Studies in History and Philosophy of Modern Physics 46 (2014) 228–238

Contents lists available at ScienceDirect

Studies in History and Philosophy of Modern Physics

journal homepage: www.elsevier.com/locate/shpsb

The epistemology of climate models and some of its implications for climate science and the philosophy of science

Joel Katzav

Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form
18 February 2014
Accepted 2 March 2014
Available online 5 April 2014

Keywords:
Climate change
Climate models
Model assessment
Scientific models
Confirmation
Climate Science

ABSTRACT

I bring out the limitations of four important views of what the target of useful climate model assessment is. Three of these views are drawn from philosophy. They include the views of Elisabeth Lloyd and Wendy Parker, and an application of Bayesian confirmation theory. The fourth view I criticise is based on the actual practice of climate model assessment. In bringing out the limitations of these four views, I argue that an approach to climate model assessment that neither demands too much of such assessment nor threatens to be unreliable will, in typical cases, have to aim at something other than the confirmation of claims about how the climate system actually is. This means, I suggest, that the Intergovernmental Panel on Climate Change's (IPCC's) focus on establishing confidence in climate model explanations and predictions is misguided. So too, it means that standard epistemologies of science with pretensions to generality, e.g., Bayesian epistemologies, fail to illuminate the assessment of climate models. I go on to outline a view that neither demands too much nor threatens to be unreliable, a view according to which useful climate model assessment typically aims to show that certain climatic scenarios are real possibilities and, when the scenarios are determined to be real possibilities, partially to determine how remote they are.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Modern Physics*

1. Introduction

Climate model assessment often aims to teach us to what extent models capture existing knowledge about the climate system. For example, it might involve assessing how well models simulate observed precipitation patterns. But climate model assessment also often aims to compare model simulations with observations in order to learn something new about the climate system. For example, it might involve comparing simulations with observations in order to assess model predictions of precipitation patterns for the coming fifty years or model-based explanations of the causes of climate change. My question here is what the primary target claims of useful climate model assessment are when such assessment aims to produce new knowledge about the climate system.

I outline and examine five answers to my question. For reasons of brevity, I will present these answers as views about climate model assessment, taking it as read that they are views about climate model assessment that aims to produce new knowledge about the climate system. The first two views I examine, *the standard view* and *the adequacy-for-purpose view*, are the views that are most prominent in the emerging philosophy literature on assessing climate models. According to the standard view, climate model assessment aims primarily to confirm climate models. Lloyd (2009 and 2010), I will argue, assumes a version of the standard view in arguing for the high probability of the claim, put forward in the Intergovernmental Panel on Climate Change fourth report (IPCC-AR4) (Solomon et al., 2007), that anthropogenic greenhouse gas emissions were responsible for most of the observed, post-1950 global warming.[1] The adequacy-for-purpose view, which is suggested by Parker (2009), tells us that useful climate model assessment primarily aims to assess climate model adequacy for specific purposes. On this view, once assessment establishes which purposes models are adequate for, we can select which model results can be trusted and thus learn about the climate system.

E-mail address: j.k.katzav@tue.nl

[1] A similar claim is made in the IPCC fifth assessment report (IPCC, 2013).

http://dx.doi.org/10.1016/j.shpsb.2014.03.001
1355-2198/

The third and the fourth views I examine are *the conservative view* and *the Bayesian view*. The conservative view, I will argue, is inevitably at the heart of actual climate model assessment whenever (empirical) model success is taken to confirm claims about how the climate system actually is. This view recognises that there are substantial limitations to our ability to assess what known model (empirical) limitations imply about the accuracy of uncertain climate model assumptions and results, where such results include model-based predictions and model-based explanations. It nevertheless allows us to assign uncertain model assumptions and results some initial, presumptive degree of confirmation given available model successes. This presumptive degree of confirmation is, however, then weakened or strengthened in light of background knowledge. The Bayesian view is an application of a Bayesian theory of confirmation. According to the Bayesian view, useful assessment of climate models primarily aims to use Bayes' theorem in order to provide probabilistic confirmation of some climate model assumptions and results.

The fifth view I examine, *the possibilist view*, builds on a variety of positions outlined by philosophers and climate scientists (see, e.g., Stainforth, Allen, Tredger, & Smith, 2007; Betz, 2010; Katzav, Dijkstra, & Jos de Laat, 2012). On the possibilist view, useful climate model assessment does not primarily aim to teach us something about how the climate system actually is but, rather, primarily aims to teach us something about how it might be. More precisely, it tells us that useful climate model assessment primarily aims to show that uncertain climate model assumptions and results, and some modifications of uncertain climate model assumptions and results, describe real possibilities. When this is shown, on the possibilist view, assessment further aims partially to determine how remote the described possibilities are.

Extending an argument of Parker's (2009) will indicate that known climate model error is too pervasive to allow climate model confirmation to be of use. In particular, the confirmation of climate models cannot, contrary to Lloyd, be used to support claims about the causes of climatic phenomena. The adequacy-for-purpose view avoids assuming that climate models can be confirmed, but makes assessing adequacy-for-purpose hypotheses so demanding that, in typical cases, it largely begs the question of their truth. The conservative view avoids the idea that climate models can be confirmed and, further, is not too demanding. Nevertheless, it turns out to be unreliable because of the relative ease with which it allows confirmation to be achieved. The Bayesian view is, in actual usage, also unreliable and, in any case, has a presupposition that is typically too demanding to be fulfilled in practice. The possibilist view neither demands too much nor threatens to be unreliable.

My broad conclusion is that, in order to avoid both the threat of unreliability and of being too demanding, climate model assessment should not primarily comprise adequacy-for-purpose assessments or primarily aim directly to confirm claims about how the climate system actually is. It should primarily aim at more modest targets of assessment, for example, at determining how the climate system might really be. Moreover, much epistemology of science that purports to provide a general understanding of science, including Bayesian and other approaches to confirmation, is largely irrelevant to the assessment of climate models.

Sections 2 and 3 consider the standard and adequacy-for-purpose views. Section 4 considers the conservative and Bayesian views. Section 5 discusses the possibilist view. Section 6 provides the concluding discussion.

2. Confirming climate models

According to the standard view, useful climate model assessment primarily aims to confirm climate models. Further, while confirming a model may comprise confirming its truth, it need not involve anything so ambitious. On the standard view, a necessary and sufficient condition for confirming a model is the confirmation of the conjunction of those of its assumptions that are responsible for its success. A model assumption is one of those that are responsible for a model's success if, roughly and following Psillos (1999), simply suspending the assumption means that the model can no longer generate its success and, further, no other available assumption can be used instead to generate the success.[2]

Lloyd is committed to a version of the standard view. Her focus (Lloyd, 2010, *p*. 971) is on the confirmation of global climate models (GCMs), which are highly complex climate models that are characterised by their coupling of explicit representations of the ocean to explicit representations of the atmosphere. Since her discussion dates from 2010, her examples of GCMs are drawn from those used for IPCC-AR4. According to Lloyd (2010, *pp*. 979–982), a particularly important form of confirmation of these and other GCMs results from robustness analyses as Weisberg understands such analyses.[3] According to Weisberg, a robustness theorem has the form, “Ceteris paribus, if [common core (causal) structure] obtains, then [robust property] will obtain” (Weisberg, 2006, *p*. 738). We can establish the truth of such theorems by running simulations of some property of a target system using models that differ systematically, but share common causal assumptions about the target system. If the simulations make it very likely that the property will obtain, the robust theorem's truth is established. One can then, according to Weisberg, confirm the common causal assumptions of the models by, e.g., observing the relevant property (Weisberg, 2006, *p*. 739). Climate model assessment is naturally thought of as making use of robustness analysis since it often relies on multiple, differing models and, when it does, involves taking inter-model agreement to warrant increased confidence in model results (Lloyd, 2010; Parker, 2011; Katzav et al., 2012). Lloyd's example of a confirmed, robust theorem from climate science is the following one: “Ceteris paribus, if [Greenhouse gases relate in lawlike interaction with the energy budget of the earth] obtains, then [increased global mean temperature] will obtain” (Lloyd, 2010, *p*. 980). This theorem is supposedly established by the agreement of IPCC-AR4 GCMs about the causal role of anthropogenic greenhouse gases. Given the theorem, the GCMs' successful simulation of 20th century warming and related climatic phenomena supposedly confirms what the models tell us about the role of anthropogenic greenhouse gases.

Now, if we construe robustness analysis as Weisberg does, then taking climate models to be confirmed by such analyses involves, at least, taking the truth of those of their assumptions that are responsible for their successes to be confirmed and thus involves a commitment to their being confirmed in the sense specified by the standard view. For those model assumptions that are responsible for climate model successes will have to be among those assumptions that are shared by the models used to generate robust results. In addition, if we succeed in confirming the truth of the conjunction of those assumptions that are responsible for model

[2] One can reformulate the five views that I will outline so that they are in accord with the view, espoused by Giere (2010) and others, that models are defined or specified by complex hypotheses but are not themselves such hypotheses. The claim that it is sometimes useful to confirm climate models can, for example, be expressed in the following way: each climate model is defined/specified by a complex hypothesis about the climate system and it is sometimes useful to have as a target of confirmation a model's defining/specifying hypothesis. I will, in order to keep things simple, stick to speaking of the confirmation of climate models.

[3] Lloyd does not state that the form of confirmation she is concerned with is the kind that useful assessment *primarily* aims at. Whether it *is* is tangential to the worries I raise about her position, however. Moreover, as we will see, if climate models were confirmed in the sense she implies they are confirmed, further climate model assessment would appear to be secondary.

successes, we also acquire warrants for the uncertain implications of the conjunction. So further assessment of our models would then appear to be secondary.

Parker (2009, *p*. 235) worries that some of what Lloyd writes, in particular Lloyd's claim (Lloyd, 2009, *pp*. 214 and 216) that GCMs have been confirmed, in fact suggests a commitment to the view that climate model truth can be confirmed. As Parker observes, climate models are known to be false (Parker, 2009, *p*. 235). Parker's observation, however, does not affect weaker versions of the standard view, that is, versions according to which we can confirm a climate model by confirming less than its truth. Moreover, given Lloyd's reliance on robustness analysis, it seems that Lloyd can reasonably be thought of as presupposing that confirming GCMs only involves confirming the truth of their shared assumptions and results, and thus involves confirming less than model truth.

Lloyd's position is nevertheless problematic. GCMs incorporate highly corroborated theoretical physics, e.g., the Navier–Stokes equations, as well as observational information. And I have pointed out elsewhere (Katzav, 2013a) that those GCM assumptions that are responsible for GCM successes include some of their highly corroborated component claims as well as known to be false claims. Claims from these classes are, I have argued, not possible targets for confirmation in tests of climate models (Katzav, 2013a, *p*. 17). I concluded that, therefore, inference to the best explanation does not provide us with the warrants we have in light of climate model successes. For inference to the best explanation warrants, at least, all those of a model's assumptions that are responsible for its successes (Katzav, 2013a, *p*. 19). Now, even if some model components that are responsible for a climate model's success cannot be confirmed because they are already well established, we may be able to confirm the conjunction of those assumptions that are responsible for its success. So my argument regarding well-established claims is tangential to our concerns here. Still, my point that false claims are responsible for the successes of GCMs implies that we cannot, as Lloyd implies we can, confirm the truth of the conjunction of assumptions that are responsible for such a model's success.[4]

Arguably, not all false GCM assumptions are responsible for such models' successes. Fig. 1 allows us to see that successful simulation of global mean surface temperature (GMST) by IPCC-AR4 GCMs is not needed for all the successes these models exhibit. We can see that most of the GCMs simulate GMST poorly. But the GCMs that simulate GMST poorly share in many of the major successes of all the GCMs, e. g., in successfully simulating GMST trends. So some inaccurate assumptions of the GCMs that simulate GMST poorly, namely, those inaccurate assumptions that are responsible for the poor simulation of GMST, are arguably irrelevant to some of the shared GCM successes. Still, some false assumptions that are shared by all the GCMs in question play a crucial role – the models would not run without them – in generating model successes. For example, all the GCMs overestimate the extent to which the greenhouse effect of water vapour – the most important greenhouse gas – amplifies the warming effect of anthropogenic greenhouse gases (Sun, Yu, & Zhang, 2009). Similarly, the GCMs share biases in their simulations of internal variability – that is of climatic changes due to the climate's own dynamics rather than to external forcing factors such as solar radiation and anthropogenic greenhouse gases (Valdes, 2011; Ruiz-Barradas, Nigam, & Kavvada, 2013).

One might suggest, in response, that theoretical knowledge sufficed to show that although the GCMs did rely on false assumptions in generating their successes, correcting these assumptions would not substantially affect the successes. If so, then while GCMs were not able to generate their successes without relying on the relevant assumptions, one might think that theory allows us to see that these assumptions are not responsible for the successes. This may well be possible for some of the shared biases of the GCMs at hand, but it is not possible with respect to all these biases. We often lack established, quantitative theory that allows concluding that such biases are irrelevant to GCM successes. The above two examples are cases in point. Sun et al., for example, claim no more than that the "common biases revealed… may not necessarily be carried over to the simulated global warming" (Sun, Yu, & Zhang, 2009, *p*. 1287).

Things are, in fact, worse than I have just been suggesting. We will often have good reason to conclude that some shared climate model assumptions are wrong without being able to identify which are wrong.[5] For example, Wallace, Fu, Smoliak, Lin, and Johanson (2012, *pp*. 14340–1) attribute the shared bias they uncover in simulations of patterns of observed warming mainly to limited ability to simulate the dynamics relating to internal variability. But they recognise that the bias may be due to limited ability to simulate the role of anthropogenic factors in the climate system. Thus, not only can the conjunction of those climate model assumptions that are responsible for GCM successes not be confirmed, we often do not even know, once we know that these assumptions include errors, where the error precisely is and thus which assumptions remain possible targets of confirmation.

A weaker version of the standard view than Lloyd's remains to be considered. It might still be suggested that it is useful to confirm that the conjunction of those assumptions that are responsible for a GCM's success is, to some specified degree or another, approximately true and thereby to confirm the model to some specified degree or another. The problem with this suggestion is that, according to it, GCM assessment merely results in information such as that the conjunction of assumptions that are responsible for the success of a GCM is moderately true, or that it is closer to the truth than corresponding conjunctions in other GCMs. Such information does not have any implications about the specific aspects of the climate system that climate scientists and others who rely on their work are interested in. For example, it does not have implications about changes in tropopause height, temperature changes over the rest of the century or the causes of changes in Arctic sea ice coverage. Envisage being told that the conjunction of assumptions that are responsible for the success of one GCM is closer to the truth than the corresponding conjunction of a GCM from an earlier generation. This information does not, since the additional truth content may be distributed across the assumptions of the newer GCM in a variety of ways, tell us how the additional truth content of the newer GCM is distributed across its assumptions. Moreover, without knowing where model assumptions have improved, we will not be able to draw new conclusions from these assumptions regarding specific aspects of the climate system.[6]

[4] Odenbaugh and Alexandrova (2011) consider the use of robustness analysis in cases where the models used include differing idealisations, and hence differing false assumptions. They note that, in such cases, we cannot examine the robustness of results to replacing all idealised, non-shared model assumptions with accurate alternatives. But, they claim, it is only if we do so that we can rightly apply robustness analysis in order to confirm shared model assumptions. Kuorikoski, Lehtinen, and Marchionni (2012) respond by rejecting this condition for the applicability of robustness analysis. I, however, am objecting to the applicability of robustness analysis when some model assumptions that are responsible for model successes, and thus some shared model assumptions, are known to be false. Kuorikoski et al. would agree that robustness analysis is not applicable in such cases.

[5] Lenhard and Winsberg (2010) argue that we will often have good reason to conclude that some climate model assumptions are wrong without being able to identify which are wrong. They do not, however, make this point specifically about shared model assumptions.

[6] Would concluding that the conjunction of assumptions that are responsible for the success of a GCM is *very* close to the truth be more useful? For there to be any chance of drawing such a strong conclusion, the background knowledge built into GCMs would have to be well beyond what is currently built into them. With

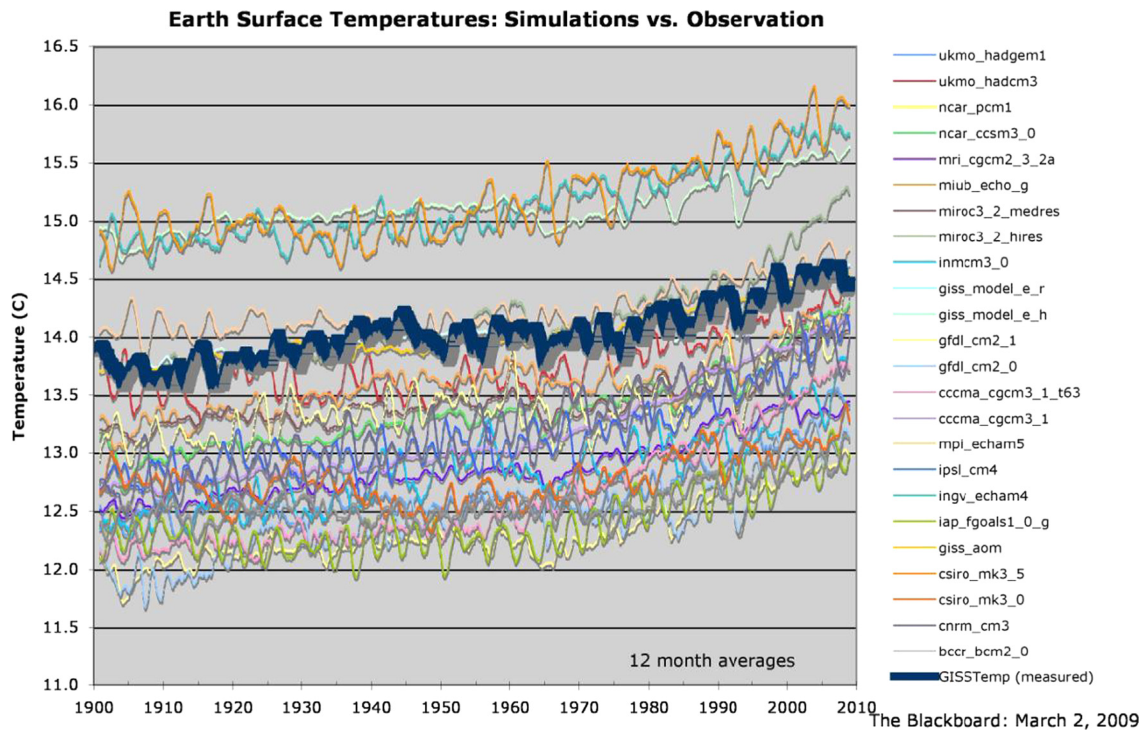

Fig. 1. Observations (thick blue line) and IPCC-AR4 Model Simulations of recent GMST (thin coloured lines). This figure was produced by Lucia Liljegren using data from the CMIP3 archive. http://rankexploits.com/musings/2009/fact-6a-model-simulations-dont-match-average-surface-temperature-of-the-earth/. Although climate scientists are familiar with the information this figure contains, it is rarely found in refereed publications. Fyfe et al. (2013) is the earliest example I have managed to locate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Adequacy-for-purpose assessments

The next view to be considered here is the adequacy-for-purpose view. According to Parker (2009), we need to recognise that climate models are representational tools and that, as such, they serve particular representational goals. Correspondingly, useful assessment of climate models will include assessing their adequacy for specific (representational) purposes, e.g., their adequacy for specific explanatory or predictive purposes.[7] Parker does not explicitly specify what characterises adequacy-for-purpose assessments as such. Alexandrova (2010) worries that, without this information, the adequacy-for-purpose view will be too broad. Indeed, if the view neither constrained the kinds of purposes that adequacy-for-purpose assessments might be concerned with nor the forms of argumentation they might require, every climate model assessment would count as an adequacy-for-purpose assessment. Even assessment that aimed to confirm the truth of a model and, thereby, to confirm its predictions would do so. As a result, the adequacy-for-purpose view would be compatible with all the other views discussed here. I accordingly begin this section by further characterising adequacy-for-purpose assessments. Once this is done, I more fully characterise the adequacy-for-purpose view and then proceed to criticise it in light of my characterisation.

One way of characterising adequacy-for-purpose assessments is by narrowing down the set of purposes that might be the goals of such assessments. Alexandrova proposes that adequacy-for-purpose hypotheses generally be understood to be very specific and very local claims, e.g., they may be about "such and such observables within such and such a range under such and such conditions" (Alexandrova, 2010, *p*. 298). But each typical climatic quantity is strongly dependent on many other such quantities. As a result, being able to predict one climatic quantity to within some margin of error typically requires being able to predict many other such quantities suitably accurately. Thus, in order for a model to be able to predict one climatic quantity of interest, the model will typically have to simulate a wide variety of climatic quantities sufficiently accurately as well as sufficiently accurately to simulate how these quantities interact. Accordingly, it seems that characterising adequacy-for-purpose assessments in terms of the narrowness of their targets is not the way to go.

Consider GCM predictions of 21st century GMST trends. For these predictions to be accurate to a specified degree, GCMs must be able to simulate 20th and 21st century climate internal variability sufficiently accurately. In addition, GCMs need to be able to simulate, with sufficient accuracy, how the various external factors influence the climate system during the period under consideration. A natural way of describing what the models need to capture overall is by saying that they need to capture many of the fundamental climatic processes and the way these interact during the 20th and 21st centuries.

To be sure, a model can sometimes have features that allow it to compensate for its limited ability to simulate some climatic quantities. An underestimate of the influence of one quantity on a prediction might, for example, be compensated for by an overestimate of the influence of another quantity. However, such compensation will, given the complicated ways in which climatic quantities interact, be partial and sufficiently accurate simulation of the involved quantities will still be required. This may be

(*footnote continued*)
such a high level of built-in knowledge, it would remain unclear what new information about specific aspects of the climate system the conclusion would licence. In any case, known GCM errors, and especially errors that result from the fact that GCMs have a horizontal resolution of several hundred kilometres and cannot explicitly represent phenomena that are smaller than this resolution, are too substantial to allow anything like the supposition that the relevant conjunctions are very close to being true.

[7] A similar position is found in Shackley (2001).

illustrated by the results from the generation of GCMs being used in the IPCC's fifth assessment report. The way in which these models combine internal variability and external forcing may differ from the way it is combined in the climate system (Fyfe, Gillett, & Zwiers, 2013). Nevertheless, the models successfully simulate 20th century GMST trends (Knutson, Zeng, & Wittenberg, 2013), thus suggesting that, to some extent, model errors cancel each other out when it comes to their implications for GMST trends. At the same time, error cancellation is limited. First, the models nevertheless do accurately simulate many aspects of internal variability. Second, inaccuracies in simulated internal variability and external forcing may well be why the models substantially overestimate the increase in GMST since 1992 (Fyfe et al., 2013).

A second way of characterising adequacy-for-purpose assessments is in terms of the kind of arguments they require. Obviously, if we have a well confirmed model from which a prediction follows, we can conclude that the model is adequate for the purpose of generating the prediction. But how do we confirm that a model is adequate for some purpose when we cannot confirm the model? Parker (2009, *p*. 237) suggests that the first requirement is that

> (i) we determine what observations are likely if the model is adequate for the purpose and whether what is actually observed fits with what we are likely to observe if the model is thus adequate.

When condition (i) has been met in the case of a given adequacy-for-purpose hypothesis and, further, fit with observations is sufficient, the hypothesis is confirmed (Parker, 2009, *p*. 237). Now, we can use condition (i) in order to characterise adequacy-for-purpose assessments. We can stipulate that the assessment of the adequacy of some model for some purpose is an adequacy-for-purpose assessment if, and only if, the basis for the assessment is provided by fulfilling condition (i) in the case of the purpose in question. What is required of such assessment is thus adequacy-for-purpose hypothesis specific argumentation. Assessing whether an adequacy-for-purpose hypothesis is indirectly confirmed via the confirmation of some model is no adequacy-for-purpose assessment of the hypothesis.

The proposal that adequacy-for-purpose assessments be characterised in terms of a kind of argument captures Parker's central idea (Parker, 2009, *p*. 238 & 242) that it will often be difficult, and that in many cases we may be unable, to develop the kinds of arguments needed in order to confirm adequacy-for-purpose hypotheses. What we need to observe in order to confirm a model is relatively straightforward. Any model result could in principle do for testing the model. But, Parker points out, if our concern is with the model's adequacy for some predictive or explanatory purpose, it will typically be far from clear which of the model's implications for currently observable quantities are possible sources of confirmation/refutation (Parker, 2009, *p*. 237, *p*. 240–8). For what the implications of a claim that a model is adequate for some predictive or explanatory purpose are for current observations is often not a straightforward matter.

Here is Parker's line of reasoning. Determining whether a model is adequate for the purpose of explaining or sufficiently accurately predicting some climatic quantity requires determining whether its simulations of relevant observed climatic quantities are sufficiently accurate and comprehensive for this purpose. Now, as I have noted, we know that simulating one quantity well will typically require simulating many other quantities well. But our knowledge of how a model's ability to simulate one quantity well to a specified degree is dependent on the accuracy of its simulations of other quantities is often limited. As a result, it will often be a challenge to figure out how an error in simulating one observed quantity might affect a models' ability to predict or explain another quantity. For example, consider the claim that some GCM is adequate for the purpose of accurately simulating late 21st century GMST trends. What the observable implications of this claim are is not straightforward. It might be reasonable to suppose that accurately simulating observed GMST trends is relevant here, but one also needs to consider how the ability to simulate other related, observable climatic quantities, such as precipitation levels or absolute GMST, might be relevant. And doing so involves taking into account complex considerations concerning the interdependence of the relevant quantities and the way GCMs simulate these quantities.

Indeed, adequacy-for-purpose assessment is even more challenging than Parker implies it is. As we have seen, assessing the adequacy of a model for a given explanatory or predictive purpose typically involves assessing its adequacy for a wide variety of other explanatory and predictive purposes. The challenges Parker points to will arise with respect to each involved assessment.

We can, finally, characterise the adequacy-for-purpose view. It is the view that useful climate model assessments are primarily adequacy-for-purpose assessments. It adds that an adequacy-for-purpose assessment is an assessment of an adequacy-for-purpose hypothesis in which the basis for the assessment is provided by fulfilling condition (i) in the case of the relevant purpose. Assessing models in the way described by the standard view is allowed on the adequacy-for-purpose view. However, given the supposed primacy of adequacy-for-purpose assessments, assessing climate models must be supposed to be of limited use.

Despite noting that adequacy-for-purpose assessment tends to be difficult, Parker thinks that such assessment is sufficiently feasible so as to allow it to comprise the core of climate model assessment.[8] She might be wrong about this, however. Perhaps the difficulties she raises for carrying out adequacy-for-purpose assessments are such that, not just often but in typical cases, they cannot be carried out. The worry that she might be wrong here is supported by my observation above that adequacy-for-purpose assessments are more difficult than she implies they are. Further, I now want to bring out yet another difficulty that arises in trying to carry out adequacy-for-purpose assessments, one that indicates that such assessments will, in typical cases, largely be question begging. If this is correct, adequacy-for-purpose assessment should not comprise the core of climate model assessment.

Adequacy-for-purpose assessments involve estimating what the degrees of accuracy of simulations of a wide variety of observed climatic quantities imply about the correctness of uncertain model assumptions and results. Partly, this is a matter of seeing how, according to the models, uncertain climatic quantities depend on observed climatic quantities. But it is also a matter of figuring out how accurate models actually are about the dependencies in question and thus of estimating what the dependencies actually are.[9] The problem is that, typically, we are only really able to provide estimates of how climatic quantities depend on each other by relying on climate models. This is especially true if the

[8] Parker (2009, *p*. 236–7) takes adequacy-for-purpose assessment to characterise assessment of climate models in general, whether this be assessment of the extent to which models capture existing knowledge or assessment of the kind I am concerned with in this paper, namely, assessment that aims to produce new knowledge about the climate system.

[9] Dependencies between simulations of observed quantities and model estimates of uncertain quantities need to be figured out because the degrees of dependency are suggestive of how relevant the accuracy of the simulations is to the accuracy of the estimates. Dependencies between the accuracy of simulations of observed quantities and the accuracy of estimates of uncertain quantities need to be figured out because we need to be sure that what the models tell us is/is not relevant to the accuracy of an estimate actually is/is not relevant.

estimates need to be, as they usually need to be, quantitative estimates of how complex climatic phenomena depend on each other, e.g., of how patterns of European precipitation change depend on Arctic sea ice extent. So, in typical adequacy-for-purpose assessments, we would have to rely on climate models to tell us what the accuracy of simulations of observed climatic quantities implies about the correctness of uncertain model assumptions and results. But in thus relying on models we would just be assuming that, for the most part, the models are correct about how the uncertain quantities we are interested in depend on observed climatic quantities. More explicitly, we would be assuming that, for the most part, the models are correct about how the uncertain quantities depend on the climate system's own dynamics and external forcing factors. But this would just be to assume that, for the most part, the models are adequate for whatever purpose they are being considered.

Suppose we are assessing whether a GCM is adequate for the purpose of predicting GMST trends over the rest of this century to a specified degree of accuracy. We will have to determine whether the model simulates the observable aspects of the climate system's own dynamics sufficiently well for it to be adequate for this purpose. Thus, for example, we will have to see whether the model simulates modes of internal variability such as the Pacific Decadal Oscillation and the Atlantic Multidecadal Oscillation sufficiently well. However, estimating how well the climate system's own dynamics needs to be simulated involves estimating how GMST trends depend on it, and GCM simulations need to be relied on in doing so. As the IPCC third assessment report states,

> [t]he instrumental record is short and covers the period of human influence and palaeo-records include natural forced variations, such as those due to variations in solar irradiance and in the frequency of major volcanic eruptions. These limitations leave few alternatives to using long "control" simulations with coupled models for the estimation of [GMST trends relating to] internal climate variability (IPCC, 2001, *p*. 56).[10]

But in relying on GCM simulations to teach us about the effects of the climate system's own dynamics we are just assuming that, largely, GCMs are correct on this matter.

One might respond, on behalf of the adequacy-for-purpose view, that relying on what climate models tell us about how some climatic quantities depend on each other is acceptable when empirical evidence supports trusting the models regarding these dependencies. Thus, we might hope that what GCMs tell us about how 21st century GMST trends depend on the climate system's own dynamics is supported by empirical studies. But many of the dependencies that need to be examined in adequacy-for-purpose assessments are dependencies about which we do not have the supporting evidence. This is so when it comes to the effects of the climate's own dynamics, including, as the above quote from IPCC (2001) indicates, its effects on GMST trends. While there are empirical studies of the effects of the system's dynamics on 20th century GMST trends, the studies often do not agree with what GCMs tell us (see, e.g., Swanson, Sugihara, & Tsonis, 2009; Wu, Huang, Wallace, Smoliak, & Chen, 2011; Tung & Zhou, 2013). Moreover, since empirically isolating the effects of internal variability is challenging, the conclusions of the relevant empirical studies are primarily taken to be suggestive about what is the case or indications of what might be the case. They are not able to offer real support to climate model adequacy-for-purpose hypotheses.

One might also respond that model assumptions can, and do, vary from climate model to climate model. As a result, one might suggest, climate model ensembles can assist in assessing what some climate model limitations imply about the accuracy of climate model results and do so without begging relevant questions. Consider again the case of simulations of 20th century GMST trends. The agreement of GCMs about these trends despite GCMs' differing estimates of 20th century GMST might suggest that the accuracy of their predictions of GMST trends for the 21st century is substantially independent of their ability to simulate GMST accurately. So, one might think, one can legitimately appeal to the ensemble results indicating the relative independence of the accuracy of simulations of GMST trends from the accuracy of simulated GMST to show that a GCM's limited ability to simulate GMST should not undermine confidence in its ability to predict 21st century GMST trends.

Appeals to inter-model differences do not, however, allow us to avoid begging relevant questions regarding adequacy-for-purpose hypotheses. Differences between climate models do not help in exploring the implications of shared model limitations. Moreover, even when climate models do allow exploring the implications of some of their limitations, they can only partially do so. In the case of the question of the relevance of the limited ability to simulate GMST to the accuracy of simulations of 21st century GMST trends, it may well be that the relative independence of the ability to simulate GMST accurately and the ability to simulate GMST trends accurately is due to shared GCM errors. The GCMs cannot, without begging the question, be appealed to in order to determine whether this is so.

A less ambitious version of the adequacy-for-purpose view might still be thought to evade the worries raised above. One could accept that confirming adequacy-for-purpose hypotheses is typically unachievable but insist that such hypotheses can, often enough, be refuted or undermined by sufficient knowledge of model limitations.[11] Indeed, we do sometimes recognise that models are too limited to be of use for some purposes, e.g., we recognise that climate models are not adequate for the purpose of predicting weather. But such cases are typically cases in which there was, at the outset, little expectation that climate models would be adequate for the relevant purposes. This paper's concern is with cases in which assessment might teach us something new about the climate system. In such cases, we will need to draw out the implications of our knowledge of the climate system in order to determine whether specific, known model limitations imply that models are not adequate for our purposes. And climate models will typically have to be relied on in doing so. In any case, if adequacy-for-purpose assessments typically yield no more than information about what models cannot do, such assessment will only exceptionally do more than suggest ways in which models might be improved. It would rarely assist in learning about the climate system. Moreover, by implication, it would rarely assist in developing policy that is based on climate model results.

Finally, the use of expert judgement to supplement the results of GCM ensembles does not affect my conclusion about the adequacy-for-purpose view. As we saw in the previous section, expert knowledge of the source of GCM errors and what these errors imply regarding model assumptions and results is, to a significant extent, limited. So experts can typically only provide partial estimates of what shared GCM biases might imply about GCMs' adequacy for purposes of interest.

[10] The control simulations appealed to are simulations of pre-industrial climate in which external forcing is constant.

[11] This view was suggested to me by Parker in conversation.

4. The conservative and Bayesian views

We have seen that, since the conjunctions of assumptions that are responsible for climate models' successes are known to be false, confirming these conjunctions cannot be a useful target of climate model assessment. For this reason, the conservative view, unlike the standard view, does not allow them to be confirmed. Model assumptions and results that can be confirmed, on the conservative view, include only model assumptions and results about which there is real uncertainty. Additional claims that can be confirmed on this view include appropriately corrected versions of known to be biased climate model results and assumptions. An appropriately corrected claim is one that has been corrected in a way that is justified by background knowledge. Assume that we have a set of model results that are known to be biased in some way, e.g., a range of predictions that is known to be biased upwards because the models fail to represent some climatic mechanism. Other things being equal, the conservative view allows confirming a correction to this range of predictions, one that shifts the range downwards in order to compensate for the known bias.

Thus far, I have characterised the conservative view in terms of what it tells us about which claims are the targets of climate model assessment. It tells us that such assessment is of all uncertain climate model assumptions and results, as well as of appropriate corrections to known to be biased climate model assumptions and results. The conservative view, however, says more and, in doing so, takes the discussion of the adequacy-for-purpose view into account. That discussion indicated that assessing climate models can, in typical cases, only involve a limited examination of what known model limitations imply about the accuracy of uncertain model assumptions and results. So, if climate model assessment is ordinarily to result in some degree of confidence in these assumptions or results, models will have to be trusted about them to some degree and trusted even though the implications of known model limitations have only been partially explored. What the conservative view accordingly tells us is that adequacy-for-purpose assessment is not required. Rather, the uncertain assumptions and results of a model are given some initial, defeasible degree of confirmation in light of the model's success. The same is true of corrections to known to be biased model assumptions and results. These initial degrees of confirmation are then supposedly weakened or strengthened in light of background knowledge. A higher degree of confirmation is given to an assessed claim when background knowledge, including knowledge of the assumptions and results of other models, supports the claim. Confirmation is reduced to the extent that the claim is undermined by background knowledge.

For example, if background knowledge tells us that our model is a relatively realistic representation of the climate system, the presumptive confirmation one of its uncertain results receives will, other things being equal, be modified upwards. For it is natural to suppose that a model that is a more realistic representation of the climate system will, other things being equal, be more trustworthy than less realistic models and thus confer a greater degree of confirmation on its uncertain assumptions and results. Similarly, if we know that a result is produced by all of our most sophisticated GCMs and, further, makes sense from a theoretical perspective, it will receive an even higher degree of confirmation. Conversely, if GCMs disagree about the result and, further, background theory does not support it, its initial, presumptive confirmation will be reduced.

The conservative approach is, given our limited ability to examine the implications of known model limitations, inevitably going to be implicit in climate model assessment that aims to establish confidence in claims about how the climate system actually is. And such assessment has been, and is, common. Thus, for example, the chapter on projections in the IPCC third assessment report takes a result to be very likely to virtually certain if it is a result that is shared by most models (IPCC, 2001, *p*. 527). It does this despite the inevitably limited exploration of the impacts of known model limitations on the accuracy of model results. Chapter 8 of IPCC-AR4 (Randall et al., 2007) builds confidence in GCMs by showing that these incorporate theoretical knowledge about, and successfully simulate important observed aspects of, the climate system. It is in light of this general confidence that GCM predictions are then assessed in Chapters 10 (Meehl et al., 2007) and 11 (Christensen et al., 2007). In these chapters, confidence in many GCM results is expressed even though, once again, exploration of the implications of known climate model limitations for the accuracy of model results was inevitably limited.

A first candidate worry about the conservative view can be extracted from Mayo's work (Mayo, 2010). She worries that views of confirmation that allow evidence to be good evidence for a theory merely because the theory better accounts for the evidence than do its rivals allow evidence that is not guaranteed to be reliable to be good. Her worry is relevant here because the conservative view does not require of good evidence that it be reliable. True, even the adequacy-for-purpose view need not fully address the issue of the reliability of adequacy-for-purpose assessments. But, as Parker makes clear (Parker, 2009, *p*. 242), the adequacy-for-purpose view can easily be supplemented so that it does address this issue. The requirement that we test adequacy-for-purpose hypotheses by examining their own empirical implications can be supplemented with the requirement that, for an adequacy-for-purpose hypothesis to pass a test, the test must provide what Mayo would call good evidence for the hypothesis. According to Mayo (1996), a test provides good evidence for a hypothesis if the hypothesis agrees with the evidence produced by the test and it is very unlikely that the hypothesis would have agreed with the evidence as well as it does were the hypothesis false. The conservative view, by contrast, cannot be adjusted so as to demand much more of evidence before it is confirmatory. To do so would be to give up on the conservative view's core assumption that uncertain climate model assumptions and results should receive an initial, presumptive degree of confirmation in light of climate model successes.

The literature does not settle the question of what is required of evidence before it is good. Indeed, what is and can be required of evidence before it is good is one of the core disagreements between approaches such as Mayo's and less demanding approaches. While Mayo sees the requirement that good evidence be guaranteed to be reliable as a virtue of her position others will tend to think this requirement is too demanding (see, for example, the discussion in Chalmers (2010) and Mayo (2010)).

However, the conservative view not only comes with no guarantee of reliability, there is in fact a real threat that it will turn out to be unreliable. There is, as we saw in Section 3, research that suggests that there are substantial errors in the uncertain assumptions and results of sophisticated GCMs, notably including empirical studies that suggest that these models do not adequately capture the impacts of the climate system's own dynamics on GMST trends. Since these studies are merely suggestive, they are not generally taken to undermine confidence in GCM assumptions or results.[12] Still, the studies show that there is a real threat that the GCMs are unreliable in shared, currently insufficiently acknowledged ways. Further, we ought, in the present context, to feel substantial discomfort about relying on approaches that

[12] See the discussion of these studies in Chapter 10 of the IPCC fifth assessment report (IPCC, 2013).

threaten to be unreliable. For many of the claims that these approaches are used to evaluate, especially those claims relating to the causes and future of climate change, are of substantial practical import. To be wrong about these claims may be costly indeed.

The undemanding nature of the conservative view makes things even worse for it. Recall that we saw, in Section 2, that the IPCC-AR4 GCMs have shared false assumptions that we have been unable to pinpoint. These shared assumptions are not epistemically distinguishable from the other shared, uncertain assumptions of the GCMs. Thus, the relatively high degree of confirmation that is accrued when the GCMs agree will extend to false model assumptions and results. It turns out, therefore, that the conservative view is not merely potentially unreliable, but is actually unreliable.

How does the Bayesian view fit into the picture here? The Bayesian view is an application of Bayesian confirmation theory to the case of climate models. On some versions of Bayesian confirmation theory, a claim is confirmed by evidence if its posterior probability, i.e., its probability in light of the evidence, is higher than its prior probability, i.e., than its probability independently of the evidence. On other, less popular, versions of Bayesian confirmation theory, confirmation of a claim by evidence amounts to the claim's posterior probability being sufficiently high. The posterior probability of a hypothesis, *h*, in light of evidence, *e*, and background knowledge, *B*, is given by Bayes' theorem in the following way:

$$P(h|e,\ B) = P(e|h,\ B)P(h|B)/P(e|B),$$

where $P(h)$ is the probability of *h* independently of *e*, $P(e \mid h, B)$—the so-called likelihood of *e*—is the probability of *e* given *h* and *B*, and $P(e)$ is the probability of *e*.

Now, the Bayesian view constrains what *h* might be in the context of climate model assessment. Given that the goal of confirming climate models is not a useful one, the Bayesian view tells us that *h* should not be identified with any climate model. And given the limited ability to assess climate model adequacy-for-purpose hypotheses, it tells us that *h* should rarely be an adequacy-for-purpose hypothesis. Rather, *h* should typically be identified with some uncertain model assumption or result, so that the Bayesian goal will be to confirm this assumption or result.

In addition, the Bayesian view incorporates a specific Bayesian definition of confirmation, and a specific approach to calculating prior probabilities and likelihood functions. This is required in order to avoid the problems faced by the standard and adequacy-for-purpose views. The result is that the Bayesian view handles these problems in much the same way as the conservative view does. The definition of confirmation the Bayesian view incorporates identifies confirmation with a high posterior probability. As a result, the Bayesian view, unlike the standard view, will avoid implying that the conjunctions of assumptions that are responsible for the successes of GCMs are confirmed by these successes. We know that these conjunctions are false and thus will assign them very low prior probabilities. So their posterior probabilities will also be low. In order to ensure that the Bayesian view is not too demanding, calculation of likelihood functions and prior probabilities will have to be carried out with only limited exploration of the implications of known climate model limitations for the accuracy of uncertain climate model assumptions and results. The probability of the evidence in light of background knowledge can, for example, simply be equated with the probability climate models give to the most clearly relevant observations. Prior probabilities can be set on the basis of expert judgement.

Bayesian confirmation theory is one of the most prominent theories of confirmation. Thus it is natural and important to try to use it in answering the question of what the target claims of useful climate model assessments might be. Unfortunately, the result of this attempt, i.e., the Bayesian view, threatens to be unreliable in much the same way as the conservative approach does and does so precisely because it is not demanding enough about examining the implications of known model limitations. Indeed, the Bayesian view will be unreliable. Consider, once again, the knowledge that there are false assumptions that are shared by IPCC-AR4 GCMs and that have yet to be identified. Since we are not in a position reliably to identify these assumptions, we cannot assign them low enough prior probabilities so as to ensure that they are not confirmed by model successes. These assumptions will, accordingly and given the confidence commonly invested in sophisticated GCMs, be confirmed along with others. Of course, we might decide to assign all uncertain GCM assumptions low prior probabilities. But then we will have to provide low probabilities to false and true GCM assumptions at the same time. Either way, our posterior probabilities will be unreliable.

There is, however, a worry about the Bayesian view that is more fundamental than the worries raised about the conservative view. The conditions for the applicability of Bayesian confirmation theory to climate model assessment are not conditions that we can usually fulfil. The conditional probability $P(e|h, B)$ is a probability that is conditional on the truth of *h* and *B*. Accordingly, if *h* is some climate model assumption, then *B* will have to comprise additional assumptions the conjunction of which may be assumed to be true and which, in conjunction with *h*, yield *e* with some degree of probability or another. But, typically, no such additional assumptions are to be had. In order to generate *e* we must typically rely on climate models along with those of their shared assumptions that are recognised to be false. The conservative view allows us to rely on sets of assumptions that include assumptions recognised to be false. Bayesian confirmation theory does not do so.[13]

One can, of course, attempt to modify the likelihood functions generated by models so as to correct for the ways in which models are known to be inaccurate. That is, one can attempt to correct models' likelihood functions so that the probabilities they provide are the same as those they would have provided had the models not included known to be false assumptions (see, e.g., Rougier, 2007). But it should by now be clear that corrections will typically have to be very partial. As I have been emphasising, we have a limited ability to correct the results of climate models in a model independent way.

5. Beyond confirmation: the possibilist view

The conservative and Bayesian views, we have seen, do not require that we comprehensively determine the implications of known climate model limitations for the accuracy of uncertain climate model assumptions and results. Unfortunately, the fact that these views are undemanding in this way threatens to make them unreliable and, indeed, they are unreliable. Moreover, since we are only partially able to determine the implications of known climate model limitations for the accuracy of uncertain model assumptions and results, any view according to which climate model assessment primarily aims to confirm claims about how the climate system actually is will threaten to be unreliable in much the same way. For any such view will have to specify which claims about how the climate system actually is are up for confirmation in climate model assessments and will have to do so given only a partial specification of which climate model assumptions are false.

[13] This is why even an undemanding variant of the Bayesian view cannot address the difficulty posed by our partial ability to explore the implications of climate model limitations in the way the conservative view does. The conservative view, as we can recall, addresses the difficulty by assigning model assumptions and results some initial degree of confidence in light of model successes.

In order to avoid this threat, the possibilist view specifies something other than the confirmation of claims about how the climate system actually is as what useful climate model assessment primarily aims at. At the same time, the possibilist view incorporates the improvements of the conservative view over the standard and adequacy-for-purpose views.

Stainforth et al. (2007) and Betz (2010) outline the view that useful assessment of climate model predictions often aims to establish that the predictions describe real possibilities. Katzav et al. (2012) suggest, in addition, that when an assessment manages to establish that predictions describe real possibilities, it may also aim to rank the possibilities as to how remote they are. The possibilist view extends Katzav et al.'s (2012) suggestion to claims other than predictions. In the spirit of the conservative view, the possibilist view tells us that useful assessment of climate models is of all their uncertain assumptions and results and, in some cases, of appropriate modifications to some of these assumptions and results. However, on the possibilist view, useful assessment does not primarily aim to confirm the assumptions, results or modifications in question. Rather, it aims to show that they describe real possibilities and, when this is shown, partially to determine how remote the described possibilities are.

Real possibilities are the kinds of states of affairs that we affirm when we affirm that it is a real possibility that the European Union will be dissolved, that fusion power will become a practical reality within one hundred years or that Neanderthals invented leather-working bone tools before modern humans did. Further, which real possibilities obtain is a time-relative issue. As time goes by, some possibilities cease to be real. We tend to use the language of real possibilities when we are talking about states of affairs the obtaining of which we are uncertain about, usually sufficiently uncertain so as to make us hesitate to assign probabilities to their obtaining. At the same time, we tend to be willing to affirm of states of affairs that they are real possibilities only when we do have a rough understanding of some of the ways in which they might have/might come about. The mere absence of knowledge that something is not the case does not make that something a real possibility.

Building on Katzav et al. (2012), a state of affairs in a target domain is here taken to be a real possibility relative to some time *t* if and only if (a) its realisation is compatible with the basic way things are in the target domain over the period during which it might be realised and (b) our knowledge at *t* does not exclude its realisation over that period. A state of affairs will be said to be compatible with the basic way things are in a target domain over a period of time if, in something like the circumstances obtaining over the period, the domain's laws and/or mechanisms, or a similar set of such laws and/or mechanisms, would bring the state of affairs about. One of the ways, and the one which will interest us below, in which a representation can capture the basic way a system is over a period of time is if it provides a rough characterisation of the central processes, along with the laws and/or mechanisms that underlie these processes, of the system over the period.[14]

Given the above definition of a real possibility, a case for thinking that it is currently a real possibility that a climatic state of affairs, *s*, will obtain/has obtained over a specified period of time can be made by arguing that (a1) *s*'s then obtaining is compatible with the basic way the climate system then is and that (b1) current knowledge does not exclude *s*'s then obtaining. We can argue for (a1) by showing that a GCM simulation of the relevant period assumes that *s* obtains during the period or yields the result that *s* then obtains and that the simulation captures the basic way the climate system is at the relevant times. (b1) can be ascertained simply by examining what our background knowledge explicitly tells us about *s*.

Consider, once again, the implication of IPCC-AR4 simulations of the 20th century climate that anthropogenic greenhouse gases were responsible for most of the observed post-1950 global warming. We can argue that this implication describes a real possibility by arguing that the GCM simulations which yield it capture the basic way the climate was in the 20th century and, further, that background knowledge does not show that it is false.

In cases where background knowledge indicates that models are only able to simulate some of the ways the climate system might really be, background knowledge may nevertheless indicate how model results can be modified so as to capture a broader range of ways it might really be. For example, our knowledge that GCMs fail to represent some factor that tends to warm the atmosphere might lead us to conclude that a broader range of positive temperature anomalies than those simulated by GCMs are real possibilities.

A crucial question is whether GCMs can indeed help to establish that some possibilities are real. Betz (2010, *p.* 96) worries that they cannot do so. Since climate models are false, the total states of affairs they represent are not real possibilities. But, Betz points out, this means that we cannot argue that since a climate model represents a real possibility, so do its predictions. My proposal, however, does not assume that climate models represent real possibilities. Climate models are only supposed to help us to show that certain states of affairs are compatible with the basic way the climate system is over relevant periods of time. So the models only need to provide us with simulations that represent the basic way the climate system is over the periods in question. And representing the basic way the climate system is over a period of time is compatible with being false to a substantial degree. It only requires representing something *like* the circumstances that obtain in the system and something *like* the way in which the system evolves. Plausibly, given the substantial knowledge built into GCMs and given the empirical successes of their simulations, their simulations often provide what is required here.

Further, the arguments according to which the conservative and Bayesian views threaten to be unreliable, never mind those for these views' actual unreliability, do not affect the possibilist view. Assessing that some climatic state of affairs is a real possibility is subject to error. But the relevant error here is that of mistakenly supposing that GCM simulations roughly characterise the central processes of the climate system over the period when the state of affairs might obtain. And the kinds of model limitations which motivated worries about the reliability of the conservative and Bayesian views do not motivate the worry that GCM simulations are less than roughly correct about these processes. A rough characterisation of a system's central processes is often rough precisely in that it does not include relevant processes and does not fully represent those it does include. The relevant GCM limitations we suspect exist involve no more than a failure to represent some relevant processes and an incomplete representation of others.

Nor is the possibilist view too demanding. In order to argue that, relative to the present, some simulated climatic state of affairs is a real possibility, we do not have to engage in detailed estimation of what known GCM limitations imply about the accuracy of uncertain GCM assumptions and results. Determining whether background knowledge, as it is, shows that a state of affairs does not obtain is merely a matter of noting whether this knowledge

[14] My guess is that the claim that a model is roughly correct about the central processes of its target system is weaker than the claim that the conjunction of those of its assumptions that are responsible for its success is approximately true. If so, showing that a model is roughly correct about the central processes of its target system is not enough to confirm the model in the sense specified by the standard view.

has been able/unable to generate the conclusion that the state of affairs does not obtain. And determining whether a GCM simulation captures the basic way things are in the climate system is part and parcel of standard practice. It is a matter of examining the extent to which the GCM which produced the simulation incorporates relevant background knowledge as well as how well it is able to simulate observed aspects of the climate system.

Let me emphasise, however, that although the possibilist view tells us to use GCMs in order to establish that some possibilities are real, it does not tell us that GCMs should in general be used to show that some possibilities are not real. Inferring that some possibility is not real from the fact that GCMs fail to simulate it does threaten to be unreliable. As a variety of authors have pointed out, existing GCMs are aptly described as 'best guess' models (see, e.g., Stainforth et al., 2007; Katzav, 2013b). They are, despite their differences, similar attempts to model the climate system rather than the result of an attempt to produce models that collectively explore all the theoretical possibilities that are compatible with our knowledge of how the system is. Accordingly, existing GCMs do not allow exploring all the theoretical possibilities that are compatible with our knowledge of the basic way the climate system actually is. Yet some of these unexplored possibilities may turn out to be real ones. Indeed, we can recall, the possibilist view takes this into account in that it allows, when background knowledge supports doing so, modifying model results so as to broaden the range of real possibilities they represent.

I have considered what role GCMs might have in assessing whether their assumptions and results describe real possibilities. I still need to say something about how to rank real possibilities as to how remote they are. A real possibility's remoteness depends on the number of conditions that need to obtain for it to be realised as well as on how these conditions depend on each other. Since GCMs can help us to investigate the conditions for the realisation of real possibilities, one might be tempted to use GCMs to assess how remote real possibilities are. For example, one might be tempted to argue that, since a certain effect is a real possibility but one that GCM simulations suggest is rare, the effect is a remote possibility. However, GCM errors about the number of relevant conditions required for some really possible climatic state of affairs to be realised, or about how these conditions depend on each other, threaten to result in errors in GCM-based estimates of how remote a possibility the state of affairs is. Here too, then, worries about unreliable assessments suggest limiting appeals to GCM results.

Nevertheless, we can sometimes provide a reliable, comparative ranking of the remoteness of some of the real possibilities GCMs teach us about. Consider cases in which GCM results indicate that it is a real possibility that the anthropogenic influence on the climate system will have some effect *x*. In such cases, we can conclude that *x* is a real possibility that is less remote than it would have been in the absence of anthropogenic influences. For the absence of anthropogenic influences would just mean the obtaining of fewer of the conditions that are conditions for the obtaining of *x*. For example, GCM simulations of the possible effects of increases in anthropogenic greenhouse gases on the climate system suggest that one such effect might be the shutoff of the Atlantic Thermohaline Circulation in the 22nd century (Lenton et al., 2008). Assume that we can conclude, on the basis of the GCM results, that such a shutdown is a real possibility. We can then also conclude that it is less remote than it would have been in the absence of the influence of anthropogenic greenhouse gases on the climate system.

It is worth noting, before closing this section, that the possibilist view may also be able to illuminate the assessment of simple climate models. Consider simple energy balance models. They model the climate system solely on the basis of energy balance considerations. They do not explicitly represent any of the mechanisms that govern climatic processes or even its three-dimensional structure (North, Cahalan, & Coakley, 1981). Such models, it might accordingly seem, cannot be used to argue that some of their assumptions or results describe real possibilities. But the parameters of simple climate models can be, and often are, set by climate scientists to fit the settings of related GCM parameters. In such cases, one can view some of the results of the simple models as proxies for those of GCMs (Randall et al., 2007, Section 8.8.1) and, accordingly, as candidates for being evaluated as to whether they are real possibilities.

6. Concluding discussion

Climate model error, we have seen, extends to model assumptions that are responsible for model successes. As a result, contrary to the standard view, climate models are not useful targets of confirmation. In particular, aiming to confirm climate models cannot, as Lloyd presupposes it can, establish claims about the causes of climate change. The adequacy-for-purpose view makes, because it commits us to using climate models to assess the implications of known climate model limitations, assessment of whether a model is adequate-for-purpose impossible in typical cases. The conservative approach, which I have suggested is inevitably going to be the main approach used when climate model assessment aims to confirm claims about the climate system, represents partial progress. It does not target the confirmation of models and is not too demanding. It is, however, unreliable. The Bayesian view would, if it were used, be unreliable in much the same way and, in any case, usually cannot be used because we do not usually know how to calculate the likelihood functions that it requires we calculate. The possibilist approach has the merits of the conservative approach, but avoids the threat of unreliability.

What are the general implications of my discussion for climate model assessment? In typical cases, what climate model assessment aims to establish should not be model adequacy-for-purpose or the direct confirmation of claims about how the climate system actually is. Rather, it should concern more modest targets, targets such as establishing that certain possibilities are real. This goes contrary to standard IPCC practice, which has as an important focus the use of climate models in order to establish confidence in claims about how the climate system actually is.

What are the implications of my discussion for the philosophy of science in general? As just noted, a precondition for applying Bayesian theories of confirmation in climate model assessment does not usually obtain. Further, errors in climate model assumptions turn out to be too pervasive to allow the useful application of many familiar forms of ampliative inference in such assessment. For many familiar forms of ampliative inference tell us that the conjunctions of those model assumptions that are responsible for model successes are confirmed by such successes. We have seen this in the case of robustness analysis and in the case of inference to the best explanation. The same can be seen, however, when it comes to hypothetico-deductive inference, inference to the most likely cause and inferences warranted by relative likelihood views of confirmation. Even more broadly, epistemologies of science that tell us that what scientific assessment focuses on is confirmation either threaten to be, or are, unreliable in the context of climate model assessment. In attempting to illuminate such assessment, we should accordingly focus on views which, like the possibilist view, focus on more modest targets of assessment.

Acknowledgment

Thanks to Wendy Parker and to the anonymous referees for their comments on this paper.

References

Alexandrova, A. (2010). Adequacy-for-purpose: The best deal a model can get. *The Modern Schoolman*, *LXXXVI*, 295–300.

Betz, G. (2010). What's the worst case? The methodology of possibilistic prediction. *Analyse & Kritik*, 87–106.

Chalmers, A. (2010). Can scientific theories be warranted?. In: D. G. Mayo, & A. Spanos (Eds.), *Error and inference* (pp. 58–72). Cambridge: Cambridge University Press.

Christensen, J. H., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., et al. (2007). Regional climate projections. In: S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge and New York: Cambridge University Press.

Fyfe, J. C., Gillett, N. P., & Zwiers, F. W. (2013). Overstimating global warming over the last 20 years. *Nature Climate Change*, *3*, 767–769.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, *172*(2), 269–281.

IPCC (2001). Climate change 2001: The scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change. In: J. T. Y. Houghton, D. J. Ding, M. Griggs, P. J. Noguer, X. van der Linden, K. Dai, Maskell, & C. A. Johnson (Eds.), Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

IPCC (2013). Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. In: T. F. Stocker, D. Qin, G. -K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Katzav, J. (2013a). Hybrid models, climate models and inference to the best explanation. *British Journal for the Philosophy of Science*, *64*(1), 107–129.

Katzav, J. (2013b). Severe testing of climate change hypotheses. *Studies in History and Philosophy of Modern Physics*.

Katzav, J., Dijkstra, H. A., & Jos de Laat, A. T. J. (2012). Assessing climate model projections: State of the art and philosophical reflections. *Studies in History and Philosophy of Modern Physics*, *43*(4), 258–276.

Knutson, T. R., Zeng, F., & Wittenberg, A. T. (2013). Multi-model assessment of regional surface temperature trends: CMIP3 and CMIP5 20th century simulations. *Journal of Climate*, http://dx.doi.org/10.1175/JCLI-D-12-00567.1 (early online release).

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2012). Robustness analysis disclaimer: please read the manual before use! *Biology and Philosophy*, *27*(6), 891–902.

Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Modern Physics*, *41*(3), 253–262.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., & Rahmstorf, S. (2008). Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences*, *105*(6), 1786–1793.

Lloyd, E. A. (2009). Varieties of support and confirmation of climate models. *Proceedings of the Aristotelian Society Supplementary Volume*, *83*(1), 213–232.

Lloyd, E. A. (2010). Confirmation and robustness of climate models. *Philosophy of Science*, *77*(5), 971–984.

Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.

Mayo, D. G. (2010). Can scientific theories be warranted with severity?. In: D. G. Mayo, & A. Spanos (Eds.), *Error and inference* (pp. 28–87). Cambridge: Cambridge University Press.

Meehl, G. A., Stocker, T. F., Collins, W. D., Friedlingstein, P., Gaye, A. P., Gregory, J. M., et al. (2007). Global climate projections. In: S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge and New York: Cambridge University Press.

North, G. R., Cahalan, R. F., & Coakley, J. A., (1981). Energy balance climate models. *Reviews of Geophysics and Space Physics*, *19*, 91–121.

Odenbaugh, J., & Alexandrova, A. (2011). Buyer beware: robustness analyses in economics and biology. *Philosophy and Biology*, *26*, 757–771.

Parker, W. S. (2009). Confirmation and adequacy-for-purpose in climate modelling. *Proceedings of the Aristotelian Society Supplementary Volume*, *83*(1), 233–249.

Parker, W. S. (2011). When climate models agree: the significance of robust model predictions. *Philosophy of Science*, *78*(4), 579–600.

Psillos, S. (1999). *How science tracks truth*. London: Routledge.

Randall, D. A., Wood, R. A., Bony, S., Coleman, R., Fichefet, T., Fyfe, J., et al. (2007). Climate models and their evaluation. In: S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, & H. L. Miller (Eds.), *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge and New York: Cambridge University Press.

Rougier, J. (2007). Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change*, *81*(3–4), 247–264.

Ruiz-Barradas, A., Nigam, S., & Kavvada, A. (2013). The Atlantic Multidecadal Oscillation in twentieth century climate simulations: uneven progress from CMIP3 to CMIP5. *Climate Dynamics*, *41*(11–12), 3301–3315.

Shackley, S. (2001). Epistemic lifestyles in climate change modelling. In: C. Miller, & P. Edwards (Eds.), *Changing the atmosphere: Expert knowledge and environmental governance* (pp. 107–133). Cambridge Massachusetts: MIT Press.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., & Miller, H. L. (Eds.). (2007). *Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge and New York: Cambridge University Press.

Stainforth, D. A., Allen, M. R., Tredger, E. R., & Smith, L. A. (2007). Confidence, uncertainty and decision-support relevance in climate predictions. *Philosophical Transactions of the Royal Society A—Mathematical Physical and Engineering Sciences*, *365*(1857), 2145–2161.

Sun, D. Z., Yu, Y., & Zhang, T. (2009). Tropical water vapour and cloud feedbacks in climate models: A further assessment using coupled simulations. *Journal of Climate*, *22*, 1287–1304.

Swanson, K. L., Sugihara, G., & Tsonis, A. A. (2009). Long-term natural variability and 20th century climate change. *Proceedings of the National Academy of Sciences*, *106*(38), 16120–16123.

Tung, K., & Zhou, J. (2013). Using data to attribute episodes of warming and cooling in instrumental records. *Proceedings of the National Academy of Sciences*, http://dx.doi.org/10.1073/pnas.1212471110 (first published online 23.01.13).

Valdes, P. (2011). Built for stability. *Nature Geoscience*, *4*, 414–416.

Wallace, J. M., Fu, Q., Smoliak, B. V., Lin, P., & Johanson, C. M. (2012). Simulated versus observed patterns of warming over the extratropical Northern Hemisphere continents during the cold season. *Proceedings of the National Academy of Sciences*, *109*(36), 14337–14342.

Weisberg, M. (2006). Robustness analysis. *Philosophy of Science*, *73*, 730–742.

Wu, Z., Huang, N., Wallace, J., Smoliak, B., & Chen, X. (2011). On the time-varying trend in global-mean surface temperature. *Climate Dynamics*, 1–15.