The Error of Ignorance: Suspension of Judgement in Climate Science

Malou Sopcak

A Research Thesis in the Department of Philosophy

Presented in Partial Fulfillment of the Requirements

For the Degree of

Master of Arts

at Concordia University

Montréal, Québec, Canada

March 2026

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Malou Sopcak

Entitled: The Error of Ignorance: Suspension of Judgement in Climate Science

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________________ Examiner
Dr. Anna Brinkerhoff

______________________________________ Thesis supervisor
Dr. Matthew Barker

Approved by ______________________________________ Graduate Program Director
Dr. Matthias Fritsch

______________________________________ Dean of Faculty of Arts and Science
Dr. Pascale Sicotte

Abstract

The Error of Ignorance: Suspension of Judgement in Climate Science

Malou Sopcak

Arguments from inductive risk (AIRs) have widely been accepted as demonstrating that non-epistemic (e.g., social or ethical) values must or ought to play a role at core stages of scientific investigation. However, these classic versions of the AIR focus on only two options for treating hypotheses: accept or reject. In practice, scientists often exercise a third option: *formally suspend judgement* on the hypothesis. My focus is the "generalization" of the classic AIR to accommodate this third option. I argue that an adequately generalized AIR requires more careful distinctions between the epistemic (truth-oriented) and pragmatic (action-oriented) dimensions of trade-offs involved. More specifically, I distinguish between an epistemic error, *failed reticence*, as distinct from a pragmatic error, *dangerous inaction*, both associated with increased rates of suspension of judgement. Throughout, I use the case of scientists refraining from reporting numerical values for sea level rise from disintegration of the West Antarctic Ice Sheet in the Intergovernmental Panel on Climate Change (IPCC) fourth assessment report as an example where trade-offs involving these errors must be determined. I argue that applying the epistemic/pragmatic distinction to this case allows for the beginnings of a framework for evaluating decisions to suspend judgement and clarifies differing roles for scientists and policymakers in cases of uncertainty in policy-relevant science. The overarching aim of the paper is to provide considerations that enable more rational non-epistemic value influences on the selection of evidential standards in climate science.

Acknowledgements

I'm very grateful to the teachers, students, friends, and family who made this thesis possible. I thank especially my supervisor, Dr. Matthew Barker, who offered guidance from before my application to the program; I'm particularly grateful for his thoughtful comments and careful feedback. I'm grateful also to the other faculty members in the department who expanded my philosophical horizons, deepened my interests, and entertained my presence in classes I wasn't enrolled in. I thank especially Dr. Anna Brinkerhoff, Dr. Jing Hu, Dr. David Morris, Dr. Nabeel Hamid, Dr. Emily Perry, and Dr. Matthias Fritsch, as well as Dr. Eran Tal at McGill. For his mentorship and willingness to answer my many questions, I am grateful Dr. Nathan Kowalsky. I'm deeply thankful to my parents, Nicolette and Paul, for their constant encouragement and support, to Ella for always believing in me and being there to listen, and to Molly for her encouraging peptalks and sharp insights. Last but certainly not least, I have to thank the rest of my family and all of my friends for being the best team I could have.

Table of Contents

1. Introduction

Demands for objectivity in science may reasonably arise from the epistemic reliance on scientists to provide empirical information that has consequences in people's everyday lives. In climate science, these may include the requirement that the findings that inform policy action be grounded in the state of the climate, rather than resulting from political or economic motives. However, there are challenges to attaining objective scientific knowledge, as science is an inductive endeavor and therefore entails uncertainty. One prominent attempt to secure objectivity was by the value free ideal (VFI) that sought to relegate non-epistemic (often ethical, social, or political) values to the "external" stages of scientific investigation. However, with the argument from inductive risk (AIR)—first brought to the philosophy of science from statistics (Churchman 1948; Rudner 1953) and further developed with attention to details of scientific practice (Douglas 2000, 2009)—implicit non-epistemic values in the core stages of science were, and continue to be, identified, and some of these endorsed. The AIR is widely supposed to have established that non-epistemic values must or should be weighed up along with more epistemic (or truth-oriented) considerations, such as predictive accuracy, when deciding on hypotheses (Douglas 2009; Longino 1990).

However, there is of course the possibility of *misuse* of non-epistemic values in science. For instance, the manipulation of science for the pursuit of inappropriate (often economic or political) ends has led to the promotion of climate change denial and the obfuscation of the harmful effects of tobacco smoke, highlighting the importance of continuing to restrict and specify the types, roles, and stages of such value influences in scientific investigation (Oreskes and Conway 2010). We thus must navigate between the VFI's minimization of non-epistemic values at core stages, on one hand, and inappropriate uses, on the other. This motivates current research that seeks to characterize appropriate roles for values in scientific practice (Douglas 2015; Elliott and Steel 2017).

The classic AIR involves two options of hypothesis acceptance or rejection. However, in current scientific practice, scientists often exercise a third option: *formally suspend judgement on the hypothesis*. This is common in climate science, which, for various reasons, has a high level of inconclusivity (Winsberg 2012; Helgeson, Parker, and Tuana 2025). The Intergovernmental Panel on Climate Change (IPCC), in an attempt to account for inconclusivity and secure public trust in climate science, endorses high evidential standards (Elabbar forthcoming). This leads to "no assessment" labels in sections of IPCC reports, which then cannot inform policy action. In this paper, I argue that adequately generalizing the AIR to include the option of suspending judgement requires more carefully distinguishing between epistemic and pragmatic dimensions of trade-offs, and that this allows for scientists to more rationally exercise suspension of judgement in practice.

The layout of the paper is as follows. In the second section, I examine existing accounts and applications of a generalized AIR and draw on previous versions of the AIR to develop an argument for the application of an epistemic/non-epistemic (or epistemic/pragmatic) distinction to the trade-offs that emerge when scientists suspend judgement. In the third section, I clarify the epistemic and pragmatic gains and costs associated with both the classic and generalized AIR, and characterize an epistemic error, *failed reticence*, as distinct from the pragmatic error of *dangerous inaction*. I then apply this to a case study, the treatment of the West Antarctic Ice Sheet (WAIS) in the fourth IPCC assessment report, to demonstrate how distinguishing between these two errors allows for considerations to be weighed up more precisely and judgement to be suspended more rationally. In the fourth section, I discuss the importance of such a distinction for evaluating proposals to increase power of a scientific investigation. The fifth section provides a synthesis by way of a conclusion.

2. The Argument from Inductive Risk (AIR) and the Suspension of Judgement

2.1. Generalizing the Classic AIR

The AIR challenges the VFI by contending that non-epistemic values either inevitably do or ought to play a role at core stages of science. Both the VFI and AIR rely on some distinction between epistemic and non-epistemic (or pragmatic) values. Epistemic values are those that are truth-oriented or -promoting, and non-epistemic values refer to all other values—main examples in this literature are various social, ethical, and political values (Douglas 2009; Steel 2010).[1] The classic AIR can be formulated as follows:

1. If science is endemically uncertain, then scientists must make decisions about whether there are sufficient grounds to accept or reject hypotheses.[2]

2. If scientists must make decisions about whether there are sufficient grounds to accept or reject hypotheses, then this decision should involve (explicitly or implicitly) trading off between false positive and false negative error.

3. If this decision should involve (explicitly or implicitly) trading off between false positive and false negative error, then this decision also should involve a trade-off between ethically significant non-epistemic consequences associated with each type of error.

4. If this decision should involve a trade-off between ethically significant non-epistemic consequences associated with each type of error, then scientists, as the most authoritative source for empirical knowledge, are required to transparently consider non-epistemic

(ethical, social, etc.) values when determining whether there are sufficient grounds to accept or reject hypotheses.[3]

---

5. If science is endemically uncertain, then scientists, as the most authoritative source for empirical knowledge, are required to transparently consider non-epistemic (ethical, social, etc.) values when determining whether there are sufficient grounds to accept or reject hypotheses.

6. Science is endemically uncertain.

---

7. Scientists, as the most authoritative source for empirical knowledge, are required to transparently consider non-epistemic (ethical, social, etc.) values when determining whether there are sufficient grounds to accept or reject hypotheses.

Rudner's (1953) classic AIR formulation begins from the claim that part of the role of the scientist is to form judgements about hypotheses in light of empirical investigation, specifically to judge that they are true ("accepting" the hypothesis) or that they are false ("rejecting" the hypothesis), which requires explicitly or tacitly selecting a level of evidential sufficiency. For example, one may be required to determine what level of evidence suffices to accept a particular hypothesis. If one chosen level is higher than another (demanding stronger evidence for acceptance), there is an increased risk of rejecting hypotheses that are true; if a chosen level is lower, there is an increased risk of accepting hypotheses that are instead false. No matter the level chosen, there is some "inductive risk" in any non-certain setting (Hempel 1965; Douglas 2000).

The paradigmatic AIR case involves the setting of a threshold for statistical significance: since type-I error (false positive, or mistakenly accepting a false claim) and type-II error (false negative, or mistakenly rejecting a true claim) are systematically linked by their mathematical framing, to minimize one type of error *just is* to be more permissive of the other type. Proponents of the AIR then propose that the setting of evidential standards inevitably does or ought to depend, at least partly, on the severity of consequences associated with each error and thus must or should involve the influence of non-epistemic values (Rudner 1953; Douglas 2000). They reason that in inductive settings, the risk of error cannot be avoided and to decrease the risk of one sort of error is to increase the risk of the other; we must or should decide which sort of error to minimize, and this decision should depend on the severities of different consequences associated with each type of error. For example, in the assessment of the toxicity of a chemical, a false positive error may consist in claiming a non-toxic chemical to be toxic (prompting overregulation), and a false negative error may consist in claiming a non-toxic chemical to be toxic (prompting underregulation) (Douglas 2000). According to the AIR, responsible management of inductive risk requires factoring in the weight of costs to public health vs. excess costs to industries, and this requires appeal to non-epistemic values.

The classic AIR trade-off between type-I and type-II error is complicated when the suspension of judgement becomes a third option (Steel 2016; Wilholt 2013, 2016). Daniel Steel (2016) characterizes a generalized AIR that includes the suspension of judgement in terms of ignorance. According to Steel, there are two ways to remain ignorant about a hypothesis: to be in error (i.e., by rejecting a true hypothesis or accepting a false hypothesis) or to suspend judgement (2016). In this version, scientists' decisions to accept, reject, or suspend judgement on a hypothesis involve a nonnegligible risk of ignorance (Steel 2016). As a result, suspending judgement can be an "epistemic failure," even if no error in the classic sense was made (706).

Torsten Wilholt (2013, 2016), on the other hand, focusses on the trade-offs that emerge when the suspension of judgement becomes an option, arguing that an epistemic trade-off between epistemic reliability and epistemic power emerges. When scientists suspend judgement, they increase epistemic reliability in the sense that they reduce the chances of both false positive error and false negative error; these chances are reduced because there is not a positive or negative claim that could be in error. However, this comes at the expense of a reduction of epistemic power, which involves the ability to generate a potentially true conclusion, whether positive or negative (2013). When this ability is reduced via suspension of judgement, there is a corresponding reduction in the rate of "definitive" findings (2013, 244). That the resources allocated to a particular investigation are limited illustrates that power has both epistemic and pragmatic dimensions—as Wilholt notes, "[t]he aim of getting at the truth implies not only that we want reliable results, but also that we want results" (2016, 9). Method selection is underdetermined by reliability, and a trade-off between epistemic power and epistemic reliability must be determined for an appropriate distribution of epistemic risks (Wilholt 2016).

Ahmad Elabbar (forthcoming) applies Wilholt's (2013, 2016) characterization of a reliability/power trade-off to argue for the variation of evidential standards according to non-epistemic values. According to Elabbar, lowering evidential standards allows the IPCC to counteract imbalances of epistemic power disproportionately affecting data poor regions by indirectly increasing the rate of findings, which are critical for policy action (forthcoming, 19). However, as Elabbar acknowledges, varying evidential standards to increase power may lead to an "intolerable risk of acting on error" that may lead a region to expend limited resources, which in turn may leave the region in a worse position to cope with changes (17). As a result, he restricts the proposed variation to "some cases" where findings "would still produce epistemic goods" (Elabbar forthcoming, 17). This suggests the need for a framework for assessing the appropriate balance between reliability and power. Similarly, there has been a call for a methodological framework for determining an appropriate balance between reticence and judgement in science (Peacock 2018). In this paper, I suggest that carefully distinguishing between epistemic and pragmatic gains and costs should constitute the beginning of such a framework.

2.2. A Generalized AIR in Terms of Error

In my formulation of a generalized AIR, I begin from the classic formulation, and draw on Steel's (2016) generalized argument in terms of ignorance rather than error, and on Wilholt's (2013, 2016) account of a reliability/power trade-off for a generalized AIR argument in terms of error. I formulate the argument as follows:

1. If science is endemically uncertain, then scientists must make decisions about whether there are sufficient grounds to render a judgement (accept or reject) or suspend judgement about hypotheses.

2. If scientists must make decisions about whether there are sufficient grounds to render a judgement (accept or reject) or suspend judgement about hypotheses, then this decision should involve (explicitly or implicitly) trading off between decreased reliability and failed reticence.[4]

3. If this decision should involve (explicitly or implicitly) trading off between decreased reliability and failed reticence, then this decision also should involve a trade-off between ethically significant non-epistemic consequences associated with each type of error.

4. If this decision also should involve a trade-off between ethically significant non-epistemic consequences associated with each type of error, then scientists, as the most authoritative source for empirical knowledge, are required to transparently consider non-

epistemic (ethical, social, etc.) values when determining whether there are sufficient grounds to render a judgement (accept or reject) or suspend judgement about hypotheses.

---

5. If science is endemically uncertain, then scientists, as the most authoritative source for empirical knowledge, are required to transparently consider non-epistemic (ethical, social, etc.) values when determining whether there are sufficient grounds to render a judgement (accept or reject) or suspend judgement about hypotheses.

6. Science is endemically uncertain.

---

7. Scientists, as the most authoritative source for empirical knowledge, are required to transparently consider non-epistemic (ethical, social, etc.) values when determining whether there are sufficient grounds to render a judgement (accept or reject) or suspend judgement about hypotheses.

This argument closely resembles the classic AIR formulation above, and the role for non-epistemic values in the conclusion remains the same. The differences lie in the addition of the option to suspend judgement, which leads to an epistemic trade-off between decreased reliability and failed reticence. This trade-off is modified from Wilholt's (2013) characterization of a trade-off between epistemic reliability and epistemic power. While the power that is in tension with reliability is an important epistemic consideration, such power may also have a pragmatic dimension. For example, in the variation of evidential standards, when the focus is on the rate of conclusions generated without sufficient epistemic consideration, a more pragmatic notion of power is adopted. Similarly, when the reasons for such a focus are practical, there is a pragmatic aspect. For example, a growing concern over the irreproducibility of scientific results has been attributed to an excessive rate of false positives due to "current scientific practices [that] create strong incentives to publish statistically significant (i.e., "positive") results" (Head et al. 2015).

Though Wilholt does not discuss statistical power, statistical testing can be used to provide clear examples of the epistemic and pragmatic dimensions of power. Statistical power is a mathematical factor that depends on the relationships between specific variables, namely sample size and variability, the significance criterion (α), and the population effect. Thus, modifying these variables leads to changes to statistical power. The increase of sample size to improve statistical power is an example of an epistemically motivated increase of power. On the other hand, alterations of the alpha value can come from concern about which error type to minimize (epistemic concerns) or out of pragmatic concerns (whether appropriate or inappropriate) to obtain significant results. While pragmatic concerns may legitimately influence the significance level chosen, p-hacking is considered a form of bias involving an illegitimate influence, and involves selectively reporting significant results because they are significant (Head et al. 2015). One example of p-hacking is the decision to terminate experimentation when a significant result is obtained *because* a significant result is obtained (Head et al. 2015). Pursuing statistically significant results to further one's career or reputation is not an epistemically adequate reason for a pragmatic increase of an investigation's power.

In the following sections, I argue for the distinction between epistemic and pragmatic dimensions of risks emerging in the trade-offs associated with the generalized AIR. I also argue that attending to this distinction allows for a more rational weighing of the trade-offs in practice, particularly in the evaluation of proposals for increasing power of a scientific investigation. According to both the classic and generalized AIR, scientists are required to transparently consider non-epistemic values when determining whether there are sufficient grounds to accept, reject, or suspend judgement about hypotheses. However, restrictions guiding the appeal to non-epistemic values are important in addition to transparency about their inclusion to prevent their misuse, as discussed above. Douglas, Steel, and Whyte maintain a privileged position for epistemic values in scientific investigation and assert that this is essential for preserving the integrity of science. Douglas advances an account that distinguishes between direct and indirect roles for non-epistemic values, and Steel and Whyte adopt an epistemic/non-epistemic distinction to discriminate between legitimate and illegitimate value influences (Douglas 2009; Steel 2010; Steel and Whyte 2012).

A starting point for my contributions is the adoption of a version of this epistemic/non-epistemic (or epistemic/pragmatic) distinction that helps separate values that are truth-oriented from those that are relevant to goals other than truth, a separation that helps identify, understand, assess, and make recommendations about value influences in scientific investigation (Steel 2010). I argue that this is helpful, as it clarifies the trade-offs at stake, which in turn highlights the roles for both scientists and policymakers in the report and application of policy-relevant science. Before returning to this conclusion, I more carefully examine the trade-offs emerging in a generalized AIR in terms of gains and costs in the next section.

3. Epistemic and Pragmatic Dimensions of a Generalized AIR

3.1. Gains and Costs in the Classic and Generalized AIR

According to the argument above, I characterize a generalized AIR as involving an epistemic trade-off between reliability and failed reticence. In what follows, I distinguish this trade-off from a pragmatic trade-off between wasted resource-use and dangerous inaction. I propose that *failed reticence* is the epistemic error associated with decreased epistemic power and suggest that this is distinct from the pragmatic error of dangerous inaction also associated with decreased power. The notion of failed reticence also isolates and specifies the epistemic error picked out by climate scientist James Hanson's conception of "scientific reticence" (2007).

Failed reticence can by illustrated by the example of the use of bands of agnosticism in the setting of statistical significance levels. I will continue to discuss failed reticence in terms of bands of agnosticism, but these can be understood as a metaphor or interpretation that would need to be specified in particular contexts, depending on the statistical framework used. Typically when using a p-value significance testing framework, an alpha value is selected to set the evidential threshold that determines whether a result can be characterized as “significant” or not. This value is commonly 0.05. Statistical methods then involve the calculation of a p-value that may fall below the alpha-value (e.g., 0.049), allowing for the acceptance of the hypothesis of interest (and rejection of the null hypothesis), or above (e.g., 0.051), leading to the rejection of the hypothesis of interest. However, to formalize suspending judgement, a band of agnosticism may be introduced, for example from 0.03 to 0.06. As a result, judgements are mandated with p-values below 0.03 (accept) and with those above 0.06 (reject), but when p-values are in between these values the scientist formally suspends judgement. When judgement is suspended on a hypothesis that is true or false, the error of failed reticence is committed: either the scientist fails to judge a true hypothesis to be true or fails to judge a false hypothesis to be false. The wider the band of agnosticism, the greater the chance of this error. However, committing this error may be advisable in some cases, as it helps prevent (via suspension of judgement) two other kinds of error, namely, accepting a false hypothesis and rejecting a true one.[5]

Specifying the epistemic error of failed reticence as distinct from a pragmatic error associated with the suspension of judgement allows for the separation of two errors that are often treated interchangeably, namely failed reticence and dangerous inaction, respectively. While failed reticence is an epistemic cost, *dangerous inaction* is the pragmatic cost of delaying policy action to gather more evidence, and thereby allowing a potentially harmful situation to progress without attempting to counteract it. To summarize, increasing the number of cases in which one suspends judgement by widening the band of agnosticism leads to an increased epistemic risk of failed reticence *and* an increased pragmatic risk of dangerous inaction. Both of these costs are associated with the favouring of reliability over power in the reliability/power trade-off.

Similarly, it is possible to distinguish epistemic and pragmatic costs associated with power. As in the favouring of epistemic reliability, there is both a possible epistemic error and a possible pragmatic error associated with the favouring of power: increased (type-I and/or type-II) error and *wasted resource-use*, respectively. Wasted resource-use refers to the misdirection of monetary and other resources toward ineffective policy measures, which, in the case of climate science, may leave a country or region with fewer resources to adapt to harmful changes (Elabbar forthcoming). These risks are increased when suspension of judgement is decreased by (via one statistical means or another) narrowing the band of agnosticism, which is associated with the favouring of power over reliability.

This characterization of gains and costs is summarized in the two tables below. The first summarizes the gains and costs associated with the classic AIR, and the second summarizes the gains and costs associated with a generalized AIR where the suspension of judgement is an option.

Table 1. Gains and costs associated with the type-I/type-II error trade-off that arises in the classic AIR.

	High alpha value (demand weaker evidence)		*Low alpha value (demand stronger evidence)*	
	Gains	**Costs**	**Gains**	**Costs**
Epistemic	Decreased type-II (false negative) error	Increased type-I (false positive) error	Decreased type-I (false positive) error	Increased type-II (false negative) error

Pragmatic	Decreased dangerous inaction	Increased wasted resource-use	Decreased wasted resource-use	Increased dangerous inaction

The classic AIR is expressed in terms of error, specifically, type-I (false positive) and type-II (false negative) error. The classic formulation of the AIR has been one of the most influential arguments favouring the influence of some non-epistemic values in the core stages of science, which may be partly attributable to the simplicity of the trade-off involved. Wilholt's trade-off between reliability and power, on the other hand, consists in a more complex trade-off associated with two epistemic virtues, rather than types of error. This is seen below in Table 2, where I identify the distinct gains and costs associated with a generalized AIR, applying the epistemic/pragmatic distinction as discussed above.

Table 2. Gains and costs associated with the reliability/power trade-off arising in a generalized AIR.

	Narrow band of agnosticism (low suspension of judgement—power is favoured)		*Wide band of agnosticism (high suspension of judgement—reliability is favoured)*	
	Gains	**Costs**	**Gains**	**Costs**
Epistemic	Decreased failed reticence	Increased type-I and/or type-II error	Decreased type-I and/or type-II error	Increased failed reticence
Pragmatic	Decreased wasted resource-use Decreased dangerous inaction	Increased wasted resource-use Increased dangerous inaction	Decreased wasted resource-use	Increased dangerous inaction

With the widening of a band of agnosticism, there is a shift from favouring power to favouring reliability by increasing the suspension of judgement. The increased complexity of the generalized AIR can be observed in the gains and costs associated with it compared to the classic AIR. Notably, in Table 2 (bottom left quadrant) we see that the pragmatic gains and costs associated with decreasing power via narrowing the band of agnosticism (wasted resource-use and dangerous inaction) are unlike in Table 1, in that these gains and costs may both increase *and* decrease simultaneously. This happens in cases where the narrowing of the band of agnosticism entails both increased chances of accepting hypotheses (the narrowing involves making the acceptance threshold less stringent), and increased chances of rejecting them (the narrowing also involves making the rejection threshold less stringent).

For instance, suppose researchers switch from a band of agnosticism stretching from alpha-values between 0.01 and 0.08, to a narrower band residing between 0.03 and 0.06. This increases the chance of accepting hypotheses because 0.03 is larger than 0.01, e.g., when the study result is a p-value of 0.02, researchers previously would have suspended judgment because 0.02 *does not* reside below the 0.01 acceptance threshold, but will now accept the hypothesis because 0.02 *does* reside below the new 0.03 acceptance threshold. But the chance of rejecting hypotheses is also increased because, on the other side of the band, 0.06 is smaller than 0.08, e.g., when the result is a p-value of 0.07, researchers previously would have suspended judgment because 0.07 does not reside above the 0.08 rejection threshold, but will now reject the hypothesis because 0.07 *does* reside above the new 0.6 rejection threshold. Now on one hand, the increased chance of *acceptance* both increases the chance of accepting true hypotheses *and* the chance of accepting false ones; the former of those, in turn, decreases the chance of *wasting resources*, while the latter simultaneously increases such chance. On the other hand, the simultaneous increase in chance of *rejection* that comes with the narrowing band of agnosticism increases both the chance of rejecting false hypotheses *and* the chance of rejecting true ones; the former of those, in turn, decreases the chance of *dangerous inaction*, while the latter simultaneously increases such chance.

The widening of the band of agnosticism does not lead to an increase and decrease of both gains and costs in all cases; this depends on the location of the band of agnosticism. For example, if the researchers select a narrower band stretching from 0.01 to 0.06 rather than 0.01 to 0.08, the chance of accepting hypotheses remains unchanged while the chance of rejecting hypotheses is increased. As a result, the chances of pragmatic costs associated with acceptance (wasted resource-use) also remain the same, while the chances associated with rejection (dangerous inaction) are increased (i.e., the chance of dangerous inaction is both increased and decreased, depending on whether a true hypothesis or a false hypothesis is rejected).

3.2. Failed Reticence in the Treatment of The West Antarctic Ice Sheet in the Fourth IPCC Assessment Report

The treatment of ice sheet loss from disintegration of the West Antarctic Ice Sheet (WAIS) in the IPCC fourth assessment report (AR4) illustrates the trade-offs at stake when scientists decide to suspend judgement. In this section, I examine it as a case where a trade-off between reliability and failed reticence must be determined.

In the "Summary for Policymakers" in AR4, the group of scientists tasked with projecting sea level rise created a table with numerical projections of sea level rise according to several warming scenarios (Table 3 below). These projections were based on global climate models (Keohane et al. 2014). Sea level rise from ice sheet loss occurs by two mechanisms, melting and sliding (or "rapid dynamical changes"). Sliding ice then either melts or forms an iceberg, thus contributing to sea level rise. However, current climate models only effectively predict melting (O'Reilly et al. 2012). As a result, IPCC authors decided to exclude numerical estimates of ice sheet loss due to sliding (seen in the note at the top of the last column: "Model-based range excluding rapid dynamical changes in ice flow"). This decision effectively excluded sea level rise from WAIS disintegration in the numerical estimates, as rapid dynamical loss is the predominant mode of ice sheet loss in the low Antarctic temperatures. The exclusion was concerning, as scientists began to raise concerns about the stability of the WAIS in the 1960s, and "relatively rapid loss" was already being observed in some parts (O'Reilly et al., 2012). Complete disintegration of the WAIS would lead to an estimated 3.3-6.0m sea level rise (O'Reilly et al. 2012).

Table 3. Table of projected sea level rise (m) in the 21st century based on scenarios of temperature increase (°C), published in the Summary for Policymakers of the IPCC Fourth Assessment Report (2008) (original table caption included).

Table SPM.3. *Projected global average surface warming and sea level rise at the end of the 21st century. {10.5, 10.6, Table 10.7}*

	Temperature Change (°C at 2090-2099 relative to 1980-1999)[a]		Sea Level Rise (m at 2090-2099 relative to 1980-1999)
Case	**Best estimate**	***Likely* range**	**Model-based range excluding future rapid dynamical changes in ice flow**
Constant Year 2000 concentrations[b]	0.6	0.3 – 0.9	NA
B1 scenario	1.8	1.1 – 2.9	0.18 – 0.38
A1T scenario	2.4	1.4 – 3.8	0.20 – 0.45
B2 scenario	2.4	1.4 – 3.8	0.20 – 0.43
A1B scenario	2.8	1.7 – 4.4	0.21 – 0.48
A2 scenario	3.4	2.0 – 5.4	0.23 – 0.51
A1FI scenario	4.0	2.4 – 6.4	0.26 – 0.59

Table notes:
[a] These estimates are assessed from a hierarchy of models that encompass a simple climate model, several Earth System Models of Intermediate Complexity and a large number of Atmosphere-Ocean General Circulation Models (AOGCMs).
[b] Year 2000 constant composition is derived from AOGCMs only.

In this case, each reported value can be understood as a particular hypothesis, and the decision to not report or to report the value as the decision to suspend judgement or instead to make a judgement that is conditional on the scenario indicated at the start of the corresponding row in the table. The suspension of judgement involved in the exclusion of numerical projections of sea level rise from WAIS loss comes at the cost of epistemic power. In light of the generalized AIR, this looks like a case in which the decision to suspend judgement does and should stem from reasoning through a trade-off involving epistemic reliability versus epistemic power.

The scientists involved generally appear to believe that the AR4 decision to exclude values from rapid dynamical changes in flow was made on epistemic grounds. They justified the decision in various ways, with one scientist notably stating, "if the science doesn't allow you to give a number that means anything then you have no business giving it" (O'Reilly et al. 2012, 724). This statement includes underlying assumptions about the nature of evidence required, particularly what is required for a number to mean something and involves a favouring of epistemic reliability over epistemic power. In other IPCC working groups, however, projections were sometimes made with little consensus and uncertain data. Moreover, the decision to exclude the values was criticized for inadequately communicating the state of scientific knowledge (Keohane et al. 2014). Relatedly, O'Reilly et al. (2012) argue that the "textual caveat" about the exclusion of numerical estimates for dynamical ice sheet loss was "overshadowed by the numbers, in the minds of many readers" (2012, 711). This comes from a common assumption about the nature of evidence—numerical values are considered a stronger form of evidence, leading readers to defer to the numerical projections and ignore the textual caveat. Together, these concerns suggest that, consistent with AIR assumptions, there may not be an optimal and purely epistemic way to report or not report the future of sea level rise.

Sea level rise is a concern to policymakers for many reasons, including that risk of flooding has led states and municipalities to rely on IPCC projections for infrastructure planning (Keohane et al. 2014). Keohane et al. (2014) claim that the decision to omit numerical estimates of sea level rise attributed to WAIS disintegration in the "IPCC assessment led to confusion and may have inhibited effective planning by coastal communities" (344). This can be analyzed according to my Table 2 above as reflecting two distinct errors: first, the epistemic error of failed reticence in the report leading to confusion and, second, a pragmatic error of dangerous inaction or ineffective planning, both of which are associated with excessive suspension of judgement given the context. In the next section, I argue that distinguishing between these errors clarifies considerations in the evaluation of proposals for the increase of power.

In the WAIS case, increasing epistemic power would lead to a report of some value for sea-level rise due to WAIS loss. This may come at the cost of decreased reliability (increased type-I and/or type-II error), as the true value of sea-level rise is more likely to deviate from the reported value. However, this may be a case where this is the preferable side to err on, as it may give policymakers a more comprehensive understanding of the risk of sea level rise, constituting an epistemic good. More specifically, it may improve the accuracy of understanding magnitude estimates associated with the risks, and concurrent decreases to reliability of probability estimates may be mitigated by a sensitivity analysis, as will be further discussed below. Keohane et al. (2014) propose the incorporation of "subjective expert judgement based on all available evidence" for the generation of a projection that would allow policymakers to "more accurately grasp the risks and the unknowns" (347). Additionally, according to O'Reilly et al. (2012), this could include publishing a range of possible numbers and scenarios, "highlighting [the scientists'] lack of consensus and the high uncertainty surrounding the rapid disintegration of

WAIS" (725). This may resemble Table 3 above, but perhaps with an added column that reports values from WAIS disintegration, stating high uncertainty. The New York City Panel on Climate Change (NPCC) 2009 projections of sea level rise will be examined as an example of the adoption of such a proposal in the next section.

4. Implications for Evidential Standards

4.1. A More Complete Typology of Error

Thus far, I have argued for the application of epistemic/pragmatic distinctions to previous formulations of a generalized AIR and proposed a more complete typology of error. My aims in doing so were two-fold. First, I hoped to contribute to the theoretical discussion of AIRs and further develop the consensus that non-epistemic values should play roles at core stages of the scientific process. My second aim was to highlight factors that may guide changes to evidential standards in practice by allowing considerations to weighed up more effectively. In what follows, I argue that applying my clarified epistemic/pragmatic distinctions aid in the evaluation of legitimate and illegitimate non-epistemic value influences and provide insight into the roles of scientists and policymakers in making non-epistemic value judgements. In the remaining sections, I continue to use the case of reports of sea level rise and examine an alternative approach used in the 2008 New York City Panel on Climate Change (NPCC) that incorporated numerical values for sea level rise from WAIS disintegration in contrast to IPCC AR4.

The table of gains and costs I characterized as emerging in a generalized AIR (Table 2) integrates Steel's (2016) characterization of a generalized AIR in terms of ignorance and Wilholt's (2013, 2016) account of an epistemic trade-off between reliability and power. This allows for a fuller account that retains the advantages of both Steel's and Wilholt's account, and explicitly identifies the gains and costs at stake. Steel's generalized AIR account has the advantage of outlining when scientists ought to favour power, and identifying that the suspension of judgement may, in certain cases, constitute an "epistemic failure" (2016, 706). For policy to be appropriately science-based, according to Steel, scientists must report *informative*, rather than conclusive, findings. This requires that scientists communicate the current state of scientific knowledge to the best of their abilities; as Steel puts it, "[j]ustification in this context entails that scientific knowledge provides reasons for preferring the selected course of action to alternatives" (707). Refraining from making and reporting a particular claim leaves policymakers without scientific guidance for policy decisions relating to the claim.

Wilholt's (2013, 2016) account has the advantage of characterizing the trade-off emerging with the option to suspend judgement. While Steel (2016) emphasizes that the suspension of judgement may constitute an epistemic failure, Wilholt's characterization notes that for a method to be appropriately geared toward the truth, both power and reliability must be considered, and that these are often in tension with each other (2013). The trade-off is epistemically underdetermined and requires an appeal to non-epistemic values. Rather than attempting to determine whether a particular method or finding is sufficiently informative, this characterization highlights the risks at stake in any such decision. Improved understanding of the epistemic considerations at stake by expression as a trade-off can also make required value judgements more transparent, as required by Douglas (2009). My application of an epistemic/pragmatic distinction supplements Wilholt's characterization of a reliability/power trade-off by both acknowledging that pragmatic factors may be involved in such a trade-off, and isolating the epistemic dimension in terms of failed reticence.

4.2. Evaluating Proposals for the Increase of Power

Attending as I have to the epistemic/pragmatic distinction in a generalized AIR can be helpful for the evaluation of proposals for the increase of epistemic power (and correspondingly, reduction of the risk of failed reticence). To evaluate O'Reilly et al's (2012) suggestion to publish a (greater) range of possible values of sea level rise, I will examine the alternative method of the 2008 NPCC that generates a "rapid ice-melt sea level rise" scenario by considering contributions to sea level rise not modelled by global climate models (GCMs). These contributions are estimated by "extrapolation of recent accelerating rates of ice melt from the Greenland and West Antarctic Ice sheets," as well as considering paleoclimate studies (NPCC, 18). Thus, while the numerical projections of the IPCC in AR4 were derived only from GCMs, the 2008 NPCC included an extra row in their table with this rapid ice-melt sea level rise scenario with numerical projections derived from two additional sources (Table 4).

Table 4. The 2008 New York Panel on Climate Change (NPCC) table of sea level rise projections (inches) according to four methods. (Original table caption: "Total Sea Level Rise Projections in Inches for the New York City Region for Four Methods.")

Average (minimum to maximum)	2020s	2050s	2080s	2090s[1]
IPCC Global Estimate + Local Subsidence	NA[2]	NA[2]	NA[2]	(10.4 to 23.4)[5]
IPCC-adapted Methods for the NYC Region	3.7 (1.4 to 5.5)[3]	9.7 (5.0 to 13.6)[3]	17.8 (9.3 to 25.6)[3]	22.2 (14.9 to 30.0)[6]
Rahmstorf/Horton Method + Local Subsidence	4.9 (3.7 to 6.2)[4]	13.1 (10.0 to 16.6)[4]	24.6 (18.2 to 31.6)[4]	28.1 (22.6 to 33.7)[7]
Rapid Ice-Melt Sea Level Rise	~4 to 10[8]	~17 to 30[8]	~37 to 59[8]	~48-70[9]

1 Shown below if a comparison for the 2090s of the four methods, with the baseline years and selection of extreme values calibrated to the IPCC AR4.

2 IPCC projections not available for this time period.

3 The first number shown is the mean based on seven available GCMs and three available emissions scenarios (B1, A1B, and A2). Base period is 2000-2004. The numbers in parentheses represent the minimum and maximum values, respectively, based on the seven available GCMs and three available emissions scenarios.

4 The first number shown is the mean based on eleven available GCMs and three available emissions scenarios. Base period is 2000-2004. The numbers in parentheses represent the minimum and maximum values, as described in Horton et al. 2008.

5 IPCC (2007) projection is from a different set of GCMs. Mean is not available; numbers represent the 5-95% range. Local subsidence is from Peltier (2007), as it is for the other two methods. Meltwater is included in a similar way in both rows 2 and 3. For easy comparison with rows 3 and 4, IPCC AR4 results are only shown for three emissions scenarios (B1, A1B, and A2).

6 Base period is calibrated to 1980-1999 (rather than 2000-2004); the numbers in parentheses represent the second smallest and second largest values, respectively, based on seven available GCMs and three available emissions scenarios; these two values facilitate comparison with the IPCC AR4 methods shown in row 3, since (as in the IPCC AR4) 5 percent of values are higher and 5 percent are lower.

7 Base period is calibrated to 1980-1999 (rather than 2000-2004); the numbers in parentheses represent the third smallest and third largest values, respectively, based on the eleven available GCMs and three available emissions scenarios; these values facilitate comparison with the IPCC AR4 methods shown in row 3, since 6 percent of values are higher and 6 percent are lower.

8 Numbers shown represent GCM minimim and maximum values across two rapid ice-melt component ranges shown in Table 11.

9 Base period is calibrated to 1980-1999 (rather than 2000-2004); the numbers represent the second smallest and second largest values, respectively, based on the seven available GCMs and three emissions scenarios calculated using the two rapid ice-melt component ranges.

The NPCC is an advising body to the mayor of New York City and the task force responsible for considering climate change and adaptation, specifically regarding infrastructure (NPCC, 2008). It resembles the IPCC in both function and method, as can be seen from the adaptation of IPCC estimates for the region in the first two rows (Table 4). The decision of the NPCC to include additional factors into projections of sea level rise arose from the recognition that factors not modeled by GCMs contribute to sea level rise, and out of concerns that "the IPCC approach to sea level rise may substantially underestimate the range of possible increases" (NPCC, 13). However, due to the high level of uncertainty associated with these methods in the rapid ice-melt sea level rise scenario, NPCC authors did not assign either a quantitative probability or a qualitative likelihood (Appendix C, 52). They did note, however, that these values are comparable to other recent estimates (citing Pfeffer et al. 2008 and Grinsted et al.

2009). These factors are an example of the manner in which including estimates of additional contributions to sea level rise may provide a more comprehensive understanding of risk. In addition to improving understanding of the magnitude estimates, labelling the estimates as having high uncertainty allows for a sensitivity analysis that helps to preclude worries about accuracy.

Consistently with other versions of the AIR, I argue that permissible increases of power must be epistemically adequate, and that such increases are called for when favoured by the balance of non-epistemic values. In such cases, scientists have both epistemic and pragmatic reasons to incorporate sources such as those examined by the NPCC into their estimates. While excluding such estimates altogether may improve reliability, the appropriate balance between reliability and power is epistemically underdetermined, and pragmatic reasons may favour the path of reporting these additional possibilities. While both epistemic and pragmatic reasons play an important role in the selection of the evidential standard, these must be evaluated as conceptually distinct to first determine whether the inclusion of an additional scenario of numerical estimates is supported by necessary epistemic reasons.

Steel's (2016) generalized AIR characterization also requires respecting epistemic constraints and may be consistent with Wilholt's (2013) account. An aim of my account is to make explicit factors that allow the generalized AIR to be more compelling to scientists. For example, distinguishing epistemic and pragmatic errors allows for the concerns of scientists committed to providing high quality empirical knowledge to be addressed. The claim of the scientist working on the sea level rise in IPCC AR4, that "if the science doesn't allow you to give a number that means anything then you have no business giving it" (qtd. in O'Reilly et al. 2012, 724), involves assumptions about the amount and type of evidence required to generate a numerical projection, as "giv[ing] a number" requires the adoption of evidential standards, which according to the AIR cannot be derived from the empirical evidence alone. However, the claim also reflects an important concern about unduly increasing power. Prizing apart the epistemic and pragmatic risks associated with increasing power allows for an initial rejection of certain increases of power at the expense of reliability. More specifically, it isolates epistemic considerations in the adoption of evidential standards, while simultaneously acknowledging a role for non-epistemic values in the adoption of such standards. When a scientist gives epistemic reasons for not including certain numerical estimates (as the scientist quoted above does), a trade-off of epistemic risk allows for a response with epistemic reasons.

A generalized AIR that incorporates an epistemic/pragmatic distinction can be used to argue that 2008 NPCC report's predictions of sea level rise may be preferable to the report of estimated sea level rise in IPCC AR4. AIRs are an influential group of arguments for the permissibility of incorporating non-epistemic values into core stages scientific investigation, and such accounts are often phrased in terms of influence that respects epistemic constraint. In Steel (2016), non-epistemic values ought to influence scientific processes "in a manner that is compatible with the epistemic integrity of science" (712). However, this may be a bit of a non-starter to scientists like the one quoted above, who perhaps have not been convinced that non-epistemic values can influence core scientific processes in a manner that preserves the epistemic integrity of science. While philosophers of science have generally moved away from the VFI, related assumptions are still common throughout the sciences. Characterizing the trade-off in terms of error highlights, as Wilholt (2013) argues, that such decisions are epistemically underdetermined, and thus may make the appeal to non-epistemic reasons more plausible to scientists.

4.3. Evaluating Informativeness and Definitiveness of Findings

While according to Steel's (2016) generalized AIR the inclusion of additional numerical estimates could be accounted for by claiming that the estimates reported in IPCC AR4 were insufficiently informative, an epistemic/pragmatic distinction may help to provide common ground in disagreements about what constitutes informativeness. More specifically, it complements Steel's informativeness criterion that requires that scientists report findings that are informative rather than conclusive. It does this by providing criteria for the evaluation of whether a particular manner of increasing power is informative or not. From this, it is possible to conclude, for example, that if the epistemic reasons are sufficiently strong for an increase of power, then such an increase is permissible. In such a case, the risk of committing the error of failed reticence is sufficiently weighty in the trade-off with type-I and/or type-II error.

That previous estimates of sea level rise, particularly by the IPCC, have been conservative and underestimated actual changes is an epistemic reason for the inclusion of values derived from sources other than GCMs. Additionally, the NPCC rapid ice-melt scenario "takes into account the potential for a substantial increase in the rate of melting based on recent observations of accelerated icemelt, new scientific understanding of icemelt dynamics, and paleoclimate studies" (NPCC, 52), and is not merely an addition of an arbitrary number out of concern for the consequences associated with sea level rise. Pragmatic reasons, such as concerns about the consequences of sea level rise for infrastructure, for example, ought to be secondary to such epistemic reasons in the selection of an evidential standard. Thus, attending to an epistemic/pragmatic distinction clarifies a factor promoting informativeness, namely the epistemic character of the manner in which power is increased.

While it remains a challenge in practice to evaluate precisely where a method of increasing power falls on the epistemic/pragmatic spectrum, the conceptual distinction is helpful as a starting point of an account of how pragmatic factors do and ought (and ought not) to impinge on weighing the epistemic dimensions of the reliability/power trade-off. The concern of the scientists in IPCC AR4 can be analyzed as a concern about pragmatically driven increases of power that lead to a variation of evidential standards to increase the rate of findings, which in turn may lead to undesirably high error rates. Had previous IPCC estimates of sea level rise not systematically erred on the side of underestimation, the epistemic reasons to include sources beyond GCMs in the 2008 NPCC report may have been weaker. Providing a more complete typology of error prevents conflation of epistemic and pragmatic notions of power, as it allows for the evaluation of whether it is failed reticence (epistemic) or dangerous inaction (pragmatic) that is being weighed against decreased reliability. Pragmatic risks may be relevant to the selection of an evidential standard, but ought not override epistemic considerations.

In addition to clarifying ambiguities in the informativeness criterion, the distinction may clarify Wilholt's characterization of power as the "rate at which a scientific investigation delivers definitive results" (244). Wilholt (2013) adopts the notion of power from Alvin Goldman (1992), but with a shift from characterizing it as a rate of *true* findings to a rate of *definitive* findings. This slight distancing from truth introduces some ambiguity and appears to allow for the incorporation of a more pragmatic notion of power. The position of the scientist quoted above can be characterized as asserting that no matter the weight of consequences associated with dangerous inaction, these cannot be weighed directly against the epistemic (type-I and type-II) errors associated with decreased reliability.

Additionally, I suggest that the epistemic/pragmatic distinction is helpful in that it provides grounds for evaluating and criticizing the conventional standards that are necessary to strike a balance between reliability and power on Wilholt's (2013) account. More specifically, the error of failed reticence may articulate one problem with current evidential standards in climate science, as scientists and philosophers raise concerns about rampant suspension of judgement in climate science leading to policy inaction (Hanson 2007; Brysse et al. 2013; Lloyd et al. 2021). Elabbar (forthcoming) distinguishes between a "bias charge" (toward false negatives) and a "reticence charge" related to rampant suspension of judgement (7). Steel raises a concern about the sufficiency of conventional standards in his discussion of the levels of organization at which non-epistemic values ought to play a role (2016). I believe a similar concern about accepted conventions about methodological standards is evident in criticisms about conservatism and bias toward false negatives in climate science. Such conventions can be criticized as permitting excessive rates of the epistemic error of failed reticence.

The application of an epistemic/pragmatic distinction also suggests different roles for scientists and policymakers in making non-epistemic value judgements, which are required of those in both roles on this account. The role of the scientist is to ensure that the findings reported allow the ensuing policy to be science-based, and the role of the policymaker is then to consider additional pragmatic factors in determining the most appropriate course of action. In the case of the WAIS, this may require expanding sources evaluated for the generation of numerical projections of sea level rise beyond model projections, while the policymaker may then evaluate the risks associated with dangerous inaction and, for example, decide to refrain from building certain infrastructure in risky locations due to additional pragmatic considerations.

5. Conclusion

In this paper, I argued that adequately characterizing a generalized AIR requires clarifying distinct epistemic and pragmatic dimensions of trade-offs that emerge when suspension of judgement becomes an option in the treatment of hypotheses. This, in turn, allows for clarification of the considerations at stake, allowing for these factors to be weighed up more rationally. Using the case of the IPCC treatment of sea level rise from WAIS loss, I distinguished the epistemic risk associated with decreased power—failed reticence—from the pragmatic risk associated with decreased power—dangerous inaction. This account may be used to provide an initial characterization of the cases in which evidential standards may be varied to increase power according to non-epistemic values, as proposed by Elabbar (forthcoming). These distinct risks have differing implications for scientists and policymakers, which are highlighted by distinguishing epistemic and pragmatic dimensions of trade-offs that emerge when judgement is suspended on hypotheses.

References

Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3 (2): 207–20. https://doi.org/10.1007/s13194-012-0062-x.

———. "Why the Argument from Inductive Risk Doesn't Justify Incorporating Non-Epistemic Values in Scientific Reasoning," in *Current Controversies in Values and Science*, eds. Elliott, Kevin C., and Daniel Steel. 1st ed. Routledge.

Brysse, Keynyn, Naomi Oreskes, Jessica O'Reilly, and Michael Oppenheimer. 2013. "Climate Change Prediction: Erring on the Side of Least Drama?" *Global Environmental Change* 23 (1): 327–37. https://doi.org/10.1016/j.gloenvcha.2012.10.008.

Churchman, C. West. 1948. "Statistics, Pragmatics, Induction." *Philosophy of Science* 15 (3): 249–68. https://doi.org/10.1086/286991.

Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. https://doi.org/10.1086/392855.

———. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. https://doi.org/10.2307/j.ctt6wrc78.

———. 2015. *Values in Science*. Edited by Paul Humphreys. Vol. 1. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199368815.013.28.

———. 2017. "Why Inductive Risk Requires Values in Science," in *Current Controversies in Values and Science*, eds. Elliott, Kevin C., and Daniel Steel. 1st ed. Routledge. https://doi.org/10.4324/9781315639420.

Elabbar, Ahmad. Forthcoming. "Varying Evidential Standards as a Matter of Justice." *The British Journal for the Philosophy of Science*, August, 727429. https://doi.org/10.1086/727429.

Goldman, Alvin. 1992. *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, Cambridge, MA: MIT Press.

Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions. 2015. "The Extent and Consequences of P-Hacking in Science." *PLOS Biology* 13 (3): e1002106. https://doi.org/10.1371/journal.pbio.1002106.

Hansen, J E. 2007. "Scientific Reticence and Sea Level Rise." *Environmental Research Letters* 2 (2): 024002. https://doi.org/10.1088/1748-9326/2/2/024002.

Helgeson, Casey, Parker, Wendy & Tuana, Nancy. Forthcoming. "How Uncertainty Interacts with Ethical Values in Climate Change Research." In Linda Mearns, Chris Forest, Hayley Fowler, Robert Lempert & Robert Wilby (eds.), *Uncertainty in Climate Change Research: An Integrated Approach*. Springer.

Hempel, Carl G. 1965. "Science and Human Values", in *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: The Free Press, 81–96.

Intergovernmental Panel on Climate Change. 2007. "Summary for Policymakers." *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.

Keohane, Robert O., Melissa Lane, and Michael Oppenheimer. 2014. "The Ethics of Scientific Communication under Uncertainty." *Politics, Philosophy & Economics* 13 (4): 343–68. https://doi.org/10.1177/1470594x14538570.

Lloyd, Elisabeth A., Naomi Oreskes, Sonia I. Seneviratne, and Edward J. Larson. 2021. "Climate Scientists Set the Bar of Proof Too High." *Climatic Change* 165 (3–4): 55. https://doi.org/10.1007/s10584-021-03061-9.

Longino, Helen. 2017. "Values, Heuristics and the Politics of Knowledge." *Scientiae Studia* 15 (1): 39. https://doi.org/10.11606/51678-31662017000100003.

New York Panel on Climate Change. 2009. "Climate Risk Information." Report for the Mayor's Office of Long Term Planning and Sustainability, 17 February.

O'Reilly, Jessica, Naomi Oreskes, and Michael Oppenheimer. 2012. "The Rapid Disintegration of Projections: The West Antarctic Ice Sheet and the Intergovernmental Panel on Climate Change." *Social Studies of Science* 42 (5): 709–31. https://doi.org/10.1177/0306312712448130.

Oreskes, Naomi, and Erik M. Conway. 2022. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Climate Change*. Paperback edition, Nachdruck. Bloomsbury.

Peacock, Kent A. 2018. "A Different Kind of Rigor: What Climate Scientists Can Learn from Emergency Room Doctors." *Ethics, Policy & Environment* 21 (2): 194–214. https://doi.org/10.1080/21550085.2018.1509483.

Rahmstorf, Stefan. 2007. "A Semi-Empirical Approach to Projecting Future Sea-Level Rise." *Science* 315 (5810): 368–70. https://doi.org/10.1126/science.1135456.

Rooney, Phyllis. 2017. "The Borderlands Between Epistemic and Non-Epistemic Values," in *Current Controversies in Values and Science*, eds. Elliott, Kevin C., and Daniel Steel. 1st ed. Routledge. https://doi.org/10.4324/9781315639420.

Rudner, Richard. 1953. "The Scientist *Qua* Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6. https://doi.org/10.1086/287231.

Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77 (1): 14–34. https://doi.org/10.1086/650206.

———. 2015. “Acceptance, Values, and Probability.” *Studies in History and Philosophy of Science Part A* 53 (October):81–88. https://doi.org/10.1016/j.shpsa.2015.05.010.

———. 2016. “Climate Change and Second-Order Uncertainty: Defending a Generalized, Normative, and Structural Argument from Inductive Risk.” *Perspectives on Science* 24 (6): 696–721. https://doi.org/10.1162/POSC_a_00229.

———. “Qualified Epistemic Priority: Comparing Two Approaches to Values in Science,” in *Current Controversies in Values and Science*, eds. Elliott, Kevin C., and Daniel. 1st ed. Routledge. https://doi.org/10.4324/9781315639420.

Steel, Daniel, and Kyle Powys Whyte. 2012. “Environmental Justice, Values, and Scientific Expertise.” *Kennedy Institute of Ethics Journal* 22 (2): 163–82. https://doi.org/10.1353/ken.2012.0010.

Wilholt, Torsten. 2013. “Epistemic Trust in Science.” *The British Journal for the Philosophy of Science* 64 (2): 233–53. https://doi.org/10.1093/bjps/axs007.

———. 2016. “Collaborative Research, Scientific Communities, and the Social Diffusion of Trustworthiness.” In *The Epistemic Life of Groups*, edited by Michael S. Brady and Miranda Fricker, 218–34. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198759645.003.0012.

Winsberg, Eric. 2012. “Values and Uncertainties in the Predictions of Global Climate Models.” *Kennedy Institute of Ethics Journal* 22 (2): 111–37. https://doi.org/10.1353/ken.2012.0008.

Endnotes

[1] The argument does not require an exacting and context-independent distinction between epistemic and non-epistemic values. While the classification of values may be ambiguous and context-dependent (Rooney, 2017; Steel, 2010), a conception of certain values as truth-oriented vs. those oriented toward values other than truth may be central to empirical science. For example, the manipulation of scientific investigation to acquire results that coincided with political and economic ends in the denial of the harmful effects of tobacco smoke and climate change can be attributed to inappropriate non-epistemic value influences (Oreskes and Conway, 2010).

[2] There are two ways to read the "must" in this premise: one is more modal—that it is part of the definition of the scientist that they must make these decisions (for this claim see Rudner, 1953), and the other is normative—that scientists ought to make such decisions (for this version see Douglas, 2009). For an objection to both interpretations, see Betz (2013), and for a response to Betz's objection, see Steel (2016).

[3] This requirement is moral on Douglas' (2009) account but could be understood more broadly as a rational or social requirement. There have been objections to versions of this premise (see Betz, 2017); for a defense, see Douglas (2017).

[4] The notion of failed reticence is clarified in later sections.

[5] Two characterizations of failed reticence are possible. The first is as an epistemic error about an epistemic issue (i.e., one commits the error when one fails to make judgements that one is in a position to make), and the second is as an epistemic error about a metaphysical issue (i.e., one fails to make a judgement on any true or false claim). The problem with the first option is that it appears to undermine the starting assumptions for the AIR, namely that evidential sufficiency cannot be determined based on epistemic considerations alone. As a result, my characterization will be as an epistemic error about a metaphysical issue. A concern related to this second option is the implication that one would be in error at all times, since there are an infinite number of truths one is not accepting (or falsehoods one is not rejecting) at any given time. There are two options to navigate this: either to bite the bullet and accept that one is constantly in error, but only the ones we are in a position to address are relevant, or to characterize failed reticence as only applying to the claims one is in a position to address. While I have a slight preference for the second, I will not further develop either characterization in this paper.