Allocating confirmation with derivational robustness

Aki Lehtinen

28th December 2015

Abstract

Robustness may increase the degree to which the robust result is indirectly confirmed if it is shown to depend on confirmed rather than disconfirmed assumptions. Although increasing the weight with which existing evidence indirectly confirms it in such a case, robustness may also be irrelevant for confirmation, or may even disconfirm. Whether or not it confirms depends on the available data and on what other results have already been established.

keywords: robustness, indirect confirmation, climate models, diagrams

Paper forthcoming in Philosophical Studies

1 Introduction[1]

> 'When you have eliminated the impossible, then whatever remains, however improbable, must be the truth. It may well be that several explanations remain, in which case one tries test after test until one or other of them has a convincing amount of support.' (from Sherlock Holmes, The Adventure of the Blanched Soldier, Arthur Conan Doyle, 1926).

A result is said to be derivationally robust if it can be derived from several sets of assumptions. Derivational robustness concerns derivational relationships between assumptions and results, and as such it does not consist in collecting new evidence. Orzack and Sober (1993) thus argue that the robustness of a result is irrelevant for evaluating its confirmation status. Their argument is correct insofar as only direct evidence for the robust result is taken into consideration,

[1]Acknowledgement: This research was financially supported by the Academy of Finland. Thanks for Till Grüne-Yanoff, Jaakko Kuorikoski, Chiara Lisciandra, Uskali Mäki, Caterina Marchionni, Jani Raerinne and Jacob Stegenga for their comments on various versions of this paper. I also wish to thank Jouni Räisänen for directing me to some relevant climate modelling literature, and Jonah Schupbach for the Sherlock Holmes quote. Finally I want to thank an anonymous reviewer for exceptionally useful comments on the manuscript at the final stages. The usual disclaimers apply.

but real models often have a large number of results that may confirm the robust result *indirectly*. Evidence is indirect with respect to a given result if the result does not imply the evidence but yet the evidence confirms it (see, in particular, Laudan and Leplin 1991). More recently, various contributors have argued that robustness does not confirm (Houkes & Vaesen 2012; Odenbaugh 2011; Odenbaugh & Alexandrova 2011).

The aim in this paper is to show that derivational robustness may increase the degree to which existing pieces of evidence indirectly confirm a result. Here is a rough outline of the argument. Given that models always incorporate auxiliaries that are known to be false, modellers typically modify and refine their models so as to see whether such auxiliaries are driving the results (Wimsatt 1981; Kuorikoski et al. 2010). As a result they spawn families of models with partly overlapping sets of assumptions. Individual members of a family of models typically share a set of assumptions that is sometimes called the common *core* (e.g., Levins 1993; Raerinne 2013). If a result is robust, only the assumptions that overlap between the models could be needed for its derivation, and the other assumptions are thus dispensable. Showing the derivational robustness of a result confirms it if the confirmatory power of the existing positive evidence on an initial version of the model can be allocated to the core, and the robust result is shown to depend on the confirmed core rather than the disconfirmed auxiliary assumptions. The argument is thus based on combining robustness and indirect confirmation such that the evidential boost from old evidence is shown to bear more heavily on those parts of the models that are needed for deriving the robust result: confirmatory evidence may bear more heavily on the robust result if it is shown to be derivable from the same assumptions as the robust result.

If robustness confirms, it is not some murky non-empirical kind of confirmation that it provides. Only empirical evidence can ultimately do the confirming. Nevertheless, I will use expressions such as 'the robustness of a result confirms'. In the context of this paper robustness never establishes a new link between confirming evidence and a result[2]. This means that the basic structure of indirect confirmation must always already be in existence, and robustness can only strengthen the links between the components in such a structure. The notion of confirmation is thus inevitably incremental. The expression 'the robustness of a result confirms' means that demonstrating the robustness of a result *increases* its indirect confirmation.

Making a convincing case for confirmation through increasing the weight of existing evidence is most transparent in an example in which some results are initially disconfirmed. The example from climate science I discuss below is of a model that has initially both confirmed and disconfirmed results. Showing that some particular auxiliaries were responsible for the disconfirmed result in the initial model, and that those auxiliaries are not needed to derive the robust result decreases the weight with which disconfirming evidence disconfirms the robust result. Showing the robustness of a result thus increases confirmation (of

[2] I discuss such a case in another paper (XXXX)

the robust result) if it shows that the pre-existing indirect confirmatory evidence can be allocated to the robust result, and if the indirect disconfirmatory evidence can be isolated from it and allocated to the false auxiliaries.

Demonstrating how robustness increases indirect confirmation is rather complicated because the inferences depend on which results have already been derived from the various models, and must proceed by comparing the internal structure of several models. I therefore begin Section 2 with a diagrammatic presentation of modellers' inference-making. In Section 3 I present an example from climate modelling in order to provide at least one scientific context in which robustness may confirm. The example is described in a very thin manner in order to focus on the logic of the inferences. Furthermore, I use various kinds of counterfactual scenarios in order to study the logic of robustness and indirect confirmation. The extant literature on robustness is rich in real-world examples but lacks formal representations. It is my conviction that a diagrammatic, more formal approach is justified by the fact that the controversy over whether robustness could possibly confirm has not been resolved despite a large number of recent contributions (see also Justus 2012; Kuorikoski et al. 2012; Weisberg 2013; Woodward 2006).

Strengthening the connection between the core, the robust result and a confirmed result is not sufficient for confirmatory robustness because confirmation is impossible without empirical data. If there are no empirical data, or if demonstrating the robustness of a result fails to allocate the confirmatory boost to the core and the robust result, robustness does not confirm. Whether or not robustness confirms is thus a context-specific matter. Furthermore, robustness may even disconfirm if the assumptions that were responsible for the robust result are needed for deriving the disconfirmed result, but irrelevant for the confirmed results. Section 4 is devoted to examining this context dependence. The intention is to explain why the controversy has not been resolved thus far: depending on what else the modellers already know, demonstrating the robustness of a result may strongly confirm, weakly confirm, confirm so weakly as to be practically moot, not confirm at all, or disconfirm. If different philosophers have come across different cases, it is no wonder that they have developed conflicting views about robustness.

2 Preliminaries: representing inferences in families of models with diagrams

In order to study the derivational relationships between modelling assumptions and results derivable from them, I will write $M_1 = (A_1A_2A_3A_4A_5A_6) \vdash R_1$ when the assumptions $A_1, A_2, ..., A_6$ of model M_1 jointly entail result R_1. In most of what follows, I disregard results derivable from a single assumption. This is not to deny that individual assumptions often have such consequences, nor that there may well be direct evidence for such consequences. Direct support for individual assumptions yields an additional channel of indirect confirmation that may or may not be consistent with the kind of indirect confirmation considered in this paper. I briefly discuss such cases in Section 3.3 but do not attempt to give an exhaustive account of them.

Models often have assumptions that are redundant for deriving various results. This is why there is a crucial difference between belonging to a set of assumptions that jointly entail a result, and being necessary for deriving a result. For example, suppose that model M_1 is modified such that $M_1^{'}$ employs A_7 instead of A_2 for deriving R_1: $M_1^{'} = (A_1 A_7 A_3 A_4 A_5 A_6) \vdash R_1$. Such a demonstration of the robustness of R_1 shows that A_2 cannot be necessary for deriving it. I indicate this by writing $A_2 \nvdash_c R_1$. The subscript 'c' derives from the notion of a *content part* (Gemes 1993). The point of using it is to indicate that there is a difference between taking part in an entailment and being genuinely needed for a result: $A_2 \nvdash_c R_1$ means that A_2 cannot be confirmed by R_1. One way of formally demonstrating that this is indeed the case is via the notion of a content part. For this paper, however, it is not necessary to know the details of content parts,[3] and it is sufficient to use the rough idea that if an assumption is not necessary for deriving a result, it cannot be confirmed by evidence for *that* result.

The basic structure of indirect confirmation can be represented as follows. A piece of evidence E_1 indirectly confirms result R_M (written '$E_1 c_i R_M$') if E_1 is implied by result R_1, and model M_1 entails both R_1 and R_M:

$$E_1 c_i R_M \text{ because } \begin{array}{cc} M_1 \vdash & R_1 \\ \top & \top \\ R_M & E_1 \end{array} \tag{1}$$

Typically, however, M_1 contains a large number of assumptions and, at least initially, it is not known whether or not the confirmed result R_1 depends on the assumptions that are necessary for deriving the result R_M. I have substituted a model M_1 for the general theory T in Laudan & Leplin's (1991) account. In their example, the theory of continental drift (T) implies the hypothesis that the magnetic poles have changed their location (and polarity) (R_1) as well as the hypothesis that the climate on any piece of soil has changed over time (R_M). Thus the variably aligned streaks in the lava at the bottom of the ocean (E_1) confirm the magnetic pole hypothesis (R_1), and by way of confirming the general theory T also indirectly the variable climate hypothesis (R_M). A model, however, has a lot of internal structure, and given that the components always incorporate false auxiliaries, it is unclear which ones are genuinely needed to obtain which results. It may be that model components that are needed for deriving the confirmed result R_1 are not needed for deriving R_M. If they are not, E_1 does not confirm R_M indirectly, but if they are, E_1 does indeed indirectly confirm R_M. Let us describe the uncertainty concerning which model components are needed by writing $\Diamond_\top$ underneath each assumption in M_1 to indicate that the assumption could be needed for deriving R_M.

[3]Schurz (1991) and Sprenger (2011) found alternative ways of expressing similar ideas.

$$\begin{array}{llllllll} M_1 = & (A_1 & A_2 & A_3 & A_4 & A_5 & A_6) & \vdash R_1 \\ & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \\ & R_M & R_M & R_M & R_M & R_M & R_M & \end{array}$$

A similar uncertainty concerns R_1:

$$\begin{array}{llllllll} M_1 = & (A_1 & A_2 & A_3 & A_4 & A_5 & A_6) & \vdash R_M \\ & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \Diamond_{\top} & \\ & R_1 & R_1 & R_1 & R_1 & R_1 & R_1 & \end{array}$$

Derivational robustness may increase indirect confirmation if R_1 is shown to depend on the same assumptions as R_M. As I have shown, deriving $M_1^{'} = (A_1 A_7 A_3 A_4 A_5 A_6) \vdash R_1$ implies that $A_2 \nvdash_c R_1$. Given that the confirmatory boost from E_1 on R_1 remains the same as before, and that each of the remaining assumptions were used in deriving R_1 from $M_1^{'}$, they must be more likely to be needed for deriving R_1 than before. Let us indicate this with '$+_{\top c}$', and present the epistemic situation (Achinstein 2001) as follows:

$$\begin{array}{lllllll} M_1^{'} = & (A_1 & A_7 & A_3 & A_4 & A_5 & A_6) \\ & +_{\top c} & & +_{\top c} & +_{\top c} & +_{\top c} & +_{\top c} \\ & R_1 & & R_1 & R_1 & R_1 & R_1 \end{array}$$

The diagrams thus depict epistemic situations which specify what the modellers know about the derivational relationships and the available evidence. They tell us how the modellers see the situation and are in this sense subjective. The empty space underneath A_7 indicates that it cannot be required for deriving R_1. Modellers must believe that the remaining assumptions are more likely to be needed for deriving R_1 because the diagrams depict the modellers' epistemic situation rather than what logically omniscient (Garber 1983) modellers would believe. As we will now see, deriving further results may change the modellers' epistemic situation in such a way that some of these remaining assumptions are no longer taken to be needed for deriving R_1.

Thus far there is no increase in confirmation because a confirmed result is merely derived in another model. However, if R_M is now derived from $M_1^{'}$, it becomes more indirectly confirmed because it is shown to depend on the same assumptions as the confirmed result.

$$\begin{array}{llllllll} M_1^{'} = & (A_1 & A_7 & A_3 & A_4 & A_5 & A_6) & \vdash R_1 \\ & +_{\top c} & & +_{\top c} & +_{\top c} & +_{\top c} & +_{\top c} & \\ & R_M & & R_M & R_M & R_M & R_M & \end{array}$$

In terms of indirect confirmation (1), the change can be expressed by noting that the link from R_1 to R_M has become stronger:

$$\begin{array}{cc} (A_1 A_3 A_4 A_5 A_6) + \vdash_c & R_1 \\ +_{\top c} & \top \\ R_M & E_1 \end{array} \tag{2}$$

Note that the confirmatory boost is not very strong because it has only been shown that assumptions A_2 and A_7 cannot be responsible for the two results. It continues to be perfectly possible that the indirect confirmation from E_1 to R_M is entirely removed. This happens, for example, if modellers show that $(A_1 A_3 A_4) \vdash R_M$ and $(A_5 A_6) \vdash R_1$. All the confirmatory boost from E_1 would go to A_5 and A_6, but since these assumptions would be irrelevant for deriving result R_M, R_M would obtain no indirect confirmation from E_1. Suppose now, however, that $M_1^{''}$ is used to derive R_1 and R_M.

$$\begin{array}{rcl} M_1^{''} = & (A_1 \quad A_7 \quad A_8 \quad A_9 \quad A_{10} \quad A_{11}) & \vdash R_1 \\ & \stackrel{+}{\top}{}_c & \top \\ & R_M & E_1 \end{array}$$

It is now possible to conclude that $A_3 \nvdash_c R_1$, $A_4 \nvdash_c R_1$,... and also that $A_3 \nvdash_c R_M$, $A_4 \nvdash_c R_M$,... This is where Sherlock Holmes' point about eliminative induction referred to at the beginning of this article becomes relevant: all the other assumptions have been eliminated, and only A_1 could possibly be responsible for R_1 and R_M. If none of A_2, A_3,..., A_{11} can be confirmed by E_1, but E_1 confirms each collection of assumptions in models M_1, $M_1^{'}$ and $M_1^{''}$to exactly the same degree, then it must be the case that A_1 is more strongly confirmed than before. At the same time, only A_1 could be necessary for R_M. Hence R_M is more indirectly confirmed than before.

One might argue that it is too quick to conclude that only A_1 could be responsible for R_1 and R_M: It is possible that A_2 was responsible for R_1 in M_1 and A_7 in $M_1^{''}$. There is a quick response to this. It is possible to derive e.g., $M_1^{'''} = (A_1 A_3 A_8 A_9 A_{10} A_{11}) \vdash R_1$. This would show that A_2 or A_7 could not be responsible for R_1. However, it is worth pointing out that Holmesian reasoning exploits Mill's method of agreement, and this method is known to yield false conclusions if the modellers either misidentify the alternative assumptions, or if deriving the results requires a combination of several assumptions rather than a single one.

Suppose, therefore, that there is a misidentified assumption X that implies A_3 and A_4 ($X \vdash A_3 A_4$), as well as A_7 and A_8 ($X \vdash A_7 A_8$). It would then be possible, for example, for X rather than A_1 to be responsible for R_1 (via $A_3 A_4$ in M_1 and $M_1^{'}$, and via $A_7 A_8$ in $M_1^{''}$), and for A_1 to be responsible for R_M. However, it is possible to find out assumptions like X. If the modellers would find such an assumption, they would realise that they were mistaken about which assumptions are needed to derive R_1. Furthermore, it is also possible to drop A_1 from the model to see if it still implies R_1 and R_M. I discuss the possibility that combinations of assumptions are responsible for the results later in Section 4. Let us now consider how such ideas could be applied to an example from climate-change modelling.

3 A climate example

3.1 Surface temperature and the extent of sea ice

Lloyd (2010) discusses an ensemble of 14 models in the fourth IPCC Assessment Report (AR4, 2007). I have chosen two such models that have some overlapping assumptions but give different results:

$$M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_1, R_2 \tag{3}$$

$$M_2 = (A_1 A_3 A_5 A_7 A_8) \vdash R_3, R_4 \tag{4}$$

R_1, R_2, R_3, and R_4 denote fingerprints, i.e., computer-generated results from simulations that can be compared with empirical data so as to evaluate climate models (see Parker 2010a). For example, let R_1 denote the Global Mean Surface Temperature (GMST) in the past, R_2 the extent of Arctic summer sea ice, and R_3 the height of the tropopause[4]. There is no need to give an interpretation of R_4 for the analysis that follows. Let A_2 and A_4 stand for modules used for modelling the atmosphere and the extent of sea ice, respectively. For the present purposes it is not necessary to provide interpretations of all the assumptions, but let A_1 stand for the core assumption of greenhouse gas forcing.[5][6] The rest of the assumptions can be taken to represent various auxiliaries such as the parameterizations for cloud formation or vegetation. For example, A_3 could stand for a parametrization for cloud formation in one model, and A_7 in another. Climate modellers study ensembles of models because they know that some of the auxiliaries they use are clearly false. The purpose of ensemble modelling is to see whether the same results may be derived from different sets of auxiliaries (see Parker 2010b).[7]Let R_M denote a fingerprint that indicates a rise in GMST *in the future*. By definition, there cannot be direct evidence for a future temperature. Yet, R_M is clearly the most interesting climate variable.

At this point I should perhaps acknowledge some limitations of my analysis. Real general circulation models are extremely complex. They may contain hundreds of thousands of lines of computer code. My highly idealised presentation only picks up some of the elements. I also distort a large number of their features. My excuse for making such simplifications and idealizations is that they allow me to focus on the logic of robustness, and should thus be evaluated with the purpose of the analysis in mind. The question is, would de-idealising

[4]The tropopause is the point where air ceases to cool with height, and becomes almost completely dry.

[5]Climate forcing is defined as the difference of insolation (sunlight) absorbed by the Earth and energy radiated back to space. Greenhouse gas forcing is thus a measure of the influence of such gases in altering the balance of incoming and outgoing energy in the Earth-atmosphere system.

[6]Climate modellers often refer to the 'physical core' when discussing the partial differential equations from fluid mechanics and thermodynamics. It should be clear that the 'core' in this paper does not refer to that.

[7]Lloyd (2015) argues for confirmatory robustness, using GMST as a case study. She emphasises the confirmatory value of the variety of evidence for the different individual assumptions.

this or that change the conclusions concerning how robustness affects the indirect confirmation of results? For example, I assume that R_1 derived from one model is exactly the same result as R_1 derived from another model. In reality, however, the results are expressed in real numbers, and deciding that the divergence is small enough to justify talking about the 'same result' requires judgement. Instead of one number, R_M is usually calculated for several different emission scenarios, and it corresponds to several different concepts of temperature (transient and equilibrium, for example). However, for the present purposes, there is some support for regarding these results as sufficiently similar to justify calling them robust:

> 'Models are unanimous in their prediction of substantial climate warming under greenhouse gas increases, and this warming is of a magnitude consistent with independent estimates derived from other sources, such as from observed climate changes and past climate reconstructions' (Randall et al. 2007, p. 601).

The models used in the AR4 were not able to predict the downward trend in the extent of Arctic summer sea ice: the ice was melting too slowly in the models (R_2) compared to the evidence ($\sim E_2$). The modellers thus obtained direct supporting evidence E_1 (on the GMST) for R_1 and E_3 (on the height of the tropopause) for R_3, and disconfirming evidence $\sim E_2$ (decrease in the extent of Arctic summer sea ice) and $\sim E_4$ for R_2 and R_4, respectively. Let us assume that $R_2 \vdash E_2$, and $\sim R_2 \vdash \sim E_2$ and similarly for R_4 and E_4.

$$\begin{array}{cccc} M_1 \vdash R_1, & R_2 & M_2 \vdash R_3, & R_4 \\ \top & \not\top & \top & \not\top \\ E_1 & \sim E_2 & E_3 & \sim E_4 \end{array} \tag{5}$$

Each model thus had some empirical merits and weaknesses, but it was difficult to tell which assumptions were responsible for which result, and thus how much confirmatory or disconfirmatory weight the various pieces of empirical evidence carried for the various results. It was not clear, either, exactly which assumptions were confirmed or disconfirmed, and thus whether other models that also included some of the assumptions included in M_1 and M_2 were indirectly confirmed or disconfirmed by the data.

Deriving

$$M_1 \vdash R_M \tag{6}$$

shows that

$$E_1 c_i R_M \qquad \text{because} \qquad \begin{array}{cc} A_1 A_2 A_3 A_4 A_5 A_6 \vdash & R_1 \\ \top & \top \\ R_M & E_1 \end{array} \tag{7}$$

However, (6) does not increase the indirect confirmation of R_M at all because $\sim E_2$ also indirectly disconfirms $(d_i)\, R_M$.

$$(\sim E_2) d_i R_M \qquad \text{because} \qquad \begin{array}{cc} A_1 A_2 A_3 A_4 A_5 A_6 \vdash & R_2 \\ \top & \not\top \\ R_M & \sim E_2 \end{array} \tag{8}$$

It is not yet known whether indirect confirmation of R_M by E_1 is stronger or weaker than indirect disconfirmation of R_M by E_2. As noted, the strength of indirect confirmation depends on how closely R_M and R_1 are related to a core and thereby to each other. The more likely it is that the same assumptions are mainly responsible for the confirmed results (e.g., R_1) and the robust result R_M, the more likely it is that the indirect confirmation the data (E_1) conferred on those assumptions also flows downward to the robust result R_M.

If R_M is shown to be robust by deriving

$$M_2 \vdash R_M, \tag{9}$$

when the modellers already know that $M_1 \vdash R_M$, the implication is that assumptions A_7 and A_8, as well as A_2, A_4 and A_6 are irrelevant for R_M, such that $A_7 \nvdash_c R_M$, $A_8 \nvdash_c R_M$, and so on. It was known even before establishing such a robust theorem that the two models (M_1 and M_2) shared assumptions A_1, A_3, and A_5. Given that (6) had already been established, the effect of (9) can be depicted as follows.

$$(A_1 A_3 A_5) + \vdash_c R_M. \tag{10}$$

In other words, deriving (9) indicates that $A_1 A_3 A_5$ are more likely to be needed for deriving result R_M. Unfortunately, although this inference provides information about which results depend on which assumptions, it does not increase indirect confirmation at all because, although ($A_1 A_3 A_5$) are all likely to be needed to derive R_M, this set of assumptions is no more strongly confirmed than the other assumptions in M_1 and M_2. The fact that multiple models predict R_M in itself – as critics often put it - provides no confirmation.

A quarter of the models in the ensemble employed in the fifth IPCC report (AR5, 2013) were able to predict the sea-ice extent correctly. Let M_5 stand for such a successful model ($M_5 \vdash \sim R_2$). In what follows, I will ignore the AR5 models that do not predict the sea-ice extent correctly, and only look at how the successful model M_5 would affect the confirmation of the robust prediction R_M. I will show that deriving $\sim R_2$ from M_5 is crucial for indirectly confirming R_M, but that it is confirmed only if models M_1, M_2 and M_5 all predict global warming in the future R_M, as well as correctly explain observed trends in the GMST (E_1) and the height of the tropopause (E_3), and a host of other kinds of climatic evidence. I will start by analysing the impact of these other kinds of climatic evidence. The modellers showed that

$$M_5 = (A_1 \tilde{A}_2 A_3 \tilde{A}_4 A_5 A_6) \vdash R_1. \tag{11}$$

$\tilde{A}_2$ and $\tilde{A}_4$ denote the new atmospheric and sea-ice components. The direct support from E_1 on R_1 can now also be taken to indirectly confirm R_M. $M_5 \vdash$

R_1 shows that R_1 can be derived from similar assumptions $(A_1A_3A_5)$ as the robust result R_M even if the modellers had not yet derived R_M from M_5. Given (6) and (9), this derivation alone indirectly confirms R_M. By way of clarification, it should be mentioned that some but not all assumptions in M_1 and M_5 are indirectly confirmed by the derivation because, given (3), (11) means that $(A_1A_3A_5A_6)+\vdash_c R_1$. The robustness of R_M (i.e., 6 and 9 together) has become confirmatory after all, because now it shows that all of $(A_1A_3A_5)$ are likely to be needed for deriving both R_M and R_1. Note, however, that A_6 is always involved in deriving R_1, but not in deriving R_M. This means that, in principle at least, it is possible that A_6 alone is sufficient for deriving R_1. Deriving R_1 from A_6 without using A_1,A_3 and A_5 ($A_6 \vdash R_1$) would thus remove such indirect confirmation on R_M entirely.

Had (6) and (11) but not (9) been derived it would have been clear that there was an overlap between models M_1 and M_5, and that one of the directly confirmed results (R_1) was also robust. However, (6) and (11) without (9) could not have confirmed R_M because, for all the modellers knew before deriving (9), A_2A_4 might have been needed for R_M and R_2 but not for R_1, and $(A_1A_3A_5A_6)$ might have been needed for R_1 but not for R_M and R_2.

$$M_1 = (A_1\mathbf{A_2}A_3\mathbf{A_4}A_5A_6) \vdash \mathbf{R_M}, R_1, \mathbf{R_2} \tag{3,6}$$
$$M_5 = (A_1\tilde{A}_2A_3\tilde{A}_4A_5A_6) \vdash R_1 \tag{11}$$

Had this turned out to be the case, (11) would not have confirmed R_M. This would have been the case if they had derived $A_2A_4 \vdash R_M$ and $A_2A_4 \vdash R_2$. Demonstrating the robustness of R_M by deriving it from M_2 or M_5 showed that neither A_2 nor A_4 was needed for deriving R_M, and that $A_1A_3A_5$ were more likely to be needed.

$$M_2 = (A_1A_3A_5A_7A_8) \vdash R_M \tag{9}$$

(11) now increases the indirect confirmation of R_M by E_1 by way of making it more likely that $A_1A_3A_5$ are all necessary for deriving R_1. In other words, had $M_1 \vdash R_M$ and $M_5 \vdash R_1$ already been at hand, demonstrating the robustness of R_M by deriving $M_2 \vdash R_M$ or

$$M_5 = (A_1\tilde{A}_2A_3\tilde{A}_4A_5A_6) \vdash R_M \tag{12}$$

would have confirmed R_M because such demonstrations make it more likely that the same set of assumptions is required for the robust result R_M and the confirmed result R_1. Inference (12) together with (9) means that the better atmospheric ($\tilde{A}_2$) and sea-ice ($\tilde{A}_4$) modules were not really needed for deriving the prediction of future warming. The structure of indirect confirmation is given by (13):

$$E_1 c_i R_M \qquad \text{because} \qquad \begin{array}{cc} A_1A_3A_5+\vdash_c & R_1 \\ +\top c & \top \\ R_M & E_1 \end{array} \tag{13}$$

(13) shows that a confirmed result R_1 is closely related to the assumptions that are responsible for the robust result R_M, but although R_2 was never derived from M_5, (13) does not rule out the possibility that $A_1A_3A_5$ might also be necessary for its derivation. Thus, $M_5 \vdash R_2$ continued to be possible, and such a derivation would indeed have ruined the indirect confirmatory benefits that the modellers thought the previous derivations conferred on R_M. The epistemic situation would have been as follows.

$$E_2 d_i R_M \quad \text{because} \quad \begin{array}{ccc} A_1A_3A_5 + \vdash_c & R_1 & R_2 \\ +\top_c & \top & \not\top \\ R_M & E_1 & \sim E_2 \end{array} \tag{14}$$

E_2 would thus have indirectly disconfirmed R_M. The only difference from the original situation would have been that (11) would have shown that R_1 cannot confirm A_2 or A_4.

However, as noted earlier, a quarter of the models included in the 5th IPCC report correctly predict the ice coverage, and the modellers were thus able to isolate the robust result R_M from the indirect disconfirmation of R_2 by deriving

$$M_5 = (A_1\tilde{A}_2A_3\tilde{A}_4A_5A_6) \vdash\sim R_2 \tag{15}$$

Such results mean that the assumptions required for deriving the disconfirmed result R_2 (at least A_2 and A_4) are different from those required for deriving the robust result R_M and the confirmed result R_1. The modellers were now more certain that E_2 did not compromise the indirect confirmation of R_M by E_1.

$$\begin{array}{ccc} A_1A_3A_5 + \vdash_c & R_1 & \nvdash R_2 \\ +\top_c & \top & \not\top \\ R_M & E_1 & \sim E_2 \end{array} \tag{16}$$

However, they were not able to pinpoint a change in a single assumption that would account for the improved predictions: they attributed the change both to atmospheric ($\tilde{A}_2$) and sea-ice components ($\tilde{A}_4$) in the models (see Sect 9.4.3. in Flato and Marotzke 2013).

One could argue that deriving the correct result $M_5 \vdash\sim R_2$ is what really makes the difference here rather than the robustness of R_M or R_1. This is not quite correct, however. Had the modellers derived $M_5 \vdash\sim R_2$ when (3)-(6) were available but (9)-(11) were not, they would have learned that $A_1A_3A_5A_6$ could not be necessary for deriving result R_2 because they would have been used in deriving both R_2 and $\sim R_2$. Even if result R_M had been derived from M_1, it would only have been weakly indirectly confirmed by E_1 because $A_1A_3A_5A_6$ might have been needed for deriving R_1 but not for deriving R_M:

$$M_5 = (A_1\tilde{A}_2A_3\tilde{A}_4A_5A_6) \vdash\sim R_2 \tag{15}$$
$$M_1 = (A_1\mathbf{A_2}A_3\mathbf{A_4}A_5A_6) \vdash R_1, \mathbf{R_2} \tag{3}$$
$$M_1 = (A_1\mathbf{A_2}A_3\mathbf{A_4}A_5A_6) \vdash \mathbf{R_M} \tag{6}$$

In this epistemic situation it is possible that R_M depends mostly on A_2 and A_4, and that the disconfirmed result R_2 depends on these same assumptions. In other words, without the robustness of R_M, deriving the correct result $M_5 \vdash \sim R_2$ is not alone sufficient for confirming R_M. Showing the robustness of R_M by deriving (9)

$$M_2 = (A_1 A_3 A_5 A_7 A_8) \vdash R_M \tag{9}$$

revealed that A_2 and A_4 cannot be needed for R_M. Now, (15) together with (9), show that almost all the assumptions used for deriving $\sim R_2$ $(A_1 A_3 A_5 A_6)$ are also needed for R_M $(A_1 A_3 A_5)$. The robustness of R_M is thus one crucial element in establishing that the robust result R_M is unaffected by disconfirmatory results obtained for the AR4 models, and that it does have strong indirect confirmatory support from E_1 through R_1. It is, of course, true that $\sim R_2$ and the evidence for it $\sim E_2$ now also indirectly confirm R_M:

$$\sim E_2 c_i R_M \text{ because } \begin{array}{lcl} A_1 A_3 A_5 & + \vdash_c & \sim R_2 \\ +\top c & & \top \\ R_M & & \sim E_2 \end{array} \tag{17}$$

However, without the robustness of R_M this inference cannot be made. Note, finally, that deriving (12)

$$M_5 = (A_1 \tilde{A}_2 A_3 \tilde{A}_4 A_5 A_6) \vdash R_M \tag{12}$$

after having derived

$$M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_M \tag{6}$$

shows that deriving R_M from climate models does not depend on de-idealising them by finding better auxiliaries $(\tilde{A}_2)$ for modelling sea ice. Yet, this de-idealisation is not irrelevant for the truth of R_M because it increases its indirect confirmation. In conclusion, showing the robustness of the result R_M confirms it if its derivation is shown to depend on confirmed rather than disconfirmed assumptions.

3.2 Duhem-Quine problems in climate modelling

Given Lenhard and Winsberg's (2010) argument that the Duhem-Quine problem is particularly vicious in climate modelling, I am obliged to discuss their argument here. They point out that improvements in climate models are usually attained by modifying various collections of assumptions they call modules. The fact that the various modules interact with each other implies that the net effect of a module is testable only by the overall outcomes of the entire general circulation model. 'And so when some new elements are added to a model, and improve model performance, it is often impossible to know if this happens because what has been added has goodness-of-fit on its own, or merely because, in combination with the rest of the model, what is achieved on balance is an improvement' (p. 257). Their argument is a variant of the possibility I mentioned earlier: it is possible that deriving a result requires the combination of several assumptions. Let us incorporate their argument into the framework. It cannot be concluded from a comparison of (3) and (15),

$$M_1 = (A_1A_2A_3A_4A_5A_6) \vdash R_2 \quad (3)$$
$$M_5 = (A_1\tilde{A}_2A_3\tilde{A}_4A_5A_6) \vdash\sim R_2 \quad (15)$$

that $\tilde{A}_2$ and $\tilde{A}_4$ in themselves have a better goodness-of-fit because it is possible that $\tilde{A}_2$ and $\tilde{A}_4$ generate the better results ($\sim R_2$) only together with A_1, A_3, A_5 or A_6 or some combination of them.[8] However, this is a problem that can be solved, at least in principle. If modellers could derive

$$\begin{aligned} M_{51} &= (\tilde{A}_1\tilde{A}_2A_3\tilde{A}_4A_5A_6) \vdash\sim R_2 \\ M_{52} &= (\tilde{A}_1\tilde{A}_2\tilde{A}_3\tilde{A}_4A_5A_6) \vdash\sim R_2 \\ M_{53} &= (A_1\tilde{A}_2\tilde{A}_3\tilde{A}_4A_5A_6) \vdash\sim R_2 \\ M_{54} &= (\tilde{A}_1\tilde{A}_2\tilde{A}_3\tilde{A}_4\tilde{A}_5A_6) \vdash\sim R_2, \end{aligned}$$

and so on, in other words, if they exhaustively tried each possible combination of assumptions, and if the same result$\sim R_2$ continued to emerge from each model, then they could be fairly certain that $\tilde{A}_2$ and $\tilde{A}_4$ were responsible for $\sim R_2$. If there is reason to suspect that there are important mutual dependencies between the various assumptions, every possible combination must be tried before it could be said that $\tilde{A}_2$ and $\tilde{A}_4$ are responsible for $\sim R_2$. For example, it is possible, even after deriving the aforementioned results, to derive

$$M_{55} = (\tilde{A}_1\tilde{A}_2\tilde{A}_3\tilde{A}_4\tilde{A}_5\tilde{A}_6) \vdash R_2.$$

Such a result would mean that $\tilde{A}_2$ and $\tilde{A}_4$ generate $\sim R_2$ as long as A_6 rather than $\tilde{A}_6$ is used, or perhaps $\tilde{A}_2$ and $\tilde{A}_4$ generate $\sim R_2$ as long as they are not combined with $\tilde{A}_1$ and $\tilde{A}_6$ or,... This is why every possible combination of assumptions must be tried. Although such exhaustive testing of auxiliaries may sound like Odenbaugh and Alexandrova's (2011) requirement of 'absolute robustness', the motivation in this case is somewhat different: Odenbaugh and Alexandrova require that the 'true' auxiliary be among the ones that the modellers try, but there is no similar requirement here.

However, given that real climate models include thousands rather than dozens of assumptions, the number of simulations that would be needed to span *every* combination of assumptions is huge. In principle, then, the performance of climate models that is evaluated by means of inferences, such as $M_5\vdash\sim R_2$, is sufficient for allocating at least some confirmation, and robustness is needed for such inferences. In practice, however, the complexity of climate models

[8]Lenhard and Winsberg recognise the 'lucky' possibility of testing the auxiliaries in isolation. If climate modellers are able to derive $(A_2A_4)\vdash R_2$ and $(\tilde{A}_2\tilde{A}_4) \vdash\sim R_2$, they could conclude that $\tilde{A}_2$ and $\tilde{A}_4$ have a better goodness-of-fit than A_2 and A_4.

implies that the modellers usually have to settle for considerably less than an exhaustive examination of the various possible combinations of assumptions. If the Duhem-Quine problem is particularly vicious in climate modelling, it is because of the sheer number of assumptions that have to be made and tested rather than because it is impossible in principle to allocate confirmation on individual assumptions.

3.3 Direct evidence for the assumptions

What if there is direct evidence for some of the assumptions? Consider, for example, a situation in which a piece of evidence, $E_{\tilde{A}2}$, directly confirms the new sea ice module $\tilde{A}_2$ ($\tilde{A}_2 \vdash E_{\tilde{A}2}$) (but not the old one A_2: $A_2 \nvdash_c E_{\tilde{A}2}$). Such evidence would, of course, bolster the confirmation of the new sea-ice module, but this confirmation could not flow to the robust result R_M because it is known that it ($\tilde{A}_2$) was not necessary for deriving it ($\tilde{A}_2 \nvdash_c R_M$). However, consider a scenario in which there is direct confirming evidence E_{A1} for A_1: $A_1 \vdash E_{A1}$. The structure of indirect confirmation is now simpler than before. Demonstrating the robustness of R_M by deriving, for example, (12)

$$M_5 = (A_1 \tilde{A}_2 A_3 \tilde{A}_4 A_5 A_6) \vdash R_M$$

after having derived

$$\begin{array}{lll} M_1 = & (A_1 & A_2 A_3 A_4 A_5 A_6) \vdash R_M \\ & \top & \\ & E_{A1} & \end{array}$$

shows that

$$E_{A1} c_i R_M \text{ because } \begin{array}{l} A_1 A_3 A_5 \ \vdash E_{A1} \\ +_{\top c} \\ R_M \end{array} .$$

Given that A_1 alone already entails E_{A1}, the link between them cannot become stronger due to robustness, but the link between A_1 and R_M can, and direct evidence for the assumptions also indirectly confirms the robust result R_M. Note, however, that if the direct evidence $\sim E_{A1}$ were to disconfirm A_1, the disconfirmation would also be strengthened by robustness ($\sim E_{A1} d_i R_M$). If A_1 simultaneously has indirectly confirming evidence as in (18),

$$\begin{array}{lllll} M_1 & = & (A_1 & A_2 A_3 A_4 A_5 A_6) \vdash & R_1 \\ \top & & \not\top & & \top \\ R_M & & \sim E_{A1} & & E_1 \end{array} \quad (18)$$

the modellers would have confirming indirect evidence (E_1) and disconfirming direct evidence ($\sim E_{A1}$) for this assumption. Whether such conflicts could be resolved by deriving further robustness results is an interesting question, but I will not address it here in order not to introduce further complexities.

4 The context dependence of confirmation via robustness

As shown in the previous section, demonstrating the robustness of R_M by deriving (9) confirmed it indirectly only when the modellers had also derived (11). In this section I consider several different orders in which the various results could be derived. The purpose of such counterfactual exercises is to show that demonstrations of robustness may confirm weakly or strongly, not confirm at all, or even disconfirm, depending on what else is known about the derivational relationships and the data. Each of the subsections that follow considers a different counterfactual order of deriving the results.

4.1 The robustness of a directly empirically confirmed result may confirm another result

If the modellers have (9) and (3) but not (6) at hand, does (11) confirm R_M? Perhaps it does, but very weakly (cf. Parker 2011). The following results would then be available:

$$M_1 = (A_1A_2A_3A_4A_5A_6) \vdash R_1, R_2 \tag{3}$$
$$M_5 = (A_1\tilde{A}_2A_3\tilde{A}_4A_5A_6) \vdash R_1 \tag{11}$$
$$M_2 = (A_1A_3A_5A_7A_8) \vdash R_M \tag{9}$$

The robustness of R_1 (i.e. deriving 11 when 3 is available) shrinks the set of assumptions that could be necessary for deriving R_1 from $A_1A_2A_3A_4A_5A_6$ to $A_1A_3A_5A_6$. If these same assumptions are responsible for R_M, confirmation may flow downwards to R_M. However, it is not known before the derivation of result (6): $M_1 = (A_1A_2A_3A_4A_5A_6) \vdash R_M$ that it is the confirmed assumptions that are really needed for result R_M: it is possible that R_M depends mainly on A_7 and A_8, and that R_1 depends on $A_1A_3A_5A_6$.

If (6) is indeed available, part of the indirect confirmation of R_M depends on the robustness of R_1. This is an example of a more general possibility deriving from the fact that robustness may confirm indirectly: if a result (here R_1) has direct confirming evidence (here E_1), although its robustness cannot confirm itself (as Orzack and Sober 1993 argue), it may confirm other results (here R_M). The reason for this is that the robustness of R_1 allows the modeller to shrink the set of assumptions that could be needed for deriving it: deriving (11) when (9), (3) and (6) are available increases the indirect confirmation of R_M by showing that A_2 and A_4 cannot be responsible for the confirmed result R_1 or R_M. However, if (6) is not available, the modellers can only conclude that A_2 and A_4 cannot be responsible for the confirmed result R_1. The degree to which demonstrating the robustness of a confirmed result (R_1) confirms R_M thus depends on whether or not R_M itself is robust.

4.2 There can be indirect confirmation without robustness, but it is fragile

Suppose now that the modellers' background knowledge is radically less extensive than before. Results (3-8) are not available, and the modellers start by deriving (9).

$$M_2 = (A_1 A_3 A_5 A_7 A_8) \vdash R_M \tag{9}$$

They then derive (11) and also show that R_1 is directly supported by E_1:

$$\begin{array}{ll} M_5 = (A_1 \tilde{A}_2 A_3 \tilde{A}_4 A_5 A_6) & \vdash R_1 \\ & \quad \top \\ & \quad E_1 \end{array} \tag{19}$$

Then, given the similarity in structure between models M_2 and M_5, (19) shows that R_M is indirectly confirmed by E_1 insofar as $A_1 A_3 A_5$ are needed for deriving R_M. Deriving R_1 from M_5 (i.e., 11) may indirectly confirm R_M because it shows that new confirming evidence E_1 is relevant for assumptions $A_1 A_3 A_5$, and these same assumptions are used for deriving R_M.

The epistemic situation can now be depicted as follows.

$$\begin{array}{llllll} M_2 = (A_1 & A_3 & A_5 & A_7 & A_8) & \vdash R_M \\ \quad\ \Diamond_\top & \Diamond_\top & \Diamond_\top & \Diamond_\top & \Diamond_\top & \\ \quad\ R_M & R_M & R_M & R_M & R_M & \end{array} \tag{9}$$

$$\begin{array}{llllllll} M_5 = (A_1 & \tilde{A}_2 & A_3 & \tilde{A}_4 & A_5 & A_6) & \vdash R_1 \\ \quad\ \Diamond_\top & \Diamond_\top & \Diamond_\top & \Diamond_\top & \Diamond_\top & \Diamond_\top & \quad \top \\ \quad\ R_1 & R_1 & R_1 & R_1 & R_1 & R_1 & \quad E_1 \end{array} \tag{19}$$

Note that this indirect confirmation (i.e., $E_1 c_i R_M$) does not require the robustness of any result (recall the assumption that (3), and (6) are not available).[9]

[9] An anonymous reviewer suggested that there is no indirect confirmation here without robustness on the grounds that merely sharing some elements is not sufficient. I see some merit in this interpretation, but I am not quite able to present it as my own. It is true that without robustness, we do not know which assumptions are responsible for which results. However, even though it may be so extremely weak as to be unrecognisable for the modellers, there is some indirect confirmation from E_1 on R_M merely due to the fact that they are, in principle, able to grasp the structure of indirect confirmation because the links from E_1 to R_M via A_1, A_3 and A_5 exist. I agree that whether or not there is confirmation without robustness is to be decided on the basis of the modellers' epistemic situation, i.e., what they know about evidence and derivational relationships. If they cannot even grasp that there could be indirect confirmation on R_M in the initial situation without robustness, then there would indeed be no confirmation before robustness. The reviewer suggests considering a case in which A_1, A_3 and A_5 are just mathematical simplifications. If they are, then the analysis below shows

One possible way of representing this is as follows:

$$E_1 c_i R_M \text{ because } \begin{array}{lccccccc} M_5 = & (A_1 & \tilde{A}_2 & A_3 & \tilde{A}_4 & A_5 & A_6) \vdash & R_1 \\ & \Diamond\top & & \Diamond\top & & \Diamond\top & & \top \\ & R_M & & R_M & & R_M & & E_1 \end{array} \tag{19}$$

However, it is fragile in the sense that there are several ways in which it could be removed. For example, A_6 or $\tilde{A}_2$ or $\tilde{A}_4$ rather than $A_1 A_3 A_5$ could be needed for deriving R_1, and if they are, R_M is not indirectly confirmed by E_1. Similarly, A_7 and A_8 could be needed for deriving R_M, and if they are, E_1 does not indirectly confirm R_M at all. Let us now see how the robustness of either result may rule out such possibilities. Deriving (3)

$$M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_1, R_2 \tag{3}$$

in this epistemic situation implies that $\tilde{A}_2$ and $\tilde{A}_4$ are not relevant for R_1: $\tilde{A}_2 \nvdash_c R_1, \tilde{A}_4 \nvdash_c R_1$, and we obtain the 'downward' inference

$$\begin{array}{lccccccc} M_5 = & (A_1 & \tilde{A}_2 & A_3 & \tilde{A}_4 & A_5 & A_6) \vdash & R_1 \\ & +\top_c & & +\top_c & & +\top_c & +\top_c & \top \\ & R_1 & & R_1 & & R_1 & R_1 & E_1 \end{array} \tag{19}$$

Deriving (6)

$$M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_M \tag{6}$$

implies $A_7 \nvdash_c R_M, A_8 \nvdash_c R_M$, and the downward inference

$$\begin{array}{lcccccc} M_2 = & (A_1 & A_3 & A_5 & A_7 & A_8) \vdash & R_M \\ & +\top_c & +\top_c & +\top_c & & & \\ & R_M & R_M & R_M & & & \end{array} \tag{9}$$

These derivations indicate that what might be really necessary for R_M is $A_1 A_3 A_5$, and for R_1 this set could be $A_1 A_3 A_5 A_6$. Further results such as

$$M_6 = (A_1 A_3 A_5 A_{10} A_{11}) \vdash R_1 \tag{20}$$

could indirectly confirm R_M even further. (20) implies $A_6 \nvdash_c R_1$ and

$$\begin{array}{lccccccc} M_5 = & (A_1 & \tilde{A}_2 & A_3 & \tilde{A}_4 & A_5 & A_6) \vdash & R_1 \\ & +\top_c & & +\top_c & & +\top_c & & \top \\ & R_1 & & R_1 & & R_1 & & E_1 \end{array} \tag{19}$$

---

that the simplifications are indirectly confirmed due to robustness. If there were independent disconfirming evidence for the simplifications, one might have to conclude that robustness leads one astray. In real life, however, modellers know that A_1 represents greenhouse gas forcing, for example. In such circumstances, it would seem odd to deny that there is any indirect confirmation on R_M already in the initial situation.

Note, however, that without (20), $M_6 \vdash R_M$ would be very weakly confirmatory because it would not rule out the possibility that R_1 mainly depends on A_6, which has been shown to be irrelevant for R_M. To summarise the lessons from this order of deriving the results, the point is that there may be indirect confirmation of results without robust results, but robustness strengthens it by ruling out some of the possible ways in which it could be removed.

4.3 Only empirical evidence confirms

Let us look at yet another order. If the modellers start with (11), (3) and (9) rather than (19), and then derive the aforementioned robustness results (i.e., 20, 12 and 6), results R_M and R_1 are not yet confirmed because as long as (5) or (19) are not available, there is no evidence to confirm any model, only results R_M, R_1 and R_2. This observation constitutes a second sense in which the fact that multiple models predict R_M in itself – as critics often put it - provides no confirmation.

$$M_5 = (A_1 \tilde{A}_2 A_3 \tilde{A}_4 A_5 A_6) \vdash R_1 \quad (11)$$
$$M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_1, R_2 \quad (3)$$
$$M_6 = (A_1 A_3 A_5 A_{10} A_{11}) \vdash R_1 \quad (20)$$
$$M_5 = (A_1 \tilde{A}_2 A_3 \tilde{A}_4 A_5 A_6) \vdash R_M \quad (12)$$
$$M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_M \quad (6)$$
$$M_2 = (A_1 A_3 A_5 A_7 A_8) \vdash R_M \quad (9)$$

I agree with the critics here. As noted earlier, even though modellers are justified in thinking that the robustness of the result R_M makes the robust theorem (ceteris paribus, $A_1 \vdash R_M$) more firmly established insofar as A_1 is indeed involved in all the derivations, it would be inappropriate to say that the theorem is more strongly *confirmed* than before because more is known about the derivational relationships. Although strengthening the theorem justifiably increases the modellers' *confidence* in it, only empirical evidence can ultimately do the confirming. However, this is not to say that the modellers are not justified in having increased confidence in the theorem if it is shown to be robust. Furthermore, the robustness of the result may become confirmatory rather than merely confidence-increasing as soon as there is some evidence that bears on it. Let us now see how this happens.

Deriving (19) brings evidence into the family. The previously established robustness results would now show that this evidence indirectly confirms the core assumptions $A_1 A_3 A_5$ and thereby the robust result R_M. (11) and (3) now show that E_1 indirectly confirms assumptions A_1, A_3 and A_5 rather than A_2, A_4. Given (3), (19) implies that $\tilde{A}_2$ and $\tilde{A}_4$ could not have been responsible for R_1 ($\tilde{A}_2 \nvdash_c R_1, \tilde{A}_4 \nvdash_c R_1$), and that A_1, A_3, A_5 and A_6 are more likely to be responsible for it

$$\begin{array}{lcccccccl} M_5 = & (A_1 & \tilde{A}_2 & A_3 & \tilde{A}_4 & A_5 & A_6) & \vdash R_1 & \\ & +\!\top_c & & +\!\top_c & & +\!\top_c & +\!\top_c & \top & \quad (19) \\ & R_1 & & R_1 & & R_1 & R_1 & E_1 & \end{array}$$

(20) implies $A_6 \nvdash_c R_1$ so that (19) becomes

$$
\begin{array}{rllllll}
M_5 = & (A_1 & \tilde{A}_2 & A_3 & \tilde{A}_4 & A_5 & A_6) \vdash R_1 \\
 & \not\top_c & & \not\top_c & & \not\top_c & \top \\
 & R_1 & & R_1 & & R_1 & E_1
\end{array}
\tag{19}
$$

Given (6), $\tilde{A}_2 \nvdash_c R_M, \tilde{A}_4 \nvdash_c R_M$, and (12) can be written as follows:

$$
\begin{array}{rllllll}
M_5 = & (A_1 & \tilde{A}_2 & A_3 & \tilde{A}_4 & A_5 & A_6) \vdash R_M \\
 & \not\top_c & & \not\top_c & & \not\top_c & \not\top_c \\
 & R_M & & R_M & & R_M & R_M
\end{array}
\tag{12}
$$

Finally, given (12), (9) can now be written as follows:

$$
\begin{array}{rlllll}
M_2 = & (A_1 & A_3 & A_5 & A_7 & A_8) \vdash R_M \\
 & \not\top_c & \not\top_c & \not\top_c & & \\
 & R_M & R_M & R_M & &
\end{array}
\tag{9}
$$

(19) and (9) now show that R_M depends on the same set of assumptions as R_1, and the earlier demonstrations of robustness have become confirmatory. It should now be clear why there has been so much confusion about whether or not robustness confirms: the proponents may have thought that showing that the core is more likely to be responsible for the robust result means that the robust theorem is also more likely to be true: this is what the justifiably increased confidence in the theorem might lead us to believe if we are not too picky about linking 'confirmation' to empirical evidence. Critics such as Orzack and Sober (1993), Forber (2010) and Calcott (2011), on the other hand, reserve this honorific term for empirical confirmation. As I have shown, if there is the right kind of indirect evidence, the increased knowledge of derivational relationships (i.e., the robustness of the result) increases the confirmation of the result but not of the theorem. Robustness may thus provide empirical (but indirect) confirmation. Nevertheless, the confirmation increases if and only if the theorem is more firmly established (or 'more likely to be true') in virtue of the derivational relationships: they show that the robust result depends on the core rather than the auxiliaries. In other words, although strengthening the robust theorem does not yet *mean* that the robust result is confirmed, it is a necessary condition for it to be indirectly confirmed due to robustness if there is only old evidence. As I have shown, strengthening the robust theorem is not a sufficient condition for increasing indirect confirmation of the robust result: one also has to be able to show that the core, rather than some auxiliaries, is responsible for the confirmed results.

Whereas the semantics of the the term 'confirmation' may not be particularly important, there is another group of critics who deny that robustness could even justifiably increase confidence (in the theorem or the result). Note that what has been indirectly confirmed is not the core alone but rather the core supplemented with auxiliaries A_3 and A_5. Recall that the core consists of assumption A_1.

Suppose now that A_3 and A_5 are known to be false. Joel Katzav (2013; 2014) argues that the empirical success of climate models cannot confirm the truth of a conjunction of assumptions if that conjunction contains elements that are known to be false.

I have just shown that $A_1A_3A_5$ have been confirmed relative to what the modellers knew at the time of the fourth IPCC report. This does not mean, however, that the conjunction is true in some absolute sense. Indeed, the probability that a conjunction containing idealizations is true is very close to zero, and because the conjunction of assumptions in General Circulation Models will always contain a large number of idealizations, it will always remain very close to zero. I agree with Katzav that the components that can be confirmed are those about which there is genuine epistemic uncertainty. Thus the Navier-Stokes equations, for example, are not confirmed by the success of climate models. The most important uncertainty concerns whether the results depend on the various components that are known to be false.

Let us disregard the complexity of climate modelling for the time being. As a matter of logic, we could end up in a situation in which A_1 (greenhouse gas forcing) is the only model component that is involved in all derivations of R_M. For example, model M_3 could be developed as follows:

$$M_3 = (A_1A_{12}A_{13}A_{14}A_{15}) \vdash R_M, R_1 \tag{21}$$

Such results would mean that the robust theorem (ceteris paribus, $A_1 \vdash R_M$) would be well established because, given results like (12) this could be written as follows:

$$\begin{array}{rlllllll} M_3 = & (A_1 & A_{12} & A_{13} & A_{14} & A_{15} & A_{16}) & \vdash R_1 \\ & +\top_c & & & & & & \;\;\top \\ & R_M & & & & & & \;\;E_1 \end{array} \tag{22}$$

In reality, however, it is impossible to obtain results like (21) because climate models contain a huge number of assumptions, and some of them cannot be left out of the model. For example, radiation from the sun S, volcanoes V, changes in vegetation G, and so on, must be modelled in one way or another. Yet, as Hegerl et al. (2007) point out, it has been amply demonstrated that it is impossible to generate R_1 without A_1. Thus there are plenty of results of the form:

$$M_4 = (SVGA_2A_{13}A_4A_6) \nvdash R_1 \tag{23}$$

$$M_{41} = (A_1SVGA_2A_{13}A_4A_6) \vdash R_1 \tag{24}$$

Taken together, such results mean that R_M is confirmed because the logical links between the relevant items in the indirect confirmation diagram have been strengthened:

$$\begin{array}{lll} (A_1 \; . \; . \; . \; . \;) & + \vdash_c & R_1 \\ +\top_c & & \top \\ R_M & & E_1 \end{array} \tag{25}$$

R_M is thus indirectly more confirmed than before because the robustness of R_M and R_1 strengthen the robust theorem (ceteris paribus, $A_1 + \vdash_c R_M$) and the link between CO2 forcing and the evidence on past temperatures ($A_1 + \vdash_c R_1 \vdash E_1$). Odenbaugh and Alexandrova (2011) argue that robustness only confirms if we can exhaustively list all the possible auxiliaries, try them all, and be assured that the true auxiliary is among the ones we have tried. Katzav's argument is a variant of Odenbaugh and Alexandrova's 'absolute robustness' argument: it is based on the idea that something cannot be confirmed at all as long as it is known that that something may continue to contain some falsity. This argument is correct for absolute but not for relative (or 'incremental') confirmation because, albeit demonstrating the irrelevance of A_2 and A_4 does not demonstrate the truth of R_M, it does *increase* the confirmation of R_M. I agree that the robustness of R_M does not mean that it is true. Perhaps it is not even very likely, but the point of demonstrating the robustness of results is not to claim that the true auxiliaries have been found but rather to show that some assumptions that are known to be false do not affect the conclusions. As I have shown, it is not necessary to establish the truth of $\tilde{A}_2$ and $\tilde{A}_4$ to confirm R_M.

4.4 Indirect disconfirmation

Let us now consider some counterfactual scenarios that would have defeated the indirect confirmation of R_M, or even disconfirmed it. Suppose now that all the previously mentioned results are available except (20) and (21). Let us repeat them here in a concise manner.

$$\begin{array}{l} M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_1, R_2, R_M \\ M_2 = (A_1 A_3 A_5 A_7 A_8) \vdash R_3, R_4, R_M \\ M_5 = (A_1 \tilde{A}_2 A_3 \tilde{A}_4 A_5 A_6) \vdash R_1, \sim R_2, R_M \\ R_1 \vdash E_1, \sim R_2 \vdash \sim E_2, R_3 \vdash E_3, R_4 \vdash E_4 \end{array}$$

Suppose, then, that the modellers derived further robustness results:

$$M_7 = (A_1 A_3 A_{12} A_{13}) \vdash R_M, \sim R_1. \qquad (26)$$

Although such results have not been actually derived, it would be logically possible to do so, and it would certainly be a nightmare if the modellers thought that the robust result R_M was empirically well established on the basis of earlier results. This derivation would show that the robust result R_M depended on the core, and that although result R_1 could not be derived from A_6 alone, it might be a necessary factor. The robust result R_M would no longer be indirectly confirmed by E_1 because it would have been shown to be independent of A_6, and A_6 would be mainly responsible for R_1. Wimsatt (1980) demonstrated such fallibility in robustness a long time ago by way of a case study.

Things could be even worse, however. Suppose now that (3) - (6) and (9) were established:

$$M_1 = (A_1 A_2 A_3 A_4 A_5 A_6) \vdash R_1, R_2, R_M \qquad (3, 6)$$

$$M_2 = (A_1A_3A_5A_7A_8) \vdash R_3, R_4, R_M \tag{4,9}$$

$$\begin{array}{cccc} M_1 \vdash R_1, & R_2 & M_2 \vdash R_3, & R_4, \\ \top & \not\top & \top & \not\top \\ E_1 & \sim E_2 & E_3 & \sim E_4 \end{array} \tag{3}$$

but instead of (11-20), the climate modellers obtained different robustness results:

$$M_8 = (A_1A_3A_7A_8A_9) \vdash R_M, R_2 \tag{27}$$

Robustness would now disconfirm R_M :

$$\sim E_2 d_i R_M \text{ because } M_8 = \begin{array}{ccccccc} (A_1 & A_3 & A_5 & A_6 & A_9 &) \vdash & R_2 \\ \top & \top & & & & & \not\top \\ R_M & R_M & & & & & \sim E_2 \end{array}$$

Note that R_1 is not derived from M_8. Because assumptions A_1 and A_3, which seem to be primarily involved in deriving the robust result R_M, would also be responsible for R_2, the robust result R_M would also be affected by the disconfirming evidence $\sim E_2$, which is, from the perspective of R_M, indirect. Furthermore, the direct support from E_1 to R_1 would no longer indirectly confirm the robust result R_M because the assumptions that would have been needed to derive it (A_2, A_4 or A_6, or some combination of them) would no longer be needed. Deriving the robust result R_M from M_8 rather than M_5 would show that it depended on disconfirmed rather than confirmed assumptions. The demonstration of robustness may thus also remove indirect support from a result, as in (26), or even disconfirm it, as in (27). In general, there does not seem to be any reason to think that the relevance of robustness would be different for confirming and for disconfirming results. The point of discussing such a counterfactual scenario is to show that with exactly the same available data, the derivational relationships would imply a completely different assessment of the overall performance of climate models.

5 Conclusions

Confirmatory robustness requires a complex set of conditions that involve showing that the core is needed for deriving the confirmed and the robust result, and that the core itself rather than some alternative set of assumptions is needed for both. Whether or not robustness is indirectly confirmatory depends on the available evidence and on what other derivational relationships have already been established. This means that the same demonstration of robustness may confirm, be irrelevant for confirmation, or even disconfirm the robust result, depending on what other results have been established.

Given that the confirmatory benefits from the increased weight of the confirmed consequences depend on the context, what is confirmatory for someone may not be confirmatory for someone else who does not share the relevant knowledge of the derivational relationships in a family of models. This context dependence might also explain the widespread disagreement among philosophers about the confirmatory virtues of robustness. Even though derivational robustness may confirm in the right circumstances, non-confirmatory demonstrations of robustness could be common in science.

References

[1] Achinstein, P. (2001) *The Book of Evidence*. New York; Oxford: Oxford University Press.

[2] Calcott, B. (2011). Wimsatt and the robustness family: Review of Wimsatt's Re-engineering Philosophy for Limited Beings. *Biology and Philosophy*, 26, 281-293.

[3] Flato, G. & Marotzke, J. (2013). Evaluation of climate models. In J. Bartholy, R. Vautard and T. Yasunari (Eds.), *Climate Change 2013: The Physical Science Basis*. Working Group I Contribution to the IPCC Fifth Assessment Report (AR5) Cambridge: Cambridge University Press. pp. 1-205.

[4] Forber, P. (2010). Confirmation and explaining how possible. *Studies in History and Philosophy of Science* Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 41, 32-40.

[5] Garber, D. (1983). Old evidence and logical omniscience in bayesian confirmation theory. In J. Earman (ed.) *Testing Scientific Theories Minnesota*: Minnesota University Press.

[6] Gemes, K. (1993). Hypothetico-deductivism, content, and the natural axiomatization of theories. *Philosophy of Science*, 60, 477-487.

[7] Hegerl, G. C., et al. (2007). Understanding and Attributing Climate Change. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and 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, 663-745.

[8] Houkes, V. & Vaesen, K. (2012). Robust! - Handle with care. *Philosophy of Science*, 79, 345-364.

[9] Justus, J. (2012). The Elusive Basis of Inferential Robustness. *Philosophy of Science*, 79, 795-807.

[10] Katzav, J. (2013). Hybrid models, climate models, and inference to the best explanation. *The British Journal for the Philosophy of Science*, 64, 107-129.

[11] Katzav, J. (2014). The epistemology of climate models and some of its implications for climate science and the philosophy of science. *Studies in History & Philosophy of Modern Physics*, 46, 228-238.

[12] Kuorikoski, J., Lehtinen, A. & Marchionni, C. (2010). Economic modelling as robustness analysis. *British Journal for the Philosophy of Science*, 61, 541-567.

[13] Kuorikoski, J., Lehtinen, A. & Marchionni, C. (2012). Robustness analysis disclaimer: Please read the manual before use! *Biology and Philosophy*, 27, 891-902.

[14] Laudan, L. & Leplin, J. (1991). Empirical equivalence and underdetermination. *The Journal of Philosophy*, 88, 449-472.

[15] Lenhard, J. & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Science* Part B: Studies in History and Philosophy of Modern Physics, 41, 253-262.

[16] Levins, R. (1993). A Response to Orzack and Sober: Formal Analysis and the Fluidity of Science. *Quarterly Review of Biology*, 68, 547-555.

[17] Lloyd, E. A. (2010). Confirmation and Robustness of Climate Models. *Philosophy of Science*, 77, 971-984.

[18] Lloyd, E. A. (2015). Model robustness as a confirmatory virtue: The case of climate science. *Studies in History and Philosophy of Science* Part A, 49, 58-68.

[19] Odenbaugh, J. (2011). True Lies: Realism, Robustness, and Models. *Philosophy of Science*, 78, 1177-1188.

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

[21] Orzack, S. H. & Sober, E. (1993). A critical assessment of Levins's The strategy of model building in population biology (1966). *Quarterly Review of Biology*, 68, 533-546.

[22] Parker, W. S. (2010a). Comparative Process Tracing and Climate Change Fingerprints. *Philosophy of Science*, 77, 1083-1095.

[23] Parker, W. S. (2010b). Predicting weather and climate: Uncertainty, ensembles and probability. *Studies in History and Philosophy of Modern Physics*, 41, 263-272.

[24] Parker, W. S. (2011). When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science*, 78, 579-600.

[25] Raerinne, J. (2013). Robustness and sensitivity of biological models. *Philosophical Studies*, 166, 285-303.

[26] Randall et al. (2007). Climate models and their evaluation. In S. Solomon et al. (ed.) *Climate Change 2007: The Physical Science Basis.* Cambridge and New York: Cambridge University Press, 589-662.

[27] Schurz, G. (1991). Relevant deduction. *Erkenntnis*, 35, 391-437.

[28] Sprenger, J. (2011). Hypothetico-Deductive Confirmation. *Philosophy Compass*, 6, 497-508.

[29] Weisberg, M. (2013) *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

[30] Wimsatt, W. C. (1980). Randomness and Perceived Randomness in Evolutionary Biology. *Synthese*, 43, 287-329.

[31] Wimsatt, W. C. (1981). Robustness, reliability and overdetermination. In M. B. Brewer, and B. E. Collins (Eds.), Scientific Inquiry and the Social Sciences San Francisco: Jossey-Bass, 124-163.

[32] Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology*, 13, 219-240.