Synthese
DOI 10.1007/s11229-017-1617-3

S.I. : EVIDENCE AMALGAMATION IN THE SCIENCES

Ethics of the scientist qua policy advisor: inductive risk, uncertainty, and catastrophe in climate economics

David M. Frank[1,2]

Received: 13 February 2017 / Accepted: 4 November 2017

Abstract This paper discusses ethical issues surrounding Integrated Assessment Models (IAMs) of the economic effects of climate change, and how climate economists acting as policy advisors ought to represent the uncertain possibility of catastrophe. Some climate economists, especially Martin Weitzman, have argued for a precautionary approach where avoiding catastrophe should structure climate economists' welfare analysis. This paper details ethical arguments that justify this approach, showing how Weitzman's "fat tail" probabilities of climate catastrophe pose ethical problems for widely used IAMs. The main claim is that economists who ignore or downplay catastrophic risks in their representations of uncertainty likely fall afoul of ethical constraints on scientists acting as policy advisors. Such scientists have duties to honestly articulate uncertainties and manage (some) inductive risks, or the risks of being wrong in different ways.

Keywords Climate change · Economics · Catastrophic risk · Inductive risk · Environmental ethics · Research ethics

1 Introduction

How should climate economists acting as policy advisors deal with the uncertain possibility of climate catastrophe? Martin Weitzman and others have urged that avoiding

✉ David M. Frank
dfrank4@utk.edu; david.moorfield.frank@gmail.com

1 Department of Philosophy, University of Tennessee, Knoxville, TN, USA

2 Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA

Published online: 17 November 2017

catastrophe should frame climate economists' welfare analysis.[1] On this precautionary approach, we should think about climate policy as paying to reduce the risk of catastrophe as opposed to "smoothing consumption" over time. This paper discusses ethical assumptions that justify this approach and argues that Weitzman's "fat tail" probabilities of climate catastrophe pose an ethical problem for widely used Integrated Assessment Models (IAMs) of the economic effects of climate change.[2] Economists who ignore or downplay catastrophic risks in their representations and amalgamations of uncertainty may fall afoul of ethical constraints on scientists acting as policy advisors. Such scientists have duties to honestly articulate uncertainties and manage (some) inductive risks, or the risks of being wrong in different ways. However, considerations related to Weitzman's own "dismal theorem" show that this precautionary approach has unique ethical limitations as well, especially due to the difficult threshold analysis required by this approach in setting the "value of statistical civilization".

In Sect. 2, I present Weitzman's argument against commonly used IAM damage functions, showing how he motivates his precautionary approach. Here I discuss his "dismal theorem," which shows that fat-tail catastrophes can dominate climate cost-benefit analysis at the limit, driving willingness to pay to reduce the risk of catastrophe to infinity or a high upper bound. In Sect. 3, I offer an argument that, since scientists working as policy advisors have duties to honestly articulate uncertainties and manage inductive risks, or the risks of being wrong in various ways, they may be ethically obligated to include fat-tail catastrophic risks in their economic analyses of climate change, since these models have ethically significant consequences when used in policy-making or governmental cost-benefit policy analysis. In Sect. 4, I discuss objections and limitations.

2 Climate economics and "fat tail" climate catastrophes

Most contemporary discussion of ethics in climate economics has focused on two families of ethical issues: intertemporal discounting (intergenerational justice) and cost allocation for mitigation and adaptation (intragenerational justice). These issues of equity arise in the context of IAMs used to determine the "optimal" price for carbon by estimating the marginal social cost of emissions, for example Nordhaus's DICE and RICE models (Nordhaus 2013). These models use the Ramsey equation to determine the discount rate, which, as Jamieson (2014, pp. 118–122) and others have demonstrated, encodes discounted utilitarian value judgments about intertemporal and intratemporal inequality, as well as risk, a third important area of ethical concern and the focus of this paper.[3] Outside the context of the Ramsey equation, IAMs deal

[1] See Weitzman (2007, 2009, 2012, 2014). Throughout, *welfare analysis* refers to economic cost-benefit analysis and its components, e.g. attempts to estimate the social cost of carbon.

[2] The definition of 'fat tail' used by Weitzman (2009) is that a probability density function has a fat tail if the tail approaches 0 more slowly than exponentially, i.e. when its moment-generating function is infinite.

[3] The Ramsey equation explicitly encodes value judgments about intertermporal equity insofar as it attempts to determine the optimal social discount rate for welfare analysis. It also implicitly encodes value judgments about intratemporal equity because the parameter for the elasticity of marginal utility of consumption, often interpreted as a measure of inequity aversion, could be applied to the intratemporal case. But as Frisch (2017)

with risk by estimating economic damages of climate change and parametrizing risk aversion. This paper focuses here, specifically on the "damage functions" of IAMs. Figure 1 below depicts the default damage functions of three main IAMs, which were used by the US Federal Government's Interagency Working Group under the Obama administration to estimate the cost of carbon for regulatory rulemaking cost-benefit analysis (Interagency Working Group 2010; Greenstone et al. 2011).

IAMs generally assume that our economies will continue to grow and the economic future will be roughly like the economic past. However, it is also possible that anthropogenic climate change will lead to radically disruptive, catastrophic scenarios, e.g. massive population reductions due to emerging diseases, resource scarcity, or the collapse of ecosystems, the end of human societies as we know them, perhaps even human extinction. For example, one catastrophic scenario is explored in Oreskes and Conway's (2014) speculative future history, where climate change leads to the "collapse of western civilization." Leaving out details, it is widely assumed that warming of 6 or 7 °C would make such catastrophic outcomes more likely than 2 °C.[4]

The idea that increasing warming will lead to increasing economic damages is built into the damage functions, which take information about climate change and economic growth under various scenarios and map degrees warming onto economic damages for welfare analysis (Fig. 1). Weitzman (2009) argues that commonly used damage functions do not sufficiently take into account catastrophic possibilities if they extrapolate deterministically from damage estimates at low amounts of warming. These extrapolations are based on functional forms (quadratic, exponential) chosen largely for two reasons: they fit the assumption that damages will increase faster than linearly in temperature, and they are analytically tractable. No deeper scientific rationale for the shape of these functions has been proposed (Weitzman 2009, p. 16; Pindyck 2013). Furthermore, except for the PAGE IAM, these damage functions do not explicitly (probabilistically) represent any relevant uncertainties.

Of course, uncertainties surrounding IAMs come from multiple sources, since there are several inferences that must be made in an IAM to get to an estimate of economic damages. On the natural scientific side, one must infer warming from atmospheric CO_2; on the social scientific side, damages from warming. On the former, uncertainties include those surrounding the climate sensitivity parameter, defined as the amount of warming expected given a doubling of atmospheric CO_2, as well as possible "tipping points" beyond which the climate system might respond in discontinuous and potentially disastrous ways. These unlikely but high-impact scenarios include runaway warming due to massive methane release or collapse of the Atlantic Meridional

Footnote 3 continued
points out, IAMs that simply ignore intratemporal inequality (i.e. global wealth inequality and resulting unequal vulnerability) by modeling society as a series of representative consumers, thereby arguably encode non-egalitarian value judgments as well, since decisions made on their basis would likely tend to maintain such inequalities.

[4] Determining which outcomes count as "catastrophic" is as much a value judgment as determining what counts as "dangerous anthropogenic interference with the climate system." For the purposes of this paper, I need not precisely define 'catastrophe.' The kind of outcomes that are relevant to Weitzman's argument, involving "unlimited downside exposure," collapse of civilization, etc., would be agreed by all reasonable parties to be catastrophic.

Table 1 From Wagner and Weitzman (2015, p. 54)

CO_2e concentration (ppm)	500	550	600	650	700	750	800
Median temperature increase °C	2.2	2.5	2.7	3.2	3.4	3.7	3.9
Chance of > 6 °C increase (%)	1.2	3	5	8	11	14	17

Overturning Circulation. On the latter, social scientific side, uncertainties include the role of technology in adaptation and mitigation, the substitutability of natural and human-made capital, and effects of extreme warming on resource scarcity and global security.

First, consider uncertainties surrounding climate sensitivity, or more generally the relationship between emissions and overall warming. According to Weitzman's recent amalgamation of the IPCC's climate models, even if the world stabilized at 550 ppm CO_2e, there would still be a significant, non-zero probability ($\sim 3\%$) of exceeding 6 °C (11 °F) warming (see Table 1 below, reproduced from Wagner and Weitzman 2015). By 700 ppm, the chance is 11%. However, as I discuss below, Weitzman (2009) argues that even if one were to use a relatively thin-tailed probability density function (PDF) for climate sensitivity (e.g. the Normal distribution), uncertainties about feedbacks like those mentioned above (methane release, massive failure of carbon sinks, etc.) should "fatten" the tails of this PDF if they were incorporated in a Bayesian fashion. This resulting PDF is much more uncertain about the reaction of the climate to emissions.

Second, consider uncertainties surrounding the damage function itself, which takes degrees warming and maps it onto economic damages, usually in terms of loss of global GDP. (Put aside normative limitations of focusing on GDP as the sole measure of economic damages). Figure 1 illustrates existing disagreement amongst climate economists; for example, up to about 3.5 °C warming, the most optimistic of the damage functions (FUND) lies outside of the 95th percentile of damages predicted by PAGE. Furthermore, many different damage functions can be fit to existing data that have widely divergent results at higher degrees of warming. This is demonstrated by Weitzman (2012, pp. 234–235), who compares the popular quadratic damage function with one that is more "reactive" at higher degrees warming. As Frisch (2013) argues, the deepest problem with the damage functions on offer is that, if they are based on empirical data at all, they must use data on the effects of very small climate changes around recent averages on economic variables, to infer damages at much greater climate changes with greater variability. Even the most advanced recent work on this topic, for example Burke et al. (2015), only use data from recent years (in their case, 1960–2010), comparing individual countries' economic output in cooler-than-average vs. warmer-than-average years. Using this data set to extrapolate out of sample introduces significant uncertainty.

Given the inherent uncertainties of such extrapolation, one might reasonably worry whether we can honestly infer from damages at 1 or 2 °C to damages at 6°. Indeed, Tol (2012), who is relatively optimistic about the economic effects of climate change, attempted to represent existing uncertainty in damage functions by estimating a PDF for damages by "vote counting" from existing studies, and found that greater than

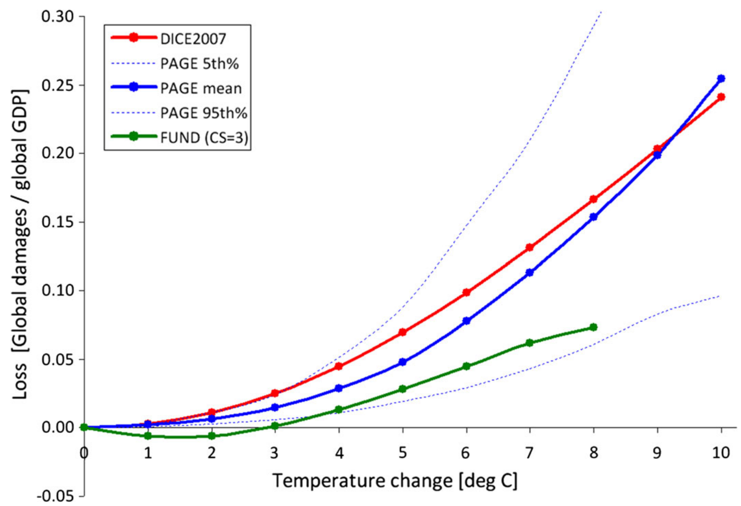

Fig. 1 Damage functions for three widely used IAMs, from Interagency Working Group (2010). Note that only PAGE's function incorporates probabilistic uncertainty

7 °C of warming could lead to a "total welfare loss," (97) obviously much greater than losses predicted by the functions shown in Fig. 1. Weitzman's "dismal theorem" (2009) shows that for a simplified case, high-impact, low-probability catastrophes in the "fat tail" of a PDF for damages include the possibility of "unlimited downside exposure" or the complete elimination of consumption or GDP, and can thus dominate cost-benefit analysis (CBA).[5]

As mentioned above, Weitzman's central claim is that this "immense cascading of huge uncertainties" surrounding the inferences in IAMs (the climate sensitivity parameter, economic damages, etc.) ought to be explicitly represented in Bayesian fashion, resulting in a fat-tailed posterior PDF of what he calls "welfare sensitivity" to climate change (2009, p. 6). Putting the argument informally, the idea is that "fat tails conjoined with fat tails beget yet-fatter tails." (2009, p. 4) More formally, the posterior-predictive distribution Weitzman derives for this idealized welfare sensitivity random variable for a simple case is the Student-t distribution, which has an undefined or infinite expectation. Thus deep uncertainty generates a (negative) infinite expectation in an expected utility calculation, so considerations of catastrophic risk can dominate CBA of climate change for sufficiently risk-averse decision-makers. The absurd result of the dismal theorem is that a risk-averse decision-maker should be willing to pay an unlimited amount of money at the margin to avoid catastrophe ("unlimited downside exposure" due to the collapse of human civilization, etc.).[6]

Weitzman's result poses technical and ethical problems for fat-tailed CBA (of climate change and other social decisions with existential implications), as well as ethical problems for commonly used IAMs. Weitzman suggests avoiding the absurd result of

[5] As Nordhaus (2012) points out, this follows from Weitzman's definition of a "fat tailed" distribution as one whose moment-generating function is infinite, i.e. its expectation is infinite.

[6] Technically, the decision-maker's utility function exhibits "Constant Relative Risk Aversion".

the dismal theorem requires setting an upper bound on the disutility of catastrophe akin to the commonly used "value of a statistical life" (VSL) in CBAs involving mortality risks. VSL, a controversial economic construct, represents the (monetary) value of saving a human life. It is often inferred from surveys or economic behavior involving mortality risks and is commonly used in climate economics, health economics, and other areas that involve risks of death. One might "scale up" a suitable VSL to come up with a finite societal willingness to pay to avoid total loss. Weitzman interprets this theoretical upper bound as an analogous "value of statistical civilization as we know it" (2009, p. 6). While he is doubtful that any particular value could be given an ethically sound rationale, it seems likely that a wide range of reasonable values could be contemplated, and quite possible that even the higher estimates for aggressive mitigation costs would be lower than society's aggregate willingness to pay to avoid the loss of human civilization. However, setting such an economic "value of statistical civilization" inherits the ethical problems inherent in VSL-like inferences, for example whether it is normatively legitimate to infer a VSL from empirically observed behavior involving mortality risks in the first place. Additional ethical complications arise due to the scale of the task, for example, whether the value of statistical civilization is merely the sum of the value of the statistical lives involved, or whether there ought to be non-additive effects. Given the sensitivity of any analysis to different values of this controversial and under-explored construct, this is a major challenge to such an analysis. On the other hand, an important ethical problem indicated for commonly used damage functions is that they risk being far too optimistic insofar as they do not adequately incorporate the uncertain fat tail risks of total catastrophe. The DICE, FUND, and PAGE damage functions with their default modeling assumptions shown in Fig. 1, for example, are clearly more optimistic than Weitzman's probabilistic model of welfare sensitivity.

In a recent study, Dietz (2011) does not attempt to estimate the VSL-like parameter directly, but follows Weitzman's advice by performing an empirical fat-tailed climate CBA with the stochastic IAM PAGE. Dietz suggests that while we might try "scaling up" the value of a statistical life to entire societies to produce this upper bound, instead his CBA works with a parameter representing the "maximum proportion of net consumption that could feasibly be lost" to bound the analysis. Thus instead of assuming that the decision maker has some finite willingness to pay to avoid total loss, it simply truncates the space of feasible outcomes to exclude total loss at some high percentage, e.g. loss of 99% of consumption relative to the baseline. Table 2 below shows Dietz's results of the marginal damage costs per ton of CO_2 (2008 $US) when this parameter is set at .99, for different (relatively low) utility discount rates and values for the elasticity of marginal utility of consumption.

These results show that for an empirical fat-tailed climate CBA, the specification of parameters in the Ramsey equation like the discount rate still have significant effects on results. Catastrophic risk does not eliminate the effect of the discount rate, especially since plausible climate catastrophes are 100 years or more away. While Weitzman is right that amalgamating uncertainty by incorporating catastrophes via fat tails dramatically increases the mean estimated social cost of carbon, the uncertainty around these values is also notably large. Considering mean values, while they are not exactly comparable since they used higher discount rates and 2007 $US for the

Table 2 From Dietz (2011, p. 537). Estimates of marginal damage costs of t/CO2 2008 $US upper bound on proportion of consumption lost at .99

Scenario	Discount rate	Elasticity of marginal utility of consumption	5% marginal damage cost	Mean marginal damage cost	95% marginal damage cost
Business-as-usual emissions (IPCC A2)	.1	1	21.13	444.65	1861.64
	1.5	3	4.67	346.21	1358.54
550 ppm stabilization (IPCC A2)	.1	1	26.70	377.61	1740.83
	1.5	3	4.49	224.41	335.10

year 2010, none of the US Government's Interagency Working Group's estimates for the same IAM (PAGE) for the year 2010 closely approach Dietz's mean values. For example, they estimate at a 2.5% discount rate the mean 550 ppm stabilization social cost of a ton of carbon in 2010 to be $42.9 (Interagency Working Group 2010, p. 26). The details and particular numbers are less important than the point that amalgamating the uncertain possibility of catastrophe via a fat-tailed damage PDF greatly increases mean estimates for the social cost of carbon, and its uncertainty, other things equal. The most optimistic IAMs, e.g. Tol's FUND model, tend to estimate the social cost of carbon much lower.

Weitzman claims that this fat-tailed approach to climate CBA is a version of (or, perhaps, is entailed by) a suitable precautionary principle. For those who do not take the precautionary principle and CBA to be mutually exclusive approaches, this makes intuitive sense, since proponents of the precautionary principle have argued that we ought to take (cost-effective) steps to avoid plausible environmental catastrophes, even when the probabilities of such catastrophes are unknown or unknowable (Steel 2015). Weitzman's approach, which still uses expected utility theory, is also precautionary insofar as it amalgamates catastrophic uncertainty into the PDF for welfare sensitivity or damages, thus increasing the decision-relevance of climate catastrophes whose probabilities are low but not precisely known. Weitzman states that,

> The general point is that [the dismal theorem] embodies a very strong form of a "generalized precautionary principle" for situations of potentially unlimited downside exposure. From experience alone one cannot acquire sufficiently accurate information about the probabilities of disasters in the bad tail to make [expected utility] independent of the VSL-like parameter—thereby potentially allowing this VSL-like-parameter aspect to dominate CBA applications of [expected utility] theory under conditions of potentially unlimited liability. (2009, p. 12)

Here Weitzman explains how his dismal theorem embodies a precautionary approach to expected utility CBA: when catastrophic loss is possible, and the probability of this loss can only be roughly represented with fat tailed PDF, the decision-maker is forced to explicitly consider a "VSL-like-parameter," i.e. their willingness to pay to avoid catastrophic loss. Thus, although the analogy is not perfect, Weitzman has framed this approach to climate CBA as "insurance" against catastrophe (Weitzman 2012). Even when couched in the terms of expected utility theory, this approach is precautionary in the way it represents catastrophic uncertainty. Other approaches attempt to take into account catastrophic possibilities in more optimistic ways. For example, the damage function of Nordhaus's DICE model incorporates economic impact of high-impact, low-probability catastrophes via a "survey of experts" (but formally rules out total loss) and the version of PAGE used by the Interagency Working Group includes a "damage sub-function" for catastrophic impacts (Interagency Working Group 2010, pp. 6–7). However, these methods do not result in nearly as high estimates for the social cost of carbon as the fat-tailed approach does; more importantly, they do not frame the analysis specifically on what we are willing to pay to avoid worst-case scenarios in the fat tails.

What might justify Weitzman's fat-tailed approach as opposed to those used, for example, by the Interagency Working Group? The next section constructs an ethical argument for Weitzman's fat-tailed precautionary approach to the amalgamation of uncertainty, by appealing to duties of scientists acting as policy advisors.

3 Ethics of climate economics: representing uncertainty and managing inductive risks

This paper assumes that use of economic CBA of climate change policy could be justified on broadly welfarist grounds, and policy-makers ought to avail themselves of such analyses for decision support.[7] I do not assume that CBA ought to determine policy; rather it should be used as one tool among many in a broader context of deliberation. The obvious use for policy makers is estimating the social cost of carbon to set a reasonable carbon tax or to estimate costs and benefits of regulations that affect emissions. The focus of this paper is the ethics of the economist's role as scientific advisor in constructing and presenting a CBA, not the decision-making role of policy-makers. I assume that policy-makers ought to follow the overwhelming consensus of climate economists in imposing *some* climate economic policy, e.g. a carbon tax or cap-and-trade scheme. (In other words, the social cost of carbon is obviously non-zero.) The question is whether and how the economist qua policy advisor ought to structure a CBA to incorporate uncertain catastrophic possibilities. In this section I explore the implications of two ethical norms: (1) scientists ought to manage so-called inductive risks by considering the non-epistemic consequences of being wrong; and (2) scientists ought to honestly and explicitly represent uncertainty for policy-makers. I go on to suggest that these (occasionally competing) ethical ideals may justify Weitzman's fat-tailed approach.

[7] Thus in this context I am putting aside, for example, the popular objections that the problem of value incommensurability dooms CBA as a policy-making tool (Anderson 1993; Steel 2015) or claims that welfare cannot be adequately represented by a CBA (see Frisch 2013, Sect. 4, and references therein).

Rudner (1953) claimed that scientists acting as advisors to decision-makers should take contextual or non-epistemic values into account when *setting the burden of proof* for scientific claims, more specifically the “alpha-level” or significance level in classical statistics. Rudner argued that the scientist qua scientist makes value judgments insofar as they incorporate the context’s *inductive risks*, or the risks of being wrong, that decision-makers implicitly take on when they act according to a classical statistical decision rule informed by the scientist’s acceptance or rejection of hypotheses. In his example, Rudner’s scientist uses a higher alpha-level (lower burden of proof) for testing the quality of belt buckles than for testing the quality of drugs, since the inductive risks in the latter context are more ethically significant. This assumes that the decision-maker will act according to the classical statistical decision rule that says: release the drugs/belt buckles just in case the hypothesis that the drugs/belt buckles are defective is rejected at the level of statistical significance set by the scientist.

Jeffrey (1956) responded in Bayesian fashion, arguing that scientists could simply hand over (posterior) probabilities to decision-makers and allow them to make decisions according to their own values and decision rules. They could also just hand over their p-values, or even raw data. Douglas (2000, 2009), Elliott (2011), Steel (2015), and others have responded to Jeffrey by arguing that *higher-order inductive risks* arise in epistemically unforced methodological choices, so even Bayesian posterior probabilities may be implicitly value-laden. For example, epistemically unforced decisions about prior distributions and parameters in a Bayesian analysis for a policy-maker raise the possibilities of underestimating or overestimating uncertainty, which might have ethically significant downstream consequences. Similarly, choices about operationalizing key concepts, characterizing ambiguous data, etc. can have downstream consequences, raising inductive risks even for a scientist using Bayesian statistical analysis.

Jeffreyans respond by claiming that instead of managing non-epistemic inductive risks, scientists ought to make all uncertainties, *including higher-order uncertainties*, explicit for the decision-maker, for example by methodological sensitivity analysis (Betz 2013). The practical challenges for this “Jeffreyan value-free ideal” of making *all* scientific uncertainties and higher-order uncertainties explicit are daunting. For example, Steele (2012) argued that scientists qua policy advisors make implicit value judgments about inductive risks if they must translate their scientific beliefs and results into another “language” that policy-makers can actually understand (e.g. from PDFs to qualitative or linguistic representation of uncertainty), since as the translation is epistemically underdetermined and these decisions might have significant downstream consequences. This kind of translation might be necessary more generally since policy makers may not understand probabilistic or other scientific representations of uncertainty, or methodological sensitivity analyses, or be able to competently use these in their decision-making in ways that are logically sound. Additionally, higher-order uncertainties associated with methodological decisions might be so pervasive that avoiding implicit inductive risk value judgments is simply impracticable.[8] Due to these and other considerations, Elliott proposed a principle he called the “no-passing-

[8] For a discussion of the limits of this Jeffreyan ideal in the context of a case study of natural scientific climate modeling, see Frank (2017).

the-buck" principle, which says that scientists have duties to manage inductive risks when "it is…socially harmful or impracticable for scientists to respond to uncertainty by withholding their judgment or providing uninterpreted data to decision makers" (2011, p. 55).

Even when *some* inductive risk management is unavoidable for the practical reasons surveyed above, it arguably remains a duty for the scientist to honestly and, when possible, *explicitly* articulate uncertainties for policy makers. Thus Rudner et al. and Jeffrey et al. can be seen as articulating somewhat distinct, and sometimes competing, duties for the scientist acting as a policy advisor. The former emphasize the scientist qua policy advisor's largely *non-epistemic* duties to consider the possible consequences of error given the practical, decision-making context; the latter emphasize the scientist qua policy advisor's largely *epistemic* duties of scientific integrity and honesty. This distinction between epistemic and non-epistemic duties (or values) is not an absolute one; for example, duties of scientific honesty and research integrity are arguably both epistemic and ethical.[9] Furthermore, the inductive risk argument shows that ethical and other non-epistemic considerations cannot be cleanly separated from the epistemology of applied sciences. However, in the context of the scientist acting as policy advisor, duties to manage inductive risks plainly appeal *primarily* to non-epistemic goals as opposed to the epistemic goals of science as an engine of reliable knowledge production. These epistemic goals include the production of true (or approximately true) well-confirmed theories with predictive and explanatory power, goals which might be pursued independently of any number of non-epistemic goals. The duty to articulate scientific uncertainties in Jeffrey's sense appeals to these epistemic goals of the scientist, who, in Jeffrey's ideal case, presents a rational decision-maker with probabilities over relevant outcomes so that she may make a decision according to her own values.

The non-epistemic duties to manage inductive risks may be derived from general duties to consider the non-epistemic consequences of our actions (Douglas 2003), while the epistemic duties may be derived either from general duties of honesty or else role obligations specific to expert policy advisors (Shamoo and Resnik 2015). These might conflict; for example, a scientist might judge herself obligated to deemphasize uncertainties when those would be misused or misinterpreted by cynical policy-makers to delay action on an issue like climate change. Depending on the nature of those uncertainties, one might think this constitutes unethical manipulation of the policy-maker. However, in complex policy-relevant fields like climate science, scientific uncertainties will abound (at the very least, plausible "unknown unknowns"), so the scientist must choose which uncertainties to make explicit, which to emphasize, and which to leave aside. This problem is especially acute given the weaponization of normal scientific uncertainty by various actors to delay policy action on important environmental problems, documented by, e.g. Oreskes and Conway (2010). In a context closer to Jeffrey's ideal case, the same scientist might make as many relevant uncertainties explicit as possible for policy-makers. In these contexts, policy-makers are able to understand and competently use explicit representations of uncertainty and are unlikely to manip-

[9] However, see Longino (1996) for an argument that this distinction ultimately breaks down. I thank an anonymous reviewer for pushing this point.

ulate them to serve predetermined policy goals (Frank 2017). Thus, in any particular context, these pro tanto duties might have to be weighed against each other, along with other relevant considerations.

These general duties might entail a more specific duty for the climate economist using an IAM to follow Weitzman's fat-tailed approach to the economics of climate catastrophe. Consider epistemic duties. First, we might weakly assume that a fat-tailed PDF for economic damages merely accurately represents the subjective beliefs of the economist about uncertainty surrounding future damages, as it does for Weitzman and others. (It does not for, e.g. Nordhaus 2012, however the stronger arguments below might apply to such economists insofar as they act as policy advisors.) Then, insofar as the scientist has a duty of subjective honesty to accurately convey their beliefs to the policy-maker, they would be (pro tanto) obligated to use this approach. The stronger, more interesting case is one where we interpret the epistemic duty as implying, at the very least, being consistent with extant evidence and meeting standards of the relevant expert communities for the probabilistic representation of uncertainty. Simplifying a bit, according to these standards, the fat-tailed PDF for damages might be epistemically permissible or epistemically impermissible. For now, assume that it is epistemically permissible according to these standards; I will deal with the objection that the fat-tailed PDF for damages is epistemically impermissible or, relatively impermissible, in the next section. Given the significant scientific uncertainties discussed in the last section, it is plausible to assume that both the fat-tailed approach and a thin-tailed approach are both epistemically permissible. This leaves considerations of inductive risk management.

According to all users of IAMs, there is significant higher-order uncertainty about these models' structure and assumptions, including *ethical* uncertainties about the appropriate discount rate, level of risk aversion, etc. (Jamieson 2014; Frisch 2013) The complex higher-order inductive risks associated with IAMs' outputs suggest that economists have some duties to manage inductive risks. Additionally, it may be harmful or impractical for economists to hand over uninterpreted results or model specifications, or to construct complex methodological sensitivity analyses. Indeed, such analyses might obscure expert knowledge or give meaningless results if they attempt to represent *too many* uncertainties. That is, Elliott's "no passing the buck" principle applies here, at least for some inductive risks. Then, if the inductive risks associated with incorrectly using a thin-tailed approach are less bad than the inductive risks of incorrectly using a fat-tailed approach, then this would provide (pro tanto) *non-epistemic* justification for using the fat-tailed approach.

To be sound, this non-epistemic justification for using the fat-tailed approach must hold that the balance of inductive risks favors the fat-tailed approach over a thin-tailed approach. This is plausible, especially given precautionary considerations, which fits with Weitzman's own precautionary interpretation of the fat-tailed approach. Consider, for example, Steel's (2015, p. 28) articulation of a relevant precautionary principle: "If a scientifically plausible mechanism exists whereby an activity can lead to catastrophe, then that activity should be phased out or significantly restricted." The scientific plausibility of climate catastrophe is a necessary condition for the *epistemic* (and thus, in this context, ethical) permissibility of using the fat-tailed approach. There are of course risks associated with any aggressive climate change mitigation policy, e.g. setting the carbon tax too high, or instituting a high carbon tax without measures to deal with its possibly regressive distributional consequences. However, these inductive risks that must be associated with the fat-tailed approach are minor compared with the inductive risks associated with a thin-tailed approach that would not necessarily justify policies that significantly reduce catastrophic risk. As Steel (2015, pp. 32–37) argues, there is no scientifically plausible "counter-scenario" in which this precautionary principle would incoherently recommend *against* aggressive mitigation due to a precautionary interest in avoiding a "mitigation catastrophe".

One might point to the "mitigation catastrophe" of Weitzman's own dismal theorem, a "counter-scenario" to climate catastrophe wherein society pays an arbitrarily large amount of money to prevent what turn out to be extremely improbable catastrophic risks. Again, the possible problem with justifying the fat-tailed approach on precautionary grounds is that the same precautionary principle may recommend against it if this counter-scenario or "mitigation catastrophe" is plausible. However, this unbounded approach of the dismal theorem would not be epistemically or ethically permissible, since (1) the counter-scenario is not scientifically plausible, since people's willingness to pay to reduce mortality risks is not unbounded; and (2) being unwilling to pay such an unlimited amount is normatively justified, given unavoidable tradeoffs with reducing other risks and pursuing other legitimate goals. While Weitzman's dismal theorem complicates the balance of inductive risks, a plausible (albeit perhaps unavoidably somewhat arbitrary) upper bound for the "value of statistical civilization" could perhaps be set such that the inductive risk of a "mitigation catastrophe" is eliminated, or outweighed by the catastrophic risk reduction measures encouraged by use of the fat-tailed approach.

4 Objections and limitations: weighing epistemic and non-epistemic values

I do not claim to have definitively established that economists ought to use Weitzman's fat-tailed approach to climate CBA. Rather I have argued that, on some plausible assumptions about the duties of scientists acting as policy advisors, and assumptions about the inductive risks involved, it may follow that other approaches (even ones that are epistemically permissible) fail to ethically manage the inductive risks at stake in the climate science-policy nexus. One might challenge the above argument above in several ways. First, consider the challenge that fat-tailed PDFs for damages are epistemically *impermissible*. According to this objection, there are epistemic standards for the representation of uncertainty such that a fat-tailed PDF for damages fails to meet those standards. Without going into too much detail, consider the reasons that the Interagency Working Group gives for *not* using Weitzman's fat-tailed approach to incorporate catastrophic outcomes. Firstly, they accurately state that there are significant "differences of opinion" (2010, p. 29) amongst economists about the appropriateness and applicability of this approach. Secondly, they point out that the risk premiums on the social cost of carbon that come out of such models are very sensitive to model assumptions. Neither of these entail epistemic impermissibility. Both claims are true of *every* IAM; so, as arguments for epistemic impermissibility they would overgeneralize. They go on to state explicitly that there is significant uncertainty associated with the extrapolation of damages at low temperature increases to high temperature increases, which just is the epistemic rationale that Weitzman gives for his fat-tailed approach. I return to a version of this objection below. Similarly, Nordhaus's (2012) empirical argument against Weitzman uses economic data from "catastrophic" events in recent economic history to argue that the fat-tailed approach would overestimate the catastrophic risks of climate change. But such an approach rests on the undefended assumption that future catastrophes fueled by climate change will be similar to those experienced in the absence of extreme climate change. This is exactly the kind of extrapolation that Weitzman's approach rejects.

Second, one might argue that instead of using the fat-tailed approach, economists acting as policy advisors should "average all plausible models" as the Interagency Working Group did with DICE, PAGE, and FUND, or give some weighted average. I do not have an argument that a Weitzman-style approach *alone* would better manage inductive risks than this approach *averaged* with other epistemically permissible IAMs. However, insofar as those other IAMs do not take the significant uncertainties about catastrophic outcomes into account, an average of these other models that excluded a fat-tailed IAM would be open to the same precautionary argument given above.

Thirdly, one might argue that balancing epistemic and non-epistemic values or duties in these contexts is much more complicated than I made it out to be above, especially since I treated epistemic permissibility as binary where epistemic permissibility (or, overall epistemic value) is a matter of degree. This objection raises deep philosophical puzzles about how (primarily) epistemic and (primarily) non-epistemic values or duties ought to be weighed and traded off in various contexts. This poses a significant problem for my argument insofar as a fat-tailed CBA might be *relatively* less epistemically valuable than some alternative CBA, even if the former is *ethically* more valuable, for example due to considerations of inductive risk and precaution. This is admittedly the weakest link in the argument. However, the massive uncertainties surrounding even the best IAMs make judgments of their relative epistemic value extremely difficult. Thus it might still be justified to regard IAMs within some reasonable range of expert opinion to be epistemically permissible, and then choose one or a subset of these on the basis of non-epistemic, precautionary considerations.

Finally, one might argue that these massive uncertainties associated with IAMs' damage functions, as well as many other model assumptions, instead of leading us to prefer a fat-tailed CBA, ought instead to lead us to conclude that *no CBA at all might be better than some CBA*. For example, Pindyck (2013) has argued against the use of IAMs, since modelers can "obtain almost any desired result because key inputs can be chosen arbitrarily" (p. 870). Weitzman also suggests this in his discussion of the fat-tailed approach:

> Perhaps in the end the climate-change economist can help most by *not* presenting a cost-benefit estimate for what is inherently a fat-tailed situation with potentially unlimited downside exposure as if it is accurate and objective—and perhaps not even presenting the analysis as if it is an approximation of something that is

> accurate and objective—but instead by stressing somewhat more openly the fact that such an estimate might conceivably be arbitrarily inaccurate depending upon what is subjectively assumed about the high-temperature damages function along with assumptions about the fatness of the tails and/or where they have been cut off. (2009, p. 18)

Here Weitzman follows Pindyck's argument that the uncertainties surrounding IAMs and their sensitivity to countless epistemically unforced assumptions should leave the economist policy advisor in the position of either emphasizing the low epistemic value of IAM results, or else avoiding IAM results in the policy context entirely. It is certainly possible that the epistemic value of IAMs is so low that non-epistemic considerations and much simpler precautionary reasoning should guide policy instead. However, on the assumption that epistemic value of IAMs meets some threshold of permissible use, my argument for using the fat-tailed approach is applicable.

While this Pindyck-style critique of IAMs has significant force, one might distinguish types of uses of IAMs. In particular, we should distinguish use of IAMs for "optimal" decision-making in the high-technocratic style of ideal classical utilitarianism, as opposed to mere *decision support* for policy-makers. Reiterating an important point from Sect. 3, CBAs should not—and due to normative and empirical uncertainties, arguably cannot—determine decisions alone; they are one tool among many in a broader framework of policy-making. At the very least, the function of such idealized decision models is to make certain features of the actual social decision explicit. In the case of Weitzman's model, we are left considering what we are willing to pay as a society to avoid climate catastrophe. Similarly, Frisch (2013, 2017) has argued forcefully that the usefulness of IAMs is limited due to multiple empirical and normative uncertainties, including uncertainties surrounding the damage function discussed above. However, he suggests,

> instead of appealing to optimization IAMs to deliver precise policy recommendations concerning the optimal and most 'efficient' path to maximizing overall utility, we could use IAMs as toy models—as extremely simplified and idealized representations of the interactions between climate and economic systems—that allow us to examine a range of plausible futures. (p. 134)

In examining these "plausible futures," we only have a rough idea of the probabilities and impacts of various climate-policy outcomes. However even toy models can give us important, albeit rough, information about the probable costs of more or less precautionary climate policy. If scientists acting as policy advisors have duties to honestly and explicitly represent uncertainties to policy-makers but also to manage inductive risks, Weitzman's approach that admits deep uncertainty about catastrophic outcomes strikes a reasonable balance between these duties. The costs of being erroneously too pessimistic in our economic models of climate catastrophe (economic costs for present and near-future generations) are more known and more manageable than the possibly catastrophic costs of being erroneously too optimistic. Indeed, Frisch concludes that Nordhaus's strategy of making relatively optimistic assumptions about the economic costs of climate change seems "astonishingly reckless." (ibid., p. 135).

5 Conclusion

I have argued that Weitzman's argument for a fat-tailed approach to climate CBA may be justified by ethical norms that apply to scientists acting as policy advisors. There are significant limitations to this argument, including the question of the balance of inductive risks and complex conflicts between epistemic and non-epistemic values. Weitzman's fat-tailed approach also raises the ethical problem of bounding the analysis to prevent the absurd result of the dismal theorem. That is, it focuses policy makers on the question of what is an acceptable cost to prevent low-probability, catastrophic collapse of civilization. In presenting climate change in such stark, existential terms, it also abstracts away from the—at least equally—important ethical issues of intra-generational and intergenerational justice. However, this argument deserves explicit articulation, especially since some major policy makers use IAMs to guide decision-making, many of which are perhaps too optimistic given existing uncertainties.

Acknowledgements Thanks to audiences and colleagues at the University of North Carolina, Chapel Hill and Duke University, and two anonymous reviewers for helping to significantly improve this paper.

References

Anderson, E. (1993). *Value in ethics and economics*. Cambridge: Harvard University Press.

Betz, G. (2013). In defense of the value free ideal. *European Journal for Philosophy of Science*, *3*, 207–220.

Burke, M., Hsiang, S. M., & Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, *527*, 235–239.

Dietz, S. (2011). High impact, low probability? An empirical analysis of risk in the economics of climate change. *Climatic Change*, *108*, 519–541.

Douglas, H. (2000). Inductive risk and values in science. *Philosophy of Science*, *67*, 559–579.

Douglas, H. (2003). The moral responsibilities of scientists: tensions between autonomy and responsibility. *American Philosophical Quarterly*, *40*(1), 59–68.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press.

Elliott, K. C. (2011). *Is a little pollution good for you? Incorporating societal values in environmental research*. New York: Oxford University Press.

Frank, D. M. (2017). Making uncertainties explicit: The Jeffreyan value-free ideal and its limits. In K. C. Elliott & T. Richards (Eds.), *Exploring inductive risk*. New York: Oxford University Press.

Frisch, M. (2013). Modeling climate policies: A critical look at integrated assessment models. *Philosophy & Technology*, *26*, 117–137.

Frisch, M. (2017). Climate policy in the age of trump. *Kennedy Institute of Ethics Journal*, *27*, 2.

Greenstone, M., Kopits, E., & Wolverton, A. (2011). Estimating the social cost of carbon for use in U.S. federal rulemakings: A summary and interpretation. Center for Energy and Environmental Policy Research Working Paper.

Interagency Working Group on Social Cost of Carbon. (2010). Technical support document: Social cost of carbon for regulatory impact analysis under executive order 12866. https://www.epa.gov/sites/production/files/2016-12/documents/scc_tsd_2010.pdf. Retrieved June 26, 2017.

Jamieson, D. (2014). *Reason in a dark time: Why the struggle against climate change failed and what it means for our future*. New York: Oxford University Press.

Jeffrey, R. C. (1956). Valuation and acceptance of scientific hypotheses. *Philosophy of Science*, *22*, 237–246.

Longino, H. (1996). Cognitive and non-cognitive values in science: Rethinking the dichotomy. In L. H. Nelson & J. Nelson (Eds.), *Feminism, science, and the philosophy of science* (pp. 39–58). Dordrecht: Kluwer.

Nordhaus, W. (2012). Economic policy in the face of severe tail events. *Journal of Public Economic Theory*, *14*(2), 197–219.

Nordhaus, W. (2013). *The climate casino: Risk, uncertainty, and economics for a warming world*. New Haven: Yale University Press.

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. New York: Bloomsbury Press.

Oreskes, N., & Conway, E. (2014). *The collapse of Western civilization: A view from the future*. New York: Columbia University Press.

Pindyck, R. S. (2013). Climate change policy: What do the models tell us? *Journal of Economic Literature*, *51*(3), 860–872.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science*, *20*, 1–6.

Shamoo, A., & Resnik, D. (2015). *Responsible conduct of research* (3rd ed.). New York: Oxford University Press.

Steel, D. (2015). *Philosophy and the precautionary principle: Science, evidence, and environmental policy*. New York: Cambridge University Press.

Steele, K. (2012). The scientist qua policy advisor makes value judgments. *Philosophy of Science*, *79*, 893–904.

Tol, R. (2012). On the uncertainty about the total economic impact of climate change. *Environmental and Resource Economics*, *53*(1), 97–116.

Wagner, G., & Weitzman, M. L. (2015). *Climate shock: The economic consequences of a hotter planet*. Princeton, NJ: Princeton University Press.

Weitzman, M. L. (2007). A review of the Stern Review on the economics of climate change. *Journal of Economic Literature*, *XLV*, 703–724.

Weitzman, M. L. (2009). On modeling and interpreting the economics of catastrophic climate change. *Review of Economics and Statistics*, *91*(1), 1–19.

Weitzman, M. L. (2012). GHG targets as insurance against catastrophic climate damages. *Journal of Public Economic Theory*, *14*(2), 221–244.

Weitzman, M. L. (2014). Fat tails and the social cost of carbon. *American Economic Review: Papers and Proceedings*, *104*(5), 544–546.