Studies in History and Philosophy of Science 100 (2023) 32–38

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

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

Expert judgment in climate science: How it is used and how it can be justified☆

Mason Majszak [a,b,*], Julie Jebeile [a,b,c]

[a] Institute of Philosophy, University of Bern, Bern, Switzerland
[b] Oeschger Center for Climate Change Research, University of Bern, Bern, Switzerland
[c] CNRM UMR 3589, Météo-France/CNRS, Centre National de Recherches Météorologiques, Toulouse, France

ARTICLE INFO

Keywords:
Expert judgment
Climate uncertainty
Epistemic opacity
Scientific models
Tacit knowledge
Intuition
Values in science
Subjectivity

ABSTRACT

Like any science marked by high uncertainty, climate science is characterized by a widespread use of expert judgment. In this paper, we first show that, in climate science, expert judgment is used to overcome uncertainty, thus playing a crucial role in the domain and even at times supplanting models. One is left to wonder to what extent it is legitimate to assign expert judgment such a status as an epistemic superiority in the climate context, especially as the production of expert judgment is particularly opaque. To begin answering this question, we highlight the key components of expert judgment. We then argue that the justification for the status and use of expert judgment depends on the competence and the individual subjective features of the expert producing the judgment since expert judgment involves not only the expert's theoretical knowledge and tacit knowledge, but also their intuition and values. This goes against the objective ideal in science and the criteria from social epistemology which largely attempt to remove subjectivity from expertise.

1. Introduction

Climate science is a complex scientific domain characterized by high uncertainty and a strong reliance on numerical modeling and data mining; such characterizations are akin to the similar scientific domains of epidemiology, aerospace and the nuclear industries. A result of this high uncertainty in climate science has been the proliferation of expert judgment throughout the scientific process, i.e. in model design, model evaluation, data interpretation, and ultimately in the quantification and communication of uncertainties to policy-makers. The ubiquitous use of expert judgment requires review and, while climate models and climate data have been widely evaluated by philosophers with respect to their adequacy in providing understanding of the past and present climate and projections of climate change, the use of expert judgment in climate science has received less philosophical attention.

The concept of expert judgment can, however, be found within several philosophical domains, like decision and social choice theory; with social epistemology being the domain generally used to evaluate expert judgment and expertise more broadly. State-of-the-art social epistemology focuses on the relationship between expert(s) and the community employing them for further decision-making. Thus, social epistemology puts forward evaluative criteria for laypeople or decision-makers to arbitrate between (sometimes divergent) opinions of experts. Those criteria are varied and can include validity of evidence, soundness of arguments, the track record and unbiasedness of experts, the degree of agreement among experts, as well as other criteria (Goldman, 2001; Martini, 2014, 2015, 2020); overall, these evaluative criteria aim to ensure that expert judgment is grounded in objective arguments and not in mere subjective beliefs or expressions of interests from experts. However, state-of-the-art social epistemology falls short in providing guidance to evaluate expert judgment in climate science since, here, expert judgment is in large part produced *by* and *for* climate scientists themselves, where the aspects of individual subjectivity, that are usually dismissed by social epistemology, are vital for the production of an expert judgment in climate science.

Thus, we first examine how expert judgment is used in climate science, along with model projections, empirical measurements, and other data analyses. In this context, experts are climate scientists

☆ N.B.: The two authors have contributed equally.

* Corresponding author. Institute of Philosophy, University of Bern, Bern, Switzerland.
E-mail addresses: mason.majszak@unibe.ch (M. Majszak), julie.jebeile@unibe.ch (J. Jebeile).

https://doi.org/10.1016/j.shpsa.2023.05.005
Received 25 August 2022; Received in revised form 8 May 2023

0039-3681/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

(including climate modelers) who can be considered as equally qualified and informed agents and who contribute in the different steps of climate expertise, starting with model building and concluding with the quantification and interpretation of uncertainty. We argue that expert judgment is used in practice as if it is epistemically equal to or superior to models, insofar as it is employed to design models and quantify uncertainty by going beyond the projections that the (necessarily imperfect) models provide (Section 2). From this, we explicate the ingredients for the production of expert judgment (Section 3), suggesting that expert judgment can extend models and quantify uncertainty because a judgment is the result of connecting multiple lines of evidence at once, going beyond any particular model. But, as we will highlight, how this connection is made in practice is difficult to document since the production of expert judgment is epistemically opaque, and probably even more opaque than computer simulations (Section 4). We finally argue that the specific status of expert judgment over models can be justified based on how much one can trust the expert(s) *and* on what makes an expert judgment particular to its individual producer. This justification is based, on the one hand, on the ability of an expert, given their education and experience, to connect the existing lines of evidence, ultimately extending beyond the current state of knowledge to produce a judgment which overcomes the uncertainty. While most epistemological approaches attempt to counteract subjectivity in expert judgment, we will show, on the other hand, that subjective intuition and to some extent expert's values bestow particularly high epistemic interest on expert judgment (Section 5). Thus, we conclude that, in climate science, the role of expertise necessarily requires the recognition of the individual, their education and abilities including the often-maligned subjective features of the individual.

2. The superior status of expert judgment over models

In this section, we examine how expert judgment is used throughout climate science and show that it is used in two key ways, i.e. to design models and to quantify uncertainty. From this we argue that in the context of climate science, expert judgment can be seen as having a superior status over models.

2.1. Design of models

In climate modeling, there is not a uniquely most adequate model of the climate system; we mean "adequate" in the sense of Parker's "adequacy for purpose" view of model evaluation (Parker, 2009, 2020). Instead, several, and at a given point in time valid, methodological choices can be made to represent the climate system and tune the model equally well. Facing this form of underdetermination, expert judgment provides the avenue for making these methodological choices.

Within *model building*, idealizations and parameterizations are necessary for the mathematical formalization and the numerical resolution of the models. There are a variety of idealizations and parameterizations that can represent the system under investigation, and choosing among them requires expert judgment (see Thompson, 2022 about the co-dependence between mathematical models and expert judgment in climate science; see also Jebeile & Crucifix, 2020). In practice, modelers have to select the relevant processes to integrate in the models under the mathematical and computational limitations, and to assess to what extent the representations of these processes are sufficiently accurate for the given purpose. Expert judgment is particularly important in the design of parameterization, where parameterization is a method of idealizing, with a simplified sub-model, the physical processes that are either too small in scale to be represented in the model or are not well known. In this case experts must weigh or balance the computational cost of a given parameterization with the accuracy of the model output (e.g. Pincus & Stevens, 2013).

After the first version of a model is constructed, *model tuning* aims to assign values to key model parameters for emergent properties to match certain observations. Since the exact values of the key parameters are not all known, and because these parameters act as surrogates for the processes that modelers haven't been able to describe explicitly in the model, expert judgment is necessary to make the decision of which parameters to tune and more specifically on which targets to tune for (Jebeile et al., 2023). The priorities of this decision-making process vary across research teams or modeling groups, and can have a genuine impact on the emergent responses, and notably the model's sensitivity, i.e. the way a model's output can vary with relatively small changes in the model's inputs (Mauritsen et al., 2012; Schmidt et al., 2017; Hourdin et al., 2017). Expert judgment is also utilized to prevent overtuning; this issue arises when models match the data too well, in a way that makes them incapable of providing genuinely novel predictions outside the range of the data.

Expert judgment is also utilized in *model evaluation,* which concerns the adequacy of models for the purpose of understanding the past and present climate and of projecting future possible climate states. This assessment is made on the basis of model performance and model agreement. Model performance is often evaluated based on certain variables that climate experts deem relevant, which include, among other things, global distributions of temperature, precipitation, and radiation. The simulated variables are compared to the available past and present observations, where agreement among models is often used as an indicator of robust results. Model performance is measured quantitatively and statistically via performance metrics whose definition is left to the judgment of experts as well (Reichler & Kim, 2008). The assessment of model performance is done through utilizing "expert judgment based on the agreement with observations of the multi-model mean and distribution of individual models around the mean, taking into account internal climate variability" (IPCC, 2013, p. 822).

2.2. Quantification of uncertainty

Expert judgment is also used as an alternative cognitive resource for *estimating model uncertainty* because the available scientific models are not considered to provide, individually, sufficient enough or reliable enough projections, or to yield, collectively, exact estimates of uncertainty associated with projections.

Uncertainty quantification results from *ensemble-based approaches*. Those approaches take multiple climate models or versions of models and run each one in an attempt to obtain a range of possible projections (Flato et al., 2013). Experts sometimes weigh the ensemble models depending on their performance scores, i.e. on their individual ability to simulate specific observed phenomena or geographical locations, instead of treating each model as equal (see e.g. Brunner et al., 2020). From that weighing, they calculate multi-model weighted means "based on how processes are implemented or based on expert judgment" (Knutti et al., 2010, p. 3). In addition, expert judgment is utilized to re-assess uncertainty ranges spanned by the spread of the multiple projections of the ensemble (Katzav et al., 2021).[1] In this context, ensembles are "ensembles of opportunity", meaning that the ensemble members are just the available models that research teams are willing to provide and that comply with the standards defined, for instance, by the Coupled Model Intercomparison Project (CMIP). Therefore, the limited sample of available models is insufficient to properly span uncertainty ranges, and additional re-assessment of those ranges requires expert judgment (Oppenheimer et al., 2016; Thompson et al., 2016).

[1] As highlighted by an anonymous reviewer this use of expert judgment, where a model estimate is adjusted by an expert, is in line with the concept of 'selective defection', which has been recognized as a fallible strategy for expertise (see chapter 2 of Bishop & Trout, 2005 for more on this concept).

In situations where physical understanding is very low, with not only many known unknowns but also many unknown unknowns, experts make *claims under high uncertainty*, about, e.g., tipping points and extreme events. Thus "some of the most important uncertainties – such as the projected surface warming – are still based on expert judgment" (Knutti et al., 2002). Expert judgment is also used for estimating the collapse of marine-based sectors of the Antarctic ice sheet (IPCC, 2013, p. 1186) and estimating effective radiative forcing due to the high levels of uncertainty surrounding effective radiative forcing because of aerosols (IPCC, 2013, p. 574). Additionally, the use of expert judgment can be seen in the ranking of tipping points' sensitivity to global warming and the associated uncertainty surrounding the underlying physical mechanisms of these tipping points (Lenton et al., 2008). This process of identifying and ranking tipping points has been done using structured elicitation protocols, where a group of experts are asked to elicit individual judgments on a set of questions according to a formalized methodology. This methodology allows each individual expert to pull from different sources of information like relevant literature or mathematical considerations in producing their individual judgment, which are then aggregated using a specific aggregation rule. What remains clear is that under high uncertainty, like in the tipping point context, experts extrapolate claims about the system under consideration or investigation from prior knowledge about well-known systems as well as mathematical considerations (Lam & Majszak, 2022). In turn, these judgments often assist in the supplementation of knowledge gained through models and measurements.

Expert judgment is not only used to quantify the uncertainty estimates of the scientific claims, but also to *communicate the degree of confidence* in these assessments. This kind of use is quite evident in the Intergovernmental Panel on Climate Change's (IPCC) reports. In the *Summary for Policy Makers*, it is stated that "human activities are estimated to have caused approximately 1.0 °C of global warming above pre-industrial levels, with a *likely* range of 0.8 °C–1.2 °C. Global warming is *likely* to reach 1.5 °C between 2030 and 2052 if it continues to increase at the current rate. (high confidence)" (IPCC, 2018, p. 3). Within this sentence, one comes across the two metrics the IPCC uses to assess uncertainty, through the calibrated language of "likelihood", seen throughout the statement in italics, and a level of confidence, provided at the end of the statement in parentheses. The likelihoods can be based on, among other things, the "elicitation of expert views" to express a probabilistic estimate of a predicted outcome in a common and well-defined language (Mastrandrea et al., 2010). Similar to the likelihoods, the confidence metric utilizes a calibrated language of five phrases, from very low to very high confidence, and is based on the synthesis of the authors' judgment about the validity of the findings in the entire statement under consideration (Mastrandrea et al., 2010). The ultimate goal of this summary is for the communication of uncertainties, in the form of these likelihoods and confidence metrics, to non-expert decision makers (Drouet et al., 2021). This calibrated language allows the complex and dense information, that comes from expertise, and is generally only privy to experts, to be accessible and understandable for policy makers and the general public at large, to assist in their decision making.

In sum, there is a strong intertwined connection between expert judgment and climate models, and the former often supplants the latter. Thus, the choices of idealizations and parameterizations in models or the re-assessment of the uncertainty quantification are done with expert judgment, where this judgment provides a check on the quantified model uncertainty and the calculated probabilities.

2.3. Explanation

In climate science, expert judgment maintains a specific status since it is used to design models and to quantify uncertainty, this status requires an explanation. We suggest that expert judgment can supplant models in practice simply because the judgments result from considerations that go beyond the content and outputs of a particular model (Mach et al., 2017). In the design of models, the choice of idealizations and parameterizations is made based on the expert's knowledge of the scope of validity and of the computational cost of a variety of considered idealizations and parameterizations (Jebeile et al., 2023). The choice of the relevant parameters to tune and the observational variables to match depend on the defined priorities and purposes — themselves influenced by the values of the modelers (Intemann, 2015; Winsberg, 2018; Parker and Winsberg, 2018). The same is true for the evaluation of models and ensembles where the choice of performance metrics and the weighing of ensemble members in order to optimize the calculation of probabilities (or means) depend on what variables are prioritized (e.g. temperature, precipitation, wind) and which geographical locations are probed more accurately.

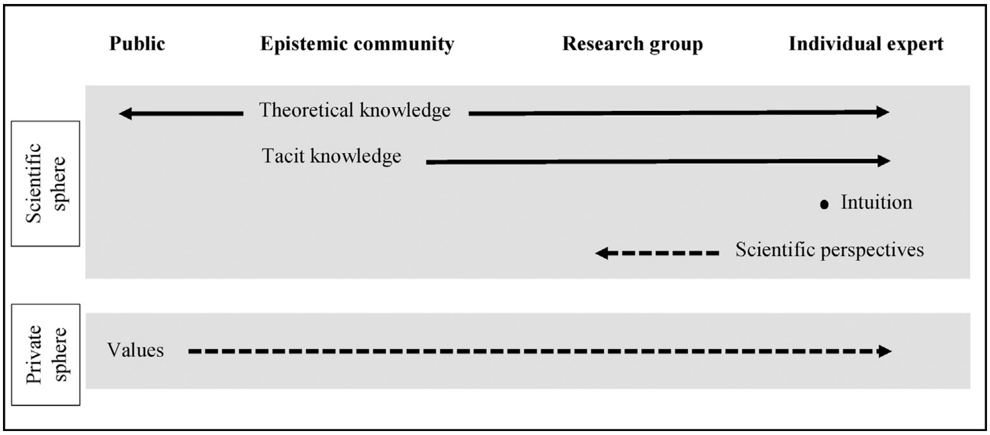
In the quantification of uncertainty, making claims under high uncertainty requires the expert's knowledge about the likely tipping points and considerations from thermodynamics for theorizing potential effects of tipping points. There is a limited and incomplete set of data about tipping points, as a result much of the information remains disjointed without clear avenues to explain how different pieces of information should fit together, resulting in high uncertainty throughout the domain. In this context, experts routinely are asked in expert elicitation protocols to bring together these different lines of evidence, i.e. model results, paleo-climate data and theoretical/mathematical considerations to overcome the epistemic gap and produce a holistic picture (Kriegler et al., 2009, p. 5044). When specifically discussing the Atlantic Meridional Overturning Circulation it has been argued that the expert elicitations are "based on the experts' synthesis of published literature and knowledge that is not explicit in the formal literature" (Zickfeld et al., 2007, p. 237). For the communication of degrees of confidence within the IPCC reports, experts evaluate the type, amount, quality, and consistency of evidence as well as the degree of agreement between the different pieces of evidence; for example, they seek the number of publications in the peer-reviewed scientific literature endorsing a given key finding. In a nutshell, one can see that in practice, experts are utilizing multiple and diverse lines of evidence to inform their judgment. Thus, expert judgment can supplant models because they are based on a broader set of background information, which the expert exploits and on which they form their judgment. Let us now try to precisely explicate the different components involved in the production of expert judgment.

3. Components of expert judgment

This section is an attempt at explicating the ingredients for the production of expert judgment, synthetized in Table 1. First, of course, an expert judgment is based on the theoretical knowledge the expert possesses on the subject matter (e.g. the Navier-Stokes equations to describe fluid dynamics in oceans and atmosphere). However, we are discussing an expert's judgment and not an expert's knowledge, thus there is more involved. An expert judgment results from an epistemic decision-making regarding, for example, whether such or such parameterization is suitable, or whether such or such uncertainty range is an adequate estimate. The formation of an expert judgment is therefore based not only on the available model outputs and empirical data, not only on the theoretical knowledge that the expert has on the subject matter, but also on unarticulated background information that enables the expert to transform their knowledge into a judgment, with that unarticulated background information being tacit knowledge and intuition (Soler et al., 2014).

Tacit knowledge is the know-how that experts have gained and keep maintaining through the long immersion and apprenticeship in their epistemic community (Collins & Evans, 2007, p. 6). Within science, it enables them to conduct an experiment, to make a calculation, to draw a data analysis, or to write and publish a scientific paper; in this context it is the set of an expert's skills that demarcates mere interactional expertise, i.e. the ability to converse about a technical matter as scientific journalists do, from contributory expertise, i.e. the ability to make genuine research contributions as researchers do (Collins & Evans, 2007). In the case of climate expertise, the required tacit knowledge should provide additionally and importantly the skill to

Table 1
Components of expert judgment. The location of the term indicates the origin of the production of the individual components. The arrow then indicates the breath of where those components can be disseminated. When the arrow is full, the dissemination is homogenous. When the arrow is dashed, the dissemination is unequally distributed.



compare and connect together model outputs, empirical data and theoretical considerations; this reflects on how much one knows the domain literature in depth and in breadth. In practice, this tacit knowledge allows the expert to build on the existing knowledge base, knowing how to extend beyond the existing set of information, by making additional connections to ultimately produce a judgment which can overcome the epistemic uncertainty. Albeit incommunicable to the outside world, tacit knowledge is assumed to be shared by all the practitioners belonging to the same epistemic community and thereby having a common scientific culture. This is akin to sharing the same natural language, where there can be multiple languages across the varied scientific domains but all those working within the same community share this common language (Collins & Evans, 2007, p. 7). It is a kind of "things you just know how to do without being able to explain the rules for how you do them" (Collins & Evans, 2007, p. 13).

On top of tacit knowledge, intuition is also necessary to form a judgment, giving the idiosyncratic dimension of the judgments produced by the individual person. Intuition is a concept that is hard to grasp (although it is an old one in the history of philosophy, see for example the *Regulae* of Descartes, 1628) as it refers to an inner mental process. It is a skill of someone that can be measured by the relevant and interesting insights or opinions a person can provide on a subject matter, but whose justification is particularly opaque. The concept of intuition is often referred to as the ability of an individual to make *good* judgments, however one must be careful that this intuition is "wisdom based on experience" and "can be gained through practice and socialization" since "this is mysterious enough" (Collins, 2010, p. 149). Like tacit knowledge, intuition is hardly communicable. Unlike tacit knowledge, this seems not something one can easily share with a group, but instead something that is very personal and individual.

Values are preferences, interests, and priorities that a scientist may have on social, political or moral matters. Examples of possible intrusion of values in climate science include focusing on scientific models that better document issues related to social injustice, or biodiversity loss. They can also be economical models including considerations related to economic growth that can influence the type of scenarios chosen for climate models runs. Additionally, they can be manifested in the choice of research topics or the locations/regions of the world which are to be investigated. The importance of values has been increasingly recognized by the philosophers of climate science and is now scrutinized by the climate science community. In the latest IPCC assessment report (AR6), it is stated that "[s]ocial values are implicit in many choices made during the construction, assessment, and communication of climate science information" (IPCC, 2021, §1.2.3.2). A person can share values with their epistemic community, but not necessarily; members of a climate science group can have different political opinions and can have different understanding on how to best serve the society regarding the values they prioritized (e.g. social justice, biodiversity, economic growth).

In turn, experts having their own intuition and values often orientate their research in accordance to their vision on what their scientific domain can and should do. Thus, theoretical knowledge, tacit knowledge, values and intuition can together ground a (new) choice of scientific perspective. Differences in these features and experiences of the expert would then result in differences in perspectives and in turn apparent disagreement among experts. However, the mere presence of disagreement should not warrant concerns of biases in individual judgments, rather we may even value these differences in perspectives for addressing some of the more complex problems within the domain (Douglas, 2008, p. 11). People privy to the climate science community know that there are prominent researchers in the field promoting different research programs and avenues which are themselves based on different scientific perspectives, promotion of, e.g., high resolution, model complexity, possibilities over probabilities, storylines, or historical approach to name just a few examples. One knows that those people do have incredibly valuable expertise on their specific perspectives, making the evaluation of their expertise a task based on the identity of experts, as individuals and unique scientific minds.

4. The production of expert judgment and opacity

How experts operate, in practice, to generate these judgments is hard to document as the production of expert judgment is partly epistemically opaque.[2] This is a strikingly common feature between expert judgment, model outputs and empirical data as they are all usually the outcome of an opaque production process. Roughly, this opaque production process means that there is no clear set of arguments or explanations available which relates the output of the process to the initial assumptions and considerations.

[2] It should also be noted that in addition to the epistemic opacity in the production of expert judgment there is additional institutional opacity, caused by the practice of how expert judgments are currently being used in the climate science domain. Given that there is not widespread use of structured elicitation protocols, with the concept of tipping points being the notable exception, there is then no concrete record of how the judgments are produced and under what conditions these judgments are being used for decision making or evaluation.

In scientific modeling, computer simulations stand as the relationship between model inputs and outputs insofar as they depart from the initial conditions and parameter values, and solve the equations encoded in the computer program. But computer simulations are opaque in that human agents cannot cognitively survey and master, from one end to the other, the entire series of the mathematical operations (Beisbart, 2021; Durán & Formanek, 2018; Humphreys, 2004, 2009). In measurements, experiments are opaque in that human agents usually ignore the physical processes or causes at work; this ignorance is often the very motivation for the experimental inquiry (Guala, 2002; Jebeile, 2017). In expert judgment, the production is opaque due to the fact that it utilizes tacit knowledge, intuition and values, and that it implies picking and connecting multiple lines of evidence in a way that is difficult (but not impossible) to explicate.

On the one hand, opacity is, in part, due to the nature of tacit knowledge on which expert judgment bears down. While the theoretical knowledge an expert possesses is explicable, as it can be found in textbooks or scientific papers for instance, tacit knowledge, on the contrary, is incommunicable as it is a skill somehow embodied in the expert's body, and thereby not reducible to epistemic propositions (Chang, 2014, p. 71; Polanyi, 1958/1962; ch. 4; Polanyi, 1967). Thus, tacit knowledge is often qualified as part of "the implicit, hidden, intuitive, and often partly unconscious dimensions of science" (Soler et al., 2014, p. 18). On the other hand, experts make their judgments based, in part, on their intuition fed by their unique experience as practitioners. This part of the judgment production is particularly opaque. Indeed, the production of an intuition is an internal, mental and thereby not entirely accessible process that operates within the mind of the expert. Intuition is the part of the expert judgment that remains non-explicable, non-reducible into proper arguments, and is therefore difficult to communicate. Due to these internal processes only being accessible to the individual expert, one cannot provide an accurate and complete account of the expert judgment, as an output, by appealing only to the scientific evidence and arguments of the matter that were used as inputs. Thus, in cases when experts disagree with each other, it might be due to them having different insights and intuition. Making values explicit is a difficult task requiring reflexivity to which scientists are usually not trained to do, and for this reason values can also be part of the opacity surrounding the production of expert judgment.

Finally, while computer simulations (or experiments) are generally thought of as mirroring or mimicking (or reproducing in the matter) the evolution or the behavior of the targeted phenomenon, the production of expert judgment is a multitasking process that can be difficult to document in detail. For instance, when re-assessing uncertainty ranges, the production of expert judgment is about confronting the ensemble-based projections with state-of-the-art understanding of future climate possibilities, weighing their likelihood based on other available climate data, or connecting unarticulated background knowledge. In a nutshell, experts are not computer programs or empirical instruments. Their inner reasoning is not reducible to an automatic procedure implementable by a machine, additionally it is not reducible to a mere physical process like a measurement either. Their reasoning, being an interior process, is opaque in that it involves theoretical knowledge, tacit knowledge, values but also intuition and insight which depend on their unique professional training and experience as scientists.

Interestingly, features of the evaluation of the opaque production processes are common in each of the three cases discussed, computer simulations, experimentation, and expert judgment. Examples of those features are calibration and robustness. Because deduction and/or direct perception cannot be warrants for the outcome of an opaque production process, a methodology is instead adopted in which calibration and robustness play an essential role. In models, parameters are tuned for the model outputs to match a set of available empirical data. In experiments, instruments are calibrated for the instrument to correctly provide well-known observations or measurements. In expert judgment, the equivalent is what is called the calibration score of experts in elicitation methods. "Calibration measures statistical likelihood, very loosely characterized as 'correspondence with reality'. In scoring calibration, each expert is regarded as a statistical hypothesis" (Cooke & Goossens, 2000, p. 12). The calibration process consists of rating the performance of an expert in providing probabilistic assessments which are confirmed by empirical data. Regarding robustness, when multiple models agree with each other, provided the models are independent from each other, their outputs are said to be robust and are often deemed reliable. The same holds when different kinds of instruments yield similar observations. The equivalence of robustness in expert judgment would be the majoritarian principle. This principle states that the more experts agree on a particular judgment, the more likely the judgment is supposed to be (Martini, 2014). Independence, in all contexts, is an important condition. If the models share common biases, which some argue could be the case given the common history of their creation and the subsequent building of models based on common code across different modeling efforts (Lenhard & Winsberg, 2010), the robustness reasoning could become invalid. Additionally, if all of the measuring apparatuses used to calculate one value were to share common biases, one would have little reason to trust the conclusions of that observational study. Likewise, if experts are all biased in the same way, there is no good reason to trust a given judgment based on agreement by the majority.

The burning question then becomes unavoidable: if the production of expert judgment is to some extent even more opaque than computer simulations and experimentation, how can we justify the critical role expert judgment plays in designing models and quantifying uncertainty in climate science? A pessimistic view would answer that expert judgment is simply a last resort: facing high epistemic uncertainty, one has no other choice than appealing to expert judgment because one would otherwise be left with unarticulated and inconsistent data. Such a pessimist view is somehow suggested by the "subsidiarity" principle in social epistemology which states that "[s]ubjective expert judgment should be invoked only to the extent that it cannot be substituted by other kinds of evidence, or mechanical/actuarial rules" (Martini, 2015, p. 393). Conversely, an optimistic view would certainly recognize that there is some quality in this expert judgment that makes them precious cognitive resources. This quality, which we will now argue for, has to be found in the competence and intuition of the experts themselves.

5. Justification in competence and intuition

Expert judgment can go beyond models as the production of a judgment involves a broader set of knowledge and know-how on the scientific domain, knowledge and know-how that extend what is in or derived from models. However, there is a clear epistemic difference between models or experiments and expert judgment, and as such the justification of the specific status of expert judgment in climate science cannot entirely rely on an examination of its production as it is in part opaque. In this last section, we argue that the justification for why expert judgment can supplant models derives from the high epistemic interest of eliciting the judgment of an expert as being exceptionally well-informed *and* having developed a unique insight on the subject matter, i.e. on the competence and intuition of an expert.

The scientific competence of an expert is the first important element of justification for why expert judgment can legitimately supplant models in climate science. The scientific competence begins with education, as experts are required to understand the theoretical foundations of the scientific domain which covers the subject matter at stake. This is the theoretical knowledge, produced over years of scientific development and discourse. But expert judgment is built on arguments that go beyond what is contained within textbooks or scientific papers. An expert also relies on their professional experience, providing them with a unique overview on the state-of-the-art research regarding the subject matter. The experts' access to and ability to understand the latest literature or the current state of the field affords them the rare viewpoint of being able to assess cutting edge research, theories and methodologies, a valuable addition to the set of theoretical knowledge an expert can utilize. The experience of the expert can also be in a practical sense, in the exercise of performing science within the domain. This results in tacit knowledge, gained through specific experiences of working with their instruments of investigation, whether that is a computer model or physical observational equipment. In a nutshell, these aspects of an expert's scientific competence, their theoretical and tacit knowledge, provide a baseline for the expert judgment as it relies on years of scientific discovery and scientific methodological progress and are thus rigorously evaluated by the scientific community. In practice, theoretical knowledge and tacit knowledge place the physical limits on the judgment of an expert, not allowing the judgment to be a wild guess or result in conclusions which are known to be physically impossible by the scientific community. However, in all cases of expert judgment there are varying levels of uncertainty. Thus, experts must rely on some component of expert judgment which goes beyond the existing set of knowledge, as there is a gap in the existing knowledge which the expert must overcome with their judgment. This is done in part by the expert's intuition, or inner reasoning; in this scientific context, intuition is born out of education and experience as an aspect of the expert's scientific competence which allows them to build on the existing knowledge base, to produce a judgment which can overcome the epistemic gap. This is also done with the expert's values which can possibly raise a problem of trustworthiness; the use of values facing epistemic uncertainty is now particularly well-documented where values serve in the definition of purposes and priorities of models, with the given underdetermination in the modeling assumptions (e.g. Jebeile & Crucifix, 2021; Parker & Winsberg, 2018).

To evaluate the subjective aspects of an expert judgment in climate science, i.e. the expert's intuition, values and scientific perspective, we suggest following a framework in line with the logic of induction. Within this logic there is an underlying assumption that a past instance of success provides an indication of future success. If this is applied to a judgment that is elicited from an expert, as is often done when utilizing the track record criteria (see Martini, 2014), then an expert who has a history of being correct, should be given a higher level of likelihood that their future judgments are correct simply because this judgment was elicited from that expert. However, we can take this logic further and apply it directly to how an expert utilizes their intuition, values and scientific perspective. Given that all elicited expert judgments involve the overcoming of an epistemic gap, all expert judgments would then have some subjective aspect as a meaningful component of producing the judgment, as this was the means for overcoming the gap. If an expert is continuously providing successful judgments, then the expert would continuously be using their scientific competencies, their knowledge and subjective aspects, to at least in part guide these judgments. From this we argue that each instance of a successful judgment provides reason to support an expert's ability and their trustworthiness.

Where by a successful judgment we do not mean that this judgment is necessarily true in some objective sense. Rather, we argue that the unique insight of experts on a subject matter constitutes the second part of the justification as it makes their individual judgments particularly valuable. In the context of expert judgment by and for climate experts, what matters is not only that an expert is competent at providing right or wrong answers (they are indeed usually asked to provide expertise in virtue of their scientific competence), but also that the expert provides adequate information relevant to a specific question. The relevance of an expert judgment then often depends on the specific scientific perspective the expert has adopted in their research career. Thus, an expert judgment is successful in this context if the judgment overcomes the epistemic gap, providing an answer to the specific question at hand based on the features of the specific individual, in line with their current state of knowledge on the topic. However, one may argue that the motivations behind an expert's choice of scientific perspective and then the subsequent use of such a perspective may constitute a bias. We don't see this type of bias as problematic, rather when one is eliciting a judgment from such an expert they need to be aware of the context within which the perspective was created, thus identifying if the values behind the judgment are legitimate or illegitimate. The criterion of demarcation between legitimate and illegitimate values goes outside the scope of this paper although it remains a very important subject; we here refer to Intemann (2015) who defines legitimate values as those which are democratically endorsed. By evaluating judgments as suggested, one can allow experts to utilize those subjective aspects which make their judgments successful while also recognizing that each expert may not be suited to provide a successful judgment under all conditions or for all questions. Experts are then indeed not interchangeable rational agents, rather their epistemic advantage to provide judgment stems not only from their specific education and professional experience, but also from the subjective features of the individual, being the intuition, the values and the scientific perspectives they have adopted.

6. Conclusion

In sum, throughout all the instances of expert judgment we identified in climate science, these judgments are used to overcome some type of uncertainty or epistemic gap, whether that be in the creation of a model or the identification of tipping points. From this we have highlighted that these judgments supplant modeling and observational data efforts by extending beyond what is captured by these processes. This then begged the question of how experts can extend beyond modeling and data. To answer this question, we turned to the production of expert judgment, illustrating that, even though the production is an opaque process, what remains clear is that experts bring together potentially disparate pieces of information and synthesize it into a single judgment. This allows the expert to go beyond what is included in a single model or a single line of observational data.

This is not where the story ends however, as we are left to answer the question of why experts are able to do this while models and even other non-experts cannot. We argued that the scientific competencies and intuitional abilities of the expert provide the means to overcome the epistemic gap and, converse to the objective ideal, the subjective elements play a vital role in this process. The individual subjective aspects of the expert's scientific capacities provide the vital and unique process for including tacit knowledge and intuition while eliciting each judgment. Thus, each expert must then retain the status of a precious cognitive resource. Ultimately, we conclude that, in climate science, experts must be recognized and evaluated on an individual level as their scientific competencies and intuitional abilities are subjective features of the individual expert which cannot be captured by evaluating the entire scientific community.

Funding

The authors are grateful to the Swiss National Science Foundation for financial support (grant PP00P1_170460).

Acknowledgments

The authors thank, for helpful comments on a previous version of this paper, the two anonymous reviewers as well as the working group on philosophy of science in Bern, especially Claus Beisbart, Matthias Egg, Andreas Freivogel, August Hämmerli, Ralf Hand, Vincent Lam, Vera Matarese and Tim Räz.

References

Beisbart, C. (2021). Opacity thought through: On the intransparency of computer simulations. *Synthese, 199*, 11643–11666.
Bishop, M. A., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment.* Oxford University PressNew York.
Brunner, L., Pendergrass, A. G., Lehner, F., Merrield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics, 11*, 995–1012.

Chang, H. (2014). In Soler, Zwart, Lynch, & Israel-Jost (Eds.), *Epistemic activities and systems of practice: Units of analysis in philosophy of science after the practice turn* (pp. 67–79) (Chapter 2).

Collins, H. M. (2010). *Tacit and explicit knowledge*. Chicago: The University of Chicago Press.

Collins, H., & Evans, R. (2007). *Rethinking expertise*. Chicago: The University of Chicago Press.

Cooke, R. M., & Goossens, L. (2000). Procedures guide for structural expert judgement in accident consequence modelling. *Radiation Protection Dosimetry, 90*, 303–309.

Descartes, R. (1628-1629). *Règles pour la direction de l'esprit*.

Douglas, H. (2008). The role of values in expert reasoning. *Public Affairs Quarterly, 22*(1), 1–18.

Drouet, I., Andler, D., Barberousse, A., & Jebeile, J. (2021). Expert reports by large multidisciplinary groups: the case of the International Panel on Climate Change. *Synthese, 199*, 14491–14508.

Durán, J., & Formanek, N. (2018). Grounds for trust: Essential epistemic opacity and computational reliabilism. *Minds and Machines, 28*(4), 645–666.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chan Chou, S., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., & Rummukainen, M. (2013). Evaluation of climate models. *IPCC*, 741–866, 2013.

Goldman, A. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research, 63*(1), 85–110.

Guala, F. (2002). Models, simulations, and experiments. In L. Magnani, & N. J. Nersessian (Eds.), *Model-based reasoning: Science, technology, values* (pp. 59–74). New York: Kluwer.

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., & Williamson, D. (2017). The art and science of climate model tuning. *Bulletin of the American Meteorological Society, 98*(3), 589–602.

Humphreys, P. (2004). *Extending ourselves. Computational science, empiricism, and scientific method*. Oxford: Oxford University Press.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese, 169*(3), 615–626.

Intemann, K. (2015). Distinguishing between legitimate and illegitimate values in climate modeling. *European Journal for Philosophy of Science, 5*(2), 217–232.

Jebeile, J. (2017). Computer simulation, experiment, and novelty. *International Studies in the Philosophy of Science, 31*(4), 379–395.

Jebeile, J., & Crucifix, M. (2020). Multi-model ensembles in climate science: mathematical structures and expert judgements. *Studies in History and Philosophy of Science Part A, 83*, 44–52.

Jebeile, J., & Crucifix, M. (2021). Value management and model pluralism in climate science. *Studies In History and Philosophy of Science Part A, 88*(August), 120–127.

Jebeile, J., Lam, V., Majszak, M., & Räz, T. (2023). Machine learning and the quest for objectivity in climate model parameterization. *Production at Climatic Change*.

Katzav, J., Thompson, E., Risbey, J., Stainforth, D., Bradley, S., & Frisch, M. (2021). On the appropriate and inappropriate uses of probability distributions in climate projections and some alternatives. *Climatic Change, 169*(1–2).

Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P. J., Hewitson, B., & Mearns, L. (2010). Good practice guidance paper on assessing and combining multi model climate projections. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, & P. M. Midgley (Eds.), *IPCC working group I technical support unitMeeting report of the intergovernmental Panel on climate change expert meeting on assessing and combining multi model climate projections*. Switzerland: University of Bern, Bern.

Knutti, R., Stocker, T., Joos, F., & Plattner, G.-K. (2002). Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature, 416*(6882), 719–723.

Kriegler, E., Hall, J. W., Held, H., Dawson, R., & Schellnhuber, H. J. (2009). Imprecise probability assessment of tipping points in the climate system. *Proceedings of the National Academy of Sciences, 106*(13), 5041–5046.

Lam, V., & Majszak, M. M. (2022). Climate tipping points and expert judgment. *WIREs Climate Change, 13*(6), Article e805.

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

Lenton, T., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the earth system. *Proceedings of the National Academy of Sciences, 105*(6), 1786–1793.

Mach, K., Mastrandrea, M. D., Freeman, P. T., & Field, C. B. (2017). Unleashing expert judgment in assessment. *Global Environmental Change, 44*, 1–14.

Martini, C. (2014). Experts in science: A view from the trenches. *Synthese, 191*, 3–15.

Martini, C. (2015). Expertise and institutional design in economic committees. *Journal of Economic Methodology, 22*(3), 391–409.

Martini, C. (2020). The epistemology of expertise. In M. Fricker, P. J. Graham, D. Henderson, & N. J. L. L. Pedersen (Eds.), *The routledge Handbook of social epistemology* (pp. 115–122). NY: Routledge (Chapter 12).

Mastrandrea, M., Field, C., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., & Zwiers, F. W. (2010). *Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. Intergovernmental Panel on climate change (IPCC)*.

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., & Tomassini, L. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems, 4*(3).

Oppenheimer, M., Little, C. M., & Cooke, R. M. (2016). Expert judgement and uncertainty quantification for climate change. *Nature Climate Change, 6*(5), 445–451.

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

Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science, 87*(3), 457–477.

Parker, W. S., & Winsberg, E. (2018). Values and evidence: How models make a difference. *European Journal for Philosophy of Science, 8*(1), 125–142.

Pincus, R., & Stevens, B. (2013). Paths to accuracy for radiation parameterizations in atmospheric models. *Journal of Advances in Modeling Earth Systems, 5*(2), 225–233.

Polanyi, M. (1958/1962). *Personal knowledge: Towards a post-critical philosophy* (corrected edition). Chicago: University of Chicago Press.

Polanyi, M. (1967). *The tacit dimension*. London: Routledge & Kegan Paul.

Reichler, T., & Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society, 89*, 303–312.

Schmidt, G., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., Molod, A., Neale, R. B., & Saha, S. (2017). Practice and philosophy of climate model tuning across six US modeling centers. *Geoscientific Model Development, 10*(9), 3207–3223.

Soler, L., Zwart, S., Lynch, M., & Israel-Jost, V. (Eds.). (2014). *Science after the practice turn in the philosophy, history, and social Studies of science* (1st ed.). Routledge.

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

Thompson, E. (2022). *The co-development of models with expert judgement suppresses model diversity and underestimates risk*. https://doi.org/10.21203/rs.3.rs-234517/v1. Manuscript submitted for publication.

Thompson, E., Frigg, R., & Helgeson, C. (2016). Expert judgment for climate change adaptation. *Philosophy of Science, 83*(5), 1110–1121.

Winsberg, E. (2018). *Philosophy and climate science*. Cambridge University Press.

IPCC. (2021). In V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment Report of the intergovernmental Panel on climate change [Masson-Delmotte*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

IPCC. (2018). Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways. [V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.)] *In the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*.

Zickfeld, K., Levermann, A., Morgan, M., Kuhlbrodt, T., Rahmstorf, S., & Keith, D. (2007). Expert judgements on the response of the Atlantic meridional overturning circulation to climate change. *Climatic Change, 82*, 235–265.