmental governance, nature conservation areas, cultural landscapes, parks, sacred places
MET	**MET→ CUL:** needs, constraints, supply of valued goods, effects of technological innovations, monitoring, observation	**MET→ MET:** interlinkage of systems of infrastructure, supply chains, demographic change, agriculture, material economics	**MET→ ENV:** Greenhouse gas emissions, land-use change, extraction of resources, chemical pollution and wastes, footprints
ENV	**ENV→ CUL:** Environmental embedding and foundations of culture, observation, monitoring, cultural ecosystem services	**ENV→ MET:** Climate impacts, resource flows, provisioning and regulating ecosystem services	**ENV→ ENV:** atmosphere-ocean-land couplings, geophysics, biogeochemistry, ecological networks, supporting ecosystem services

Figure 2. Taxonomic matrix for classifying directed interactions between subsystems in models of the World-Earth system. This 3×3 classification system builds upon the taxonomy of three classes for subsystems introduced in Sect. 2. The unshaded matrix elements (here containing examples of interactions) correspond to the interaction arrows drawn between the three subsystem taxa shown in Fig. 1. Shaded elements correspond to self-interactions. The examples for directed interaction mechanisms given in the matrix elements are indicative and based on our particular areas of research.

All subsections below provide (i) a general description of the interaction taxa with some examples, and (ii) a summary of how these interactions are represented in current models.

Furthermore, possible extensions of our taxonomic approach to classify feedback loops and more complex interaction networks between subsystems are discussed (Sect. 3.10). We acknowledge that finding a conceptualisation that is satisfactory for all purposes is unlikely, but our particular pragmatic taxonomy can be useful for constructing models of the World-Earth system. It has already proven fruitful in the development of the copan:CORE open World-Earth modelling framework (Donges et al., 2020) by guiding the choice of process classes and entities that can be described in the framework as well by defining the coupling interfaces of model components that can be integrated using copan:CORE.

3.1 ENV → ENV: Biophysical Earth system self-interactions

This taxon encompasses interactions between biophysical subsystems of the type studied in current process-detailed Earth system models such as those in the CMIP5 model ensemble (Taylor et al., 2012) used in the ~~International Panel on Climate Change (IPCC)~~ IPCC reports (Stocker et al., 2013). For example, this includes modelled geophysical fluxes of energy and momentum between atmosphere and ocean, interactions between land vegetation, atmospheric dynamics and the hydrological cycle, or, more generally, exchanges of organic compounds between different compartments of biogeochemical cycles (excluding human activities here).

A detailed representation of these biophysical interactions is largely missing so far in current first attempts at modelling social-ecological dynamics at the planetary scale (e.g. Kellie-Smith and Cox (2011); Heck et al. (2016)). However, emerging socio-hydrological (Di Baldassarre et al., 2017; Keys and Wang-Erlandsson, 2017) and agent-based land-use dynamics models at regional scales (Arneth et al., 2014; Rounsevell et al., 2014; Robinson et al., 2017) include some processes involving interactions between biophysical subsystems such as the atmosphere, hydrological cycles and land vegetation.

3.2 ENV → MET: Climate impacts, provisioning and regulating ecosystem services, etc.

This taxon describes modelled interactions through which biophysical subsystems exert an influence on socio-metabolic subsystems. Relevant examples in the context of global change in the Anthropocene include the impacts of climate change on human societies (Barros et al., 2014) such as damages to settlements, production sites and infrastructures and supply chains (Otto et al., 2017), impacts on agriculture or human health, but also provisioning and regulating ecosystem services such as resource flows (Millennium Ecosystem Assessment, 2005).

Some of these interactions such as climate change impacts are now being included in IAMs (a prominent example being the DICE model, Nordhaus (1992)) and stylised models (for example Kellie-Smith and Cox (2011) and Sect. 4), but there remain challenges, e.g. in estimating damage functions and the social cost of carbon (Nordhaus, 2017). Influence from weather and climate on agriculture are studied on a global scale using model chains involving terrestrial vegetation models such as LPJ (Sitch et al., 2003) and agricultural economics models such as MAgPIE (Nelson et al., 2014). As another example, models of the distribution of vector-born diseases such as Malaria are employed to assess the impacts of climate change on human health (Caminade et al., 2014).

3.3 ENV → CUL: observation, monitoring, cultural ecosystem services, etc.

This taxon contains modelled interactions through which the state of the biophysical environment directly influences socio-cultural subsystems. ~~For example, these~~ These links can be mediated through the observation, monitoring and assessment of environmental change from local to global scales (e.g., chemical pollution, deforestation or rising greenhouse gas concentrations in the atmosphere) by social actors that in turn are processed by public opinion formation and policy making in socio-cultural subsystems (Mooney et al., 2013). The links described by the ENV → CUL taxon also relate to cultural identity connected to the environment, sense of place (Masterson et al., 2017), and more generally what has been described as cultural ecosystem services (Millennium Ecosystem Assessment, 2005).

~~Beckage et al. (2018) for examplemodel~~ For example, Beckage et al. (2018) have modelled the effect of changes in extreme events resulting from climate change on risk perception of individuals. Changes in risk perception may result in changes in emission behaviour given the perceived behaviour of others (social norms) and structural conditions in society, thus feeding back on future climate change.

ENV → CUL also play a role in regional-scale models of poverty traps where decline in natural capital reduces traditional ecological knowledge as a form of cultural capital (Lade et al., 2017b), or in models of human perceptions of local scenic beauty in policy contexts (Bienabe and Hearne, 2006). At the moment, most models deal with these interactions only at a sub-global level. But there is increasing recognition of the need for the more dynamic understanding that formal modelling can provide of such complex psychologically and culturally mediated aspects of human behavior in the Anthropocene (Schill et al., 2019).

3.4 MET → MET: economic and socio-metabolic self-interactions

This taxon describes modelled interactions between MET subsystems that connect the material manifestations and artefacts of human societies. Examples include the energy system driving factories, supply chains connecting resource extractors to complex networked production sites or machines constructing infrastructures such as power grids, airports and roads.

Certain processes involving such interactions, e.g. links between the energy system and other sectors such as industrial production, are represented in IAMs in an abstracted, macroeconomic fashion. There exist also agent-based models resolving the dynamics of supply chains that allow to describe the impacts of climate shocks on the global economy in much more detail (e.g. Otto et al. (2017)). Another class of examples are population models that may include factors such as the influence of income on fertility (Lutz and Skirbekk, 2008). However, to our best knowledge, process-detailed models of the socio-industrial metabolism (Fischer-Kowalski and Hüttler, 1998; Fischer-Kowalski, 2003) or the technosphere (Haff, 2012, 2014) comparable in complexity to biophysical Earth system models have not been published so far.

3.5 MET → ENV: greenhouse gas emissions, land-use change and biodiversity loss, impacts on other planetary boundary processes, etc.

This taxon encompasses modelled influences exerted by socio-metabolic subsystems on the biophysical environment including various forms of the "colonisation of nature" (Fischer-Kowalski and Haberl, 1993). Prominent examples in the context of global change and sustainability transformation include human impacts on the environment addressed by the planetary boundaries framework (Rockström et al., 2009a, b; Steffen et al., 2015) such as anthropogenic emissions of greenhouse gases (Stocker et al., 2013), nitrogen and phosphorous, other forms of chemical pollution and novel entities (e.g., nano particles, genetically engineered organisms), land-use change and induced biodiversity loss, exploitation and use of natural resources (Perman, 2003). This taxon also includes various forms of the conversion of energy and entropy fluxes in the biophysical Earth system by human technologies such as harvesting of renewable energy by wind turbines and photovoltaic cells (Kleidon, 2016) or different approaches to geoengineering (Vaughan and Lenton, 2011).

The interactions described by the MET → ENV are central in IAM and ESM studies of the global environmental impacts of human activities in the Anthropocene such as anthropogenic climate change as driven by greenhouse gas emissions and land-use change (Barros et al., 2014; Edenhofer et al., 2014). The latter two key processes are also frequently included in emerging studies of planetary social-ecological dynamics using stylised models (Kellie-Smith and Cox, 2011; Anderies et al., 2013; Heck et al., 2016; Heitzig et al., 2016; Lade et al., 2017a; Nitzbon et al., 2017).

3.6 MET → CUL: needs, constraints, etc.

This taxon describes modelled influences and constraints imposed upon socio-cultural dynamics by the material basis of human societies (socio-metabolic subsystems). These include, for example, the effects, needs and constraints induced by the biophysical "hardware" that runs socio-cultural processes: infrastructures, machines, computers, human bodies and brains, and associated availability of energy and other resources. It also includes the effects of technological evolution, revenues generated from economic activity, supply of valued goods, e.g. on opinion formation and behavioural change in the socio-cultural domain, or the consequences of change in demographic distribution of pressure groups on political systems and institutions.

As a recent example, the Beckage et al. (2018) model mentioned above (Sect. 3.3) has one parameter to reflect structural constraints in society that affects the degree to which emission behaviour can be changed. MET → CUL links also appear in models of resource use in social-ecological systems, where social learning of harvesting effort depends on the harvest rate ~~(Wiedermann et al., 2015; Barfuss et al., 2017)~~ (Wiedermann et al., 2015; Barfuss et al., 2017; Geier et al., 2019) and fish catches influence perceptions about the state of the fishery (Martin and Schlüter, 2015; Lade et al., 2015), or in models of economic impacts on individual voting behaviour (Lewis-Beck and Ratto, 2013).

3.7 CUL → CUL: socio-cultural self-interactions

This taxon contains modelled self-interactions between subsystems in the socio-cultural domain that have been described as parts of the ~~Noösphere~~ noosphere (Vernadsky, 1929/1986), the global subject (Schellnhuber, 1998), or the mental component of the Earth system (Lucht and Pachauri, 2004). Examples include the interaction of processes of opinion dynamics and preference formation on social networks, governance systems and underlying value systems (Gerten et al., 2018) as well as interactions between different institutional layers such as governance systems, formal and informal institutions ~~(Williamson, 1998)~~(Williamson, 1998; Ott

Some of these processes related to human behaviour and decision making (Müller-Hansen et al., 2017) have already been studied in models of social-ecological systems on local and regional scales (Schlueter et al., 2012; Schlüter et al., 2017) and have been modelled in various fields ranging from social simulation to the physics of social dynamics (Castellano et al., 2009). However, they are so far largely not included in IAMs of global change or stylised models of planetary social-ecological systems (Verburg et al., 2016; Donges et al., 2017a, b).

3.8 CUL → ENV: environmental governance, nature conservation areas, social taboos, sacred places etc.

This taxon encompasses modelled influences that socio-cultural subsystems exert on the biophysical environment. An example for such a class of interactions is environmental governance realized through formal institutions (Ostrom et al., 2007; Folke et al., 2011), where a piece of land is declared as a nature protection area excluding certain forms of land-use which has a direct impact on environmental processes there. Similarly, nature protection areas for biodiversity conservation have been represented in marine reserve models (Gaines et al., 2010). Another related example for CUL → ENV links are nature-related values and informal institutions such as respecting sacred places in the landscape and following social taboos regarding resource use (Colding and Folke, 2001). ~~While such direct~~ Different forms of environmental governance have been modelled via so-called decision or sustainability paradigms (Schellnhuber, 1998; Barfuss et al., 2018; Heitzig et al., 2018).

Direct CUL → ENV links ~~may be implemented in models, they~~ arguably cannot be found in the real world, ~~since in that case~~ in that socio-cultural influences on environmental processes must be mediated by their physical manifestations in the socio-metabolic domain (e.g. in the case of nature protection areas through the constrained actions of resource users, government enforcement efforts and infrastructures such as fences).

~~An example for~~ However, such direct CUL → ENV links ~~are nature protection areas for biodiversity conservation in marine reserve models (Gaines et al., 2010).~~ may be implemented in models, even on the global scale, such as in trade-off assessments of multiple land-uses (e.g. Boysen et al. (2017); Phalan (2018)).

3.9 CUL → MET: socio-economic policies and governance choices, value-driven consumption, etc.

Finally, this taxon contains modelled links pointing from socio-cultural to socio-metabolic subsystems. Examples include socio-economic policies and governance choices such as taxes, regulations or caps that influence the economy (e.g. carbon caps or taxes in the climate change mitigation context) or demographics (e.g. family planning and immigration policies) as well as the physical manifestations of financial market dynamics such as real estate bubbles. CUL → MET interactions also encompass the influence of cultural values, norms and lifestyles on economic demand and consumption and consequent changes in industrial production, building, transportation and other sectors.

Policy measures such as taxes, regulations or caps are much studied by IAMs of anthropogenic climate change (Edenhofer et al., 2014), while influences of value and norm change on economic activities such as general resource use ~~(Wiedermann et al., 2015; Barfu~~ fishing (Martin and Schlüter, 2015; Lade et al., 2015) has been studied in the social-ecological modelling literature, but at a mostly local to regional level.

3.10 Higher-order taxonomies of feedback loops and more complex interaction networks

Beyond the taxonomy of interactions introduced above, higher-order taxonomies could also be derived. For example, a taxonomy of feedback loops can be derived from the 3×3 taxonomy of links, leading to six taxa for feedback loops of length two in models of the World-Earth system: given a pair of interaction taxa $A \rightarrow B$ and $B \rightarrow A$, the resulting taxon for loops between A and B may be denoted as $A \circlearrowright B$. Many such feedback loops relevant for sustainability are not or only rigidly treated in current

ESMs and IAMs. For example, the ENV $\circlearrowleft$ MET feedback loop is typically not sufficiently represented in IPCC-style analyses, because the impacts of climate change on human societies are not explicitly modelled or ill-constrained in IAMs (Sect. 3.5). Furthermore, feedback loops of the type CUL $\circlearrowleft$ X, where X may be subsystems from ENV, MET or CUL are mostly missing altogether, not the least because CUL is not represented, or only fragmentarily included, in current ESMs and IAMs.

Longer and more complex paths and subgraphs of causal interactions between subsystems could be classified by further higher-order taxonomies (e.g. inspired by the study of motifs, small subgraphs, in complex network theory, Milo et al. (2002)). ~~While, this~~ This approach quickly leads to a combinatorial explosion, e.g. ~~there are already 11 distinct taxa for feedback~~ for 3-loops of the type $A \to B \to C \to A$ involving three modelled subsystems A, B, C and their interactions enumeration and counting of all possible combinations shows that there are already 11 distinct taxa for feedback loops of this kind. However, there are systematic methods available for classifying and clustering causal loop diagrams that could be leveraged to bring order into more complex models of the World-Earth system (Van Dijk and Breedveld, 1991; Rocha et al., 2015). Overall, such higher-order taxonomies could help in the design of models or model suites that can deal with different aspects of (nonlinear) interactions between World-Earth subsystems and serve as tools for understanding the emergent co-evolutionary macrodynamics.

4 An exemplary model showing complex co-evolutionary dynamics in the World-Earth system

At present, to our best knowledge, process-detailed World-Earth models that are comprehensive in the sense of the proposed taxonomies are not available~~to our best knowledge~~. Therefore, in this section, we give an illustrative example of ~~how even very simple stylised~~ a stylised World-Earth system model that covers all classes of real-world processes that appear relevant in major global feedbacks. Even such a very simple World-Earth system ~~models may already~~ model can contain a social-ecological feedback loop involving ~~most of the classes of~~ subsystem interactions introduced above (Sect. 3), and ~~lead~~ leading to a biophysical Earth system dynamics that depends crucially on a social-cultural evolution and vice versa. We also demonstrate how the taxonomies described above can be applied to classify model components and reveal the interaction structures that are implicit in the model equations. ~~In a next step, the~~ The companion paper of this article applies the taxonomies to develop a more complex illustrative World-Earth model using the copan:CORE framework (Donges et al., 2020).

The example model studied here, copan:DISCOUNT, describes a world where climate change drives a change of countries' value systems, ~~represented by~~ represented here just by the long-term discount factors their governments use in policy-making, which can be interpreted as their ~~time preference rate, i.e., their~~ relative interest in future welfare as opposed to current welfare~~, and the latter drives~~. These discount factors drive countries' emissions and thus in turn ~~drives~~ drive climate change, represented by a global atmospheric carbon stock. While the detailed description of the model's assumptions below will make clear that this causal loop involves eight of the nine interaction taxa shown in Fig. 2, the model is so designed that the description of the resulting dynamics from all these interactions can be reduced to just two ordinary differential equations, one for the fraction of "patient" countries and one for atmospheric carbon stock.

The aim of this particular model design is to show clearly that while the taxonomy developed in this paper aims at being helpful in designing and analysing World-Earth models, this does not mean the different taxa need always be easily identifiable from the final model equations.

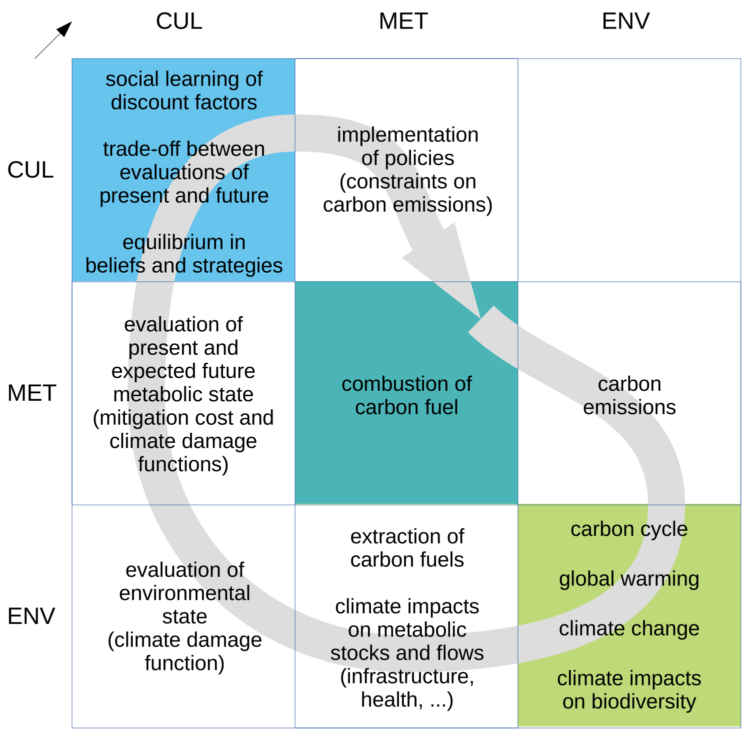
Before relating its ingredients to the introduced taxa, let us describe the model without referring to that classification. In our model, we assume that each country's metabolic activities are guided by a trade-off between the undesired future impacts of climate change caused by global carbon emissions, and the present costs of avoiding these emissions domestically. Similar to the literature on international environmental agreements and integrated assessment modelling, this tradeoff is modelled as a non-cooperative game between countries applying cost-benefit optimisation. The tradeoff and hence the evolution of the carbon stock is strongly influenced by the discount factor δ that measures the relative importance a country assigns to future welfare as compared to present welfare. The higher δ, the more a country cares about the future and the more they will reduce their emissions in order to avoid future climate impacts. While the economic literature treats δ as an exogenous parameter that has to be chosen by society (e.g., Arrow et al. (2013)), our model treats δ as a social trait that changes in individual countries over time because countries observe each other's welfare and value of δ and may learn what a useful δ is by imitating successful countries and adopting their value of δ. Because of the existence of climatic tipping points, this social dynamics does not only influence the state of the climate system but is in turn strongly influenced by it. Depending on whether the system is far from or close to tipping points, the trade-off between emissions reduction costs and additional climate damages can turn out quite differently and different values of δ will be successful.

Let us now present and decompose the model's basic causal loop in terms of the above introduced taxonomy, as shown in Fig. 3, starting in the central box. The countries' metabolisms (MET) combust carbon (MET → MET), leading to emissions (MET → ENV) that increase the global atmospheric carbon stock C (ENV), part of which is then taken up by other carbon reservoirs (ENV → ENV). C increases global mean temperature, leading to climate change (ENV → ENV) and thus to future climate impacts (i) on the countries' metabolisms (ENV → MET) and (ii) on aspects of the environment people care about, such as biodiversity (ENV → ENV → CUL). Countries evaluate these expected damages (MET → CUL; ENV → CUL) and the costs of avoiding emissions (MET → CUL), use their respective discount factors (CUL), which they learn by imitation (CUL → CUL), to assess possible domestic emissions constraints, then reach a strategic equilibrium with other countries (CUL → CUL) and implement the chosen emissions constraints (CUL → MET), this closing the long loop.

In the statistical limit of this model for a large number of countries, derived in detail in the Appendix A, this complex feedback dynamics is nicely reduced to just two equations,

$$\dot{C} = E_0 - c\,s(C)\,\phi(F) - rC, \tag{1}$$

$$\dot{F} = \ell F(1-F)[P(D(C,F)) - P(-D(C,F))], \tag{2}$$



	CUL	MET	ENV
CUL	social learning of discount factors trade-off between evaluations of present and future equilibrium in beliefs and strategies	implementation of policies (constraints on carbon emissions)	
MET	evaluation of present and expected future metabolic state (mitigation cost and climate damage functions)	combustion of carbon fuel	carbon emissions
ENV	evaluation of environmental state (climate damage function)	extraction of carbon fuels climate impacts on metabolic stocks and flows (infrastructure, health, ...)	carbon cycle global warming climate change climate impacts on biodiversity

Figure 3. Planetary social-ecological processes and interactions represented in the copan:DISCOUNT model displayed in matrix form following Fig. 2. The co-evolutionary cycle of dynamic interdependencies implemented in the model is indicated by the grey arrow.

where C is excess atmospheric carbon stock and F the fraction of "patient" countries (those that apply a large value of δ), and where

$$\begin{aligned}
s(C) &= \gamma \exp(-(C-\mu)^2/2\sigma^2),\\
\phi(F) &= F\tfrac{\alpha}{1-\alpha} + (1-F)\tfrac{\beta}{1-\beta},\\
D &= [\alpha-\beta](G - E_0\, s(C) + c\, s(C)^2\, \phi(F) - 1)\\
&\quad - \left[\tfrac{\alpha^2}{1-\alpha} - \tfrac{\beta^2}{1-\beta}\right]\tfrac{c\, s(C)^2}{2N},\\
P(D) &= \frac{1}{1+\frac{1-p_0}{p_0}\exp(-\frac{q}{p_0(1-p_0)}D)}.
\end{aligned}$$

$s(C)$ is a damage factor, $\phi(F)$ is a certain linear transformation of F, $D(C,F)$ is the utility difference between a country using discount factor α and a country using β, and $P(D)$ is a resulting imitation probability, all these derived in detail in the Appendix A. Some of the various terms in these formulas can be classified clearly as belonging to one taxon, e.g., BAU emissions E_0 belong to MET → ENV, carbon-uptake $-rC$ to ENV → ENV, and the imitation probability $P(D)$ to CUL → CUL. But others cannot, e.g., ~~the term $es(C)^2\,\phi(F)$ in~~ certain terms occurring in the formula for D ~~combines~~ combine climate damages $s(C)$ (ENV → MET → CUL) with countries' values systems, represented by $\phi(F)$ (CUL). The dynamics are governed by about a dozen parameters controlling the relative speeds and intensities of subprocesses, costs and benefits of emissions reductions, and details of the learning-by-imitation process, as described in the Appendix (Sect. A).

Let us analyse a typical dynamics of the model, shown in Fig. 4, and relate it again to our taxonomy of subsystem interactions. Consider the middle green trajectories in the lower panel starting at a low atmospheric carbon stock of $C = 1$ (fictitious units) and a medium fraction of patient countries of $F = 0.5$ (green dot). At this point, both patient and impatient countries evaluate the state of the world very similarly, hence not much imitation of discount factors happens (weak CUL → CUL dynamics), so that F may fluctuate somewhat but is not expected to change much. At the same time, as the climate damage curve (middle panel) is still relatively flat, global emissions are higher than the natural uptake rate (strong MET → ENV influence), and C is likely to increase to about 1.7 without F changing much. During this initial pollution phase, climate damages increase (the ENV → MET/CUL links becomes stronger) and the slope of the damage curve increases as more climatic tipping points are neared or crossed. This decreases the patient countries' evaluations faster than the impatient countries', hence patience becomes less attractive and countries fatalistically decrease their discount factor, so that F declines to almost or even exactly zero (the CUL → CUL dynamics becoming first stronger then weaker again) while C grows to about 3.0. In that region, most tipping points are crossed and the damage curve flattens again, causing the opposite effect, i.e., making patience more attractive. If the idea of patience has not "died-out" at that point (i.e., F is still > 0), discount factors now swing to the other extreme with F approaching unity (CUL → CUL dynamics becoming temporarily very strong), shown by one green trajectory, while emissions are first almost in equilibrium with natural carbon uptake at about $C = 3.2$ (weak MET → ENV effect) and then decline ever faster once the vast majority of countries got patient (stronger MET → ENV). This trajectory finally converges to the stable steady state at a low carbon stock of about $C = 1.5$ and $F = 1$. Note that there is also some small probability that this point is reached much faster without the long detour if the stochastic social dynamics at the starting point give patience a random advantage, as on two of the plotted trajectories.

As is typical in models with various interactions, changes in their relative interaction rates can cause highly nonlinear and even qualitative changes in model behaviour. A comparison of the top and bottom panels in Fig. 4 (see also its caption) shows that this is in particular true for World-Earth models when the rates of socio-cultural processes of the CUL → CUL type are changed (as can be claimed is indeed happening in reality since the middle of the 20th century). It should be emphasised again that these socio-cultural processes are specifically those that are least or not at all represented in current models of global change, pointing to the necessity and expected progress in understanding when including them in more comprehensive World-Earth models.

Overall, the DISCOUNT model provides a first test of the taxonomy's guiding principles. ... operative capacity - established dynamical systems methodology and macro behaviour; compatibility with diverse research fields, and compactness, since the loops and flows between taxa do not make us need to rethink the whole structure.

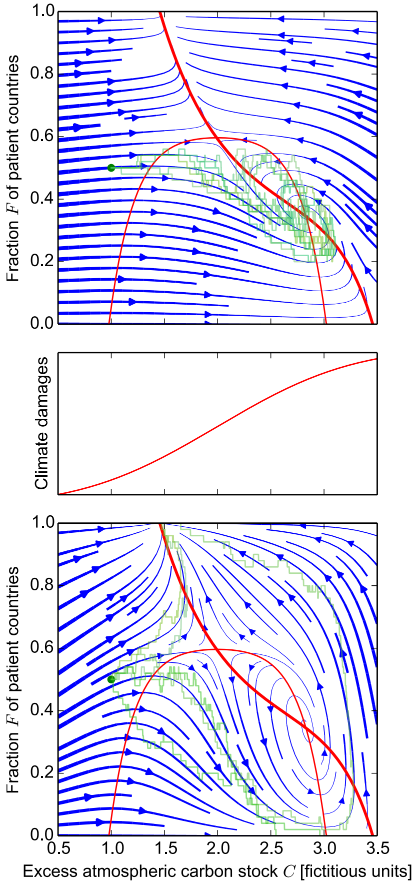


Figure 4. Typical dynamics of the copan:DISCOUNT model of the co-evolution of the global atmospheric carbon stock ~~S~~ C and the time preferences of countries, represented by the fraction F of patient countries. Of five simulated stochastic trajectories (top and bottom panel, green lines) starting at the same initial state (green dot), some will converge fast to the more desirable stable steady state at $C \approx 1.5$, $F = 1$ where climate damages (middle panel) are still relatively low, while other trajectories will approach the less desirable focus point (spiralling steady state) at $C \approx 2.8$, $F = 0.35$ where climate damages are relatively high. Depending on whether countries adjust their time preferences slowly (top panel) or fast (bottom), that focus point is either a stable attractor catching most trajectories that come near it (top) or an unstable repeller which many trajectories have to compass to approach the desirable state after a long transient detour of high damages (bottom). Blue lines show the average development represented by two ordinary differential equations (see Appendix A for details), red lines are the corresponding nullclines (thin: $\dot{F} = 0$, thick: ~~$\dot{S} = 0$~~$\dot{C} = 0$), and their other intersection at $C \approx 2$, $F \approx 0.6$ is a saddle point. Parameters: $E_0 = 1.6$, $c = 1$, $r = 0.45$, $l = 0.2$ (top) or 1.3 (bottom), $\gamma = 1.1$, $\mu = 2$, $\sigma = 1$, $\beta = 0.1$, $\alpha = 0.5$, $G = 2$, $N = 50$, $p_0 = 0.5$, $q = 3$.

5 Conclusions

In this article, we have presented a taxonomy of processes and co-evolutionary interactions in models of the World-Earth system (i.e. the planetary social-ecological system). For reasons of compactness and compatibility with existing research fields and methodologies we have proposed three taxa for modelled subsystems, and furthermore described a classification of modelled interactions between subsystems into nine taxa. We have illustrated the clarity that this taxonomic framework confers, using a stylised model of social-ecological co-evolutionary dynamics on a planetary scale that includes explicitly socio-cultural processes and feedbacks.

We argue that a relatively simple taxonomy is important for stimulating the discourse on conceptualisations of the World-Earth system as well as for operational model development as is illustrated by the work reported in the companion paper (Donges et al., 2020). It can also help communication by providing ~~and~~ an organisational scheme and a shared vocabulary to refer to the different components that need to be brought together. However, we acknowledge that alternative, more detailed taxonomies can be beneficial in more specialised settings, e.g. ecological processes are now subsumed in the biophysical taxon, but it may be useful to distinguish them from the geophysical for a clearer understanding of interactions with the socio-metabolic taxon. In ~~these cases~~other contexts, it may be useful to establish a socio-epistemic taxon separate from the socio-cultural taxon for describing subsystems, processes and interactions involving, for example, symbolic representations and transformations of knowledge through science and technology (Renn, 2018). Along these lines, our framework may be helpful as a blueprint for constructing ~~alternative~~ such alternative, possibly more detailed taxonomies.

Throughout the paper, we have illustrated the taxonomic framework using examples of subsystems, processes and interactions that are already represented in mathematical and computer simulation models in various disciplines. We have not attempted to provide a comprehensive classification of such modelling components that would be relevant for capturing future trajectories of the World-Earth system in the Anthropocene. Neither have we addressed dynamics beyond the reach of current modelling capabilities, such as long-term evolutionary processes acting within the biophysical taxon or broad patterns and singularities in the dynamics of technology, science, art and history (Turchin, 2008). But we have shown the merits of epistemological pluralism, to enable productive dialogue and interaction between the diversity of World modelling approaches and the biophysical Earth representations that exist and that have agency in a Latourian sense, e.g. through the IPCC processes.

Applying the proposed taxonomy reveals relevant directions in the future development of models of global change to appropriately represent the dynamics of the planetary social-ecological system in the Anthropocene. Regarding the sticky problem of representing causality in such a complex system, every possible contributory model is a Pandora's box out of which theoretical controversies and cross-disciplinary battles emerge. The taxonomy outlined here at least partly illuminates what is in this box, making it easier to have more open discussions among modellers about their theories and hypotheses about causality.

While current Earth System Models focus exclusively on representing biophysical subsystems and interactions and Integrated Assessment Models capitalise on those in the socio-metabolic taxon, socio-cultural subsystems and processes such as the dynamics of opinions and social networks, behaviours, values and institutions and their feedbacks to biophysical and socio-metabolic subsystems remain largely uncovered in planetary-scale models of global change. Integrating these decisive dynamics in World-Earth Models is a challenging, but highly promising research programme comparable to the development of biophysical Earth system science in the past decades following the foundational blueprints of Bretherton et al. (1986, 1988). Following this track will help to develop models that go beyond a climate-driven view of global change and to bridge the "divide" that keeps being spotlighted as the problematic hyphen in prevalent social-ecological/human-nature/etc system concepts. It will also contribute to a deeper understanding of the functioning of the complex World-Earth system machinery in the Anthropocene and promises to yield important insights on well-designed policy interventions to foster global sustainability transformation, build World-Earth resilience and avoid social-ecological collapse.

Acknowledgements. This work has been carried out within the framework of PIK's project on Coevolutionary Pathways in the Earth system (COPAN). It was supported by the Stordalen Foundation via the Planetary Boundary Research Network (PB.net), the Earth League's EarthDoc programme, the Leibniz Association (project DOMINOES), the Federal Ministry for Education and Research (BMBF, project GLUES), Humboldt University, Berlin, the Swedish Research Council Formas (Project Grant 2014-589), a core grant to the Stockholm Resilience Centre by Mistra, the Heinrich Böll Foundation, the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC grant agreement no. 283950 SES-LINK~~and funding from the European Research Council (ERC)~~) and under the European Union's Horizon 2020 ~~research and innovation programme (~~Research and Innovation Programme (ERC grant agreement No. 682472 – MUSES and ERC grant agreement No. 743080 ERA). We are thankful to the participants of the LOOPS 2014 workshop on "Closing the loop – Towards co-evolutionary modelling of global society-environment interactions", held at Kloster Chorin, Germany and the LOOPS 2015 workshop on "From limits to growth to planetary boundaries: defining the safe and just space for humanity" held in New Forest, Southampton, UK for inspiring discussions that sparked this paper. Ilona M. Otto, Marc Wiedermann and Finn Müller-Hansen are acknowledged for helpful comments on ideas presented in this paper. We are grateful for the excellent quality of creative stimulus prompted by the comments of Carsten Herrmann-Pillath and an anonymous reviewer.

References

Anderies, J. M., Carpenter, S., Steffen, W., and Rockström, J.: The topology of non-linear global carbon dynamics: from tipping points to planetary boundaries, Environmental Research Letters, 8, 044 048, 2013.

Arneth, A., Brown, C., and Rounsevell, M.: Global models of human decision-making for land-based mitigation and adaptation assessment, Nature Climate Change, 4, 550–557, 2014.

Arrow, K. J., Cropper, M. L., Gollier, C., Groom, B., Heal, G. M., Newell, R. G., Nordhaus, W. D., Pindyck, R. S., Pizer, W. A., Portney, P. R., Sterner, T., Tol, R. S. J., and Weitzman, M. L.: How Should Benefits and Costs Be Discounted in an Intergenerational Context?, 2013.

Barfuss, W., Donges, J. F., Wiedermann, M., and Lucht, W.: Sustainable use of renewable resources in a stylized social–ecological network model under heterogeneous resource distribution, Earth System Dynamics, 8, 255, 2017.

Barfuss, W., Donges, J. F., Lade, S. J., and Kurths, J.: When optimization for governing human-environment tipping elements is neither sustainable nor safe, Nature communications, 9, 1–10, 2018.

Barfuss, W., Donges, J. F., Vasconcelos, V. V., Kurths, J., and Levin, S. A.: Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse, Proceedings of the National Academy of Sciences, 117, 12 915–12 922, 2020.

Barrett, S.: Self-enforcing international environmental agreements, Oxford Economic Papers, 1994.

Barros, V., Field, C., Dokken, D., Mastrandrea, M., Mach, K., Bilir, T., Chatterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., and White, L., eds.: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.

Beckage, B., Gross, L., Lacasse, K., Carr, E., Metcalf, S., Winter, J., Howe, P., Fefferman, N., Franck, T., Zia, A., Kinzig, A., and Hoffman, F.: Linking models of human behaviour and climate alters projected climate change, Nature Climate Change, 8, 79–84, 2018.

Bienabe, E. and Hearne, R. R.: Public preferences for biodiversity conservation and scenic beauty within a framework of environmental services payments, Forest Policy and Economics, 9, 335–348, 2006.

Biggs, R., Schlüter, M., Biggs, D., Bohensky, E. L., BurnSilver, S., Cundill, G., Dakos, V., Daw, T. M., Evans, L. S., Kotschy, K., et al.: Toward principles for enhancing the resilience of ecosystem services, Annual Review of Environment and Resources, 37, 421–448, 2012.

Boysen, L. R., Lucht, W., and Gerten, D.: Trade-offs for food production, nature conservation and climate limit the terrestrial carbon dioxide removal potential, Global Change Biology, 23, 4303–4317, 2017.

Bretherton, F. P. et al.: Earth System Science. Overview, Tech. rep., National Aeronautics and Space Administration, Washington, DC, 1986.

Bretherton, F. P. et al.: Earth System Science: A Closer View, Tech. rep., National Aeronautics and Space Administration, Washington, DC, 1988.

Brondizio, E. S., O'brien, K., Bai, X., Biermann, F., Steffen, W., Berkhout, F., Cudennec, C., Lemos, M. C., Wolfe, A., Palma-Oliveira, J., et al.: Re-conceptualizing the Anthropocene: A call for collaboration, Global Environmental Change, 39, 318–327, 2016.

Brugger, J., Feulner, G., and Petri, S.: Baby, it's cold outside: Climate model simulations of the effects of the asteroid impact at the end of the Cretaceous, Geophysical Research Letters, 44, 419–427, 2017.

Brundtland, G. H.: Report of the World Commission on Environment and Development: Our common future, United Nations, 1987.

Budyko, M. I., Ronov, A. B., and Yanshin, A. L.: History of the Earth's atmosphere, Springer, Berlin, 1987.

Calder, M., Craig, C., Culley, D., de Cani, R., Donnelly, C. A., Douglas, R., Edmonds, B., Gascoigne, J., Gilbert, N., Hargrove, C., et al.: Computational modelling for decision-making: where, why, what, who and how, Royal Society Open Science, 5, 172 096, 2018.

Calvin, K. and Bond-Lamberty, B.: Integrated human-earth system modeling—state of the science and future directions, Environmental Research Letters, 13, 063 006, 2018.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A. M., Morse, A. P., Colón-González, F. J., Stenlund, H., Martens, P., and Lloyd, S. J.: Impact of climate change on global malaria distribution, Proceedings of the National Academy of Sciences, 111, 3286–3291, 2014.

Castellano, C., Fortunato, S., and Loreto, V.: Statistical physics of social dynamics, Reviews of Modern Physics, 81, 591, 2009.

Charney, J., Quirk, W. J., Chow, S.-h., and Kornfield, J.: A comparative study of the effects of albedo change on drought in semi–arid regions, Journal of the Atmospheric Sciences, 34, 1366–1385, 1977.

Colding, J. and Folke, C.: Social taboos: “invisible” systems of local resource management and biological conservation, Ecological Applications, 11, 584–600, 2001.

Crutzen, P. J.: Geology of mankind, Nature, 415, 23–23, 2002.

Cumming, G. S. and Peterson, G. D.: Unifying research on social–ecological resilience and collapse, Trends in Ecology & Evolution, 32, 695–713, 2017.

Dearing, J. A., Wang, R., Zhang, K., Dyke, J. G., Haberl, H., Hossain, M. S., Langdon, P. G., Lenton, T. M., Raworth, K., Brown, S., et al.: Safe and just operating spaces for regional social-ecological systems, Global Environmental Change, 28, 227–238, 2014.

Di Baldassarre, G., Martinez, F., Kalantari, Z., and Viglione, A.: Drought and flood in the Anthropocene: feedback mechanisms in reservoir operation, Earth System Dynamics, 8, 225–233, 2017.

Donges, J. F., Donner, R. V., Marwan, N., Breitenbach, S. F., Rehfeld, K., and Kurths, J.: Non-linear regime shifts in Holocene Asian monsoon variability: potential impacts on cultural change and migratory patterns, Climate of the Past, 11, 709–741, 2015.

Donges, J. F., Lucht, W., Müller-Hansen, F., and Steffen, W.: The technosphere in Earth System analysis: A coevolutionary perspective, The Anthropocene Review, 4, 23–33, 2017a.

Donges, J. F., Winkelmann, R., Lucht, W., Cornell, S. E., Dyke, J. G., Rockström, J., Heitzig, J., and Schellnhuber, H. J.: Closing the loop: Reconnecting human dynamics to Earth System science, The Anthropocene Review, 4, 151–157, 2017b.

Donges, J. F., Heitzig, J., Barfuss, W., Wiedermann, M., Kassel, J. A., Kittel, T., Kolb, J. J., Kolster, T., Müller-Hansen, F., Otto, I. M., Zimmerer, K. B., and Lucht, W.: Earth system modeling with endogenous and dynamic human societies: the copan:CORE open World–Earth modeling framework, Earth System Dynamics, 11, 395–413, 2020.

Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., and Minx, J., eds.: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.

Farmer, J. D. and Foley, D.: The economy needs agent-based modelling, Nature, 460, 685–686, 2009.

Fischer-Kowalski, M.: On the history of industrial metabolism, Perspectives on industrial ecology, 2, 35–45, 2003.

Fischer-Kowalski, M. and Erb, K.-H.: Epistemologische und konzeptuelle Grundlagen der sozialen Ökologie, Mitteilungen der Österreichischen Geographischen Gesellschaft, 148, 33–56, 2006.

Fischer-Kowalski, M. and Haberl, H.: Metabolism and colonization. Modes of production and the physical exchange between societies and nature, Innovation: The European Journal of Social Science Research, 6, 415–442, 1993.

Fischer-Kowalski, M. and Hüttler, W.: Society’s metabolism, Journal of Industrial Ecology, 2, 107–136, 1998.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of Climate Models, book section 9, pp. 741–866, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.020, www.climatechange2013.org, 2013.

Flato, G. M.: Earth system models: an overview, Wiley Interdisciplinary Reviews: Climate Change, 2, 783–800, 2011.

Folke, C.: Resilience: The emergence of a perspective for social–ecological systems analyses, Global Environmental Change, 16, 253–267, 2006.

Folke, C., Jansson, Å., Rockström, J., Olsson, P., Carpenter, S. R., Chapin III, F. S., Crépin, A.-S., Daily, G., Danell, K., Ebbesson, J., et al.: Reconnecting to the biosphere, Ambio, 40, 719–738, 2011.

Folke, C., Biggs, R., Norström, A. V., Reyers, B., and Rockström, J.: Social-ecological resilience and biosphere-based sustainability science, Ecology and Society, 21, 2016.

Gabriel, M.: Warum es die Welt nicht gibt, Ullstein, 2013.

Gaines, S. D., White, C., Carr, M. H., and Palumbi, S. R.: Designing marine reserve networks for both conservation and fisheries management, Proceedings of the National Academy of Sciences, 107, 18 286–18 293, 2010.

Ganopolski, A., Winkelmann, R., and Schellnhuber, H. J.: Critical insolation–CO2 relation for diagnosing past and future glacial inception, Nature, 529, 200–203, 2016.

Garrett, T. J.: Long-run evolution of the global economy – Part 2: Hindcasts of innovation and growth, Earth System Dynamics, 6, 673–688, doi:10.5194/esd-6-673-2015, http://www.earth-syst-dynam.net/6/673/2015/, 2015.

Geier, F., Barfuss, W., Wiedermann, M., Kurths, J., and Donges, J. F.: The physics of governance networks: critical transitions in contagion dynamics on multilayer adaptive networks with application to the sustainable use of renewable resources, The European Physical Journal Special Topics, 228, 2357–2369, 2019.

Gerten, D., Schönfeld, M., and Schauberger, B.: On deeper human dimensions in Earth system analysis and modelling, Earth System Dynamics Discussions, 2018, 1–22, 2018.

Gregory, J. M., Jones, C., Cadule, P., and Friedlingstein, P.: Quantifying carbon cycle feedbacks, Journal of Climate, 22, 5232–5250, 2009.

Haff, P.: Technology and human purpose: the problem of solids transport on the earth's surface, Earth System Dynamics, 3, 149–156, 2012.

Haff, P.: Humans and technology in the Anthropocene: Six rules, The Anthropocene Review, 1, 126–136, 2014.

Hamilton, C.: Getting the Anthropocene so wrong, The Anthropocene Review, 2, 102–107, 2015.

Harfoot, M. B., Newbold, T., Tittensor, D. P., Emmott, S., Hutton, J., Lyutsarev, V., Smith, M. J., Scharlemann, J. P., and Purves, D. W.: Emergent global patterns of ecosystem structure and function from a mechanistic general ecosystem model, PLoS biology, 12, e1001 841, 2014.

Heck, V., Donges, J. F., and Lucht, W.: Collateral transgression of planetary boundaries due to climate engineering by terrestrial carbon dioxide removal, Earth System Dynamics, 7, 783–796, 2016.

Heitzig, J., Kittel, T., Donges, J. F., and Molkentin, N.: Topology of sustainable management of dynamical systems with desirable states: from defining planetary boundaries to safe operating spaces in the Earth system, Earth System Dynamics, 7, 21–50, 2016.

Heitzig, J., Barfuss, W., and Donges, J. F.: A thought experiment on sustainable management of the earth system, Sustainability, 10, 1947, 2018.

Jarvis, A. J., Jarvis, S. J., and Hewitt, C. N.: Resource acquisition, distribution and end-use efficiencies and the growth of industrial society, Earth System Dynamics, 6, 689–702, doi:10.5194/esd-6-689-2015, http://www.earth-syst-dynam.net/6/689/2015/, 2015.

Jentoft, S., van Son, T. C., and Bjørkan, M.: Marine protected areas: a governance system analysis, Human Ecology, 35, 611–622, 2007.

Kates, R. W., Clark, W. C., Corell, R., Hall, J. M., Jaeger, C. C., Lowe, I., McCarthy, J. J., Schellnhuber, H. J., Bolin, B., Dickson, N. M., Faucheux, S., Gallopin, G. C., Grübler, A., Huntley, B., Jäger, J., Jodha, N. S., Kasperson, R. E., Mabogunje, A., Matson, P., Mooney, H., III, B. M., and andUno Svedin, T. O.: Sustainability Science, Science, 292, 641–642, 2001.

Kellie-Smith, O. and Cox, P. M.: Emergent dynamics of the climate–economy system in the Anthropocene, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 369, 868–886, 2011.

Keys, P. W. and Wang-Erlandsson, L.: On the social dynamics of moisture recycling, Earth System Dynamics Discussions, 2017, 1–25, 2017.

Kleidon, A.: Thermodynamic foundations of the Earth system, Cambridge University Press, 2016.

Lade, S. J., Niiranen, S., Hentati-Sundberg, J., Blenckner, T., Boonstra, W. J., Orach, K., Quaas, M. F., Österblom, H., and Schlüter, M.: An empirical model of the Baltic Sea reveals the importance of social dynamics for ecological regime shifts, Proceedings of the National Academy of Sciences, 112, 11 120–11 125, 2015.

Lade, S. J., Donges, J. F., Fetzer, I., Anderies, J. M., Beer, C., Cornell, S. E., Gasser, T., Norberg, J., Richardson, K., Rockström, J., and Steffen, W.: Analytically tractable climate-carbon cycle feedbacks under 21st century anthropogenic forcing, Earth System Dynamics Discussions, 2017, 1–23, 2017a.

Lade, S. J., Haider, L. J., Engström, G., and Schlüter, M.: Resilience offers escape from trapped thinking on poverty alleviation, Science Advances, 3, e1603 043, 2017b.

Latour, B.: Facing Gaia: Eight lectures on the new climatic regime, John Wiley & Sons, 2017.

Lenton, T., Schellnhuber, H., and Szathmary, E.: Climbing the co-evolution ladder, Nature, 431, 913, 2004.

Lenton, T. M. and Latour, B.: Gaia 2.0, Science, 361, 1066–1068, 2018.

Lenton, T. M., Pichler, P.-P., and Weisz, H.: Revolutions in energy input and material cycling in Earth history and human history, Earth System Dynamics, 7, 353–370, 2016.

Leontief, W. W.: Quantitative input and output relations in the economic systems of the United States, The Review of Economic Statistics, 18, 105–125, 1936.

Lewis, S. L. and Maslin, M. A.: Defining the anthropocene, Nature, 519, 171–180, 2015.

Lewis-Beck, M. S. and Ratto, M. C.: Economic voting in Latin America: A general model, Electoral Studies, 32, 489–493, 2013.

Lovelock, J. E.: Geophysiology, the science of Gaia, Reviews of Geophysics, 27, 215–222, 1989.

Lovelock, J. E. and Margulis, L.: Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis, Tellus, 26, 2–10, 1974.

Lucht, W. and Pachauri, R.: The mental component of the Earth system, in: Earth system analysis for sustainability, edited by Schellnhuber, H.-J., Crutzen, P., Clark, W., Claussen, M., and Held, H., Dahlem Workshop Reports, pp. 341–365, Cambridge University Press, 2004.

Lutz, W. and Skirbekk, V.: Low fertility in Europe in a global demographic context, in: Demographic Change and Intergenerational Justice, pp. 3–19, Springer, 2008.

Martin, R. and Schlüter, M.: Combining system dynamics and agent-based modeling to analyze social-ecological interactions—an example from modeling restoration of a shallow lake, Frontiers in Environmental Science, 3, 66, 2015.

Masterson, V., Stedman, R., Enqvist, J., Tengö, M., Giusti, M., Wahl, D., and Svedin, U.: The contribution of sense of place to social-ecological systems research: a review and research agenda, Ecology and Society, 22, 2017.

Mengel, M., Nauels, A., Rogelj, J., and Schleussner, C.-F.: Committed sea-level rise under the Paris Agreement and the legacy of delayed mitigation action, Nature Communications, 9, 601, 2018.

Millennium Ecosystem Assessment: Ecosystems and human well-being, Island Press Washington, DC, 2005.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U.: Network motifs: simple building blocks of complex networks, Science, 298, 824–827, 2002.

Mooney, H. A., Duraiappah, A., and Larigauderie, A.: Evolution of natural and social science interactions in global change research programs, Proceedings of the National Academy of Sciences, 110, 3665–3672, 2013.

Morton, T.: Hyperobjects: Philosophy and Ecology after the End of the World, U of Minnesota Press, 2013.

Müller-Hansen, F., Schlüter, M., Mäs, M., Donges, J. F., Kolb, J. J., Thonicke, K., and Heitzig, J.: Towards representing human behavior and decision making in Earth system models–an overview of techniques and approaches, Earth System Dynamics, 8, 977–1007, 2017.

National Research Council: Models in Environmental Regulatory Decision Making, The National Academies Press, Washington, DC, doi:10.17226/11972, https://www.nap.edu/catalog/11972/models-in-environmental-regulatory-decision-making, 2007.

Nelson, G. C., Valin, H., Sands, R. D., Havlík, P., Ahammad, H., Deryng, D., Elliott, J., Fujimori, S., Hasegawa, T., Heyhoe, E., et al.: Climate change effects on agriculture: Economic responses to biophysical shocks, Proceedings of the National Academy of Sciences, 111, 3274–3279, 2014.

Nitzbon, J., Heitzig, J., and Parlitz, U.: Sustainability, collapse and oscillations in a simple World-Earth model, Environmental Research Letters, 12, 074 020, 2017.

Nordhaus, W. D.: An optimal transition path for controlling greenhouse gases, Science, 258, 1315–1319, 1992.

Nordhaus, W. D.: Revisiting the social cost of carbon, Proceedings of the National Academy of Sciences, p. 201609244, 2017.

Ostrom, E.: A general framework for analyzing sustainability of social-ecological systems, Science, 325, 419–422, 2009.

Ostrom, E., Janssen, M. A., and Anderies, J. M.: Going beyond panaceas, Proceedings of the National Academy of Sciences, 104, 15 176–15 178, 2007.

Otto, C., Willner, S. N., Wenz, L., Frieler, K., and Levermann, A.: Modeling loss-propagation in the global supply network: The dynamic agent-based model acclimate, Journal of Economic Dynamics and Control, 83, 232–269, 2017.

Otto, I. M., Donges, J. F., Cremades, R., Bhowmik, A., Hewitt, R. J., Lucht, W., Rockström, J., Allerberger, F., McCaffrey, M., Doe, S. S., et al.: Social tipping dynamics for stabilizing Earth's climate by 2050, Proceedings of the National Academy of Sciences, 117, 2354–2365, 2020a.

Otto, I. M., Wiedermann, M., Cremades, R., Donges, J. F., Auer, C., and Lucht, W.: Human agency in the anthropocene, Ecological Economics, 167, 106 463, 2020b.

Perman, R.: Natural resource and environmental economics, Pearson Education, 2003.

Phalan, B. T.: What have we learned from the land sparing-sharing model?, Sustainability, 10, 1760, 2018.

Purves, D., Scharlemann, J. P., Harfoot, M., Newbold, T., Tittensor, D. P., Hutton, J., and Emmott, S.: Ecosystems: time to model all life on Earth, Nature, 493, 295, 2013.

Raworth, K.: A safe and just space for humanity: can we live within the doughnut, Oxfam Policy and Practice: Climate Change and Resilience, 8, 1–26, 2012.

Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, Bulletin of the American Meteorological Society, 89, 303–311, 2008.

Renn, J.: The Evolution of Knowledge: Rethinking Science in the Anthropocene, HoST-Journal of History of Science and Technology, 12, 1–22, 2018.

Robinson, D. T., Di Vittorio, A., Alexander, P., Arneth, A., Barton, C. M., Brown, D. G., Kettner, A., Lemmen, C., O'Neill, B. C., Janssen, M., Pugh, T. A. M., Rabin, S. S., Rounsevell, M., Syvitski, J. P., Ullah, I., and Verburg, P. H.: Modelling feedbacks between human and natural processes in the land system, Earth System Dynamics Discussions, 2017, 1–47, 2017.

Rocha, J. C., Peterson, G. D., and Biggs, R.: Regime shifts in the Anthropocene: drivers, risks, and resilience, PLoS One, 10, e0134 639, 2015.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., et al.: A safe operating space for humanity, Nature, 461, 472–475, 2009a.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., et al.: Planetary Boundaries: Exploring the Safe Operating Space for Humanity, Ecology & society, 14, 32, 2009b.

Rockström, J., Gaffney, O., Rogelj, J., Meinshausen, M., Nakicenovic, N., and Schellnhuber, H. J.: A roadmap for rapid decarbonization, Science, 355, 1269–1271, 2017.

Rounsevell, M., Arneth, A., Alexander, P., Brown, D., de Noblet-Ducoudré, N., Ellis, E., Finnigan, J., Galvin, K., Grigg, N., Harman, I., et al.: Towards decision-based global land use models for improved understanding of the Earth system, Earth System Dynamics, 5, 117–137, 2014.

Schellnhuber, H.-J.: Discourse: Earth system analysis – The scope of the challenge, in: Earth system analysis: Integrating science for sustainability, edited by Schellnhuber, H.-J. and Wenzel, V., pp. 3–195, Springer, Berlin, 1998.

Schellnhuber, H. J.: Earth system analysis and the second Copernican revolution, Nature, 402, C19–C23, 1999.

Schill, C., Anderies, J. M., Lindahl, T., Folke, C., Polasky, S., Cárdenas, J. C., Crépin, A.-S., Janssen, M. A., Norberg, J., and Schlüter, M.: A more dynamic understanding of human behaviour for the Anthropocene, Nature Sustainability, 2, 1075–1082, 2019.

Schlueter, M., McAllister, R., Arlinghaus, R., Bunnefeld, N., Eisenack, K., Hoelker, F., MILNER-GULLAND, E., Müller, B., Nicholson, E., Quaas, M., et al.: New horizons for managing the environment: A review of coupled social-ecological systems modeling, Natural Resource Modeling, 25, 219–272, 2012.

Schlüter, M., Baeza, A., Dressler, G., Frank, K., Groeneveld, J., Jager, W., Janssen, M. A., McAllister, R. R., Müller, B., Orach, K., et al.: A framework for mapping and comparing behavioural theories in models of social-ecological systems, Ecological Economics, 131, 21–35, 2017.

Schneider, S. H., Miller, J. R., Crist, E., and Boston, P. J., eds.: Scientists Debate Gaia: The Next Century, MIT Press, Cambridge, Massachusetts, 2004.

Seitzinger, S. P., Gaffney, O., Brasseur, G., Broadgate, W., Ciais, P., Claussen, M., Erisman, J. W., Kiefer, T., Lancelot, C., Monks, P. S., et al.: International Geosphere–Biosphere Programme and Earth system science: three decades of co-evolution, Anthropocene, 12, 3–16, 2015.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M. T., et al.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global Change Biology, 9, 161–185, 2003.

Steffen, W., Crutzen, P. J., and McNeill, J. R.: The Anthropocene: are humans now overwhelming the great forces of nature, AMBIO: A Journal of the Human Environment, 36, 614–621, 2007.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., de Vries, W., de Wit, C. A., et al.: Planetary boundaries: Guiding human development on a changing planet, Science, p. 1259855, 2015.

Steffen, W., Rockström, J., Richardson, K., Folke, C., Barnosky, A. D., Cornell, S. E., Crucifix, M., Donges, J. F., Fetzer, I., Lade, S. J., Lenton, T. M., Liverman, D., Scheffer, M., Summerhayes, C., Winkelmann, R., and Schellnhuber, H. J.: Trajectories of the Earth system in the Anthropocene, Proceedings of the National Academy of Sciences USA, 115, 8252–8259, 2018.

Steffen, W., Richardson, K., Rockström, J., Schellnhuber, H. J., Dube, O. P., Dutreuil, S., Lenton, T. M., and Lubchenco, J.: The emergence and evolution of Earth System Science, Nature Reviews Earth & Environment, 1, 54–63, 2020.

Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., eds.: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324, www.climatechange2013.org, 2013.

Strnad, F. M., Barfuss, W., Donges, J. F., and Heitzig, J.: Deep reinforcement learning in World-Earth system models to discover sustainable management strategies, Chaos: An Interdisciplinary Journal of Nonlinear Science, 29, 123 122, 2019.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, Bulletin of the American Meteorological Society, 93, 485–498, 2012.

Traulsen, A., Semmann, D., Sommerfeld, R. D., Krambeck, H.-J., and Milinski, M.: Human strategy updating in evolutionary games., Proceedings of the National Academy of Sciences of the United States of America, 107, 2962–6, 2010.

Turchin, P.: Arise 'cliodynamics', Nature, 454, 34, 2008.

UNFCCC: Paris Agreement, uNTC XXVII 7.d, 2015.

Van Dijk, J. and Breedveld, P. C.: Simulation of system models containing zero-order causal paths—I. Classification of zero-order causal paths, Journal of the Franklin Institute, 328, 959–979, 1991.

van Vuuren, D. P., Bayer, L. B., Chuwah, C., Ganzeveld, L., Hazeleger, W., van den Hurk, B., Van Noije, T., O'Neill, B., and Strengers, B. J.: A comprehensive view on climate change: coupling of earth system and integrated assessment models, Environmental Research Letters, 7, 024 012, 2012.

van Vuuren, D. P., Lucas, P. L., Häyhä, T., Cornell, S. E., and Stafford-Smith, M.: Horses for courses: analytical tools to explore planetary boundaries, Earth System Dynamics, 7, 267–279, doi:10.5194/esd-7-267-2016, 2016.

Vaughan, N. E. and Lenton, T. M.: A review of climate geoengineering proposals, Climatic Change, 109, 745–790, 2011.

Verburg, P. H., Dearing, J. A., Dyke, J. G., van der Leeuw, S., Seitzinger, S., Steffen, W., and Syvitski, J.: Methods and approaches to modelling the Anthropocene, Global Environmental Change, 39, 328–340, 2016.

Vernadsky, V. I.: The biosphere (An abridged version based on the French edition of 1929), Synergetic Press, London, 1929/1986.

Waters, C. N., Zalasiewicz, J., Summerhayes, C., Barnosky, A. D., Poirier, C., Gałuszka, A., Cearreta, A., Edgeworth, M., Ellis, E. C., Ellis, M., et al.: The Anthropocene is functionally and stratigraphically distinct from the Holocene, Science, 351, aad2622, 2016.

Watson, A. J.: Implications of an anthropic model of evolution for emergence of complex life and intelligence, Astrobiology, 8, 175–185, 2008.

Weiss, H. and Bradley, R. S.: What drives societal collapse?, Science, 291, 609–610, 2001.

Wiedermann, M., Donges, J. F., Heitzig, J., Lucht, W., and Kurths, J.: Macroscopic description of complex adaptive networks co-evolving with dynamic node states, Physical Review E, 91, 052 801, 2015.

Williamson, O. E.: Transaction cost economics: how it works; where it is headed, De economist, 146, 23–58, 1998.

Winkelmann, R., Levermann, A., Ridgwell, A., and Caldeira, K.: Combustion of available fossil fuel resources sufficient to eliminate the Antarctic Ice Sheet, Science advances, 1, e1500 589, 2015.

Yearworth, M. and Cornell, S. E.: Contested modelling: a critical examination of expert modelling in sustainability, Systems Research and Behavioral Science, 33, 45–63, 2016.

Zalasiewicz, J., Waters, C. N., Wolfe, A. P., Barnosky, A. D., Cearreta, A., Edgeworth, M., Ellis, E. C., Fairchild, I. J., Gradstein, F. M., Grinevald, J., et al.: Making the case for a formal Anthropocene Epoch: an analysis of ongoing critiques, Newsletters on Stratigraphy, 50, 205–226, 2017.

Zeebe, R. E. and Zachos, J. C.: Long-term legacy of massive carbon input to the Earth system: Anthropocene versus Eocene, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371, 20120 006, 2013.

Appendix A: The copan:DISCOUNT model

The illustrative model copan:DISCOUNT simulates the co-evolution of $C \geqslant 0$, the excess global atmospheric carbon stock above an equilibrium value that would be attained for zero GHG emissions, and the fraction $F \in [0,1]$ of the world's countries that care strongly about their future welfare. While C represents the macroscopic state of nature, F represents the macroscopic state of the global human society.

As the derivation of the model below will show, the time evolution of C and F is eventually given by Eqs. 1 and 2. Their governing parameters are business-as-usual emissions $E_0 > 0$, an abatement cost factor $c > 0$, a carbon uptake rate $r > 0$, a learning rate $\ell > 0$, a damage ~~factor~~ coefficient $\gamma > 0$, a mean tipping point location $\mu > 0$ and spread $\sigma > 0$, two candidate discount rates $0 < \beta < \alpha < 1$, an economic growth factor $G \geqslant 1$, the total number of countries $N > 0$, a curiosity parameter $0 < p_0 < 1$, and a myopic rationality parameter $q > 0$. The equations are derived by combining a standard emissions game model from the literature on international environmental agreements (Barrett, 1994) with a social imitation dynamics that governs the evolution of the countries' time discounting factors as follows.

A1 Countries, welfare

At each point in continuous time, t, a number of $N > 1$ similar countries, i, choose their individual *abatement levels* (carbon equivalents per time), $a_i(t) \geqslant 0$. Global abatement and carbon emissions per time (an interaction of type MET → ENV) are then

$$A(t) = \sum_{i=1}^{N} a_i(t), \qquad E(t) = E_0 - A(t), \tag{A1}$$

where $E_0 > 0$ are global "business-as-usual" emissions.

Country i chooses $a_i(t)$ rationally but myopically, only taking into account its own welfare in the present and in "the future" (after a fixed time interval of, say, fifty years). Its present welfare, $W_i^0(t)$, is given by some business as usual welfare, normalised to unity, minus the costs of emissions reductions (MET → CUL), which are a quadratic function of $a_i(t)$ as usual in stylised models of international environmental agreements (Barrett, 1994),

$$W_i^0(t) = 1 - \frac{a_i(t)^2}{2c/N}, \tag{A2}$$

where $c/N > 0$ is a cost parameter that is normalised with N to make the Nash equilibrium outcome (see below) independent of N.

Country i's "future" welfare (belonging to MET), $W_i^1(t)$, is a higher business-as-usual welfare given by a growth parameter $G > 1$, minus the value of additional damages from climate change caused by the present emissions, which are a linear function of $E(t)$:

$$W_i^1(t) = G - s(C(t))E(t), \tag{A3}$$

where $s(C(t)) > 0$ is a *damage factor* that depends on the current carbon stock (see below). Note that while these additional damages $s(C)E(t)$ caused by the present emissions, total damages will still be a nonlinear function of stock C since the factor $s(C)$ changes with C, representing the presence of tipping points (see below).

A2 Discounting, emissions

Since W_i^1 increases in a_i while W_i^0 decreases, choosing an optimal value for a_i involves a trade-off between present and future welfare, which we assume is done in the usual way by using some current *discount factor* $0 < \delta_i(t) < 1$ (an element of taxon CUL) that measures the relative weight of future welfare in country i's optimisation target ("utility") at time t, $U_i(t)$:

$$U_i(t) = (1 - \delta_i(t))W_i^0(t) + \delta_i(t)W_i^1(t). \tag{A4}$$

For simplicity, we assume that only two different discount factors are possible, $0 < \beta < \alpha < 1$, and call a country with $\delta_i(t) = \alpha$ "patient", so that the state of global society at time t can be summarised by the fraction $F(t)$ of patient countries:

$$F(t) = |\{i : \delta_i(t) = \alpha\}|/N. \tag{A5}$$

Given carbon stock $C(t)$ (ENV) and discount factors $\delta_i(t)$, the countries thus face a simultaneous multi-agent multi-objective optimisation problem, each i trying to optimise their utility

$$\begin{aligned} U_i(t) = {} & ((1 - \delta_i(t))) \left(1 - \frac{a_i(t)^2}{2c/N}\right) \\ & + \delta_i(t)(\left(G - s((C(t)))\right) \left(E_0 - \sum_{j=1}^{N} a_j(t))\right). \end{aligned} \tag{A6}$$

by choosing $a_i(t)$. As in the literature on international environmental agreements, e.g., Barrett (1994), we assume this is solved by making the choices independently and non-cooperatively, i.e., putting $\partial U_i(t)/\partial a_i(t) = 0$ for all i simultaneously, leading to a system of N equations whose solutions $a_i(t)$ form the Nash equilibrium choices (CUL → CUL),

$$a_i(t) = \frac{c}{N}\frac{\delta_i(t)}{1 - \delta_i(t)} s(C(t)), \tag{A7}$$

$$\begin{aligned} U_i(t) = {} & 1 + \delta_i(t)(G - E_0\, s(C(t)) + c\, s(C(t))^2 \phi(F(t)) - 1) \\ & - \frac{c}{2N}\frac{\delta_i(t)^2}{1 - \delta_i(t)} s(C(t))^2 \end{aligned} \tag{A8}$$

and the aggregate abatement (CUL → MET) and emissions

$$A(t) = s(C(t))\, c\phi(F(t)), \qquad\qquad E(t) = E_0 - A(t), \tag{A9}$$

where

$$\phi(F(t)) = F(t)\frac{\alpha}{1-\alpha} + (1-F(t))\frac{\beta}{1-\beta}. \tag{A10}$$

A3 Evolution of discount factors

While economic models treat the discount factor of a country as an exogenous parameter, we assume that the value of δ_i is a social trait that may be changed over time due to the observation of other countries' discount factors and their resulting utility (CUL → CUL). As in many models of the spread of social traits (e.g., Traulsen et al. (2010); Wiedermann et al. (2015)), we assume that each country i may adopt another country j's value of δ (social learning by imitation) and that the probability P for doing so depends on the difference between i and j's current utility, $D_{ij}(t) = U_j(t) - U_i(t)$, in a nonlinear, sigmoid-shaped fashion, with $P(D) \to 0$ for $D \to -\infty$ and $P(D) \to 1$ for $D \to \infty$. The utility difference between a country using α and a country using β is

$$\begin{aligned} D(t) = {} & [\alpha-\beta](G - E_0 s(C(t)) + cs(C(t))^2\phi(F(t)) - 1) \\ & - \left[\frac{\alpha^2}{1-\alpha} - \frac{\beta^2}{1-\beta}\right]\frac{cs(C(t))^2}{2N}. \end{aligned} \tag{A11}$$

This difference is zero iff the discounting summary statistics $\phi(F(t))$ equals

$$\phi_F(C(t)) := \frac{\frac{\alpha^2}{1-\alpha} - \frac{\beta^2}{1-\beta}}{2N[\alpha-\beta]} + \frac{E_0}{cs(C(t))} - \frac{G-1}{cs(C(t))^2} \tag{A12}$$

Since $\alpha > \beta$, we have $D(t) > 0$ iff $\phi(F(t)) < \phi_F(C(t))$, meaning that depending on the stock and the fraction of patient countries, either patience or impatience might be more attractive, so that one can expect interesting learning dynamics.

We assume that at each point in time, each country i independently has a probability rate $\ell > 0$ to perform a "learning step". If i does perform a learning step at time t, it compares its current utility $U_i(t)$ with that of a randomly drawn country j and sets its discount factor $\delta_i(t)$ to the value of $\delta_j(t)$ with a probability given by the generalised logistic function,

$$P(D_{ij}(t)) = \frac{1}{1 + \frac{1-p_0}{p_0}\exp(-\frac{q}{p_0(1-p_0)}D_{ij}(t))}, \tag{A13}$$

where $0 < p_0 < 1$ and $q > 0$ are parameters so that $P(0) = p_0$ and $P'(0) = q$.

The "curiosity" parameter p_0 ~~and~~ can be interpreted as a measure of a country's curiosity-driven exploration of a different discount factor without expecting a welfare increase. The larger p_0, the more frequent switches will occur, but in both directions between the two candidate discount rates, mainly generating more variance and fluctuations that can be seen as a form of "noise". The "myopic rationality" parameter q can ~~very roughly~~ be interpreted as ~~measures of curiosityand~~ a measure of a country's rationality, because the probability of switching to the other country's discount rate is higher if the other country has higher welfare (and zero if that is not the case) – but it is a myopic rationality, ~~respectively.~~ because the agent only takes its present welfare into account. The larger q, the faster discount factors will converge to the one currently generating the largest welfare.

To get a deterministic evolution that can be represented by an ordinary differential equation, we only track the *expected* fraction $F(t)$ of patient countries, which evolves as

$$\dot{F}(t) = \ell F(t)(1-F(t))[P(D(t)) - P(-D(t))], \tag{A14}$$

while the actual number of patient countries would follow a stochastic dynamics involving binomial distributions that converges to the above in the statistical limit $N \to \infty$. Note that $\dot{F}(t) = 0$ iff $F(t) \in \{0,1\}$ or $\phi(F(t)) = \phi_F(C(t))$.

A4 Carbon stock, damage factor

For ease of presentation, we drop the denotation of time dependence from here on. We assume that the atmospheric carbon stock evolves according to a simplistic dynamics involving only emissions and carbon uptake by other carbon stocks,

$$\dot{C} = E - rC = E_0 - cs(C)\phi(F) - rC \tag{A15}$$

with a constant *carbon uptake rate* $r > 0$ (ENV → ENV). Note that $\dot{C} = 0$ iff $\phi(F)$ equals

$$\phi_C(C) = \frac{E_0 - rC}{cs(C)}. \tag{A16}$$

In order that $C \geqslant 0$ for all times, we require that $\dot{C} \geqslant 0$ whenever $C = 0$, which is ensured by assuming that the parameters fulfil $E_0 \geqslant c\gamma \exp(-\mu^2/2\sigma^2)\phi_1$ where $\phi_1 = \alpha/(1-\alpha)$.

We further assume that $s(C)$, the value (MET → CUL; ENV → CUL) of the additional damages from climate change (ENV → MET; ENV → CUL) due to a marginal increase in emissions at an existing carbon stock C (ENV → ENV), is a positive function of C that has a unique maximum at some critical stock μ at which small changes in stock lead to large changes in damages due to the presence of tipping points. To approximate a damage function that is a sum of a number of sigmoid-shaped functions representing individual tipping points whose locations and amplitudes are roughly normally distributed, we take $s(C)$ to be Gaussian,

$$s(C) = \gamma \exp(-(C-\mu)^2/2\sigma^2), \tag{A17}$$

with parameters $\gamma > 0$, $\mu > 0$, $\sigma > 0$. This completes our derivation of the two ordinary differential equations for C and F.

A5 Steady states, stability

We can distinguish three types of steady states where $\dot{C} = \dot{F} = 0$.

(1) All countries are impatient, $F = 0$ (which implies $\phi(F) = \phi_0 := \beta/(1-\beta)$), and $(E_0 - rC)/cs(C) = \phi_0$. The latter is equivalent to $c\phi_0\gamma \exp(-(C-\mu)^2/2\sigma^2) = E_0 - rC$ which has generically one or three solutions in C with $C > 0$. If there are three, the middle one is always unstable. The others are stable iff $D < 0$.

(2) All countries are patient, $F = 1$ (which implies $\phi(F) = \phi_1$) and $(E_0 - rC)/cs(C) = \phi_1$. The latter is equivalent to $c\phi_1\gamma \exp(-(C-\mu)^2/2\sigma^2) = E_0 - rC$ which again has generically one or three solutions in C with $C > 0$. Again, if there are three, the middle one is always unstable. Again, the others are stable iff $D < 0$. The possibility of two stable states with $F = 1$, one with a small and one with a large C, indicates that even if all countries eventually become patient, this may happen too slowly to prevent a level of climate change (large A) that makes ambitious mitigation even for patient countries too costly in view of the small amount of climate damages that could then still be avoided.

(3) $0 < F < 1$ and $\phi(F) = \phi_F(C) = \phi_C(C)$. This has at most four different solutions in C with $C > 0$, to each of which corresponds at most one solution in F. We know of no simple conditions for assessing their stability but from our numerical experiments we conjecture that (i) at most one of them is stable, namely the one with the largest C, (ii) its stability depends only on the learning rate ℓ, being stable up to a critical value ℓ^*, then unstable; (iii) For $\ell < \ell^*$, it is a stable focus and the leftmost steady state with $F = 0$ is unstable. Hence at most four stable steady states can exist: at most two with $F = 1$, and either at most two with $F = 0$ or at most one with $F = 0$ plus the stable focus with $0 < F < 1$.