UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Derivational robustness and indirect confirmation

Lehtinen, Aki

2018-06
Springer Netherlands
http://hdl.handle.net/10138/235616
Lehtinen, A 2018, 'Derivational robustness and indirect confirmation', Erkenntnis, vol. 83, no. 3, pp. 539-576. https://doi.org/10.1007/s10670-017-9902-6

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Derivational robustness and indirect confirmation

Aki Lehtinen

paper published in Erkenntnis 2018, vol. 83, no. 3. pp. 539-576.

Abstract
Derivational robustness may increase the degree to which various pieces of evidence indirectly confirm a robust result. There are two ways in which this increase may come about. First, if one can show that a result is robust, and that the various individual models used to derive it also have other confirmed results, these other results may indirectly confirm the robust result. Confirmation derives from the fact that data not known to bear on a result are shown to be relevant when it is shown to be robust. Second, robustness may increase the degree to which the robust result is indirectly confirmed if it increases the weight with which existing evidence indirectly confirms it. This may happen when it strengthens the connection between the core and the robust result by showing that auxiliaries are not responsible for the result.

1. Introduction

Although the basic idea of robustness analysis was introduced long ago (Levins 1966; Wimsatt 1981), the epistemic benefits have recently evoked increasing interest among philosophers of science. Robustness is often taken to provide epistemic support because a result is more likely to be reliable if several different and mutually independent routes lead to the same conclusion. Such derivations may consist in drawing a conclusion from a set of data or from theoretical models based on various assumptions. Woodward (2006) calls the former *inferential robustness*, and the latter *derivational robustness*. A theoretical *result* is thus derivationally robust if it can be derived from several different but partly overlapping sets of modelling assumptions. This paper deals exclusively with derivational robustness.

One of the remaining disputed issues concerns whether robustness provides some confirmation of the robust result. Kuorikoski et al. (2010) note that robustness is 'not a straightforward confirmation procedure', and Forber (2010) argues that its role is to limit the set of possible alternatives *before* empirical testing begins. Orzack and Sober (1993), Odenbaugh and Alexandrova (2011), as well as Houkes and Vaesen (2012) staunchly deny that it has any confirmatory power. Weisberg (2006; 2013, pp. 167-9) claims that robustness analysis does not offer any confirmation of robust theorems, but if it is combined with 'low-level confirmation' it may 'play a role' in confirmation. According to Parker (2011, see also 2010b), one cannot infer from the robustness of a result that scientists' confidence in it should be *significantly* increased.

Derivational robustness analysis is non-empirical in the sense that it does not require the collection of new data, it is possible in principle to study robustness without ever considering any empirical evidence, and it never confirms anything in the same sense as a piece of evidence confirms a theory. There are weaker notions of confirmation to which robustness may contribute, however. This paper explores the *indirect confirmation of assumptions* (Friedman 1953; Machlup 1955; 1956) and *of results* (Nagel 1961a; Laudan and Leplin 1991) and the discovery of mathematical or logical

relationships between a theory and various pieces of evidence (Hacking 1967; Garber 1983). A hypothesis or a result is indirectly confirmed if there is evidence that confirms it even though it is not a consequence of the hypothesis.

Models are typically modified and refined, and as a result they spawn *families of models* (Ylikoski and Aydinonat 2014) with partly overlapping sets of assumptions. Individual members of a family typically share a set of assumptions that is sometimes called the *common core* (see e.g., Levins 1993; Raerinne 2013; Lloyd 2015 for some discussion), the aim being to capture the workings of a causal mechanism. The purpose of derivational robustness analysis, then, is to ascertain whether the result is driven mainly by the core assumptions or by the various *auxiliary* assumptions that were needed in formulating the model. Theorists hope to show that the important results do not depend on auxiliaries, but rather lean on the core assumptions.

I will present an example from climate-change modelling in which derivational robustness *increases* the degree of indirect confirmation of a robust result.[1] I will show that stronger confirmation may derive from the fact that some data not previously known to have a bearing on the result are now relevant. Nevertheless, derivational robustness may also confirm by 'strengthening' the connection between the core and the robust result, thereby demonstrating that the core rather than the auxiliaries is responsible for the result. The robustness of climate models may thus increase confirmation, first through increasing the relevant indirect evidence and second through increasing the weight of the existing indirect evidence for the robust result.

I refer to the former as the *argument from the variety of evidence*, and to the latter as the *argument from strengthening the indirect confirmation of a robust result*. I will show how the indirect confirmation of a robust result may be strengthened via two possible routes: *strengthening the robust theorem* or *strengthening the indirect confirmation of the core*. My aim is to analyse the similarities and differences between these two arguments. As a rough characterisation, the first argument establishes *that* there is indirect confirmation of the robust theorem from some piece of evidence, and the second one *strengthens* the links in the structure of indirect confirmation that existed before the robustness was established.

Neither of these arguments is entirely new. According to William Wimsatt (1981), robustness allows identification of the assumptions that really 'drive' the robust result (see also Staley 2004; Kuorikoski, Lehtinen, and Marchionni 2010). Elizabeth Lloyd (2009; 2010) presents the argument from the variety of evidence and provides case studies in climate research, but she does not go into the details of how this happens. Given that Parker (2009) considers Lloyd's argument insufficiently developed, and Wimsatt's argument has not been explicitly formulated in terms of confirmation, my aim is primarily to examine the logic of these two arguments very closely, and to show how they are related to indirect confirmation. Lloyd's (2015) latest account is more detailed, but it concentrates on the variety of evidence for individual assumptions in climate models, and it is not explicitly based on indirect confirmation.

[1] Knuuttila and Loettgers (2011) present a case study (on the circadian clock) that also involves the interplay of data and robustness. See Guillemot (2010) for an account of the interplay of evidence and models in climate research.

A detailed philosophical understanding of the climate case is better gleaned from the work of Lloyd (see also 2012) Parker (see also 2010b; 2013), Katzav (2013; 2014) and a host of other authors.[2] I make no attempt to provide a detailed description of climate-change models either, because sorting out the complicated logic of the role of robustness in indirect confirmation would require me to resort to counterfactual scenarios. In fact, I use such scenarios to show that the two arguments are independent of each other: robustness may enlarge the set of relevant pieces of evidence without strengthening the robust theorem, and it may strengthen confirmation from old evidence without increasing the variety of evidence.

The structure of this paper is the following. Section 2 introduces the notion of indirect confirmation (2.1 and 2.2); discusses Okasha's (1997) critique (2.3) and a response to it (2.4); shows how it could be applied to modelling (2.5); explains what it means to say that robustness confirms (2.6); and discusses increasing the indirect confirmation of the core (2.7). Section 3 presents the arguments from the variety of evidence and from strengthening the indirect confirmation of a robust result in an example from climate modelling. Given that the philosophical literature on robustness has thus far not been couched in terms of confirmation theory, readers who are afraid of being alienated by the details of such discussions may skip sections 2.3, 2.4, 2.6 and 2.7. If reading Section 3 raises questions about the argument, they should consult these sections for further elaboration.

2. Indirect confirmation

2.1 Preliminaries on confirmation

I take 'x confirms y' to mean that getting to know x justifiably *increases* one's degree of belief in the truth of y (cf. Steele and Werndl 2013). The modellers' *epistemic situation* (Achinstein 2001, pp. 20-21) specifies which data are available and which derivational relationships between data and models they know about. If robustness confirms, it does so through changing the modellers' epistemic situation. Hence, confirmation is considered a 'subjective' notion in this paper. Nevertheless, when modellers obtain robust results, and perhaps even when they analyse data, they need not have the conscious aim of confirming an assumption or a result. The bulk of this paper analyses different epistemic situations by means of diagrams that represent the derivational relationships and the available data.

Demonstrations of robustness cannot change the epistemic situation of a logically omniscient agent. Hence, although the account presented here aims to be non-committal with respect to specific theories of confirmation, it is inconsistent with any theory that relies on logical omniscience. I will thus apply accounts of confirming *old evidence* to clarify what it means to say that robustness confirms (Sect. 2.6). In that robustness may only confirm by virtue of changing the modellers' epistemic situation, and that the modellers may know too much or too little about the derivational relationships and the available data, the confirmation it provides may be rather weak.[3] However, in that it is

[2] See e.g., the special issue (2010, vol. 41) on climate change in Studies in History and Philosophy of Modern Physics. Räisänen (2007) provides a non-technical introduction by a climatologist.

[3] I discuss the context dependence of confirmation via robustness further in Lehtinen (2016). I show, for example, that robustness may entirely fail to confirm even when there is indirect empirical

practically impossible to quantify the likelihood of evidence given a scientific theory or model (e.g., Schurz 2014a), the notion of confirmation is merely *qualitative*.

'Standard' accounts allow for *pseudo-confirmation*: every hypothesis with non-zero prior probability is confirmed by every piece of non-certain evidence E, if H only entails this evidence. The present account is based on ruling out pseudo-confirmation by applying accounts of *genuine* confirmation: Gemes' (1993; 1994; 2005) account of HD confirmation with 'content parts' is used to show that irrelevant conjuncts are not confirmed (Sect. 2.4), and Schurz' (2014a; 2014b, pp. 329-331) account of *genuine partial* confirmation is used to show how robustness may allocate confirmation to individual assumptions (Sect. 2.7). To start with, however, I introduce indirect confirmation.

2.2 Two kinds of indirect confirmation

A straightforward way of testing a model is to derive some predictions from it and then to see if they mesh with some data from the real world. Lloyd (2010, p. 974) calls such agreement between the model and the world the 'model fit'. If a result concerns the value of a variable such as the Global Mean Surface Temperature (GMST), for example, there is *model fit* if the model values at least roughly match the measured values (see also Parker 2011).

Evidence E is *direct* with respect to a hypothesis H (or a result R) if E is a consequence of H. It can be *indirect* in two ways (cf. Nagel 1961b, pp. 64-5). First, if a result R_M is derivable, together with other results R_1, R_2, … from a more general theory T, then direct evidence (E_1, E_2, …) for these other results counts as indirect evidence for R_M. Let c_i denote the indirect confirmation relation (read xc_iy as 'x indirectly confirms y'). For example, even though $R_M \not\vdash E_1$,

$E_1 c_i R_M$ if[4]

$$\begin{array}{c} T \vdash R_1 \\ \top \quad | \\ R_M \quad E_1 \end{array} \qquad (1)$$

Second, if a hypothesis C can be combined with various auxiliary assumptions A_1, A_2, … to entail results R_1, R_2, … then direct evidence for these results counts as indirect evidence for C (Machlup 1956). For example, even though $C \not\vdash E_1$,

$E_1 c_i C$ if

$$\begin{array}{c} (C \& A_1 \& A_2) \vdash R_1 \\ | \\ E_1 \end{array} \qquad (2)$$

Note, however, that (2) also implies $E_1 c_i A_1$ and $E_1 c_i A_2$. Let us call the first kind of indirectness (1) *indirect result confirmation* and the second kind (2) *indirect assumption confirmation*, and denote them by c_i^r and c_i^a, respectively.

---

evidence, but also that it is possible that a given initially non-confirming demonstration of robustness may become confirmatory later if the epistemic situation is modified in the right way.

[4]The '$\top$' and '$\vdash$' signs refer to the entailment relation, and the vertical line '|' to a direct model fit.

Laudan and Leplin (1991) provide an argument that exploits indirect result confirmation in going against underdetermination. Here is how it runs:

> Theoretical hypotheses H_1 and H_2 are empirically equivalent but conceptually distinct. H_1, but not H_2, is derivable from a more general theory T, which also entails another hypothesis H. An empirical consequence e of H is obtained. e supports H and thereby T. Thus e provides indirect evidential warrant for H_1, of which it is not a consequence, without affecting the credentials of H_2. Thus one of two empirically equivalent hypotheses or theories can be evidentially supported to the exclusion of the other by being incorporated into an independently supported, more general theory that does not support the other, although it does predict all the empirical consequences of the latter. (Laudan and Leplin 1991, p. 464)

Let E denote the common empirical consequence of H_1 and H_2 such that e Ï E. Let $x¢_i^r y$ denote 'x does not indirectly confirm y'. The argument could be schematically represented as follows:

$ec_i^r H_1$ because (T ├ H) but $e¢_i^r H_2$ because (T ⊬ H_2) (3)

T T T
H_1 e E

T
E

Laudan and Leplin use the theory of continental drift as an example of a general theory T. H_2 could state, for example, that the continents are immobile but were once connected with 'bridges' that have disappeared. Such a hypothesis could explain, say, the similarities in fauna and flora in the west coast of Africa and the east coast of South America (E). T implies two more specific hypotheses H_1 and H. The first is that the climate at any given current geographical location has undergone various changes throughout history (H_1), and the second is that the location of the magnetic poles, and even the direction of the magnetic field, will change over time (H). In the 1950s scientists observed that there were streaks in the volcanic lava at the bottom of the ocean. It was further established that the direction of the streaks varied depending on the timing of the eruption that produced the lava (e). This could be explained in accordance with continental drift theory: lava that cools down after an eruption aligns itself according to the positions of the then current magnetic poles (H ├e). This piece of evidence (e) for the second hypothesis (H) also indirectly confirms the first ($ec_i^r H_1$) by way of confirming the general theory of continental drift (T). Hypotheses with the same empirical consequences (E) may thus attract differing degrees of evidential support. This argument challenges the view that a theory may only be evaluated in terms of its empirical consequences.

Laudan and Leplin use indirect confirmation to argue against underdetermination, but I only need the part of their argument establishing that indirect confirmation does indeed provide confirmation, that is, the left side of (3). I need to establish the cogency of indirect confirmation because Okasha (1997) argued that the left side of (3) is problematic, and thus that indirect confirmation cannot confirm.

2.3 Okasha's critique

Laudan (1996, p 67) argues that indirect confirmation depends on the 'intuitive, uncontroversial principle that evidential support flows downward across the entailment relation'. Indeed, Hempel's (1965, p. 31) *special consequence condition* (SCC) is based on this intuition. It states that if an observation report *E* confirms a hypothesis *H*, and if *H* logically implies that *R* is true (*R* is a logical consequence of *H*), then *E* also confirms *R*. To put it simply, if a hypothesis is confirmed, its consequences are also confirmed.

As Okasha observes, however, Laudan and Leplin's argument for the indirect confirmation of results simultaneously rests on Hempel's *converse consequence condition* (CCC), which states that if E confirms H, and T ⊢ H, then E confirms T. The argument is thus based on two conditions that, if taken together, are known to yield bizarre implications such as the tacking paradox (Hempel 1945, p. 104).[5] The problem is that if E_1 confirms T, in accordance with the CCC, it is allowable to formulate a conjunction T&X when X is any utterly irrelevant proposition such as 'the moon is made of green cheese', and in accordance with the SCC, E_1 also confirms X:

$$E_1 c_i^r X \qquad \text{because} \qquad \begin{array}{c} T\&X \vdash R_1 \\ \top \quad | \\ X \quad E_1 \end{array} \qquad (4)$$

According to Okasha, then, the indirect confirmation of results is problematic because it must rely simultaneously on the SCC and the CCC, and this allows for the possibility of confirming irrelevant assumptions and results. If Okasha were right, indirect confirmation would be questionable because it would be possible to indirectly confirm just about any preposterously false hypothesis.

However, Okasha's argument is seriously misleading because the problem of tacking does not primarily derive from the incompatibility of Hempel's conditions. As Glymour (1980, pp. 133-5; 1983) suggests, the tacking-by-conjunction problem affects any confirmation theory that is based on the entailment relation. It is well known, for example, that Bayesians face the same problem when the hypothesis entails the evidence. Although Laudan notes this, he does not develop the argument fully, which is what I will attempt to do here. I will now show that this problem can be solved by replacing the entailment relation with the *content part* relation (Gemes 1993).

2.4 Content parts and natural axiomatizations

As Glymour notes, the problem with traditional HD accounts of confirmation was that they were too liberal: evidence E was taken to confirm a hypothesis H if H ⊢E as long as H was consistent and E was not tautological. Because H&X ⊢E if H ⊢E, E confirms H&X for any X. No consideration was given to whether deriving E from H&X depended on using both conjuncts. However, the tacking problem can be solved even in an HD framework by imposing further conditions on the confirmation relationship (Gemes 1993; 1994; 2005; Schurz 1991; 1994). Gemes' solution is based on replacing the notion of entailment with the content part relation, and on the idea that the theory

[5] Niiniluoto and Tuomela (1973, pp. 3-4) present essentially the same argument as Okasha but without using the term 'indirect confirmation', and Bangu (2006) re-employs the argument but without mentioning Hempel's result.

must be a *natural axiomatization*. Let α be a variable defined on the well-formed formulas (wff) of a language, and let β be a variable for the wffs and sets of wffs of the same language. The content-part relation is denoted $\beta \vdash_c \alpha$, and is defined as follows (1993, p. 481):

> α is a content part of β iff α and β are contingent, $\beta \vdash \alpha$, and there is no σ such that $\beta \vdash \sigma$, σ is stronger than α, and every atomic wff that occurs in σ occurs in α.

To say that σ is stronger than α means that $\sigma \vdash \alpha$ and $\alpha \nvdash \sigma$. In plain English, this definition states that α is a content part of β if β entails α, and in addition α is the strongest possible consequence that can be derived from β. The content-part relation is also used to define the notion of a natural axiomatization n(T):

> T' is a natural axiomatization of T iff (i) T' is a finite set of wffs such that T' is logically equivalent to T, (ii) every member of T' is a content part of T', and (iii) no content part of any member of T' is entailed by the set of the remaining members of T'.

The notion represents an attempt to express what exactly a given theory says about the world. Another option is to state that a n(T) does not contain any redundant axioms (ibid., p. 482). Condition (iii) expresses the idea that the axioms should be independent of one another. Gemes' account of HD confirmation is thus the following (ibid., p. 486):

> E HD confirms axiom A of theory T relative to background evidence B, iff E is a content part of (T&B), and there is no natural axiomatization n(T) of T such that for some subset S of the axioms of n(T), E is a content part of (S&B) and A is not a content part of (S&B).

This strengthened definition of HD confirmation disposes of the problem of tacking.[6] As Glymour (1980, pp. 30-31) notes, nobody denies the intuitive plausibility of SCC and CCC. If Hempel's conditions are reformulated with the content-part relation rather than the entailment relation, they are not inconsistent.

2.5 Indirect confirmation, models and robustness

Derivational robustness is a matter of investigating whether similar results can be *derived* from a family of models. Hence, to illustrate the modellers' inferences, models are presented as conjunctions of assumptions. For example, writing $M_1 = (A_1\&A_2\&A_3) \vdash R$ means that assumptions A_1, A_2, and A_3 are being used to derive result R from model M_1. Although accounts that emphasise the importance of inferential issues in modelling would be particularly suitable (Suárez 2004; Kuorikoski and Lehtinen 2009), this way of presenting modelling inferences does not imply a commitment to any specific interpretation of what models are. As a matter of fact, it would be possible to remove the symbols indicating models (M_1, M_2 etc.) in the analyses that follow, but they are retained because they provide notational shorthand for the various combinations of assumptions.

[6] I do not intend to argue for HD as opposed to other accounts of confirmation by applying Gemes' account, and neither did Gemes by presenting it (see e.g., Gemes 1993).

Let us now see how Gemes' account also disposes of the problem of tacking in the context of a family of models. If entailment were the relevant relationship between models and their results, then if, for example, model M_1 were to entail a result R_1, and if E_1 were to support R_1,

$$M_1=(C\&A_1\&A_2\&A_3) \vdash R_1, \qquad (2')$$
$$\qquad\qquad\qquad\qquad\quad |$$
$$\qquad\qquad\qquad\qquad\quad E_1$$

E_1 would confirm not only M_1 and all the individual assumptions in it, but also every imaginable assumption that could be tacked onto M_1 and thus also the infamous 'green cheese' model M_{gc} = $(C\&A_1\&A_2\&A_3\&X)$. Furthermore, its consequence X would be indirectly result-confirmed.

$$M_{gc}=(C\&A_1\&A_2\&A_3\&X) \vdash R_1, \qquad (2'gc)$$
$$\qquad\quad \top \qquad\qquad\quad |$$
$$\qquad\quad X \qquad\qquad\quad E_1$$

Let us apply the definition of natural axiomatization to confirming the infamous model M_{gc} = $(C\&A_1\&A_2\&A_3\&X)$ with evidence E_1. M_{gc} is not a natural axiomatization because by eliminating X we obtain a natural axiomatization of the model M_1 = $(C\&A_1\&A_2\&A_3)$ such that the set of axioms in M_1 constitute a subset of the axioms in M_{gc}, E_1 is a content part of M_1, but X is not. The irrelevant premise X is thus not indirectly confirmed (qua assumption or qua result), and the problem of tacking is solved.

Modellers typically have extensive background knowledge concerning the truth values of the various assumptions and whether they take part in deriving various results. The first kind of background knowledge may come from previous indirect *assumption* confirmations (or disconfirmations), from direct empirical tests of individual assumptions, from comparison to theory, or from intuitive judgments on the plausibility of the assumptions. Both kinds of background knowledge are crucial for judging whether the models can be taken to be reliable, plausible, or adequate for purpose.

It is important to stress, however, that the analysis of the confirmatory benefits of robustness ignores all direct confirmation or disconfirmation of assumptions unless explicitly indicated otherwise (i.e., in subsections 2.7 and 3.2). This means that, even if X were to have been re-interpreted as, say, some auxiliary in a climate model, it would have been ignored by the current analysis on the ground that X alone entails X, and direct evidence for or against any assumption is ignored. In other words, only indirect confirmation of assumptions is relevant background knowledge for the analysis, and it is relevant only insofar as such knowledge indicates that there is some uncertainty concerning the truth value of a given assumption. If modellers already know that an assumption is true or false, such an assumption cannot be confirmed or disconfirmed by means of robustness. Furthermore, ignoring direct confirmation or disconfirmation also implies that the analysis presented in this paper cannot take any position on whether the models should be taken to be reliable, plausible, or merely adequate for evaluating some individual results. Yablo (2014, pp. 100-1) distinguishes between *fully* and *basically* confirming a conjunction. In the former case E_1 would 'probabilify' (i.e.,

increase confirmation of) both the conjunction M_1 and each of its parts separately. In the latter case E_1 merely probabilifies the conjunction M_1. The (climate) models discussed in this paper are clearly far too complex and contain far too many idealisations to be fully confirmed.

Although robustness affects the overall evaluation of the models through the evaluation of the plausibility of various results and assumptions, it is not sufficient in itself to dramatically change modellers' judgments on their overall reliability. This is because a result may only become robust if it has already been derived at least once. If the result was important and there was confirming evidence for it, most of the confirmation relevant for judging the overall reliability of the models would already have been obtained from the first derivation. Nevertheless, robustness may incrementally confirm even when modellers judge the overall performance of the model ensemble to be poor. In other words, the epistemic benefits of robustness do not depend on the models being already considered reliable. If they are unreliable, the *absolute* confirmation remains low even for results that are shown to be robust.

The evaluation of models typically involves an investigation of the similarities and differences among various results in a family of models. I will be analysing epistemic situations that combine the two kinds of indirect confirmation:

$E_1 c_i^r R_M$ if

$$\begin{array}{ccc} (C\&A_1\&A_2\&A_3) & \vdash & R_1 \\ \top & & | \\ R_M & & E_1 \end{array} \qquad (2'^{M})$$

Readers are advised to study the diagrams carefully because they help to show how robustness contributes to confirmation. No formal results will be established, however. The strength with which pieces of evidence such as E_1 confirm model components such as C or conclusions such as R_M varies from one epistemic situation to another and, more importantly, depends on robustness. From now on I will take as given that E_1 *may* indirectly confirm R_M in $(2'^{M})$. It should be clear from the discussion on tacking, however, that E_1 indirectly confirms result R_M only if R_M and R_1 are mostly attributable to the same elements (such as C). The indirect confirmation of result R_M vanishes entirely if it is shown to depend on different assumptions than R_1. This is close to what confirmation theorists call mere *content-cutting* (Earman 1992, p. 98; Schurz 2014a; 2014b, pp. 320-2; Votsis 2014; Gemes undated). If modellers start with something like (2'M), there are plenty of possibilities for cutting the content of R_1 from that of R_M.

2.6 Incremental confirmation of old evidence

Let us now consider exactly what it means to say that robustness confirms when it does indeed confirm. The notion of confirmation used in this paper is *incremental*. To say that confirmation is incremental means that if something is confirmed, there is an increase in the degree to which it is confirmed, but there is no guarantee that any specific level or threshold of confirmation is reached. The level of *absolute* confirmation may thus remain rather low (e.g., Schurz 2014b, p. 318). In *Bayesian incremental confirmation* a piece of evidence E confirms hypothesis H incrementally given background beliefs B iff $P(H|E\&B) > P(H|B)$. However, unlike in such accounts, robustness increases confirmation not because a new piece of evidence is found, but rather because there is a change in the epistemic situation concerning derivational

relations. Standard accounts of old evidence (Garber 1983; Jeffrey 1983; Niiniluoto 1983) posit that deriving a confirmed result E from a hypothesis H increases the prior probability of that hypothesis so that $P(H|H \vdash E)>P(H)$. If robustness confirms, the confirmation increment resembles that in accounts of old evidence.

A *robust theorem* concerns the relationship between the joint core C and the *robust result* R_M. The general form of the robust theorem is 'ceteris paribus, if [common core structure] obtains, then [robust result] will obtain' (Weisberg and Reisman 2008; Weisberg 2006; Lloyd 2010; Houkes and Vaesen 2012). The core never implies the robust result all by itself, but rather entails it together with auxiliary assumptions. Several different sets of auxiliaries entail the robust result if combined with the core, but robust theorems are never completely cleansed of all auxiliary assumptions. Indeed, insofar as the theorem in question is robust rather than genuine, modellers are not able to tell exactly which assumptions are needed for deriving the robust result. (RR) thus provides a description of an epistemic situation with a robust result:

$$\left.\begin{array}{l} M_1 = (C\&A_1\&A_2\&A_3) \vdash R_M, R_1 \\ M_2 = (C\&A_2\&A_4\&A_5) \vdash R_M \\ M_3 = (C\&A_1\&A_6\&A_7) \vdash R_M, \end{array}\right. \qquad \text{(RR)}$$

Recall that in Laudan and Leplin's example,

$ec_i^r H_1$ because $(T \vdash H)$ but $e\not\subset_i^r H_2$ because $(T \nvdash H_2)$. (3)

$$\begin{array}{ccc} \top & \top & \top \\ H_1 & e & E \\ \top & & \\ E & & \end{array}$$

Let us now see the similarities and differences in epistemic situations between robustness (RR) and Laudan and Leplin's example. The confirming empirical evidence (e) is not entailed by hypothesis (H_1) in Laudan and Leplin's account, but this hypothesis is nevertheless indirectly confirmed because it is a consequence of a more general theory (T), which entails another hypothesis (H) that is confirmed by empirical evidence (e). The robust result R_M is indirectly confirmed because it is the joint consequence of the common core structure C in a family of models, and R_1 is confirmed and shown to depend crucially on C. This common core C in a family of models thus replaces the role of a general theory T in Laudan and Leplin's account of indirect confirmation.

Their argument supporting the claim that the changing-climate hypothesis H_1 but not the 'bridge' hypothesis H_2 is confirmed by evidence e depends crucially on the fact that H_1 but not H_2 can be derived from the more general theory T. Clearly, it cannot be that $P(H_1|H_1 \vdash e)>P(H_1)$ simply because $H_1 \nvdash e$, and e confirms H_1 only if $T \vdash H$, $T \vdash H_1$, and $H \vdash e$. In other words, streaks in lava at the bottom of the ocean (e) were considered irrelevant for the changing-climate hypotheses (H_1) until this hypothesis was shown to be a consequence of a general theory T ($T \vdash H_1$) that also explained the streaks ($T \vdash H$).

If the basic idea in accounts of old evidence is applied to Laudan and Leplin's example, an increment in confirmation that derives from coming to know derivational relationships could be written as follows:

$P(H_1|(T \vdash H)\&(H \vdash e)\&(T \nvdash H_1))>P(H_1)$.

Here $P(H_1)$ denotes the prior probability of hypothesis H_1 when the background knowledge does not include information on the derivational relationships, and $P(H_1|(T \vdash H)\&(H \vdash e)\&(T \nvdash H_1))$ denotes the prior when it does.

The account is similar in the case of robustness. Consider again (2'M):

$E_1 c_i^r R_M$ if

$$\begin{array}{ccc} & (C\&A_1\&A_2\&A_3) \vdash R_1 & \\ \top & | & \\ R_M & E_1 & \end{array} \qquad (2'^{M})$$

E_1 would indirectly confirm R_M if modellers somehow knew that E_1 could not cut the content of $(C\&A_1\&A_2\&A_3)$ in such a way that the confirmation would not reach R_M. If C alone implied R_M and R_1, they would know that such content cutting was impossible. The increment in confirmation would then be

$P(R_M|(C \vdash R_M)\&(C \vdash R_1)\&(R_1 \vdash E_1))>P(R_M)$.

The assumption throughout this paper is that modellers know that $R_1 \vdash E_1$, so that this could be simplified as

$P(R_M|(C \vdash R_M)\&(C \vdash R_1))>P(R_M)$. (ICR)

However, because robustness never establishes that C alone entails R_M or R_1, it alone never brings the modellers to the epistemic situation (ICR): it is always possible, in principle, that E_1 cuts the content of $(C\&A_1\&A_2\&A_3)$ in such a way that R_M is not confirmed at all. This would happen, for example, if it turned out that $C\&A_1 \vdash R_1$ and $A_2\&A_3 \vdash R_M$.

Let us now describe modellers' knowledge of the derivational relationship between assumptions and results as follows. Let $A_i \vdash_0 R_1$ denote the epistemic situation in which they do not know anything about the derivational relationship between A_i and R_1. In other words, in this situation the modellers think that A_i and R_1 are completely unrelated, and by definition $P(R_M|(C \vdash_0 R_M)\&(C \vdash_0 R_1))= P(R_M)$. Let $A_i \vdash_{\Diamond cp} R_1$ denote the situation in which the modellers know at least something about the derivational relationship between A_i and R_1, and as far as they know, A_i could be needed for deriving R_1. Let $+\top_c$ and $+ \vdash_c$ denote the change in the epistemic situation that robustness brings about when it shows that a specific assumption is needed for a result.

If an assumption A_i is indirectly assumption-confirmed due to the robustness of R_1 (recall that $R_1 \vdash E_1$), then the change in the modellers' epistemic situation could be represented as follows: $P(A_i|(A_i+ \vdash_c E_1))>P(A_i)$ or $[E_1]+c_i^a A_i$. Thus, if C is assumption-confirmed by R_1, $P(C|(C+ \vdash_c R_1))>P(C)$.

Suppose now that the modellers start in an epistemic situation in which R_M is not known to be robust and C is not known to be relevant to R_M or R_1 for some other reason (i.e., C is not known to take part in a derivation of R_M or R_1):

$P(R_M|(C \vdash_0 R_M)\&(C \vdash_0 R_1))= P(R_M)$.

Learning that $C+ \vdash_c R_1$ changes the epistemic situation into

$P(R_M|(C \vdash_0 R_M)\&(+C \vdash_c R_1))= P(R_M)$. (IAC)

The equality sign in this formulation makes it clear that indirect assumption-confirmation of the core (IAC) due to the robustness of R_1 is not sufficient in itself indirectly to confirm the result R_M. It will do so only if the modellers know that there is some derivational relationship between C and R_M. If $C \vdash R_M$ were also known to hold, the epistemic situation could be described as follows.

$P(R_M|(C \vdash R_M)\&(C+ \vdash_c R_1))>P(R_M)$,

However, modellers are never assumed to know that $C \vdash R_M$, but rather something weaker such as the robustness of R_M. Yet, if they start with at least some knowledge of the relationship between C and R_M,

$P(R_M|(C \vdash_{\Diamond cp} R_M)\&(C \vdash_0 R_1))= P(R_M)$,

and the robustness of R_1 then changes the epistemic situation into

$P(R_M|(C \vdash_{\Diamond cp} R_M)\&(C+ \vdash_c R_1)) > P(R_M)$, (SICC)

robustness confirms R_M indirectly and incrementally by Strengthening the Indirect Confirmation of the Core (SICC). Similarly, if they start with

$P(R_M|(C \vdash_0 R_M)\&(C \vdash_{\Diamond cp} R_1))$,

and the robustness of R_M then changes the epistemic situation into

$P(R_M|(C+ \vdash_c R_M)\&(C \vdash_{\Diamond cp} R_1)) > P(R_M)$, (SRT)

robustness confirms R_M indirectly and incrementally by Strengthening the Robust Theorem (SRT).

However, the modellers do not necessarily start with no information about the derivational relationships. If they start with

$P(R_M|(C \vdash_{\Diamond cp} R_M)\&(C \vdash_{\Diamond cp} R_1))=P(R_M)$,

robustness confirms as long as there is an increment via either SICC: $P(R_M|(C \vdash_{\Diamond cp} R_M)\&(C+ \vdash_c R_1))>P(R_M)$ or SRT: $P(R_M|(C+ \vdash_c R_M)\&(C \vdash_{\Diamond cp} R_1))>P(R_M)$.

Thus, when modellers start from some initial epistemic situation, *strengthening the indirect confirmation of a robust result* may happen through strengthening either the C-R_M link or the C-R_1 link. This explains why the strengthening argument has two possible routes: strengthening the robust theorem (SRT) and strengthening the indirect confirmation of the core (SICC).

The inequalities in SRT and SICC implicitly refer to evidence E_1 in the sense that $R_1 \vdash E_1$ is assumed to hold, and in the sense that they concern how E_1 would confirm R_M if it were to become available. They are deliberately written without mentioning E_1, however, to highlight the fact that E_1 is not necessarily assumed to be available, and robustness does not contribute to whether it is or not. Furthermore, modellers cannot learn SRT and SICC at the same time as they learn that a piece of evidence (E_1) arrives. If the evidence is already available when SRT or SICC are learned, it is old. If it (i.e., E_1) is not already available, robustness cannot confirm via E_1. However, when the evidence arrives, it then confirms R_M more when the modellers have knowledge about the derivational relations in SRT and SICC than it would in the counterfactual situation in which they do not have such knowledge. It thus *increases* the Bayesian *increment* in a counterfactual sense. This explains why studying counterfactual scenarios is not merely a convenient way of speculating: it is rather an indispensable aspect of the confirmation relation, just as in some accounts of confirmation with old evidence (Howson 1991; see also Sprenger 2015).

The aim in this subsection was to articulate what it means to say that robustness confirms, without trying to justify that it confirms. Note, however, that SRT and SICC do not rule out the kind of content-cutting that would remove the indirect confirmation of R_M altogether. For example, if robustness now shows that $P(R_M|(C \vdash_{\Diamond cp} R_M)\&(C+ \vdash_c R_1)) > P(R_M)$, it is possible at least in principle for later investigations to show that C is not needed to derive R_M or R_1. The possibility of such content-cutting would be entirely removed only if the weaker derivational relationships (including entailments) were all replaced by content part relations in the final epistemic situation:

$P(R_M|(C \vdash_c R_M)\&(C \vdash_c R_1))>P(R_M)$.

The relationships between the assumptions and the results can be conceptualised in terms of *sufficient* conditions: the set of assumptions used for deriving a result provides a set of sufficient conditions for it, and several models provide a collection of such sets of sufficient conditions. Accordingly, modellers can establish that some auxiliaries are *not* relevant for some results, but insofar as they are not able to provide necessary and sufficient conditions for a result, they are not able to conclusively establish that $C \vdash_c R_M$ or $C \vdash_c R_1$. In other words, the inequalities in SRT and SICC must hold even when the modellers have less than complete information about the derivational relationships. Logically omniscient modellers could perhaps establish that such content part relations hold, but then robustness would no longer be useful because there would be no epistemic uncertainty. The account presented here is based on the premise that approaching logical omniscience counts as progress even if it can never be reached. The point is that robustness may confirm precisely because showing that false auxiliaries are not responsible for the results *removes* some possibilities of content-cutting.

Robustness can thus confirm through strengthening the indirect assumption confirmation only if there is epistemic uncertainty concerning which results depend on which assumptions. Here the logical learning concerns the relationship between results and individual assumptions rather than scientific theories, which is why logical learning

is always weaker than entailment in the sense that modellers never learn that any A_i alone entails any result.[7]

The argument that robustness may confirm requires several further steps. The next subsection shows that robustness can confirm individual assumptions.

2.7 Strengthening the indirect confirmation of the core

I will now show how the link between C and R_1 can be strengthened via robustness. Below I set out my argument explaining why the robustness of a confirmed result (such as R_1) may increase the indirect assumption confirmation of the core. Some results depend more than others on some assumptions. A given *premise* is *irrelevant* for a given conclusion if and only if the conclusion can be derived with or without it (Schurz 1991). Gemes and Schurz' contributions formalise the idea that irrelevant premises are not confirmed even if they are included in a conjunction of premises that entail some confirmed result. Only the conjuncts that are needed for deriving a piece of confirming evidence (E_1) are indirectly assumption-confirmed by it. *Increasing* the indirect confirmation of the core requires that confirmation of a confirmed result can be allocated to individual assumptions, namely the core.

In what follows I use the expression 'set A is more likely to be needed (or necessary) for deriving result R_M' to indicate that at least some epistemic uncertainty concerning what depends on what is resolved. Showing the robustness of a result changes the modellers' epistemic situation by increasing the likelihood that a specific set of assumptions, the core rather than the auxiliaries, is needed for deriving the result. This may increase the indirect confirmation of individual assumptions such as the core. If there is an increment in confirmation of an individual assumption A_i due robustness, it is because A_i is shown to be needed for deriving E_1. When there is such an increase, it may be possible to write $P(A_i|(A_i+ \vdash_c E_1))>P(A_i)$.

Confirmation requires evidence. Robustness provides information about which assumptions are needed for which result if there is no evidence at all, but it cannot increase confirmation. Although information about derivational relationships can be used for allocating indirect confirmation only when there is at least some empirical evidence, it is worth emphasising that the derivational relationships and the confirmation relationships are separate: bringing in new knowledge about derivational relationships may justifiably increase *confidence* in the robust theorem or in the relevance of an assumption to a result, but such knowledge cannot *increase confirmation* if there is no empirical evidence.

Thus, if modellers derive (2*) and (2**),

$$M_1=(C\&A_1\&A_2\&A_3) \vdash R_1 \qquad (2^*)$$
$$M_2=(C\&A_2\&A_4\&A_5) \vdash R_1, \qquad (2^{**})$$

they know that C and A_2 are more likely to be needed for deriving R_1 but they are no more certain than before because there is no confirmatory evidence in the first place. Thus, if they learn (2**) after having learned (2*), they will not find out that, say, A_3 is no longer indirectly confirmed. They will merely learn that A_3 is not needed for deriving

[7] See Hartmann & Fitelson (2015) for an account in which old evidence confirms even in cases weaker than entailment.

R_1. Under epistemic uncertainty concerning what follows from what, it is reasonable to assume that any derivation that uses an individual assumption A_i to derive a result R increases modellers' confidence that A_i is needed.

Given that derivational and confirmatory relationships are separate, and that robustness affects only the former directly, it is possible to write $A_i+ \vdash_c R_1$ if the robustness of R_1 makes it more likely that A_i is needed for deriving R_1. One could now write $(C\&A_2)+ \vdash_c R_1$ to indicate that $(C\&A_2)$ is more likely to be necessary for deriving result R_1 when the modellers derived (2**) after having derived (2*). If C and A_2 are more indirectly confirmed due to robustness it is because $P(C|(C+ \vdash_c R_1)) > P(C)$ and $P(A_2|(A_2+ \vdash_c R_1)) > P(A_2)$ when E_1 is available. Because the derivation can be conducted before or after evidence E_1 arrives, robustness does not confirm if it is established before E_1 is available. Let us express an increment in confirmation as follows: $E_1+c_i^a C$ and $E_1+c_i^a A_2$. $E_1+c_i^a C$ is an example of *increasing the indirect assumption-confirmation of the core.*

Given that I only discuss cases with empirical evidence in this paper, it is not necessary to emphasise the strict separation between the derivational and the confirmatory relationships in all of what follows.[8] If an assumption is demonstrably not needed for deriving R_1, it cannot confirm R1. By way of an illustration, suppose that modellers start with (2'):

$$M_1=(C\&A_1\&A_2\&A_3) \vdash R_1, \quad (2')$$
$$\Big| \; E_1$$

Insofar as the modellers do not already know that some assumptions are not needed for deriving R_1, (2') initially means that all the assumptions in M_1 are indirectly confirmed by E_1: $E_1 c_i^a C$, $E_1 c_i^a A_1$, $E_1 c_i^a A_2$, and $E_1 c_i^a A_3$. To say that an assumption A_i is indirectly confirmed by some piece of evidence thus does not mean that it is true, or even close to being true. Indeed, modellers' background knowledge often indicates that some specific auxiliaries are clearly false. Such auxiliaries thus have *direct* disconfirming evidence. Suppose, for example, that the modellers knew that A_3 was false. The epistemic situation could be described as follows:

$$M_1=(C\&A_1\&A_2\&A_3) \vdash R_1, \quad (2'f)$$
$$\top \quad \Big| \; E_1$$
$$\sim E_{A3}$$

E_{A3} is a piece of direct evidence concerning A_3. If the modellers are interested in the truth value of A_3 they should weigh the direct disconfirming evidence against the indirect confirming evidence. They may well end up with the judgement that the direct disconfirming evidence clearly outweighs the indirect confirming evidence. Suppose they then learn that (2'') holds.

$$M_2=(C\&A_2\&A_4\&A_5) \vdash R_1 \quad (2'')$$
$$\Big| \; E_1$$

[8] See Hands (2016) for a study of robustness with virtually no empirical evidence.

After having learned (2''), A_1, A_3, A_4, and A_5 are no longer indirectly confirmed by E_1: $E_1¢_i^a A_1$, $E_1¢_i^a A_3$,… Schurz (2014a) argues that an increase in H's probability P(H|E) > P(H) spreads from H to an E-transcending content part A of H (P(A|E) > P(A)) only if A is necessary within H to make E highly probable. In other words, there exists no conjunction H* of content elements of H that makes E at least equally probable (P(E|H*)≥P(E|H)) but does not entail A. Schurz thus provides a Bayesian analogue to the claim that irrelevant conjuncts are not confirmed. In this example, the only E_1-transcending content parts of M_2 that can be necessary to make E_1 highly probable are C and A_2.

Deriving (2'') must increase the indirect confirmation of the assumptions (C and A_2) that were involved in both derivations, i.e. assumptions at the intersection of the two sets. This reasoning is based on Mill's method of agreement: C and A_2 are the only components that the two models share, thus it seems natural to assume that they explain the similarity in results.

Just like all applications of Mill's methods, such inferences are fallible. If, for example, modellers misidentify the assumptions in the models, or if deriving the results requires a combination of several assumptions rather than just one, they may be led astray. Given my focus on the case with robust results that require combinations of assumptions (Lehtinen 2016; see also Lisciandra 2017), I only discuss misidentification here. Suppose, for example, that A_3 and A_4 are not genuinely different assumptions, but that they rather share a common component A_{34}, and distinguishing components A_3' and A_4': $A_3=A_{34}\&A_3$' and $A_4=A_{34}\&A_4$'. This would mean that the models should have been written as follows: $M_1=(C\&A_1\&A_2\&A_{34}\&A_3')$ and $M_2=(C\&A_2\&A_{34}\&A_4'\&A_5)$, and that the increased indirect confirmation would be spread among a larger number of assumptions: $E_1+¢_i^a C$, $E_1+¢_i^a A_2$ and $E_1+¢_i^a A_{34}$. Thus, insofar as A_{34} would not turn out later to be irrelevant for deriving R_1, the increment in the indirect assumption confirmation of C and A_2 would be smaller due to robustness. This would mean that robustness analysis would be less *complete* in the sense that more shared assumptions would remain. Such failure to identify common assumptions in models is particularly dangerous because the modellers will think that C is more indirectly assumption-confirmed by E_1 than it ought to be.

3. Increasing confirmation through robustness

3.1 An example from climate modelling

Let us now consider the climate models that Lloyd (2010) discusses. Let C stand for 'increase in greenhouse gases' and let R_M stand for 'increased GMST in the future'.[9]

[9] 'Future temperature increase' and 'increase in greenhouse gases' may refer to various things but the details are not needed in this paper. There are different scenarios of future CO2 emissions and various ways to conceptualise future temperature increases. Equilibrium Climate Sensitivity (ECS) determines the long-term equilibrium warming response to stable atmospheric composition, but does not account for vegetation or ice-sheet changes. Transient Climate Response (TCR) is a measure of the magnitude of transient warming while the climate system, particularly the deep ocean, is not in equilibrium; and Transient Climate Response to Cumulative CO2 emissions (TCRE) is a measure of the transient warming response to a given mass of CO2 injected into the atmosphere, and combines information on both the carbon cycle and climate response. TCR is estimated with *high confidence* to be *likely* between 1°C and *2.5*°C and extremely unlikely to be greater than 3°C (Bindoff and Stott 2013, pp. 6, 59-60).

The family of models contains some common assumptions C as well as variable further assumptions A_i, and a robust result R_M can be derived from each individual model. Lloyd's example of a robust theorem is: 'Ceteris paribus, $C \vdash R_M$'. What has been established is something like this:

$$\left.\begin{array}{l} M_1 = (C\&A_1\&A_2\&A_3) \vdash R_M \\ M_2 = (C\&A_2\&A_4\&A_5) \vdash R_M \\ M_3 = (C\&A_1\&A_6\&A_7) \vdash R_M, \end{array}\right. \quad (5)$$

and the robust theorem (henceforth abbreviated as 'cp, $C \vdash R_M$') is taken to follow from such derivations. Lloyd expresses the robust theorem as follows. 'Ceteris paribus, if [Greenhouse gases relate in lawlike interaction with the energy budget of the earth] obtains, then the [increased global mean temperature] will obtain' (p. 980). The mechanism responsible for the increase in temperature is the following. Some of the radiation from the sun reaches the earth's surface and heats it. The radiation is then re-emitted from the surface in all directions. Some of the thus re-emitted radiation is absorbed by the greenhouse gases in the atmosphere, resulting in a higher surface temperature than there would be in the absence of such gases. The reason why there is more radiation coming into the earth than exiting it is that the incoming radiation from the sun has a shorter wavelength than that which is re-emitted from the ground. The radiation from the sun goes through the greenhouse gases into the Earth, but due to the longer wavelength only some of it exits the atmosphere.

C refers to greenhouse-gas forcing in climate models in the past and in the future. Let R_T denote the increase in the GMST in the past. One could also formulate another robust theorem (cp, $C \vdash R_T$) concerning the causes of the historical development. A *robust property* or a *result* (R_M or R_T) is different from a *robust theorem*. Let us imagine that the theorem is false, in other words that CO2 emissions do not increase the temperature. It is clear in that case that evidence for R_T cannot confirm C even if both R_T and C were true. Neither can evidence for R_T confirm the robust *theorem* itself.

Proponents of robustness argue that the robustness of R_M increases modellers' confidence in the robust theorem: (5) makes the robust theorem more plausible. Because direct evidence for (R_T) does not fit the bill, it would be inappropriate to call *this* increase in confidence an increase in *confirmation* because it derives purely from derivational relationships. Yet, as I explained in (Lehtinen 2016), although the derivational relationships do not yet constitute confirmation, they are a necessary but not sufficient condition for confirming the theorem. The main point in this paper is to establish whether R_M, the robust result of increased GMST *in the future*, is confirmed. I deliberately selected a case in which there cannot be direct evidence for the robust result R_M to show that the confirmation must be based on indirect rather than direct evidence.

Climate models within an 'ensemble' (see e.g., Parker 2010b; 2013; Tebaldi and Knutti 2007; Knutti et al. 2010) differ in terms of their auxiliaries, of which the 'parameterizations' are the most important (see Gramelsberger 2010). They concern processes that cannot be explicitly represented, either because of their complexity (e.g., biochemical processes of vegetation) or because the discretised model equations in the simulations cannot resolve the spatial and/or temporal scales on which they occur (e.g., cloud formation and turbulence) (Flato and Marotzke 2013, p. 9). A_3 could thus

represent one way in which cloud formation is parameterized, and A_5 another; A_1 could represent one way in which vegetation is parameterized; A_2 could be one way in which turbulence is parameterized, and A_6 another, and so on. These auxiliaries are usually considered problematic because the parameterized factors are known to affect the climate. Indeed, the very idea of studying ensembles rather than single models in isolation is based, in part, on the realisation of their problematic nature. The direct disconfirming evidence for the auxiliaries is not explicitly represented in the diagrams, however, to avoid clutter.

The climate displays 'internal variability': the weather system is intrinsically chaotic. Climate models typically study 'forcing', in other words how changing various relevant factors affects the climate. Typical forcing factors include both anthropogenic (greenhouse gases such as CO2 and aerosols) and natural (solar and volcanic) elements.

Before proceeding to analyse Lloyd's model ensemble, let me briefly comment on the idealisations and abstractions used in depicting the various examples described in this paper. The point is to highlight the logic of robustness and indirect confirmation by stripping away various kinds of detail. The strategy is to see the difference that robustness makes by counterfactually investigating how the epistemic situation would change if a result were shown to be robust.

When climate modellers run computer simulations with general circulation models, they obtain a mass of results from a single simulation. This reflects the fact that they involve large numbers of auxiliaries and physical theories. A single model thus typically looks something like this:

$$M_{11}=(C\&A_1\&A_2\&A_3\&A_{11}\&A_{12}\&A_{13},\ldots) \vdash R_T, R_H, R_P, R_R,\ldots \qquad (5')$$

R_T could stand for the temperature in history for example, R_H the height of the tropopause, R_P patterns of precipitation, R_R pressure, and so on. If the model is also used to predict future weather, it also generates a set of predictions for various emission scenarios R_{C1T}, R_{C1H},…,R_{C2T}, R_{C2H}, and so on. When another model M_2 is simulated, the large number of results give ample opportunity to strengthen the robust theorem:

$$M_{12}=(C\&A_2\&A_4\&A_5\&A_{14}\&A_{15}\&A_{16},\ldots) \vdash R_T, R_H, R_P, R_R,\ldots \qquad (5'')$$

However, the results on a given variable are never numerically identical in two different models, and even a single model yields numerically different results with different initial conditions. Thus, establishing that a result derived from model M_1 counts the same as the corresponding result derived from M_2 requires a judgment concerning how large a divergence in results is acceptable. In what follows, complications concerning several emission scenarios and the numerical non-identity of results will be completely ignored, and the results from the various models will be assumed to be the same. The reason for using such an idealisation is that if the numerical difference is large enough, the results are no longer robust. However, if the results from climate models are not robust, then they obviously cannot be confirmed *by robustness*. Given that climate scientists rather than philosophers are in the best position to evaluate the degree to

which the results are indeed robust, I will leave it to them.[10] Yet, a philosophical account of the *consequences of* robustness cannot do otherwise but assume that the results are indeed robust.

As will soon become clear, the argument from the variety of evidence requires at least some model-specific results, and indeed, some results are not derived in every model. For example, Australians might derive results for surface seawater temperature in the Tasman sea, whereas Europeans might be concerned with the extent of summer sea ice in the Arctic.

In principle, one could conduct the analysis with a representation like this:

$$\begin{array}{ccccccccccccccccc} (M_1 \vdash & R_T, & R_H, & R_P, \ldots, & R_1), & (M_2 \vdash & R_T, & R_H, & R_P, \ldots, & R_2), & (M_3 \vdash & R_T, & R_H, & R_P, \ldots, & R_3) \\ \top & | & | & | & | & \top & | & | & | & | & \top & | & | & | & | \\ R_M & E_T & E_H & E_P & E_1 & R_M & E_T & E_H & E_P & E_2 & R_M & E_T & E_H & E_P & E_3 \end{array} \qquad (5E)$$

where, for example, E_T is the evidence on GMST in history, E_H for the height of the tropopause and E_P for patterns of precipitation; R_1 is a result concerning the surface seawater temperature in the Tasman sea and E_1 confirming evidence for it; R_2 a result concerning the extent of summer sea ice in the Arctic and E_2 the corresponding evidence, and so on.[11] However, the multiple robustness of the models with respect to R_T, R_H and R_P would merely confuse the reader at this point. I will thus depict three members from the family of models in Lloyd (2010) such that only model-specific results and evidence in favour of them are explicitly represented.

$$\begin{array}{ccccccc} (M_1 \vdash & R_1), & (M_2 \vdash & R_2), & (M_3 \vdash & R_3) \\ \top & | & \top & | & \top & | \\ R_M & E_1 & R_M & E_2 & R_M & E_3 \end{array} \qquad (6)$$

This representation of Lloyd's model ensemble abstracts from the robustness of confirmed results (R_T, R_H, R_P) to concentrate on the effects of evidence for model-specific confirmed results (R_1, R_2 and R_3) on the robust result about the future temperature R_M. The abstraction is made for expositional purposes only, and I will briefly return to the more realistic representation (5E) at the end of the paper.

Climate models are confirmed when their consequences are shown to fit with the various pieces of evidence. If such confirmation also accrues to the robust result R_M, it could be said that this result is indirectly confirmed by these pieces of evidence. Had R_M not been derived from several models, it obviously could not have been indirectly confirmed by the models of which it is not a consequence. Thus, had the modellers derived these results instead,

[10] Climate modellers appear to think that R_M is robust, however: '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).

[11] The history of climate science involves adding various elements to a model that becomes larger and larger (see Edwards 2010 for an extensive history of climate science). For example, the coupling of models of the sea and the climate was a major break-through. Modules for vegetation and sea ice, for example, were then added. One could thus also interpret (6) as the result of successive models. If M_1 had been the first model in time, M_2 the second, and so on, reality would have been described by $M_3 \vdash R_1, R_2, R_3$ and $M_2 \vdash R_1, R_2$ but $M_2 \nvdash$, and $M_1 \vdash R_1$ but $M_1 \nvdash R_2, R_3$.

$M_1 = (C\&A_1\&A_2\&A_3) \vdash R_M, R_1$
$M_2 = (C\&A_2\&A_4\&A_5) \vdash R_2, M_2 \nvdash R_M$ (7)
$M_3 = (C\&A_1\&A_6\&A_7) \vdash R_3, M_3 \nvdash R_M,$

E_1 would have confirmed R_M indirectly only through R_1, and the modellers would have known that E_2 and E_3 do not confirm R_M indirectly. Here, E_2 and E_3 indirectly confirm the assumption of a common core C more than the auxiliaries, but R_M is indirectly confirmed only by E_1.

As a weaker argument, had the climate modellers not known whether R_M could be derived from M_2 or M_3, they would only have known that E_1 indirectly confirmed R_M through R_1. In other words, if they only knew that

$$\begin{array}{cccccc} (M_1 & \vdash & R_1), & (M_2 & \vdash & R_2), \; (M_3 \vdash R_3), \\ | & & | & | & & | \\ R_M & & E_1 & E_2 & & E_3 \end{array} \qquad (8)$$

they would not know that E_2 and E_3 also indirectly confirmed R_M. Derivational robustness may thus broaden the range of empirical evidence that can be brought to bear on the robust result: only by showing that R_M is robust, in other words that it can be derived from M_2 (M_3), can one claim that E_2 (E_3) confirms R_M indirectly. Let $[E_i]c_i^r R_M$ denote the epistemic situation in which E_i indirectly confirms R_M, and $+[E_i, E_j]c_i^r R_M$ the epistemic situation in which E_j adds a variety of evidence for R_M. If modellers start with (8) and then derive (6), the epistemic situation changes from

$[E_1]c_i^r R_M$ to $+[E_1, E_2, E_3]c_i^r R_M$, (IVE)

and robustness confirms R_M because it Increases the Variety of Evidence (IVE) for it.

I will now scrutinise the logic of the overall argument more closely, referring to two counter-arguments that could be presented against my analysis.

3.2 The climate sceptics' argument and a variety of evidence

I have argued thus far that Laudan's general theory and the core assumptions in a family of models have similar functions in the sense that both may be employed in indirectly confirming results derived from them. As I have also shown, the example combines elements from both kinds (assumptions and results) of indirect confirmation. One might argue that combining the two kinds of indirect confirmation is not legitimate because, although E_2 and E_3 confirm some sets of assumptions, it is not clear exactly which assumptions are primarily responsible for the robust result, or whether they are the ones that are primarily confirmed by E_2 and E_3. Climate sceptics in particular, presumably, are willing to argue that the indirect confirmation from E_2 and E_3 should not be allocated primarily to C, or alternatively that since they take the robust theorem (cp, $C \vdash R_M$) to be false (and not confirmed by E_2 and E_3), the confirmation does not flow down to R_M.

Climate models do, of course, incorporate general physical theories such as the Navier-Stokes equations, but they are not in themselves sufficient for deriving R_1 or R_M.[12] In contrast, the inference from the theory of continental drift to magnetic alignment and the climate hypothesis involves auxiliary hypotheses, which are either trivially true (e.g., 'The climate on some piece of soil on the earth depends on its location') or already known to be rather well confirmed (e.g., 'Occasionally, lava erupts from volcanoes'). Hence, the general theory of continental drift rather than the auxiliaries is obviously responsible for the magnetic alignment and the changing climate.

To further investigate the role of the common core, let us consider the analysis that would have had to be conducted had there been no assumptions shared by all models. Assume that, instead of (7), the following models and results had been established.

$$\begin{aligned} &M'_1 = (A_1 \& A_2 \& A_3) \vdash R_M, R_1, \\ &M'_2 = (A_4 \& A_5 \& A_6) \vdash R_2, M'_2 \nvdash R_M \text{ and} \\ &M'_3 = (A_7 \& A_8 \& A_9) \vdash R_3, M'_3 \nvdash R_M. \end{aligned} \qquad (7\tilde{})$$

Now E_2 and E_3 could not indirectly confirm R_M because, given the way in which this epistemic situation is specified, the modellers would not have had any idea that the three models were connected to each other. If the robustness of R_M were now demonstrated they would end up with (6˜):

$$\begin{array}{ccccccccc} M'_1=(A_1 \& A_2 \& A_3) \vdash R_1, & & M'_2= (A_4 \& A_5 \& A_6) \vdash R_2, & & M'_3= (A_7 \& A_8 \& A_9) \vdash R_3 \\ \top & | & \top & | & \top & | \\ R_M & E_1 & R_M & E_2 & R_M & E_3 \end{array} \qquad (6\tilde{})$$

E_2 now indirectly confirms R_M in model M'_2, and E_3 indirectly confirms R_M in model M'_3. The argument from the variety of evidence for robustness thus holds even if the climate sceptics were right in arguing that the robust theorem is not confirmed. All this argument requires is at least some degree of independence of evidence (cf. Justus 2012; Kuorikoski and Marchionni 2016; Schupbach forthcoming). $(E_1 \& E_2) c_i R_M > E_1 c_i R_M$ only if the pieces of evidence are independent with respect to confirmation of the robust result (i.e. $(E_1|E_2) c_i R_M = E_1 c_i R_M$ and $(E_2|E_1) c_i R_M = E_2 c_i R_M$ (see Fitelson 2001)). The example of lava streaks and climate variability implies that such independence may be attained at least in some cases in which indirect confirmation is relevant.

The argument from the variety of evidence is sufficient to provide a convincing response to one specific argument against the idea that robustness may confirm. Many who argue against its confirmatory virtues admit that it would be confirmatory if the modeller could ascertain that he or she had tested for the robustness of each possible alternative auxiliary assumption, and if there were good grounds for thinking that the true assumption was among those that had been tried.[13] It is then pointed out that this

[12] The Navier-Stokes equations belong to what modellers often refer to as the 'physical core' of climate models. In this paper, however, the 'core' merely refers to CO2 forcing. Katzav (2013) argues that these equations cannot be confirmed because we know them to be true already (see Yablo 2014, p. 101 for a more general claim to this effect).

[13] See Houkes and Vaesen (2012), Odenbaugh (2011), Odenbaugh & Alexandrova (2011) and Woodward (2006). See Katzav (2013; 2014) for a version of this criticism that is specifically targeted at climate models. Kuorikoski, Lehtinen & Marchionni (2012) provide a rejoinder to Odenbaugh & Alexandrova's version of this argument.

is usually impossible. According to the critics, if the new models are just as idealised as the old ones, the robustness of the result does not resolve the worries because idealisations have merely been traded for idealisations. According to the most extreme form of this argument, the conclusion is that robustness can never confirm anything.

Given that *each* demonstration of robustness increases indirect confirmation in the example I discuss above, there is no need to ascertain that one has gone through all the possible alternatives. If one retains the earlier interpretations of the various auxiliaries instead of exhaustively studying all parameterizations, M'_1 and M'_2 merely test two different parameterizations for cloud formation A_3 and A_5, and two different parameterizations for turbulence A_2 and A_6. The confirmatory virtues of robustness are shown to be independent of whether the possible auxiliary assumptions can be exhaustively listed or whether the remaining assumptions are acknowledged to include known falsities, and the extreme form of the argument is shown to be untenable. The 'true' auxiliary does not need to be among those that have been tried because the point of robustness is to show their irrelevance rather than their truth.

There is a grain of truth in the critics' argument, however. Explaining what it is necessitates a return to the discussion on the indirect confirmation of the core. Let us recall what is written in Section 2.7 about the possibility of misidentifying assumptions. Insofar as being able to list all the possible auxiliaries helps in guaranteeing that assumptions are not being misidentified, robustness arguments are less prone to error in allocating confirmation, and the core assumptions may be more indirectly confirmed simply because robustness analysis is more complete. Similarly, knowing that one assumption is true may further help in identifying which assumptions drive the results, but otherwise the truth value of a demonstrably irrelevant assumption has no bearing at all on how robustness indirectly confirms (assumptions and/or results). Of course, such truth values do matter in the evaluation of the overall credibility of models, but as mentioned, this is a different matter, and one that is ignored in this paper. Finally, not being able to test every possible auxiliary with respect to robustness merely means that the robustness analysis is incomplete. Whether the modellers manage to list all the possible assumptions and test them for robustness, and whether the truth is among them are thus all relevant in terms of *how much* robustness may indirectly confirm. Unfortunately, however, the critics seem to have thought that such knowledge also constitutes a necessary condition for any epistemic benefit from robustness.

Had the critics formulated the argument by noting that robustness cannot conclusively show the truth, or even always establish a high probability of a result, they would have been right. Unfortunately, however, many of them have drawn the stronger conclusion that robustness cannot be used in an evaluative way at all, and have neglected the possibility that it may confirm in the sense of justifiably *increasing* the modellers' degree of belief in the truth of the result.

The most recent IPCC report confidently argues the following.

> The instrumental records associated with each element of the climate system are generally independent … and consequently joint interpretations across observations from the main components of the climate system increases the confidence to higher levels than from any single study or component of the climate system. The ability of climate models to replicate

> observed changes (to within internal variability) across a wide suite of climate indicators also builds confidence in the capacity of the models to simulate the earth's climate. (Bindoff and Stott 2013, p. 10-68)

Bindoff and Stott's argument, which is based on a wealth of evidence (see also Lloyd 2012), is relevant to confirming the robust prediction R_M because each confirming piece of evidence is relevant, even if it is impossible to check whether the true auxiliaries are among those that the climate modellers have tried, and even if climate models undoubtedly continue to contain false auxiliaries.

The more closely the models approach unanimity in their predictions about future climate change, the weaker is the case of the climate sceptic. Were climate models not robust with respect to this prediction[14], sceptics could accept results and evidence such as R_2, R_3, E_2, and E_3, but note that emissions policy does not need to change because there is nothing to show any potential effect on the future climate. On the other hand, if the real situation were described by something like (6˜) rather than (6), the sceptic could continue to oppose measures to control emissions because the connection between greenhouse gases and future temperatures would not have been established. Indeed, (6˜) raises the question of why R_M is a consequence of a set of dissimilar models (cf. Houkes and Vaesen 2012; Stegenga 2012). Thus, although the argument from the variety of evidence is cogent in showing that robustness may confirm, it is insufficient in itself to justify a strict emissions policy.

Given that many results have already been derived from earlier models, finding entirely new indirect evidence may be a rare occurrence. Consequently, the argument from the variety of evidence *alone* may not be empirically very important. Furthermore, given that there is no core in (6˜), let alone a general theory T as in (1), one might argue that the indirect confirmation of results is very weak: (6˜) raises a question rather than providing a solid argument for indirect confirmation. Therefore, if the argument from the variety of evidence is stripped to its bare bones as in (6˜), it is not a very strong argument for the confirmatory virtues of robustness. However, it continues to be relevant because, as I show in the next section, a given demonstration of robustness may well strengthen the indirect confirmation of a robust result and increase the relevant evidence simultaneously. Then the argument from the variety of evidence is reinforced by the strengthening of the robust theorem, and it depends on the epistemic situation (8) rather than the problematic (6˜).

The climate sceptics' argument gains credibility if there is a plausible theoretical alternative to C, in other words if something other than greenhouse gases is responsible for the confirmed results R_1, R_2, and R_3. Take, for example, the alternative theory T' that the observed temperature increase (E_1) may be attributable to changes in the radiation activity of the sun S (or alternatively, to internal variability). Given that the radiation hypothesis also needs auxiliaries to derive R_1, what could be available is something like model $M_S = (S\&A_7\&A_8 \vdash R_1)$.

Indirect confirmation through robustness does not necessarily mean that modellers are able to pinpoint one assumption that is responsible for the robust result. Climate models

[14] As I have shown (see fn. 9), there are slightly different estimates of the sensibility of the climate to different forms of forcing.

always include radiation from the sun S and a host of other variables - U (aerosols), V (volcanoes), and so on - meaning that they have thus far been represented in an extremely simplified way. Indeed, as noted already, the mechanism responsible for the increase in temperature involves radiation from the sun. Suppose that instead of (5), the following results are available:

$$\left.\begin{array}{l} M_{11} = (C\&S\&U\&V\&A_1\&A_2\&A_3) \\ M_{12} = (C\&S\&U\&V\&A_2\&A_4\&A_5) \\ M_{13} = (C\&S\&U\&V\&A_1\&A_6\&A_7) \end{array}\right\vdash R_M, \qquad (5')$$

E_1, E_2, and E_3 would continue to indirectly confirm R_M if an equivalent of (6) were to hold for such models:

$$\begin{array}{ccccccc} (M_{11} \vdash R_1), & & (M_{12} \vdash R_2), & & (M_{13} \vdash R_3) \\ \top & | & \top & | & \top & | \\ R_M & E_1 & R_M & E_2 & R_M & E_3. \end{array} \qquad (6')$$

The conjunction C&S&U&V would be indirectly confirmed just as C was in the above example. Such a lack of independence between the models would mean, however, that although some auxiliary hypotheses could be ruled out as causes of global warming, it would still not be clear exactly what caused it because many candidates would remain. If this were how things stood, nobody would know what was responsible for R_M merely from looking at the data and the derivational relationships. These observations together with the huge number of assumptions in global climate models could help to explain why modellers may justifiably continue to think that the absolute confirmation of R_M is not very strong.

Let us consider a pessimistic scenario. The modellers' background knowledge indicates that all models in an ensemble omit or very poorly represent a process they believe is important for accurately predicting the value R_M. Suppose also that all the models in the ensemble give the same prediction R_M, and that they manage to yield correct results on other aspects of the system, as in (6'). Now, however, each model also contains a large number of auxiliaries that are known to be false, and that are included in every model: A_{x1}, A_{x2},…A_{xn}.

$$\left.\begin{array}{l} M_{11} = (C\&S\&U\&V\&A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}\&A_1\&A_2\&A_3) \\ M_{12} = (C\&S\&U\&V\&A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}\&A_2\&A_4\&A_5) \\ M_{13} = (C\&S\&U\&V\&A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}\&A_1\&A_6\&A_7) \end{array}\right\vdash R_M, R_1 \qquad (5^x)$$

R_M continues to be indirectly confirmed by the other results. Now, however, robustness analysis is very far from being complete. Obviously, the further it is from being complete, the more likely it is that the inference that the confirmed results R_1, R_2 and R_3 indirectly confirm R_M is erroneous: if the modellers find it plausible that some assumption in $A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}$ could be responsible for R_M, they will think that R_M is very weakly confirmed by R_1, R_2 and R_3. Thus, if they have strong reasons to believe that some assumptions in $A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}$ rather than C is responsible for R_M, they may think that R_M is rather weakly confirmed in an *absolute* sense. They can learn more by deriving further results with and without the various components. If they are right about some of $A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}$ being responsible for R_M, they should be able to derive results that indicate this kind of relationship. Something like this, perhaps:

$(C\&S\&U\&V\&A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}\&A_3) \vdash R_M, R_1$
$(C\&S\&U\&V\&A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}\&A_1) \vdash R_M, \nvdash R_1$
$(C\&S\&U\&V\&A_{x1}\&A_{x2}\&,\ldots,\&A_{xn}\&A_2) \vdash R_M, \nvdash R_2,$

Given that global models are huge, they carry an enormous number of assumptions. Hence, any judgment concerning how strongly R_M is confirmed in an absolute sense requires an enormous amount of information on the assumptions needed for deriving various results, and on the empirical evidence of those results. I have almost nothing to say in this paper about how strongly R_M is confirmed in such an absolute sense. Climate scientists may judge how many false assumptions enter all the models, and how seriously false they are, for example. I only show how robustness is used in the modelling process.

Suppose now that modellers show R_M to be a robust result (5), and that the evidence indirectly confirms this, even though the alternative theory T': $S \vdash R_1$ is true and the anthropogenic hypothesis ($C \vdash R_M$ and $C \vdash R_1$) is false. The false theory could survive for a while, but sooner or later the modellers would be able to construct models that do not include the core C but would predict the data E_1, E_2, and E_3. As a matter of fact, however, climate modellers have repeatedly tested models with and without anthropogenic forcing, and have found that natural (solar and volcanic) forcing alone is incapable of explaining the observed temperature rise between 1860 and 2010.[15] The real situation could thus be depicted as follows.[16]

$(S\&V\&A_1\&A_2\&A_3) \nvdash R_1$
$(S\&V\&A_2\&A_4\&A_5) \nvdash R_1$ (5'')
$(S\&V\&A_1\&A_6\&A_7) \nvdash R_1.$

Taking aerosols into account yields an even worse fit because their net effect is commonly taken to be cooling. M_S should also have some independent evidence that cannot be counted in favour of the anthropogenic hypothesis. In other words, if (6) has already been established, the alternative theory T' should be able to explain not just E_2, and E_3, but also some further results - E_4, E_5 and so on - which the anthropogenic hypothesis cannot explain.

Given that Lloyd (2015) discusses confirmatory robustness in climate models, I am obliged to discuss the similarities with and differences from her account. The similarity lies in the fact that the robust result R_M is indirectly confirmed because it is the joint consequence of the common core structure C in a family of models, and other consequences of this family are confirmed. However, she emphasises the direct support

[15] See e.g., Bindoff and Stott (2013, FAQ 10.1). The report uses these terms to discuss the possibility that another alternative account might explain the observed global warming, the idea that internal variability alone is sufficient: '…we conclude that it is *virtually certain* that internal variability alone cannot account for the observed global warming since 1951' (p. 22). See Parker (2010a) for a philosophical analysis of fingerprint results from attribution studies that derive results like (5'').

[16] Despite such results, the attribution to greenhouse gases is not perfect because climate-simulation models have hundreds of thousands of lines of computer code, and some parts of it have remained the same for decades. Insofar as all the code is not checked, it is still possible in principle that errors in it could generate the confirmed results.

for the various individual assumptions rather than the results derivable from the models. Recall that A_3 and A_5 denote different parameterizations of cloud formation. Let E_{A3} and E_{A5} denote directly confirming pieces of evidence for them. If one disregards the fact that Lloyd also considers results derived from the whole model, her argument could be presented as follows: Suppose that modellers start with (L):

$$\begin{array}{lll} M_1=(C\&A_1\&A_2\&A_3), & M_2=(C\&A_2\&A_4\&A_5). & \text{(L)} \\ \top & \top & \top \\ R_M & E_{A3} & E_{A5} \end{array}$$

and then derive (L')

$$\begin{array}{llll} M_1=(C\&A_1\&A_2\&A_3), & M_2=(C\&A_2\&A_4\&A_5). & & \text{(L')} \\ \top & \top \quad \top & & \top \\ R_M & E_{A3} \quad R_M & & E_{A5} \end{array}$$

Insofar as A_3 (A_5) alone is sufficient for deriving E_{A3} (E_{A5}), this piece of evidence could not indirectly confirm C, but it can indirectly confirm R_M. Deriving R_M from M_2 does indeed indirectly confirm R_M because it shows that a new piece of evidence (E_{A5}) is relevant for R_M. Note that in the epistemic situation (8),

$$\begin{array}{llllll} \mathbf{M_1=(C}\&A_1\&A_2\&A_3) \vdash R_1, & M_2=(\mathbf{C}\&A_2\&A_4\&A_5) \vdash \mathbf{R_2}, & M_3=(\mathbf{C}\&A_1\&A_6\&A_7) \vdash R_3, & \\ \top & | & \mathbf{|} & | \quad (8) \\ \mathbf{R_M} & E_1 & \mathbf{E_2} & E_3 \end{array}$$

the structure of indirect confirmation already exists, albeit in a very weak form: there is a link (in bold) from E_2 to C in model M_2 and a link from C to R_M in model M_1. In contrast, (L') literally creates the structure for indirect confirmation (of results) precisely because E_{A5} does not indirectly confirm C. In principle, then, there is Lloydian confirmation from directly supported auxiliaries to the robust result.

However, if there is a derivation that shows R_M to be derivable without A_i, it shows that A_i cannot be needed for deriving R_M (this will be denoted as $A_i \nvdash_c R_M$). Alternative parameterizations such as A_3 and A_5 are usually assumed to be mutually incompatible assumptions (Parker 2006). This diminishes the force of Lloyd's account considerably because such assumptions cannot be necessary for deriving R_M, either. This, in turn, implies $A_3 \nvdash_c R_M$ and $A_5 \nvdash_c R_M$, and consequently, neither of them confirms R_M.

On the other hand, I do not see any particular reason why one should restrict an account of confirmatory robustness to cases in which the auxiliaries are incompatible: direct evidence supporting any assumption that always takes part in deriving R_M indirectly confirms R_M. Similarly, direct evidence that is contrary to any assumption that always takes part in deriving R_M indirectly disconfirms R_M. This is just stating the obvious fact that the overall credibility of R_M depends on the truth of all the assumptions of climate models.

3.3 Strengthening the robust theorem

Whereas climate critics must be able to show why each climate model yields the prediction of a future increase in temperature (which they think is erroneous), it may be

obvious to someone who already accepts the anthropogenic hypothesis that all climate models incorporating greenhouse-gas forcing imply this prediction (cp, $C \vdash R_M$). If greenhouse gases increase the absorption of longwave radiation, and if this is responsible for the increase in temperature, then it should be responsible for such increases in the future as well.

Another possible counter-argument is thus that robustness is not really needed for indirectly confirming R_M because E_2 and E_3 indirectly confirm it irrespective of whether it is a robust result or only derivable from M_1. This is because evidence E_2 and E_3 indirectly confirms C even in diagram (8), and E_2 and E_3 thus indirectly confirm R_M as soon as one shows it to be derivable from M_1 (but of course, (7) cannot hold). In contrast to the previous counter-argument, this one presupposes that it is possible to allocate the indirect confirmation from E_2 and E_3 to C rather than the auxiliary assumptions, even without robustness.

This argument rests on the assumption that one already knows that C implies R_M. Let us consider what would happen if the interpretations of C and R_M were modified such that C *alone* entailed R_M. For example, if the modellers first knew

$$\begin{array}{l} M_1=(C\&A_1\&A_2\&A_3) \vdash R_1, \\ \quad\top \qquad\qquad\quad | \\ \quad R_M \qquad\qquad E_1 \end{array} \tag{9}$$

and then derived

$$\begin{array}{l} M_2=(C\&A_2\&A_4\&A_5) \vdash R_2, \\ \quad\top \qquad\qquad\quad | \\ \quad R_M \qquad\qquad E_2 \end{array} \tag{9'}$$

E_2 would increase the indirect confirmation of R_M because the epistemic situation would change from

$$E_1\ c_i^a\ C \quad \top_{R_M} \qquad \text{to} \qquad +[E_1,E_2]\ c_i^a\ C. \quad \top_{R_M} \tag{10}$$

(10) shows that (9') increases the indirect confirmation of R_M, but this increase is to be attributed solely to the increase in the relevant evidence. It is thus not the robustness that changes the epistemic situation and confirms here, but rather deriving the new result R_2. This counterfactual example shows that robustness is not necessary for increasing the variety of evidence.

What, exactly, is the role of the robustness of R_M in the inference that E_2 and E_3 indirectly confirm it in (6)? I have described climate modellers as if they first demonstrated the robustness of R_M and then moved from the epistemic situation (5) to (6) by way of deriving the new results R_2 and R_3 and finding evidence E_2 and E_3 for them. Let us consider how the example would have had to be analysed if the order of deriving various results had been different. Suppose that the modellers were only concerned at first with modelling the historical development of the climate. They would have started by deriving

$M_1=(C\&A_1\&A_2\&A_3) \vdash R_1.$ (8a)
$\quad\quad\quad

be derived without C. Finally, deriving R_M from M_3 establishes a further increase in indirect result confirmation: $A_2 \not\vdash_c R_M$, and

$$\begin{array}{ccccccc} M_1=(C\&A_1\&A_2\&A_3) & \vdash & R_1, & M_2=(C\&A_2\&A_4\&A_5) & \vdash & R_2, & M_3=(C\&A_1\&A_6\&A_7) \ \vdash R_3 \\ \quad +\top_c & & | & \quad +\top_c & & | & \quad +\top_c \qquad\qquad | \ (6) \\ \quad R_M & & E_1 & \quad R_M & & E_2 & \quad R_M \qquad\qquad E_3 \end{array}$$

In short, this derivation establishes that the previously established indirect assumption confirmation from E_1, E_2, and E_3 to C also accrues to the robust result R_M.

I have already shown that if the three models are entirely independent of each other as in (6˜), derivational robustness confirms merely through increasing the variety of evidence but not through strengthening the robust theorem. C is more strongly confirmed indirectly in (6) than any of the assumptions in (6˜) because each of E_1, E_2, and E_3 indirectly confirms it. Of course, this indirect *assumption* confirmation has nothing to do with the robustness of the result R_M. Yet, deriving R_M from M_2 and M_3 shows that the core assumptions C rather than some auxiliary assumptions A_i are primarily responsible for the robust result R_M. (6) confirms R_M more strongly than (8) because it makes the connection between the robust result and the confirmed core 'tighter' in showing that (7) is no longer possible. In short, (6) confirms R_M through strengthening the robust theorem.

In contrast, if one starts with

$$\begin{array}{ccccccc} M'_1=(A_1\&A_2\&A_3) & \vdash & R_1, & M'_2=(A_4\&A_5\&A_6) & \vdash & R_2, & M'_3=(A_7\&A_8\&A_9) \vdash R_3, \qquad (6\tilde{}\tilde{}) \\ \top & & | & & & | & \qquad\qquad | \\ R_M & & E_1 & & & E_2 & \qquad\qquad E_3 \end{array}$$

and then derives (6˜)

$$\begin{array}{ccccccc} M'_1=(A_1\&A_2\&A_3) & \vdash & R_1, & M'_2=(A_4\&A_5\&A_6) & \vdash & R_2, & M'_3=(A_7\&A_8\&A_9) \vdash R_3, \qquad (6\tilde{}) \\ \top & & | \ \top & & & | \ \top & \qquad\qquad | \\ R_M & & E_1 \quad R_M & & & E_2 \quad R_M & \qquad\qquad E_3 \end{array}$$

robustness does not strengthen a robust theorem because there is none, but it shows how new evidence is relevant. Given the difference between (6) and (6˜), one might want to argue that moving from (8) to (6) does not provide a case with new relevant evidence because the structure for indirect confirmation already exists: recall that E_2 and E_3 indirectly confirm assumption C, and the link from C to R_M exists, even before R_M is shown to be robust:

$$\begin{array}{ccccccc} \mathbf{M_1=(C}\&A_1\&A_2\&A_3) & \vdash & R_1, & M_2=(\mathbf{C}\&A_2\&A_4\&A_5) & \vdash & \mathbf{R_2}, & M_3=(\mathbf{C}\&A_1\&A_6\&A_7) \ \vdash \mathbf{R_3} \\ \top & & | & & & \mathbf{|} & \qquad\qquad \mathbf{|} \ (8) \\ \mathbf{R_M} & & E_1 & & & \mathbf{E_2} & \qquad\qquad \mathbf{E_3} \end{array}$$

However, one could imagine circumstances in which the link between C and R_2 or R_3 is far from obvious when modellers face an epistemic situation such as (8). It may then come as a genuine surprise that E_2 and E_3 can be brought to bear on R_M. In such cases, it seems acceptable to say that robustness increases the set of relevant pieces of evidence for a result, even though the structure for indirect confirmation already exists

before its robustness is established: if the links constituting the indirect confirmation structure (e.g., R_M-C-R_2-E_2 in 8) are so weak that the modellers do not recognise them, then deriving (6) after (8) does increase the variety of evidence. Whether the effect of robustness is to be conceptualised as an increase in the variety of evidence or as a strengthening of the robust theorem depends on the empirical details of the case, and on the individual modellers' epistemic situation: recall that climate sceptics believe the C-R_M link is very weak, whereas most IPCC reports give the impression that it is rather strong. This means that climate sceptics would conceptualise this example in terms of increasing the variety of evidence, and if they were to remain genuinely unconvinced about anthropogenic climate change they could claim, for example, that the real situation is closest to (6').

On the other hand, insofar as one thinks that the original epistemic situation (8) does not already create the structure for indirect confirmation, robustness confirms R_M via two routes in (6): by strengthening the robust theorem and increasing the relevant evidence. Pirtle et al. (2010, p. 355) seem to take the fact that 'some features, such as forcing from the main greenhouse gases are shared across all models' as an argument against the confirmatory power of robustness. If I am right about why and how robustness confirms, such a lack of independence is necessary for the second kind of indirect confirmation through strengthening the robust theorem, and it reinforces rather than weakens the indirect confirmation of the robust result.

Robustness could confirm even if there were no new relevant data. To explain this more clearly, I will consider a variation of the example in which there is only old evidence. Here, unlike in the previous example, robustness increases the indirect assumption confirmation. Suppose now that the modellers start with $M_1 \vdash R_M$. They then derive (8a')

$$\underset{\top}{M_1=(C\&A_1\&A_2\&A_3) \vdash R_1}, \qquad \underset{|}{} \qquad (8a')$$
$$R_M \qquad E_1$$

and (8b')

$$M_2=(C\&A_2\&A_4\&A_5) \vdash \underset{|}{R_1}. \qquad (8b')$$
$$E_1$$

They then show the robustness of R_M by deriving $M_2 \vdash R_M$ so that they end up with (8').

$$\underset{+\top_c \; +\top_c \quad | \quad +\top_c \; +\top_c \quad |}{M_1=(C\&A_1\&A_2\&A_3) \vdash R_1, \; M_2=(C\&A_2\&A_4\&A_5) \vdash R_1} \qquad (8')$$
$$R_M \quad R_M \quad E_1 \quad R_M \; R_M \quad E_1$$

The robustness of R_M now strengthens the indirect confirmation of R_M by E_1 by strengthening the robust theorem ($C+ \vdash_c R_M$). The robustness of R_1 strengthens the *indirect assumption confirmation of the core*, (i.e., the link between C and the confirmed result R_1: $C+ \vdash_c R_1$), and thereby also strengthens the indirect confirmation of R_M by E_1. The core can thus be assumption-confirmed by deriving previously confirmed results from it as in (8), and by robustness as in (8').

The new epistemic situation could also be described as follows:

$$\begin{array}{l}(C\&A_2)+ \vdash_c R_1. \\ \quad {}^{+}\top^{c} \quad\ \ | \\ \quad R_M \quad\ \ E_1\end{array} \qquad (11)$$

It is also possible to recall SICC and SRT from section 2.7 and write:

$P(R_M|(C\&A_2)+ \vdash_c R_M\&(C\&A_2)+ \vdash_c R_1) > P(R_M)$.

Orzack and Sober (1993) argue that the robustness of a directly confirmed result (say R_1) does not add any confirmation to *this* result over and above what the direct evidence (here E_1) already confers. They are right, but robustness may increase the indirect assumption-confirmation of the core and thereby increase the indirect confirmation of *another* result (R_M) if it, too, is robust (i.e., if it can be derived from both M_1 and M_2). Let us thus consider another counterfactual order of deriving results to show that demonstrating the robustness of a result for which there is direct confirming evidence may also confirm. Suppose that modellers start with (8a').

$$\begin{array}{l}M_1=(C\&A_1\&A_2\&A_3) \vdash R_1. \\ \quad\ \top \quad\ \ | \\ \quad R_M \quad\ \ E_1\end{array} \qquad (8a')$$

Deriving (2''*)

$$M_2= (C\&A_2\&A_4\&A_5) \vdash R_M \qquad (2''*)$$

confirms R_M only insofar as the assumptions needed to derive R_1 are also needed to derive R_M. Then, deriving R_1 from M_2 also leads us to (8'). This derivation confirms R_M by making it more likely that C is really needed for deriving R_1. Note that deriving R_1 from M_2 increases the indirect assumption confirmation of the core (i.e., $E_1 + c_i^a C$) when (8a') and (2''*) are already available, but this increase would not confirm R_M if the robust theorem had not been strengthened earlier by demonstrating the robustness of R_M. Here, demonstrating the robustness of a directly confirmed result R_1 indirectly confirms *another* result R_M, but if and only if that other result is robust as well.

This observation gives good reason to briefly revisit the more realistic representation of climate-model ensembles, one in which they generate large numbers of results:

$$\begin{array}{l}(M_1 \vdash R_T,R_H,R_P,\ldots,R_1),\ (M_2 \vdash R_T,R_H,R_P,\ldots,R_2),\ (M_3 \vdash R_T,R_H,R_P,\ldots,R_3), \\ \top \quad | \ | \ | \quad | \quad \top \quad | \ | \ | \quad | \quad \top \quad | \ | \ | \quad | \quad (5E) \\ R_M \quad E_T\ E_H\ E_P \quad E_1 \quad R_M \quad E_T\ E_H\ E_P \quad E_2 \quad R_M \quad E_T\ E_H\ E_P \quad E_3\end{array}$$

Insofar as the confirmed results really are *multiply robust* in this way, they all contribute to indirectly confirming R_M.

Strengthening the robust theorem may increase the indirect confirmation of R_M from a given piece of evidence irrespective of whether the evidence has been made relevant (i.e., new) by demonstrating the robustness of R_M, or whether it was known to be relevant (i.e., old) beforehand. The variety of evidence can be increased without

strengthening the robust theorem, and the robust theorem may be strengthened without adding relevant data. Thus, although both arguments seem to apply to some climate models, they are *almost* independent of each other. The reason why this independence claim must be qualified is the difference between (L') and (8) I noted earlier: (L') literally creates the structure for indirect result confirmation, whereas it exists in a weak form in (8).

If increasing the variety of evidence occurs without a core, it is very weak. However, as I have shown, it may occur simultaneously with strengthening the robust theorem. Indeed, the latest IPCC report describes the situation as follows:

> Human influence has been detected in the major assessed components of the climate system. Taken together, the *combined evidence increases the level of confidence in the attribution of observed climate change*, and reduces the uncertainties associated with assessment based on a single climate variable.[17] From this combined evidence it is virtually certain that human influence has warmed the global climate system. (Bindoff and Stott 2013, p. 7, modified emphasis)

If CO2 increases in the atmosphere are attributable only to human influence, Bindoff and Stott's claim is that C is indirectly confirmed not merely by E_1 but also by E_T, E_H, E_P, E_2, E_3 and so on. They thus argue that the variety of evidence increases confidence in the hypothesis that C is the cause of R_1. However, I have shown that a variety of evidence can be increased without robustness. Is this the case here? If robustness were not involved, for example, one could describe the claim as follows.

$$\begin{array}{ccccccccc} M_1=(C\&A_1\&A_2\&A_3) & \vdash & R_1, & M_2=(C\&A_2\&A_4\&A_5) & \vdash & R_2, & M_3=(C\&A_1\&A_6\&A_7) & \vdash & R_3 \\ & \top & & & & & & & \\ & R_T & & & & & & & \end{array} \quad (12)$$

and

$$\begin{array}{ccccccccc} M_1=(C\&A_1\&A_2\&A_3) & \vdash & R_1, & M_2=(A_2\&A_4\&A_5) & \nvdash & R_2, & M_3=(A_1\&A_6\&A_7) & \nvdash & R_3 \\ & \top & & & & & & & \\ & R_T & & & & & & & \end{array} \quad (12')$$

It is easy to see that although C is indirectly confirmed by E_2 and E_3 in such attribution studies, R_T is not, unless it is a robust consequence of the models:

$$\begin{array}{ccccccccc} M_1=(C\&A_1\&A_2\&A_3) & \vdash & R_1, & M_2=(C\&A_2\&A_4\&A_5) & \vdash & R_2, & M_3=(C\&A_1\&A_6\&A_7) & \vdash & R_3 \\ +\top_c & & & +\top_c & & & +\top_c & & \\ R_T & & & R_T & & & R_T & & \end{array} \quad |(12^{bs})$$

Here $[E_2,E_3]c_i^r R_T$, just as Bindoff and Stott claim. In other words, Bindoff and Stott combine the argument from the variety of evidence with the argument from strengthening the indirect confirmation of the core and of the robust theorem.

[17] They thus argue that variety of evidence allocates the confirmation to C.

4. Conclusions

Even though derivational robustness analysis is usually considered a non-empirical research strategy, I have attempted to develop an account of the way in which it could be used together with empirical data to increase the confirmation of model results, and provide an example in which confirmation occurs via robustness: directly confirmed results for climate-change models indirectly confirm their robust prediction of increased temperatures in the future because the confirmed results can only be generated from greenhouse-gas forcing. However, although the analysis in this paper shows *how* this robust result is indirectly confirmed, the exact strength with which the robust prediction is confirmed is best left for climatologists to determine.

Robustness may be necessary for establishing that some data confirm indirectly, and it may increase the weight of the existing indirect confirmatory evidence for the robust result. It may thus increase the variety of evidence or the weight of existing confirmatory evidence, depending on the modellers' epistemic situation. The robustness of a directly confirmed result may increase the indirect assumption confirmation of the core because it allows for the contrastive distribution of the confirmation provided by a given piece of evidence to individual assumptions. By the same token, it may show that the core rather than the auxiliaries is necessary to derive the robust result.

However, robustness does not confirm anything 'by itself': only empirical evidence can do this. Hence, strengthening the robust theorem may always increase the modellers' confidence in it, but confirms it only if it changes the epistemic situation in the right way. The epistemic situation must always include a robust result, and another result for which there is direct evidence that is transferred to the robust result as indirect evidence. Increasing the weight of existing evidence must also involve strengthening the robust theorem or increasing the indirect confirmation of the core.

I have also identified various possible features of the epistemic situations that weaken or remove the indirect confirmation that robustness confers. This may happen specifically if there is no core, if there is no empirical evidence or if the derivations fail to allocate confirmation to the core. More generally, it could occur if there is the kind of content-cutting that destroys the indirect confirmation relation, if the core alone is already known to imply the robust result, if the robust result is already known to be true, if the set of common assumptions is so large that it remains difficult to establish the robust theorem, or if modellers misidentify assumptions. However, it would be difficult even to formulate let alone understand these qualifications without an account of why and how robustness may confirm.

Although the degree to which modellers can test possible auxiliaries for robustness affects the probability of misidentification and thereby possibly the completeness of robustness, knowing that the 'true' assumption is among the tested ones and being able to test all possible auxiliaries are not necessary conditions for confirmation through robustness. If robustness confirms indirectly in the way I claim it does, there may well be an increase in confirmation even when only some auxiliaries have been tried. However, the larger the set of auxiliaries that remain common to all derivations of a result, the higher is the risk of their being responsible for the results, and of misidentifying assumptions. Misidentifying assumptions is different from the other qualifications because, by assumption, modellers do not recognise such failures. They

are in the position to evaluate the degree to which the other qualifications are relevant in any specific case, but failing to identify common assumptions in models means that they have systematically inflated confidence in allocating assumption confirmation to the core: they will think that the core is excessively assumption-confirmed.

Acknowledgements

For their comments on earlier drafts of this paper and discussions about these issues, I would like to thank Till Grüne-Yanoff, Jaakko Kuorikoski, Chiara Lisciandra, Caterina Marchionni, Ilkka Niiniluoto, Jani Raerinne, Jouni Räisänen, Jonah Schupbach, Jacob Stegenga, the anonymous reviewers, and participants in the following conference series: PSA, Models and Simulations, INEM, and a workshop on Robustness in Helsinki 2014. Given the complexity of the issues, the present account may still suffer from various weaknesses for which they could not be held responsible.

References

Achinstein, P. (2001). *The book of evidence.* (New York; Oxford: Oxford University Press)

Bangu, S. (2006). Underdetermination and the argument from indirect confirmation. *Ratio: An International Journal of Analytic Philosophy*, 19, 269-277.

Bindoff, N. L. & Stott, P. A. (2013). Detection and attribution of climate change: from global to regional. (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)* (pp. 1-132). Cambridge: Cambridge University Press.)

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory.* (Cambridge, Mass ; London: MIT Press)

Edwards, P. N. (2010). *A vast machine: computer models, climate data, and the politics of global warming.* (Cambridge, Mass. ; London: MIT Press)

Fitelson, B. (2001). A Bayesian account of independent evidence with applications. *Philosophy of Science*, 68, 123-140.

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.

Friedman, M. (1953). The methodology of positive economics. *Essays in positive economics* (pp. 3-43). Chicago: University of Chicago Press.)

Garber, D. (1983). Old evidence and logical omniscience in Bayesian confirmation theory. (In J. Earman (ed.) *Testing scientific theories* (pp. 99-132). Minnesota: Minnesota University Press.)

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

Gemes, K. (1994). A new theory of content I: Basic content. *Journal of Philosophical Logic*, 23, 595-620.

Gemes, K. (2005). Hypothetico-deductivism: Incomplete but not hopeless. *Erkenntnis*, 63, 139-147.

Gemes, K. (undated). Carnap-confirmation, content-cutting & real confirmation manuscript, Birkbeck College, Oxford:

Glymour, C. N. (1980). *Theory and evidence.* (Princeton: Princeton University Press)
Glymour, C. N. (1983). Discussion: Hypothetico-deductivism is hopeless. *Philosophy of Science,* 72, 322-325.
Gramelsberger, G. (2010). Conceiving processes in atmospheric models—general equations, subscale parameterizations, and 'superparameterizations'. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics,* 41, 233-241.
Guillemot, H. (2010). Connections between simulations and observation in climate computer modeling. Scientist's practices and "bottom-up epistemology" lessons. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics,* 41, 242-252.
Hacking, I. (1967). Slightly more realistic personal probability. *Philosophy of Science,* 34, pp. 311-325.
Hands, D. W. (2016). Derivational robustness, credible substitute systems and mathematical economic models: the case of stability analysis in Walrasian general equilibrium theory. *European Journal for Philosophy of Science,* 6, 31-53.
Hartmann, S. & Fitelson, B. (2015). A new Garber-style solution to the problem of old evidence. *Philosophy of Science,* 82, 712-717.
Hempel, C. G. (1945). Studies in the logic of confirmation (II.). *Mind,* LIV, 97-121.
Hempel, C. G. (1965). *Aspects of scientific explanation, and other essays in the philosophy of science.* (New York: Free Press)
Houkes, V. & Vaesen, K. (2012). Robust! - Handle with care. *Philosophy of Science,* 79, 345-364.
Howson, C. (1991). The 'old evidence' problem. *The British Journal for the Philosophy of Science,* 42, 547-555.
Jeffrey, R. C. (1983). Bayesianism with a human face. (In J. Earman (ed.) *Testing scientific theories* (pp. 133-156). Minneapolis: University of Minnesota Press.)
Justus, J. (2012). The elusive basis of inferential robustness. *Philosophy of Science,* 79, 795-807.
Katzav, J. (2013). Hybrid models, climate models, and inference to the best explanation. *The British Journal for the Philosophy of Science,* 64, 107-129.
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.
Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate,* 23, 2739-2758.
Knuuttila, T. & Loettgers, A. (2011). Causal isolation robustness analysis: the combinatorial strategy of circadian clock research. *Biology and Philosophy,* 26, 773-791.
Kuorikoski, J. & Lehtinen, A. (2009). Incredible worlds, credible results. *Erkenntnis,* 70, 119-131.
Kuorikoski, J., Lehtinen, A. & Marchionni, C. (2010). Economic modelling as robustness analysis. *British Journal for the Philosophy of Science,* 61, 541-567.
Kuorikoski, J., Lehtinen, A. & Marchionni, C. (2012). Robustness analysis disclaimer: Please read the manual before use! *Biology and Philosophy,* 27, 891-902.
Kuorikoski, J. & Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science,* 83, 227-247.

Laudan, L. (1996). *Beyond positivism and relativism: theory, method, and evidence.* (Boulder Colorado: Westview)
Laudan, L. & Leplin, J. (1991). Empirical equivalence and underdetermination. *The Journal of Philosophy,* 88, 449-472.
Lehtinen, A. (2016). Allocating confirmation with derivational robustness. *Philosophical Studies,* 173, 2487-2509.
Levins, R. (1966). The strategy of model building in population biology. *American Scientist,* 54, 421-431.
Levins, R. (1993). A response to Orzack and Sober: Formal analysis and the fluidity of science. *Quarterly Review of Biology,* 68, 547-555.
Lisciandra, C. (2017). Robustness analysis and tractability in modeling. *European Journal for Philosophy of Science,* 7, 79-95.
Lloyd, E. A. (2009). Varieties of support and confirmation of climate models. *Proceedings of the Aristotelian Society,* LXXXIII, 213-232.
Lloyd, E. A. (2010). Confirmation and robustness of climate models. *Philosophy of Science,* 77, 971-984.
Lloyd, E. A. (2012). The role of complex empiricism in the debates about satellite data and climate models. *Studies in History and Philosophy of Science,* 43, 390-401.
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.
Machlup, F. (1955). The problem of verification in economics. *Southern Economic Journal,* 22, 1-21.
Machlup, F. (1956). Rejoinder to a reluctant ultra-empiricist. *Southern Economic Journal,* 22, 483-493.
Nagel, E. (1961a). *The structure of science: problems in the logic of scientific explanation.* (London: Routledge & Kegan Paul)
Nagel, E. (1961b). *The structure of science: Problems in logic of scientific explanation.* (New York: Hartcourt, Brace & World)
Niiniluoto, I. (1983). Novel facts and Bayesianism. *The British Journal for the Philosophy of Science,* 34, 375-379.
Niiniluoto, I. and Tuomela, R. (1973). *Theoretical concepts and hypothetico-inductive inference.* (Dordrecht: D. Reidel)
Odenbaugh, J. (2011). True lies: Realism, robustness, and models. *Philosophy of Science,* 78, 1177-1188.
Odenbaugh, J. & Alexandrova, A. (2011). Buyer beware: robustness analyses in economics and biology. *Biology and Philosophy,* 26, 757-771.
Okasha, S. (1997). Laudan and Leplin on empirical equivalence. *The British Journal for the Philosophy of Science,* 48, 251-256.
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.
Parker, W. S. (2006). Understanding pluralism in climate modeling. *Foundations of Science,* 11, 349-368.
Parker, W. S. (2009). Confirmation and adequacy-for-purpose in climate modelling. *Proceedings of the Aristotelian Society,* LXXXIII, 233-249.
Parker, W. S. (2010a). Comparative process tracing and climate change fingerprints. *Philosophy of Science,* 77, 1083-1095.
Parker, W. S. (2010b). Predicting weather and climate: Uncertainty, ensembles and probability. *Studies in History and Philosophy of Modern Physics,* 41, 263-272.

Parker, W. S. (2010c). Whose probabilities? Predicting climate change with ensembles of models. *Philosophy of Science,* 77, 985-997.
Parker, W. S. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science,* 78, 579-600.
Parker, W. S. (2013). Ensemble modeling, uncertainty and robust predictions. *Wiley Interdisciplinary Reviews: Climate Change,* 4, 213-223.
Pirtle, Z., Meyer, R. & Hamilton, A. (2010). What does it mean when climate models agree? A case for assessing independence among general circulation models. *Environmental Science & Policy,* 13, 351-361.
Raerinne, J. (2013). Robustness and sensitivity of biological models. *Philosophical Studies,* 166, 285-303.
Räisänen, J. (2007). How reliable are climate models? *Tellus A,* 59, 2-29.
Randall et al. (2007). Climate models and their evaluation. (In S. Solomon et al. (ed.) *Climate change 2007: The physical science basis* (pp. 589-662). Cambridge and New York: Cambridge University Press.)
Schupbach, J. N. (forthcoming). Robustness analysis as explanatory reasoning. *British Journal for the Philosophy of Science*
Schurz, G. (1991). Relevant deduction. *Erkenntnis,* 35, 391-437.
Schurz, G. (1994). Relevant deduction and hypothetico-deductivism: A reply to Gemes. *Erkenntnis,* 41, 183-188.
Schurz, G. (2014a). Bayesian pseudo-confirmation,use-novelty, and genuine confirmation. *Studies in History and Philosophy of Science,* 45, 87-96.
Schurz, G. (2014b). *Philosophy of science: a unified approach.* (New York: Routledge)
Sprenger, J. (2015). A novel solution to the problem of old evidence. *Philosophy of Science,* 82, 383-401.
Staley, K. W. (2004). Robust evidence and secure evidence claims. *Philosophy of Science,* 71, 467-488.
Steele, K. & Werndl, C. (2013). Climate models, calibration and confirmation. *British Journal for the Philosophy of Science,* 64, 609-635.
Stegenga, J. (2012). Rerum concordia discors: Robustness and discordant multimodal evidence. (In L. Soler, E. Trizio, T. Nickles and W. C. Wimsatt (Eds.), *Characterizing the robustness of science* (pp. 207-226). London: Springer.)
Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of science,* 71, 767-779.
Tebaldi, C. & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A,* 365, 2053-2075.
Votsis, I. (2014). Objectivity in confirmation: Post hoc monsters and novel predictions. *Studies in History and Philosophy of Science,* 45, 70-78.
Weisberg, M. (2006). Robustness analysis. *Philosophy of Science,* 73, 730-742.
Weisberg, M. (2013). *Simulation and similarity: using models to understand the world.* (Oxford: Oxford University Press)
Weisberg, M. & Reisman, K. (2008). The robust Volterra principle. *Philosophy of Science,* 75, 106-131.
Wimsatt, W. C. (1981). Robustness, reliability and overdetermination. (In M. B. Brewer, and B. E. Collins (Eds.), *Scientific inquiry and the social sciences* (pp. 124-163). San Francisco: Jossey-Bass.)
Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology,* 13, 219-240.

Yablo, S. (2014). *Aboutness.* (Princeton ; Oxford: Princeton University Press)
Ylikoski, P. & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology,* 21, 19-36.