# Choosing in the Dark: Incomplete Preferences, and Climate Policy

Elizabeth Baldwin*

February 15, 2018

Latest version available from elizabeth-baldwin.me.uk

**Abstract**

I consider decision-making when the stakes are high, but information is poor, and outcomes may be far from our experience. My leading example is climate change. We do not know the probabilities of diverse outcomes; we disagree about societal impatience, risk, and inequality aversion. I offer a simple model of "justifiable acts", providing maximal agreement between decision theories, and facilitating quantification of the remaining disagreement. When this disagreement is large, I characterise the choice situation as "dismal". I demonstrate that the question of climate policy is "dismal". This illuminates how subjective much of the literature on climate change economics really is, and so how poor a guide to policy this literature may form. The "dismal" framework here generalises Weitzman's (2009a) "dismal theorem", giving a broader view, which shows that it may be unnecessary or unwise to focus on highly unlikely events.

**JEL:** D81, Q54

**Keywords:** Knightian Uncertainty, Incomplete Preferences, Climate Change, Applied Decision Theory

## 1 Introduction

Sometime a policy choice may have very large-scale consequences, but we understand these very poorly. How should economics address such issues? This paper proposes a model in which we explicitly model the range of what is plausible. In the context of climate change, it shows that existing approaches have tended to only consider a very narrow range of possible welfare losses from warming. In consequence, the range of policies they prescribe are similarly conservative.

*Department of Economics and Hertford College, Oxford University, UK; elizabeth.baldwin@economics.ox.ac.uk. Acknowledgements to be completed later.

This paper's contribution is to apply a new style of decision theory to questions of this nature: incompleteness in "tastes" and "beliefs". It proposes a characterisation of when such a description matters, and discusses how to assess policy in this scenario. This characterisation applies to the economics of climate change, as is shown numerically. Different plausible choices will lead to very different stringencies of policy prescription. Elegant stylsed facts demonstrated in the literature, such as the independence of the social cost of carbon on cumulative emissions, are shown not to be robust. It takes a decision theory which distinguishes between subjective and objective information, to illuminate this.

There are reasonable scientific bounds on what the climate response to greenhouse gas emissions might be, but precise probability distributions cannot be given. Moreover, even if climatic scenarios were predictable, their economic consequences are not. Very little work has been done, or can convincingly be done, to understand the macroeconomic welfare implications of higher levels of global warming, such as 4°C. But we will see such temperature changes this century under business-as-usual emissions, and they remain a non-trivial risk even under ambitious abatement scenarios. Additionally, the long-term cost of such abatement depends on the progress of technological change, which is far from well understood. These are examples of our inadequate information on the probability distribution of states of the world. There is no unique defensible way to predict what may happen.

Moreover, experts disagree on how outcomes should be valued. We must trade off certainty against risk; economically developed nations against those most vulnerable; the present against the future. These trade-offs embody the utility value we put upon outcomes. So there is no unique defensible way to value what happens.

One may model this formally as follows: there are many uncertain "states" that drive outcomes; many plausible "beliefs" which give a probability distribution over such states; and many valuations, or "tastes", evaluating outcomes in welfare terms. In contrast with the majority of works on the economics of climate change, I assume it is not rational to restrict attention to a single taste and belief (nor to specify a single probability distribution over beliefs, for example, as this amounts to identifying a single belief). That is, I do not apply the model of "subjective expected utility".[1] Instead, I use an "incomplete" decision theory in which one act is objectively preferred over another only if it preferred under all such tastes and beliefs.[2] Not all policy choices can therefore be ranked.

This rules out some acts, but it leaves many on the table. As I show, under mild assumptions, either an act is optimal under some tastes and beliefs, or there is an objectively preferred act. I refer to the former as "justifiable acts".[3] Thus the model is equivalent one of disagreement between multiple agents, each with simple (subjective expected utility) tastes and beliefs. I call these "formal agents" as they do not represent participants in the model.

Incompleteness, when it is significant, is a critical characteristic of the problem: in my view our models should reflect and communicate the scope of what is objectively possible. Although there are many recent decision theories available to "complete" the

[1]See Anscombe and Aumann (1963).
[2]Cf. Galaabaatar and Karni (2013) and Riella (2015).
[3]Cf. Battigalli et al. (2016).

model and return us to a uniquely optimal prescribed act (see discussion in Section 1.2 below), to choose and apply such a model would be to impose subjective choices.

Moreover, as I show, if an act is prescribed by any one of a very wide class of such models, it would also be recommended by one of my simple "formal agents". And these are the only acts that these decision theories do not all reject. Thus my characterisation is as complete as can be obtained objectively.

When does this incompleteness indeed matter? The prescription of one formal agent will be chosen, so I analyse the policy preferred by one, from the point of view of another. I suggest a criterion of "dismal choices": when the change in pay-off from a small change in act is *large*, relative to some fixed benchmark, under some plausible tastes and beliefs; and moreover if this evaluation *differs* significantly from its evaluation under other plausible tastes and beliefs.

An alternative criterion I suggest is similar to Savage-style "regret".[4] Take the perspective of one formal agent, and identify the act they would choose. Then assess the loss in expected welfare they experience, due to instead enacting the policy prescribed by a different formal agent. Broadly speaking, the situation is dismal when this "disagreement regret" is large.

I show how a general climate-economy model provides an example of this structure. I numerically simulate such a model, incorporating scientific uncertainty and incompleteness in economic knowledge. Thus I show that assessments of marginal damages from climate change are indeed "dismal", for moderately high cumulative emissions.

There is also large disagreement on the welfare cost of emission reductions. However, different subjective assessments of damages tend to diverge strongly for high cumulative emissions of greenhouse gases, whereas the cost of mitigation is better constrained in a business-as-usual scenario and only becomes ambiguous when we consider low emission targets. Thus there may be some intermediate range in which incompleteness in cost-benefit analysis does not pass the "dismal" threshold. Alternatively, policy-makers may face two sources of deep uncertainty.

The "dismal" nomenclature I use is inspired by Weitzman's (2009a) "dismal theorem", and his body of work on the subject forms the clearest point of comparison with this paper.[5] Weitzman imposes a lower bound on global disutility, and shows that the expected value of policy tends to infinity as this lower bound is relaxed. He concludes, broadly, that an analysis of the effects of climate change depends on assessing the "very rare, very catastrophic" scenarios.

His results provide an example of my dismal situation: it matters, in relative and absolute terms, how we place that subjective lower bound on disutility. However, my paradigm makes clearer that the evaluation is sensitive to many subjective assumptions within the model. So, extraordinary catastrophe need not be the driver of the multiplicity of valuations, or of the potential for extreme values in cost-benefit analysis. In concrete terms: perhaps we should not focus on the very small chance of 10 or 20°C warming, since 4°C is the median temperature change at 2100 for business as usual emissions, and the effects of this could well be dramatic.

I also assess "stylised facts" which emerge from several papers in the literature. These are not robust to changes between the many plausible modelling assumptions. First,

[4] See Savage (1951).

[5] See also Weitzman (2009b,c, 2010, 2011, 2012, 2013, 2014).

climate change policy is indeed sensitive to incorporating geophysical uncertainties – but the effect of this uncertainty depends on the "damage function".[6] Second, marginal damages do indeed depend on the final policy scenario – so that delays in applying policy do mean that optimal abatement is more expensive.

Other papers have considered climate change under ambiguity. Notably Millner et al. (2013) assess the impacts of climate change using smooth ambiguity aversion in the distribution of the climate sensitivity parameter (keeping other tastes and beliefs fixed), while Xepapadeas (2012) uses robust control. Lempert (2015), Metcalf and Stock (2015), Otto et al. (2015) and Rosen (2015) have all recently emphasised the importance understanding ambiguity and the questioning of assumptions in integrated assessment models. Lempert et al. (2006), Polasky et al. (2011), Lemoine and McJeon (2013) and Weaver et al. (2013) use techniques of robust Bayesian analysis to assess policy.

More recent papers, concurrent with this, (Danan et al., 2016, Dietz and Matei, 2016, and Chambers and Melkonyan 2017) apply models of incomplete social tastes or beliefs to climate change. These are discussed in more detail in Section 7.1. However, this paper uniquely proposes an exploration of the full range of assessments that are objectively possible, and how the extent of this range undermines much accepted wisdom on climate change policy.

The paper is arranged as follows. Sections 1.1 and 1.2 provide more background, first on climate change, and then on decision theory. Section 2 introduces the model of incomplete beliefs and tastes, and shows how to apply this to climate change, fitting an integrated assessment model, of climate change and the economy, into a simple model of private benefits and public damages. Section 3 introduces "justifiable acts" and my notions of "dismal" choices and "disagreement regret", giving an approximate relationship between the latter two. Section 4 gives numerical modelling in the climate change context. Section 5 discusses the implications of modelling in this way: the comparison with Weitzman's (2009a, and subsequent) work, and problems with "stylised facts" in the wider literature. Section 6 shows that this model of this paper provides maximal agreement between a whole host of decision theories. Section 7 provides further discussion, and concludes.

## 1.1 The Climate Change Context

An assessment of the economics of climate change depends on the economic benefits pollutants bring, the damages they cause, and the wider economic outlook. Moreover it depends on our welfare interpretations of all such outcomes. There are serious problems with such a valuation, as recent literature discusses in detail.[7]

Weitzman (2007, 2009a) first raised some of these issues, focussing most explicitly on geophysical uncertainty in the long-term global mean temperature rise that should be associated with any increase on $CO_2$ concentrations. It is difficult or impossible for science to provide truly objective probability distributions (see e.g. Kunreuther et al.

[6] All effects so happen to cancel out with the function which is conventionally used but, as discussed in Section 1.1, it is only convention which dictates this functional form.

[7] See Weitzman (2007, 2009a,b,c, 2010, 2014), Hanemann (2009), Stern (2010, 2013), Pindyck (2011a,b, 2012, 2013a,b, 2017), Millner et al. (2013), Fisher and Le (2014), Heal and Millner (2014a), Kolstad et al. (2014) and Dietz and Stern (2015).

2013). But too often still, this uncertainty is captured in restrictive distributions, or not modelled at all.

Moreover, these need not be the most important sources of uncertainty. The economic damages associated to a given level of global warming are very badly understood, especially at higher temperatures. We potentially face 4–5°C warming before 2100. Integrated assessment models typically assume this would give rise to an output loss of 1–5%, a level that some leading experts consider "ludicrously small".[8] But this has not been estimated directly; it comes from extrapolations of damages estimated at much lower temperature changes, using methods that are without "theoretical or empirical foundation".[9] Despite recent efforts in improving this situation,[10] it is unlikely that damage functions can be based on empirical estimates for the full range of possible climate responses. In fact, 4–5°C warming represents a very significant climatic shift, of the same scale as the difference between the present day and the last ice age. And there is serious concern that warming of this level might be disastrous for economic growth where it is most needed.[11] These problems have led some scientists to despair of optimal outputs from integrated climate-economy models.[12]

Technology change is also far from perfectly understood. To model it accurately, one must consider serious market imperfections, as well as endogenous growth.[13] Surveys of experts help to form forward projections, but disagreement between experts can reach an order of magnitude.[14]

Moreover, the state of the wider economy determines the demand for energy inputs, and so the cost of cutting emissions to a given level. And economic growth alters the welfare implications of climate change damages: adaptation is easier for the rich, and any losses are typically not weighted so heavily. So our estimates of both climate damage and mitigation costs are further undermined by the serious limitations to our prowess in forecasting total factor productivity growth.[15]

## 1.2 Decision Theories

The standard model for rational choice under uncertainty, subjective expected utility ("SEU") implies an agent behaves as if they know exactly both the probability distribution of uncertain "states" of the world (their "beliefs"), and how to judge and compare these outcomes (their "tastes").[16] However, as argued by Gilboa et al. (2009), "having coherent beliefs that have nothing to do with evidence and data cannot be considered rational". My interest is in situations in which tastes, beliefs, or both, are not "objectively" known – in the sense that there do not exist choices for these which any reasonable agent would be convinced are correct.[17]

[8] See Stern (2013) Figure 1 and Stern (2010).

[9] Pindyck (2013b).

[10] See e.g. Burke et al. (2015).

[11] See Dell et al. (2012), and Schellnhuber et al. (2014).

[12] See Box 3.1 of IPCC (2014).

[13] See Fisher-Vanden (2008) in her introduction to a collection of papers on this subject.

[14] See Baker et al. (2009).

[15] See, e.g. Prescott (1998) and Millner and McDermott (2016).

[16] See Savage (1954) and Anscombe and Aumann (1963).

[17] I follow Gilboa et al. (2010) in making this distinction from "subjectively" rational choices, which a reasonable agent cannot demonstrate are wrong.

There are a wealth of models addressing how we should close this gap when the problem is an absence of clear, unique beliefs or tastes, and the question of how to aggregate tastes across heterogeneous individuals.[18] However, recent experimental work shows little conclusive evidence as to which is the best model to describe how individual agents *do* behave.[19] Moreover, these models do not address how agents *should* behave, especially when the agent in question is addressing questions with which they have little familiarity, and which may have large-scale and long-term impacts. Finally, even once a model has been selected, further deeply subjective choices are required in its application. Thus I do not wish to rely on any one of these models.

But indeed, why should "objectively" rational preferences be able to address our ignorance of the situation we face? Instead, I look at models which relax the Savage axiom of "completeness", that is, comparability of any two acts. As argued by Gilboa et al. (2009) and Karni (2014), completeness of preferences even at the individual level is likely to fail in unfamiliar, high-consequence situations; aggregation across a heterogeneous society adds an additional level of difficulty.

In practice, the debate on policy to combat climate change has focussed on certain key features of our "tastes" and "beliefs", about which experts deeply disagree. Our models should reflect this.

So as a basis for this paper, I use the most general existing model of incomplete tastes and beliefs: that of Riella (2015).[20] This is similar to familiar SEU, but with a set $\mathcal{U}$ of "tastes" and a set $\Sigma$ of "beliefs", as above, so that two acts can be ranked only if they are ranked in the same way under all these tastes and beliefs.

Much of the argument of this paper would be unaffected by using only incompleteness in beliefs, and some economists may consider that a decision maker ought to know their own tastes. However, "local" tastes are more consistent with the Allais paradox.[21] And incomplete tastes are more appropriate for very unfamiliar situations.[22] Even if a social planner's preference is to aggregate tastes across society, they will usually lack sufficient information to do so.

Moreover, neglecting incompleteness in tastes seems unnatural in the case of climate change economics: a great deal of discussion and disagreement has arisen from differences in opinion on one particular taste parameter, the pure rate of time preference.

---

[18]Popular examples of decision theory under ambiguity are Anscombe and Aumann (1963), Gilboa and Schmeidler (1989), Schmeidler (1989), Hansen and Sargent (2001), Ghirardato et al. (2004), Klibanoff et al. (2005) and Maccheroni et al. (2006). Etner et al. (2012) and Karni et al. (2015) give recent surveys. See Cerreia-Vioglio et al. (2014, 2015) for work on the "local" tastes of Machina (1982), and Weitzman (2001), Heal and Millner (2013, 2014b) and Millner and Heal (2014) on aggregation of the pure rate of time preference.

[19]See Etner et al. (2012).

[20]The preceding model of Galaabaatar and Karni (2013) is very similar, but concerns a strict preference relation; it is more convenient for me to work with the weak preferences of Riella (2015). See Bewley (1986, 2002) and, for example, Seidenfeld et al. (1995) and Ok et al. (2012) for earlier models, and Hill (2013) for an alternative perspective.

[21]See Machina (1982) and Cerreia-Vioglio et al. (2014).

[22]Karni (2014) argues that an agent does not know their own tastes for states of health which they have never experienced; a dramatic decline in biodiversity, for example, may also be hard to judge without experience.

# 2 The Model

## 2.1 Assumptions for Modelling with Incomplete Preferences

This paper considers decision-making with high consequences and poor information. My key example is the question of climate change policy: see Sections 1.1, 2.2 and 4. Other suitable examples could be the decision to default on sovereign debt, or the threat of the impact on Earth of a large asteroid. Indeed, the basic model of this section is far more generally relevant; the distinctions I introduce in Section 3.2 restrict attention to these drastic cases.

The model consists of set $\Phi$ of states of the world, a set $X$ of *outcomes*, and *lotteries* $\Delta(X)$ over $X$.[23] For simplicity assume that $\Phi$ is finite and that $X$ is a compact, connected subset of Euclidian space; in the case of climate change we will think of both of $\Phi$ and $X$ as products of smaller sets. Agents choose between *acts* $f \in \mathcal{F}$, which are functions $f : \Phi \to \Delta(X)$.

In the model of Anscombe and Aumann (1963), a subjective expected utility (SEU) agent behaves as if they have a real-valued affine utility function $u$ on $\Delta(X)$,[24] and a probability measure $\sigma$ on states $\Phi$, so that acts may be ranked via their expectation: $\overline{f}_{u,\sigma} := \sum_{\phi\in\Phi} u \circ f(\phi)\sigma(\phi)$. Here, $\sigma$ is defined uniquely, while $u$ is unique up to affine transformation (i.e. behaviourally unique).

However, in the situations of interest, there will not be a clearly identified unique $(u, \sigma)$: agents may disagree, and indeed individual agents may be unable to identify their own "tastes and beliefs". Instead, I assume there is a set $\mathcal{U}$ of such $u$, and a set $\Sigma$ of such $\sigma$, and a subset $\Psi \subseteq \mathcal{U} \times \Sigma$ such that for $f, g \in \mathcal{F}$,

$$f \succcurlyeq g \iff \overline{f}_{u,\sigma} \geq \overline{g}_{u,\sigma} \quad \text{for all } (u, \sigma) \in \Psi.$$

I refer to $\Psi$ as the "plausible" tastes and beliefs. If $f \succ g$ I say $f$ is "objectively" preferred to $g$.

An axiomatic approach to decision theory takes a preference relation as primitive, and derives the modelling structure. But the answer to a policy question must work in reverse, applying a modelling structure to infer preferences. And, as described in Section 1.1, there is no agreed set of social welfare tastes $u$ or beliefs $\sigma$. It is the latter fact which I take here as my starting point; see Section 6 for an alternative perspective, involving agreement between decision theories. However, axiomatisations have been recently given for very similar models.[25]

Make additional technical assumptions:

**Assumption 1.** $\mathcal{F}$ may be represented as a convex subset of $\mathbb{R}^n$ for some $n \in \mathbb{Z}_+$.

---

[23] Cf. Savage (1954), Fishburn (1970).

[24] A function $u$ is affine if $u(\alpha f + (1-\alpha)g) = \alpha u(f) + (1-\alpha)u(g)$ for lotteries $f, g \in \Delta(X)$ and $\alpha \in [0, 1]$; in practice this means that $u$ treats lotteries in the usual additive way.

[25] The closest is Riella (2015), who generalises Bewley (1986, 2002) to incorporate incomplete tastes. Galaabaatar and Karni (2013) provide the first axiomatisation in which tastes and beliefs are simultaneously incomplete, but they work with a strict preference relation. Both Riella (2015) and Galaabaatar and Karni (2013) require a finite outcome space $X$ to derive this representation from their axioms; it seems that extensions to more general cases would require extremely difficult convex analysis. The assumptions on outcomes made here are for convenience in differentiating expected values; a sufficiently fine discretisation of outcomes would give an empirically similar model.

**Assumption 2.** $\Psi$ may be represented as a convex, compact subset of $\mathbb{R}^m$ for some $m \in \mathbb{Z}_+$.

**Assumption 3.** The map $\mathbb{E} : \mathcal{F} \times \Psi \to \mathbb{R}$ given by $\mathbb{E}(f, u, \sigma) = \overline{f}_{u,\sigma}$ is continuous in all arguments and strictly quasi-concave in $\mathcal{F}$.

An important special case is when a policy variable $q \in \mathcal{Q}$ parametrises all available acts. In a simplified model, $q$ might represent total cumulative anthropogenic emissions of greenhouse gases, with $\mathcal{Q} = \mathbb{R}_+$; for a richer model, $q$ can be a vector representing greenhouse gas emissions in each period. Write the act defined by $q \in \mathcal{Q}$ as $f(q) \in \mathcal{F}$.

I often also assume that, for any $u \in \mathcal{U}$, we may write $u \circ f = w_u - D_u + B_u$, where $D_u$ and $B_u$ represent the costs and benefits of the act $f$, respectively; and $w_u$ is underlying welfare independent of the act (but still potentially dependent on the state $\phi \in \Phi$, which may include very broad issues such as the determinants of long-term economic growth).

When acts are parametrised by $q \in \mathcal{Q}$ then these become $w_u(\phi)$, $B_u(q, \phi)$ and $D_u(q, \phi)$ as functions of the parameter and state. In this case, assume that $B_u(q, \phi)$ and $D_u(q, \phi)$ are both monotone in $q$ in opposite directions.[26] As before, expectations given beliefs $\sigma$ are written $\overline{w}_{u,\sigma}$, $\overline{B}_{u,\sigma}(q)$ and $\overline{D}_{u,\sigma}(q)$.

## 2.2 Applying This Model To Climate change

It is worth explicitly putting these ideas in context. Climate change is typically modelled economically in an integrated assessment model. I now explain how the presentation of Section 2.1 encompasses such models.

As is clear from Section 1.1, a model of the economics of climate change needs layers of effects, with a sequence of causality: emissions give rise to atmospheric concentrations, which give rise to climatic changes, and hence to economic impacts. Meanwhile, savings and investment decisions, and technology change, give rise to economic output, and this output creates new emissions. In each of these stages there is uncertainty, although it is not always modelled. Typically an optimal path of consumption and emissions is found using numerical techniques.

The *states* of Section 2.1 represent possible values for features of the world and the climate-economy interaction. There are many relevant concerns, and so we consider the state space as a product space. The first coordinate might represent the "climate sensitivity" parameter; the second might index mechanisms and sensitivities for changes in precipitation; the third, potential functional forms for economic damages from changes in temperature and precipitation; the fourth, the determinants of baseline long-term economic growth.

We do not know the true vector of states, and nor is there a single objective probability for each such vector. However, *beliefs* $\sigma$ ascribe a probability to each state vector.

*Outcomes* are typically measured in integrated assessment models as mean per-capita consumption in each period, perhaps segregated by region. In principle outcomes could be person- as well as time-specific, and could represent differentiated goods (cf. Sterner and Persson 2008). Of course there will be limits to what is computationally tractable,

[26] When $q$ is interpreted as the quantity of a pollutant, it is natural that $B_u(q, \phi)$ decreases with $q$ while $D_u(q, \phi)$ increases with $q$. However, if the parameter $q$ represents the level of a carbon tax, for example, these roles will be reversed.

but I will assume that each outcome vector has components for (finitely many) time periods; the component within each time period may also be a vector.

We evaluate an outcome, or a pure lottery in outcomes, in welfare terms. Here we use a social welfare function $u$, which incorporates preferences over risk, inequality and time, and so represents our *tastes*. Recall that these are only behaviourally defined up to affine transformation. It is standard in these models to normalise by setting marginal utility in the rich world, today, to be equal to one (see e.g. Azar and Sterner 1996). Thus small changes in acts can be understood in units of money, as experienced by the rich.[27]

An *act* is a choice of behaviour in each period – specifying climate change pollution and mitigation levels, but also giving consumption, savings and redistribution. For any given state (vector) this will determine a lottery on outcomes as described above. If a different state pertains, then the same act leads to a different outcome. So the act is a function from states to outcomes.

This gives too much flexibility in acts, so I write $\mathcal{Q}$ for a set of climate change policies. For each $q \in \mathcal{Q}$, the act $f(q)$ incorporates the climate policy $q$, with other economic behaviour chosen to be optimal with respect to some baseline beliefs. Again, each $q \in \mathcal{Q}$ can be a vector in time periods, representing emissions in each period. However, when it is desirable for $\mathcal{Q}$ to be one-dimensional, I will assume that $q$ represents total cumulative emissions of $CO_2$.[28]

To identify welfare effects $B_u(q, \phi)$ and $D_u(q, \phi)$, I show in the appendix:

**Lemma 1.** *Make Assumptions 1-3. Assume that, in each period, we can identify a baseline outcome that would have obtained if climate change were not damaging and emission cuts were free (but that still potentially depends on other states such as economic growth). Assume also that the change in outcome within each period, due to both the damages from climate change, and the private benefits of emissions, may be identified. Then we can identify* welfare *effects $B_u(q, \phi)$ and $D_u(q, \phi)$ as in Section 2.1.*

Here "benefits" from emissions are the negative of the abatement cost associated with the given pollution level. Thus while benefits themselves are negative, they increase with $q$: in a simple SEU model we would equate marginal damages with marginal benefits (subject to a second-order condition).

An important output of economic analyses of climate change is the "social cost of carbon" (SCC). If $\mathcal{Q}$ represents emissions in each period, and if we assume tastes and beliefs $(u, \sigma) \in \Psi$ are correct, then the SCC is $\frac{\partial}{\partial q_1}\overline{D}_{u,\sigma}(q)$: the marginal cost of additional emissions today, in current welfare terms. (The normalisation of social welfare given above ensures that this is measured in terms of present-day money, as experienced by the rich.[29])

In a simplified model with one dimension of policy choice, in which $\mathcal{Q}$ represents cumulative emissions, then the SCC remains closely related to $\overline{D}'_{u,\sigma}(q)$ (which is now the marginal cost of additional $CO_2$ emissions distributed optimally over time). Both these

[27] If $u$ is not additively separable in time then performing this normalisation requires a choice of reference outcome; marginal welfare of the rich at the present day is set to one, *given* that this reference outcome pertains. The marginal utility may then be altered in the true outcome.

[28] See Allen et al. (2009).

[29] As noted in footnote 27, this relationship breaks down when $u$ is not additively separable over time, in which case calculating the SCC requires a re-normalisation for every scenario.

quantities depend on the assumed future trajectory of emissions and the economy.[30] As climate damages are much less sensitive to emission timing than abatement costs, which are expected to decline over time, the majority of these will be emitted relatively soon and so the SCC will be similar to $\overline{D}'_u(q)$.[31]

# 3 "Dismal" Decision Theory

## 3.1 Justifiable Acts

If preferences are incomplete, which acts could be chosen? First, reflect that acts which are dominated under $\succcurlyeq$ are never chosen. On the other hand, acts that are optimal for some $(u, \sigma) \in \Psi$ seem promising. Importantly, these are the only two categories of acts. I define (cf. Battigalli et al., 2016):

**Definition 1.** An act $f \in \mathcal{F}$ is *justifiable* if there exists $(u_0, \sigma_0) \in \Psi$ such that $\overline{f}_{u_0,\sigma_0} \geq \overline{g}_{u_0,\sigma_0}$ for all $g \in \mathcal{F}$.

**Proposition 1.** *Make Assumptions 1-3. For any act $f \in \mathcal{F}$, exactly one of the following holds:*

*(1) $f$ is justifiable;*

*(2) there exists $g \in \mathcal{F}$ such that $\overline{g}_{u,\sigma} > \overline{f}_{u,\sigma}$ for all $(u, \sigma) \in \Psi$.*

So an act is a candidate for implementation iff it is justifiable.

I may thus suppose any act which may be implemented, to have been chosen by a "formal agent" with SEU tastes and beliefs $(u_0, \sigma_0) \in \Psi$. It is then illuminating to consider this act from the perspective of a different formal agent, with tastes and beliefs $(u, \sigma) \in \Psi$. (When comparing the perspectives of multiple agents, the normalisation of tastes $u$ are crucial; see Section 3.2.1.)

These can be considered as different evaluations from different existent agents, with inter- and intra-temporal interpretations. The "$(u, \sigma)$-agent" may well consider beliefs $\sigma_0$ or tastes $u_0$ to be quite "wrong": they may consider $u_0$ to be ethically unacceptable; they may believe their information to be of better quality; they may be an individual in the future, assess decisions made now, with access to genuinely better information. We would, of course, wish policy to still be reasonably desirable with the benefit of hindsight. In particular, if $(u_0, \sigma_0)$ and $(u, \sigma)$ are considered as corresponding to two different policy-makers, it is implicitly assumed that the second cannot change the policy.[32]

However, these different choices of $(u, \sigma) \in \Psi$ need not arise from different actual agents; they also represent a single idealised social planner with incomplete preferences.

## 3.2 Characterising when Incompleteness Matters

By Occam's razor, we should not complicate a model needlessly. So I characterise when the incompleteness of the model is important.

[30] The literature is a little inconsistent, with some defining the SCC to be the shadow price on emissions, i.e. to be the SCC in the sense of this paper only when this is evaluated only along the optimal emission trajectory.

[31] See Allen et al. (2009) for the relative insensitivity of climate outcomes to the timing of emissions.

[32] This is a plausible situation for climate change, even when we consider with the benefit of hindsight, since choices made now may greatly restrict future options: see e.g. Riahi et al. (2015).

### 3.2.1 Normalisations and Reference Outcomes

To allow meaningful comparisons, we need to normalise utility functions $u$ in a consistent way.

There are two natural candidates for this normalisation. Firstly, we could assume that every plausible utility function $u : \Delta(X) \to \mathbb{R}$ has image $[0, 1]$. This is used often in pure decision theory literature. However, this normalisation is very sensitive to the nature of the worst outcome, which may be very hard to assess, even subjectively (see Section 5.1).

It seems more natural to ensure that marginal valuations are comparable, in the range of outcomes where we have the most experience (this is also the usual normalisation in the climate change context). So fix a "reference" outcome $\mathbf{x}^0 \in X$, and identify the first factor $x_1$ of $X$; assume additionally that all $u \in \mathcal{U}$ are differentiable with respect to this first factor. Stipulate $u(\mathbf{x}^0) = 0$ and $\frac{\partial}{\partial x_1} u(\mathbf{x}^0) = 1$ (cf. Section 2.2).

### 3.2.2 "Dismal" Choices and Disagreement Regret

I now provide two possible measures of the extent of incompleteness, and show how they are linked. These definitions do not provide a decision rule, but a measure of the extent of incompleteness of preferences.

In each case, I assume that $\mathcal{Q}$ parametrises the set of acts available, and that $q_0^* \in \mathcal{Q}$ is justifiable: it is optimal for the formal agent with tastes and beliefs $(u_0, \sigma_0) \in \Psi$. It will then be evaluated by a different formal agent, characterised by $(u, \sigma) \in \Psi$. (Where there are multiple $q_0^*$ optimising $\overline{f}_{u_0,\sigma_0}(q)$, assume that the $q_0^*$ chosen is the worst from the point of view of $(u, \sigma)$.)

First, models of incomplete beliefs à la Bewley (1986, 2002) are closely linked to the field of robust Bayesian statistics.[33] So it is natural, also in the extended case of incomplete tastes, to measure "robustness" of a policy choice using a "regret" analogous to that of Savage (1951). The idea is to measure the difference between the pay-off from a given strategy, and the pay-off from the strategy that would have been optimal, had unknowns been resolved. The "unknowns" in question for Savage regret are "states" of the world; for the analysis here, I introduce the assessment of regret resulting from disagreements in tastes and beliefs.[34]

**Definition 2.** The *disagreement regret* associated to $(u, \sigma)$ when the choice from $\mathcal{Q}$ has been guided by $(u_0, \sigma_0)$ from family $\mathcal{Q}$ is

$$R^{\mathcal{Q}}_{u,\sigma}(u_0, \sigma_0) := \max_{q \in \mathcal{Q}} \{\overline{f}_{u,\sigma}(q)\} - \min \left\{ \overline{f}_{u,\sigma}(q_0^*) : q_0^* \in \arg\max_{q' \in \mathcal{Q}} \{\overline{f}_{u_0,\sigma_0}(q')\} \right\}.$$

Second, I work directly with the difference between expected utilities experienced by different formal agents, from small changes in acts. Fix $S, T \gg 1$. These are our benchmarks for "large" differences (analogous to the $p$ values used to determine statistical significance).

[33] See Seidenfeld et al. (1995) and Cerreia-Vioglio et al. (2013). Even without the precise assumptions for which their results hold, the link is clear.

[34] As with Savage regret, one could specify decision rules, such as minimising the maximum $R^{\mathcal{Q}}$. Stoye (2011) shows how such decision rules may be axiomatised.

**Definition 3.** The choice between acts $\{f(q) : q \in \mathcal{Q}\}$ is $(S,T)$-*dismal* if there exist $(u,\sigma),(u',\sigma') \in \Psi$ and $q \in \mathcal{Q}$ such that $\|\nabla \overline{f}_{u,\sigma}(q) - \nabla \overline{f}_{u',\sigma'}(q)\| > S\|\nabla \overline{f}_{u',\sigma'}(q)\|$ and $\|\nabla \overline{f}_{u,\sigma}(q)\| > T$.

In particular, if $\mathcal{Q}$ is an interval in $\mathbb{R}^n$ then the conditions are simply that $|\overline{f}'_{u,\sigma}(q) - \overline{f}'_{u',\sigma'}(q)| > S|\overline{f}'_{u',\sigma'}(q)|$ and $|\overline{f}'_{u,\sigma}(q)| > T$.

Thus, a choice situation is dismal if different beliefs and tastes can give rise to assessments that are *very different* (in relative terms) and *large* (in absolute terms). The two separate criteria distinguish "dismal choice" situations either from those in which welfare implications are large, but broadly agreed upon, and from those in which precise welfare implications are contentious, but agreed to be small.

It may be useful to consider benefits and damages separately:

**Definition 4.** If the effects of $f(q)$ split into damages and benefits, then *assessments of damages are* $(S,T)$-*dismal* if there exist $(u,\sigma),(u',\sigma') \in \Psi$ such that both $\bar{D}'_{u,\sigma}(q) - \bar{D}'_{u',\sigma'}(q) > S \cdot \bar{D}'_{u',\sigma'}(q)$ and $\bar{D}'_{u,\sigma}(q) > T$.

*Dismal assessments of benefits* are defined similarly.

Since $\overline{f}'_{u,\sigma}(q) = \overline{B}'_{u,\sigma}(q) - \overline{D}'_{u,\sigma}(q)$, if the choice between acts is $(S,T)$-dismal then assessments of either damages or benefits (or both) are at least $\left(\frac{S}{2},\frac{T}{2}\right)$-dismal. The converse is not true: it is possible for large damages and benefits, and large differences in damages and benefits, to cancel out. But the question of whether assessments of damages and benefits are dismal is of independent interest (see discussion after Corollary 1). Moreover assessments of their values may be made in quite separate models, so that such cancellation may be hard to identify.[35]

### 3.2.3 The Relationship Between These Definitions

There is a close relationship between situations of a dismal choice, and a high level of disagreement regret, as follows:

Assume that $\mathcal{Q}$ is one-dimensional and that, for all fixed $(u,\sigma) \in \Psi$, we know $\bar{f}_{u,\sigma}(q)$ is three times differentiable with respect to $q$ and well approximated by its Taylor series up to degree two. Assume also that $\bar{f}''_{u,\sigma}(q) < 0$ for all $(u,\sigma) \in \Psi$. Write $q_0^* = \arg\max_{q\in\mathbb{R}} \bar{f}_{u_0,\sigma_0}(q)$ and $q^* = \arg\max_{q\in\mathbb{R}} \bar{f}_{u,\sigma}(q)$.

**Proposition 2.** *Under the assumptions above, we may approximate*

$$R^{\mathcal{Q}}(u,\sigma,u_0,\sigma_0) \approx \frac{1}{2}\bar{f}'_{u,\sigma}(q_0^*)(q^* - q_0^*) \text{ where } q^* - q_0^* \approx \frac{\bar{f}'_{u,\sigma}(q_0^*)}{\bar{f}''_{u,\sigma}(q_0^*)}.$$

Moreover, if we split the acts into their private benefits and public damages, and assume these are both thrice differentiable, then:

**Corollary 1.** *Under the assumptions above, we may approximate*

$$R^{\mathcal{Q}}(u,\sigma,u_0,\sigma_0) \approx \frac{\left(\left(\overline{D}'_{u,\sigma}(q_0^*) - \overline{D}'_{u_0,\sigma_0}(q_0^*)\right) - \left(\overline{B}'_{u,\sigma}(q_0^*) - \overline{B}'_{u_0,\sigma_0}(q_0^*)\right)\right)^2}{\overline{D}''_{u,\sigma}(q_0^*) - \overline{B}''_{u,\sigma}(q_0^*)}.$$

[35] Our best assessments of the costs of emission reductions tend to be derived from detailed models specified for this purpose: see IPCC (2014) Annex II.10.

Under the assumptions given, $\overline{D}'_{u_0,\sigma_0}(q_0^*) = \overline{B}'_{u_0,\sigma_0}(q_0^*)$, but these terms are included in Corollary 1 because Definition 4 concerns differences between evaluations of marginal damages. Similarly, although $\bar{f}'_{u_0,\sigma_0}(q_0^*) = 0$, such terms could be introduced into Proposition 2 to illustrate that the regret may be given in terms of the differences, as well as levels, as in 3.

We conclude that, if the choice situation is dismal, then disagreement regret may be "large": this criterion identifies what may be important in total welfare terms. (When the second derivatives are larger then, all else being equal, the policies recommended are more similar; but observe that, when one compares the recommendations for different $(u, \sigma)$ giving rise to different second derivatives, it is unlikely that all will remain equal with the first derivatives.)

As will be discussed in Section 4, the difference between marginal damages under different tastes and beliefs tends to increase with $q_0$, whereas the difference between marginal costs for different tastes and beliefs tends to decrease with $q_0$. On the other hand, if $q_0$ can be found such that neither assessments of damages nor benefits are dismal, evaluations of such policies will be more robust.

Corollary 1 follows trivially from Proposition 2, but illustrates the role of assessments in damages and benefits, as well as the possibility of cancellation of dismal assessments. The latter happens if damages and costs move in the same direction with changes in tastes and beliefs. Conversely, if damages and costs move in opposite directions, the effects compound (and optimal quantities are very different). But when optimal quantities are very different under different tastes and beliefs, the values of marginal damages and marginal benefits at this price are more similar. This suggests that a price-based policy would be more robust in such a scenario, and that it is worthwhile to undertake a "prices versus quantities" analysis, in the style of Weitzman (1974). I perform this in a subsequent work.

# 4 Climate change as a scenario with dismal choices

## 4.1 Incomplete Beliefs: Climate Change Damages

A simple model illustrates the importance of our incomplete knowledge on the damages from climate change. I have combined an economic model in the style of DICE 2013 (Nordhaus and Sztorc, 2013) with the "simple" climate model of Allen et al. (2009), which incorporates five stochastic scientific parameters. (See Baldwin 2014 Chapter 2, for full details). I use a wide range of plausible damage functions, which diverge once temperatures pass the threshold at which the DICE damage function is calibrated.

Index by $\alpha$ the output damages at 4°C warming. In DICE 2013, $\alpha = 4.09\%$ and the damage function is quadratic. However, as discussed in Section 1.1, it is quite plausible that a very much higher figure is appropriate. Figure 1 shows the damage functions in use, and the resulting spread of values for the present-day social cost of carbon. (The functions in Figure 1b tend to be convex; apparent concavity is due to the logarithmic axes used!)

It is not really "plausible" to assume that any one of the damage functions of Figure 1 is correct with certainty, as the beliefs of this example do. However, a range of probability densities over these functions may be plausible. The estimated SCC will be

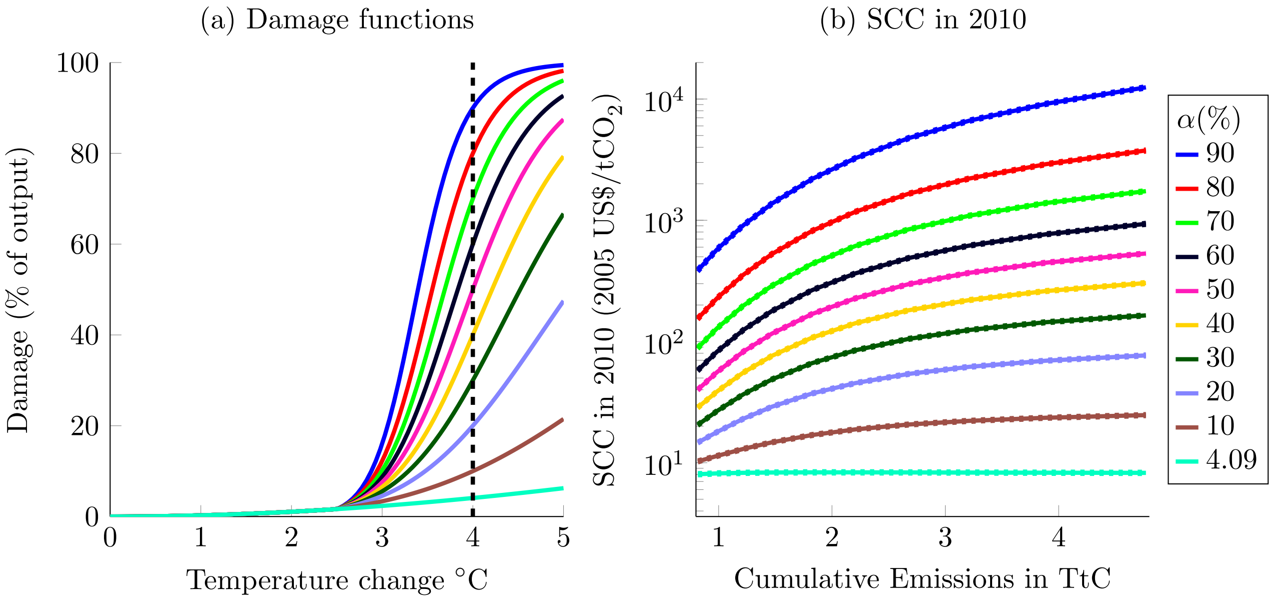


**Figure 1:** *(a) The damage functions assessed here. (b) The social cost of carbon as a function of cumulative emissions. The dotted lines denote the 95% confidence intervals of the estimators. The legends give the corresponding values of* $\alpha$*, that is, output damages from 4°C warming.*

sensitive to the probability accorded to the top end of the distribution, which it is hard to constrain in a meaningful way.

## 4.2 Incomplete tastes: the pure rate of time preference.

It is well known that different values of the pure rate of time preference $\delta$ can give rise to vast differences in the estimated social cost of carbon. One could obtain a subset of all possible damage functions $\overline{D}_{u,\sigma}(q)$ by letting $u$ determine $\delta$ and assuming tastes are otherwise fixed, along with beliefs $\sigma_0$. We obtain an expected social cost of carbon $\overline{D}'_{u,\sigma_0}(q)$ for each "plausible" choice of $u$, i.e. each plausible $\delta$. The social cost of carbon increases dramatically as $\delta \to 0$, even for moderate assumptions on the form of damages: see Tol (2009, Table 2) and Ackerman and Stanton (2012), for example.

## 4.3 Acemoglu et al. (2012) and the role of beliefs in abatement costs.

This work distinguishes clean and dirty technology. It models technological change as endogenous, and is directed between sectors on the basis of productivity, prices and market size.[36] A crucial parameter is the elasticity of substitution $\epsilon$ between the two sectors.

If $\epsilon$ is high enough, optimal policy is given by a temporary subsidy and a temporary carbon tax. If $\epsilon$ is lower, the carbon tax continues indefinitely. If the two sectors are complementary, i.e. $\epsilon < 1$, the only way to avoid environmental disaster is to bring an

[36] I focus here on the case of inexhaustible resources analysed by Acemoglu et al. (2012).

end to economic growth. Thus different beliefs on this crucial substitution parameter give rise to widely different estimates of the cost of meeting an environmental target.

## 4.4 Conclusions for Modelling Climate Change

Assume that $\mathcal{Q}$ is one-dimensional and represents cumulative emissions. Sections 4.1-4.3 substantiate my key claim:

**Claim 1.** *Assessments of damages* $\overline{D}'_{u,\sigma}(q)$ *are* $(S,T)$*-dismal for large enough* $q$ *and for* $(S,T) \leq (100, \$500)$.

*Assessments of benefits* $\overline{B}'_{u,\sigma}(q)$ *are* $(S,T)$*-dismal for small enough* $q$*, for some non-negligible* $(S,T)$.

Assessments of damages have less of a tendency to be dismal for lower temperature targets: this is clear from Figure 1. A threshold could be estimated for any $S, T$ defining the dismal situation. Conversely, benefits, i.e. the costs of emission reductions, have a less dismal nature for high $q$. Indeed, some summaries of figures suggest benefits, i.e. the cost of emission reductions, are moderate for most assessed values of $q$.[37] However, the lowest cumulative emissions regularly assessed correspond to a 2100 concentration of 450ppm $CO_2$eq, and some argue that such an emission level is still too high (i.e., implicitly, our assessment of damages from such warming should still be dismal).[38] Very tight abatement scenarios are seldom assessed, but could have dramatic short-term consequences.

The policy-maker may be trapped between two dismal assessments. But Claim 1 raises a very natural question.

**Question 1.** *Does there exist a level* $q$ *of cumulative emissions, and moderate* $(S,T) \in \mathbb{R}^2$*, such that assessments of neither damages nor benefits are* $(S,T)$*-dismal?*

If so, this presents a potential "sweet spot" to policy-makers. Such $q$ will not characterise optimal policy for all formal agents with plausible tastes and beliefs, of course, but the potential downside from such policies is better constrained and understood.

The examples given above examples rule out many plausible tastes and beliefs. Other models show the importance of considering more than one consumption good, climate damages that affect the capital stock or total factor productivity, and the importance of climatic tipping points.[39] Ideally, if several modifications seem more plausible than the baseline model, one should include them all (cf. Section 5.2.1, in which we see that the most conventional damage function masks the importance of geophysical uncertainties).

Moreover, allowing only one set of beliefs for the geophysical part of the model is inaccurate. Example 4.1 used a uniform prior distribution to convert likelihoods into probabilities. In turn, Allen et al. (2009) use log-normal parametric forms calibrated to recent data to derive the likelihoods used. Clearly, different priors and forms would alter the results.

[37] See IPCC (2014) Figure 3.4.

[38] See, e.g. Hansen and Sato (2012).

[39] See Sterner and Persson (2008), Dietz and Stern (2015) and Cai et al. (2015a,b), Lontzek et al. (2015) respectively for these points.

# 5 Implications of Modelling with Incomplete Preferences

## 5.1 Connection to Weitzman's "Dismal Theorem"

The nomenclature "dismal choice" is inspired by Weitzman's "Dismal theorem" (Weitzman, 2009a, 2014): it leads to a re-interpretation of that work's essential insight.

For expositional simplicity, I present Weitzman's ideas here as a consideration of $\overline{D}'_{u,\sigma}(q)$, where $q \in \mathcal{Q}$, which is one-dimensional, and represents cumulative emissions.[40] Weitzman imposes a lower limit on global utility, the "virtual statistical life of civilisation". This, he argues, is not well understood. I notate it as $\lambda$; it is a lower bound on tastes $u$ and so I write $u_\lambda$ for the taste function with this lower bound imposed. As he leaves the specification of $\lambda$ open, he is implicitly modelling tastes as incomplete.

Weitzman makes further assumptions, which I will denote $\sigma_0$, on the relationship between welfare damages and uncertain geophysical parameters, and on the Bayesian prior distribution for these parameters. He then argues that, however much more information we gather, we will update beliefs to some $\sigma_1$ which will satisfy

$$\lim_{\lambda \to -\infty} \overline{D}'_{u_\lambda,\sigma_1}(q) = \infty. \tag{1}$$

What does this mean? According to Weitzman (2014), it is "absurd" to say that the social cost of carbon is infinite. But in any case, since Weitzman (2009a) argues explicitly that welfare is bounded below, what is the relevance of taking the limit $\lambda \to -\infty$?

My answer is that doing so illuminates the importance of incompleteness of tastes in his model. If the lower bound on welfare were known and understood, there would be no problem: one would simply substitute in its value. But $\lambda$ is unknown, and what Equation 1 tells us is that this matters. The fact that the limit in (1) is infinite demonstrates firstly that $\overline{D}'_{u_\lambda,\sigma_1}(q)$ *depends fundamentally* on the assumed value of $\lambda$, and secondly that $\overline{D}'_{u_\lambda,\sigma_1}(q)$ can be *very large*. This, then, is exactly an instance of assessments of damages being dismal, in the sense of Definition 4: the social cost of carbon depends fundamentally on our subjective tastes and assumptions $(u, \sigma) \in \Psi$.[41]

However, in the formulation presented here, the value of $\lambda$ is clearly not the only subjective aspect to which the result is sensitive. The result depends on the particular prior beliefs $\sigma_0$. Under different $\sigma$, the result can disappear; or only hold for large values of $q$ (so that the scenarios of dismal choices are irrelevant as long as a small enough $q$ is imposed); or it can hold even for ruinously low $q$; see Millner (2013).[42] The language

[40]In fact, Weitzman (2009a) discusses the stochastic discount factor, but in later work (Weitzman, 2014) he expresses his ideas in terms of the social cost of carbon.

[41]Thinking in this way also shows clearly that we cannot, as suggested by Arrow and Priebsch (2014) and others, resolve the problem by using bounded utility functions – which are indeed required by the axioms of SEU – because we do not know where the bound lies.

[42]Weitzman's argument is only in terms of marginal willingness to pay to alleviate climate damages; to find the total willingness to pay we would have to consider also marginal costs of abatement. Millner (2013) shows that, under Weitzman's (2009a) formulation, if we can transfer utility to the future with certainty, then there exists an affordable level of mitigation which removes the "dismal" problem. However, with uncertainty in these transfers, there exist formulations such that we would willing to sacrifice all current consumption to mitigate climate change. The arguments are slightly different for

of Definition 4 allows us to identify this critical importance of $\sigma$. It is unclear how one would formulate a Weitzman-style result for this case.

Weitzman (2009a, 2014) emphasises the importance of “fat-tailed uncertainty”, under which it is the very low probability, very high impact events in the tails of these distributions which should drive our evaluations.[43] Uncertainty in the damage function encompasses the value we place on worst-case scenarios, but also what we think about the effect of around, say, 4°C warming. This outcome is not a “rare catastrophe”, but scientists’ best guess of the outcome of following business-as-usual emission scenarios.

Should one instead model with a fat tailed distribution of damage at 4°C (perhaps as a distribution over the damage functions of Section 4.1), and argue that the expectation is driven by the low-probability event that 4 degrees has a catastrophic impact? The problem with doing so is that it implies that one specific assessment of such a probability density is well-founded. There is no such assessment. So focussing on the multiplicity of plausible valuations is more accurate.

In Weitzman’s discussion, uncertainty is in future consumption under climate change. However, he focusses analysis on the “climate sensitivity” parameter, and it is this that the literature has picked up.[44] So, despite his following work discussing forms of the “damage function”, and the writing and modelling of others, uncertainty in damages is sometimes missed by those aiming to learn from his work.[45] Moreover, this is an important omission, as use of the popular “DICE” damage function happens make an integrated assessment model rather insensitive to uncertainty in climate sensitivity (see Section 5.2.1).

Finally, a difference between my formulation and that of Weitzman (2009a, 2014) is that I consider that assessments of the cost of reducing emissions may also be dismal. These combined dismal effects have important implications for the question of policy choice, as I discuss in more detail in ongoing work in this area.

## 5.2 Showing Stylised Facts are Not Robust

Economic theory is often concerned with characterising the basic shape of relationships and “stylised facts”. Many such “facts” have been derived regarding climate policy responses. Some, such as the basic principle of putting a price on carbon, supplemented with subsidies for immature low-carbon technologies, seem very robust to model specification. For example, the idea that we should stop using coal rapidly is a point of agreement between the very different approaches of Golosov et al. (2014) and McGlade and Ekins (2015). However, others have gained considerable traction despite being

---

this paper as we work with the social cost of carbon, rather than the stochastic discount rate (so we do not consider the possibility of transfers to the future with certainty) but the distinctions are as follows: we seek $q$ such that $\overline{B}'_{u,\sigma}(q) = \overline{D}'_{u,\sigma}(q) < \infty$. Such $q$ exist for some taste and belief combinations, but for others, one finds that $\overline{B}'_{u,\sigma}(q) < \overline{D}'_{u,\sigma}(q)$ for every $q > \underline{q}$, where $\underline{q}$ is a lower bound to the cumulative emission levels we can actually attain, i.e. satisfying $\lim_{q \to \underline{q}} \overline{B}_{u,\sigma}(q) = \infty$.

[43] See Weitzman (2009a, 2012, 2014) and also Pindyck (2013b, 2017).

[44] Physicists have worked with economists to question whether this is the key geophysical uncertainty; see Otto et al. (2013) and Roe and Bauman (2013).

[45] See Weitzman (2009c, 2010, 2011, 2012), Stern (2010, 2013) and Pindyck (2013b, 2017) for discussions of damages, and Dietz (2011), Ackerman and Stanton (2012), Dietz and Stern (2015) and Rezai and Van Der Ploeg (2017) for modelling using wider ranges.

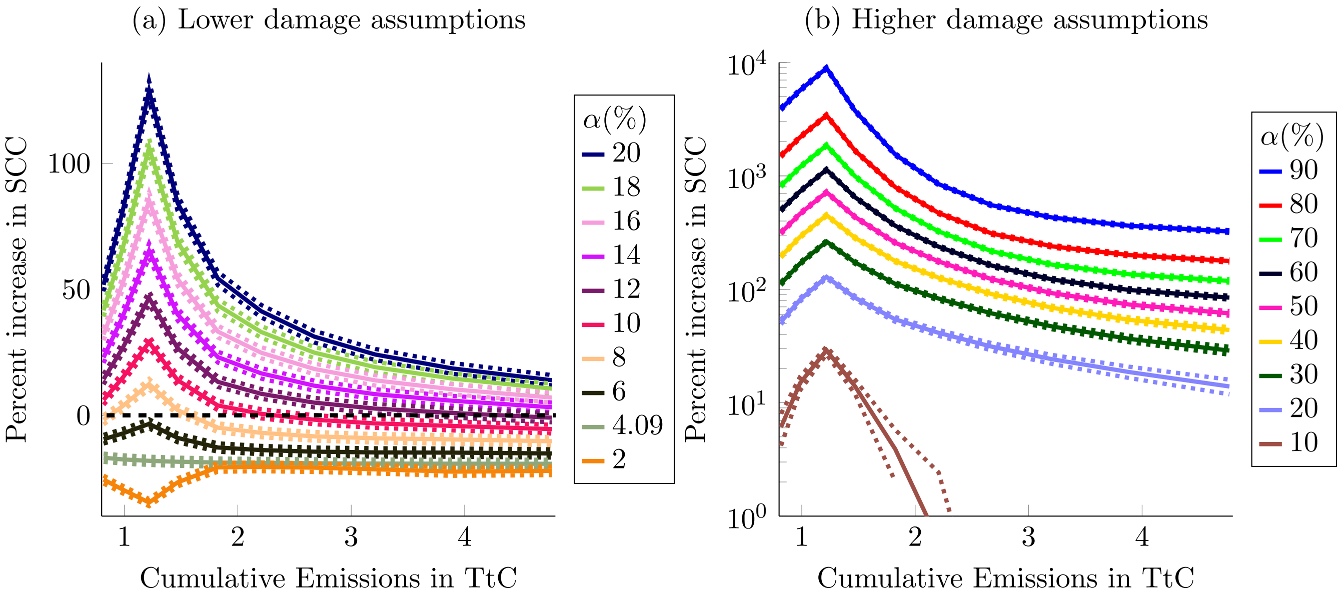


**Figure 2:** *The percent additional SCC due to using means instead of the best guesses of parameters. The legends give the damage parameter $\alpha$. The "DICE damage function" is given by $\alpha = 4.09\%$.*

dependent on a particular specification of the model.

### 5.2.1 The importance of uncertainty in climate sensitivity.

Weitzman (2009a) emphasised the importance of incorporating uncertainty in geophysical aspects of an integrated assessment model, but several studies on the economics of climate change have since dismissed it as relatively unimportant.[46] In fact, it does seem to be unimportant if one uses the DICE damage function (see, e.g. Nordhaus and Sztorc, 2013) – but not otherwise.

See Figure 2. I have taken the estimates for the SCC given in Section 4.1, which use a joint probability density function of geophysical outcomes derived from the work of Allen et al. (2009). I have compared these with estimates which use only the best guesses for the 5 uncertain parameters. The percent additional SCC in the model with uncertainty, relative to the model without uncertainty, is plotted.[47]

### 5.2.2 (In)dependence of the SCC on cumulative emissions.

Surprisingly, many conventional models of climate change derive a SCC which is independent of the emission pathway. This was first seen back in the work of Nordhaus (1993) but has been developed in many other lines of work, where it is sometimes even an

[46] See, for example, Lemoine and McJeon (2013), Golosov et al. (2014), Nordhaus (2014), Gillingham et al. (2015) and Traeger (2015). This is then used in reviews such as that by Hassler et al. (2016).

[47] The very dramatic effect of uncertainty at around 1.2TtC, as seen in Figure 2, is due to the fact that, in the deterministic case, emissions of this level give rise to warming just below 2.5°C, at which temperature the damage functions diverge.

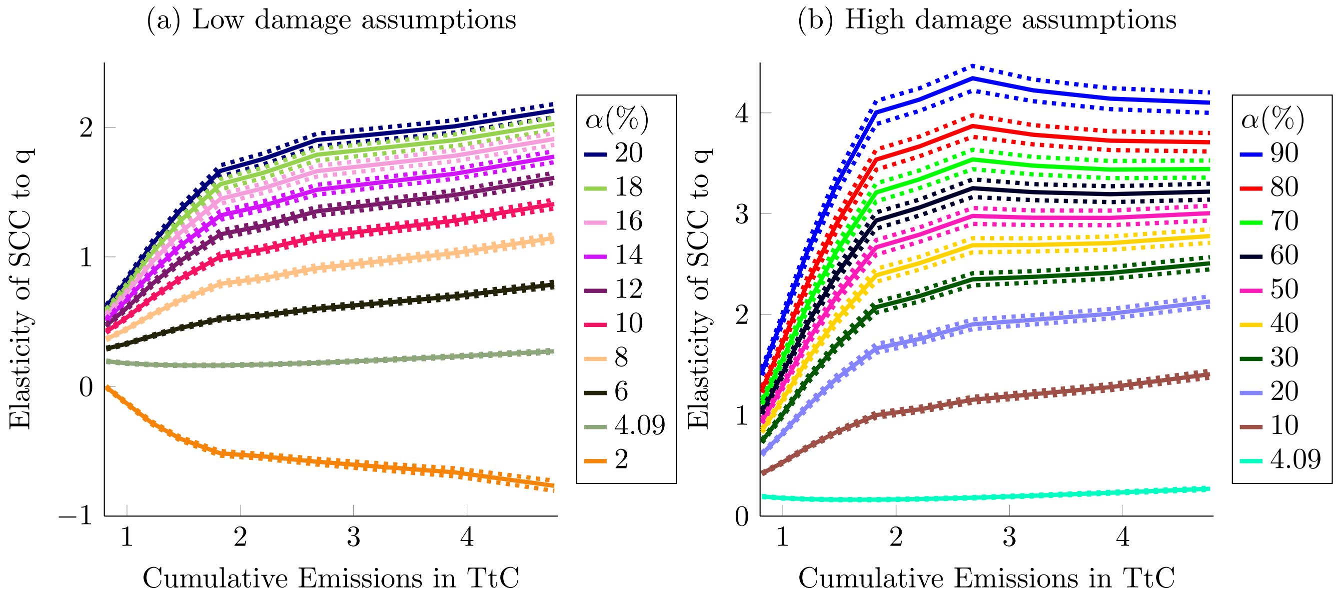


**Figure 3:** *Elasticity of the SCC to cumulative emissions $q$, plotted for multiple damage assumptions $\alpha$. The dotted lines denote the 95% confidence intervals of the estimators.*

implicit assumption.[48] But it may be seen to be an artefact of the quadratic functional form which was assumed by Nordhaus (1993), and followed by many subsequent scholars: see Figure 1(b). Figure 3 illustrates, indeed, that the SCC is reasonably elastic to cumulative emissions *unless* we assume this particular relationship.

The perceived flatness of the SCC has multiple effects in policy analysis. As is seen in the work of Nordhaus (1993, 2008, 2014), the optimal carbon tax at a given point in time is unaffected by even a very substantial delay in implementing policy. Similarly, a massive release of methane from thawing permafrost would not affect optimal policy, when the SCC is invariant in this way. Both of these cases involve a shift upward in the curve $\overline{B}'_{u,\sigma}(q)$; if $\overline{D}'_{u,\sigma}(q)$ is completely flat then doing so simply shifts the "optimal" quantity, without affecting the social cost of carbon at all.

This flatness also means that the optimal carbon price will be entirely independent of private benefits $\overline{B}_{u,\sigma}(q)$, as observed by Golosov et al. (2014).[49]

The fact that the SCC is not independent of cumulative emissions for many plausible beliefs on the structure of damages is not novel to this paper; see e.g. Hope (2008), Dietz and Stern (2015). But its prevalence in notable parts of the recent literature show it to be an important example of a non-robust stylised fact.

[48] See Nordhaus (2008, 2014), Hope (2006), Hassler and Krusell (2012), van den Bijgaart et al. (2016), Golosov et al. (2014), Hassler and Krusell (2012), Traeger (2015).

[49] Barrage (2014) performs a fairly wide sensitivity analysis of the results of Golosov et al. (2014), but unfortunately does not question the quadratic damage curve.

# 6 Extension: Multi-SEU and Agreement Across Decision Theories

Many other theories of decision-making under ambiguity derive representation theorems which effectively use SEU as a jumping-off point. In these theories, the preference relation is represented by: firstly giving a (possibly singleton) set $\mathcal{U}_0$ of utility functions, normalised as above, and a (possibly singleton) set $\Sigma_0$ of probability measures on uncertain states; and secondly providing a function $H$ on the vector of SEU evaluations $\{\overline{f}_{u,\sigma}\}_{(u,\sigma)\in\mathcal{U}_0\times\Sigma_0}$ of acts $f$. The function $H$ will be weakly increasing in the separate SEU-style evaluations individually, and strictly increasing when all strictly increase. The models of Cerreia-Vioglio et al. (2011, 2015) and Ghirardato et al. (2004) have this form, and hence so do most considered in this literature.[50]

For a particular decision problem, let $\mathcal{U}$ and $\Sigma$ be the union of all such $\mathcal{U}_0$ and $\Sigma_0$ respectively. As before, assume that $\Psi \subseteq \mathcal{U}\times\Sigma$ is a convex, compact subset of Euclidean space. The representation functions for a decision theory are then real-valued functions $H$ on $\mathbb{R}^\Psi$.

Particularly important examples are $H_{u,\sigma} : \mathbf{z} \mapsto \mathbf{z}_{u,\sigma}$, which represent SEU decision-makers who only consider one utility function and probability distribution.[51]

We may now define:

**Definition 5.** The *plausible first-order tastes and beliefs* of the model is the set $\Psi$, consisting of *plausible tastes* $u \in \mathcal{U}$ (normalised utility functions affine on $\Delta(X)$) paired with *plausible beliefs* $\sigma \in \Sigma$.

The *plausible second-order tastes and beliefs* is a set $\mathcal{H}$ of continuous functions $H : \mathbb{R}^\Psi \to \mathbb{R}$ such that $H_{u,\sigma} \in \mathcal{H}$ for all $(u,\sigma) \in \Psi$ and such that if $\mathbf{y}_{u,\sigma} \geq \mathbf{z}_{u,\sigma}$ for all $(u,\sigma) \in \Psi$ then $H(\mathbf{y}) \geq H(\mathbf{z})$; and if $\mathbf{y}_{u,\sigma} > \mathbf{z}_{u,\sigma}$ for all $(u,\sigma) \in \Psi$ then $H(\mathbf{y}) > H(\mathbf{z})$.

Suppose a decision-maker has second-order tastes and beliefs represented by $H$. To evaluate an act $f$, they first form the vector of all SEU evaluations $\{\overline{f}_{u,\sigma}\}_{(u,\sigma)\in\Psi} \in \mathbb{R}^\Psi$ of this act, and then apply the function $H$.

Note that I allow any $(u,\sigma) \in \Psi$ to define "plausible" second-order SEU preferences. That is, if $(u,\sigma) \in \Psi$ then the SEU preferences defined by $u$ and $\sigma$ are not "objectively" rejected. This assumption seems reasonable, as the purpose of decision theories with multiple beliefs or tastes is essentially to compensate for the existence a set of plausible descriptions of the world, so it would be perverse to include in such a set utilities or beliefs that are objectively falsifiable.

Now, given acts $f, g \in \mathcal{F}$, if $\overline{f}_{u,\sigma} \geq \overline{g}_{u,\sigma}$ holds for all $(u,\sigma) \in \Psi$, then by definition of $\mathcal{H}$, policy-makers using any second-order tastes and beliefs in $\mathcal{H}$ will agree that $f$ is preferred to $g$. (In this case, on may say that $f$ is *objectively* preferred to $g$). Conversely, if this fails, then then there exists some SEU decision-maker, represented by $H_{u,\sigma} \in \mathcal{H}$, who strictly prefers $g$ to $f$. So clearly:

[50] The models of Anscombe and Aumann (1963), Gilboa and Schmeidler (1989), Hansen and Sargent (2001), Klibanoff et al. (2005), Maccheroni et al. (2006) are all special cases of the model of Cerreia-Vioglio et al. (2011); the models of Arrow and Hurwicz (1972) and Schmeidler (1989) are special cases of Ghirardato et al. (2004).

[51] Similarly, if preferences are those of Gilboa and Schmeidler (1989, "maximin expected utility"), then identify some $u_0 \in \mathcal{U}$ and a convex, compact set $\Sigma_0 \subset \Sigma$, and let $H(\mathbf{z}) := \min_{\sigma\in\Sigma_0} \mathbf{z}_{u_0,\sigma}$.

**Lemma 2.** *The set of decision-makers whose preferences are represented by the set $\mathcal{H}$ of plausible second-order tastes and beliefs agree that $f$ is preferred to $g$ if and only if $\overline{f}_{u,\sigma} \geq \overline{g}_{u,\sigma}$ for all $(u, \sigma) \in \Psi$.*

Moreover, if we extend the definition of justifiable acts (Definition 1):

**Definition 6.** Say $f \in \mathcal{F}$ is *second-order justifiable* if there exists $H \in \mathcal{H}$ such that $f \in \arg\max_{g \in \mathcal{F}} H\left(\{\overline{g}_{u,\sigma}\}_{(u,\sigma)\in\Psi}\right)$.

**Proposition 3.** *An act $f \in \mathcal{F}$ is second-order justifiable iff it is justifiable.*

This attractive result implies that, to consider the scope of possible prescriptions from a range of decision theories, the model of Section 2.1 will suffice.

# 7 Discussion and Related Literature

## 7.1 Important Related Literature

Danan et al. (2016) theoretically develop decision theory with incomplete preferences, motivated by climate change. Their "unambiguous Pareto dominance" between heterogeneous tastes and beliefs, which requires a non-empty intersection between agents' beliefs, seems too demanding a criterion in the light of Claim 1 of this paper. On the other hand, their characterisation of "common-tastes unambiguous Pareto dominance" provides an alternative motivation of the model of Section 2.1.

Danan et al. (2016) also provide a result similar to my characterisation of justifiable acts (their Proposition 1). However, they assume only incompleteness in beliefs for this result, and their methods do not appear to extend, as mine do, to incomplete tastes.

Dietz and Matei (2016) find "space for agreement" on climate change, using "almost time-stochastic dominance". That is, they offer a menu of long-term atmospheric concentrations of $CO_2$, to agents with varied risk and time preferences; they find that all, except those with the most extreme risk or time preferences, choose the lowest concentration available.

However, despite allowing great flexibility in tastes, Dietz and Matei (2016) have imposed a fixed set of beliefs. In particular, they assume agents already agree that there is a moderately large risk of very extreme outcomes. It is not surprising that all would agree on climate action in this scenario. However, the existence of many experts claiming that optimal climate policy should be rather moderate (see, e.g. Hassler et al. 2016) implies that this does not describe the extent of real-world disagreement.

Chambers and Melkonyan (2017) apply the model of Bewley (1986, 2002) with incomplete beliefs. However, unlike me, they focus on his suggestion to complete the decision theory using an "inertial assumption" which leads to a bias in favour of the *status quo*. That is, when no action is preferred to continuing with the prevailing scenario, then this should prevail. Chambers and Melkonyan (2017) have used this to argue that a model of incomplete expected utility implies a weak response to climate change.

I disagree. Firstly, the "inertia assumption" of Bewley is an addition to the model: it has no objective force, and we need not accept it as normatively desirable or objectively reasonable. But more importantly, the identity of the status quo is not clear in this case! One can interpret, as Chambers and Melkonyan (2017) do, that our prevailing economic system forms a status quo. But one can also give that label to the climatic conditions under which our society developed; doing so means that the inertia assumption demands extreme cuts in emissions. I have therefore not taken this approach in assessing the implications of incomplete tastes and beliefs.

## 7.2 Models to Guide Real-World Policy

This model is not meant to be a counsel of despair for cost-benefit analysis; the model of incomplete preferences does not provide *no* information. It provides a "plausible" range. And if choices are dismal within that plausible range, then this fact is extremely important information. Moreover, it shows us where we can best re-direct our attention to analysis.

The idea that objective preferences are incomplete is especially important when the model in question is not a theoretical exploration, but a quantitative analysis to guide policy. In particular, if results are very sensitive to subjective judgements or have serious limitations to their validity, this ought to be made clear up-front. Much wider and more transparent sensitivity analyses should be available, with all modelling assumptions that cannot be substantiated open to exploration. Since the final valuation and choice will always be subjective, these questions should be a subject for public debate.

This is often the case. For example, there is no objectively "correct" level of redistributive taxation; it must be chosen according to the inequality aversion of the policy-maker and their electorate. But ideally this choice would be informed by a sound economic assessment of the dead-weight loss of taxation.

Meanwhile, such analysis illuminates the aspects of new research which would be most helpful to improve their models and narrow their range. In the climate change context, one may point to the studies of Newbold and Marten (2014), Anderson et al. (2014). For example, tests for robustness against different specifications of the damage function should be as routine as checks with different discount rates.

## 7.3 Conclusion

This paper has argued that policy choice to combat an externality can be very sensitive to subjective factors. These may be embodied in models as "tastes", or value judgements. And these may take the form of "beliefs", or probability distributions. A choice situation has been identified as "dismal" if differing tastes and beliefs give rise to very different (normalised) marginal valuations of policies. The analysis of Weitzman (2009a) is an example of a dismal choice, but I question whether he has focussed attention on the most important unknown feature and uncertain state of the situation.

An important example of such a situation is the question of climate change. I have argued that assessments of both damages from climate change, and the cost of abatement, may have these dismal features. Assessments of damages are more likely to be dismal for a lax policy choice, whereas assessments of mitigation costs may be dismal for very stringent policy. So, in particular, one should be very suspicious of the accuracy of estimates of the social cost of carbon which are evaluated along emission trajectories which may give rise to high levels of warming. And some stylised facts that emerge as more qualitative results from models are not robust to different plausible choices of tastes and beliefs. Modelling with incomplete tastes and beliefs is needed to illuminate these key features of the choice situation.

# References

D. Acemoglu, P. Aghion, L. Bursztyn, and D. Hemous. The environment and directed technical change. *American Economic Review*, 102(1):131–66, 2012.

F. Ackerman and E. A. Stanton. Climate risks and carbon prices: Revising the social cost of carbon. *Economics: The Open-Access, Open-Assessment E-Journal*, 6(2012-10), 2012.

M. Allen, D. Frame, C. Huntingford, C. Jones, J. Lowe, M. Meinshausen, and N. Meinshausen. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, 458(7242):1163–1166, 2009.

B. Anderson, E. Borgonovo, M. Galeotti, and R. Roson. Uncertainty in climate change modeling: Can global sensitivity analysis be of help? *Risk Analysis*, 34(2):271–293, 2014.

F. J. Anscombe and R. J. Aumann. A definition of subjective probability. *The Annals of Mathematical Statistics*, 34(1):199–205, 1963.

K. Arrow and M. Priebsch. Bliss, catastrophe, and rational policy. *Environmental and Resource Economics*, 58(4):491–509, 2014.

K. J. Arrow and L. Hurwicz. An optimality criterion for decision making under ignorance. In *Uncertainty and Expectations in Economics*. B. Blackwell, Oxford, 1972.

C. Azar and T. Sterner. Discounting and distributional considerations in the context of global warming. *Ecological Economics*, 19(2):169–184, 1996.

E. Baker, H. Chon, and J. Keisler. Advanced solar R&D: Combining economic analysis with expert elicitations to inform climate policy. *Energy Economics*, 31(0):S37–S49, 2009.

E. Baldwin. *Modelling Preferences in Economics*. PhD thesis, University of Oxford, 2014.

L. Barrage. Sensitivity analysis for Golosov, Hassler, Krusell, and Tsyvinski (2014): “Optimal taxes on fossil fuel in general equilibrium”. *Econometrica*, 82(1): Supplementary material, 2014.

P. Battigalli, S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci. A note on comparative ambiguity aversion and justifiability. *Econometrica*, 84(5):1903–1916, 2016.

T. F. Bewley. Knightian decision theory, Part I. Discussion Paper 807, Cowles Foundation, 1986. 25(2):79–110, 2002.

T. F. Bewley. Knightian decision theory. Part I. *Decisions in Economics and Finance*, 25(2):79–110, 2002.

M. Burke, S. M. Hsiang, and E. Miguel. Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239, Nov. 2015.

Y. Cai, K. L. Judd, T. M. Lenton, T. S. Lontzek, and D. Narita. Environmental tipping points significantly affect the cost-benefit assessment of climate policies. *Proceedings of the National Academy of Sciences*, 112(15):4606–4611, 2015a.

Y. Cai, K. L. Judd, and T. S. Lontzek. The social cost of carbon with economic and climate risks. Technical Report arXiv:1504.06909v2, arXiv.org, April 2015b.

S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio. Uncertainty averse preferences. *Journal of Economic Theory*, 146(4):1275–1330, 2011.

S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and L. Montrucchio. Ambiguity and robust statistics. *Journal of Economic Theory*, 148(3):974–1049, 2013.

S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci. A new look at local expected utility. Mimeo., 2014.

S. Cerreia-Vioglio, D. Dillenberger, and P. Ortoleva. Cautious expected utility and the certainty effect. *Econometrica*, 83(2):693–728, 2015.

R. G. Chambers and T. Melkonyan. Ambiguity, reasoned determination, and climate-change policy. *Journal of Environmental Economics and Management*, 81:74–92, Jan. 2017.

E. Danan, T. Gajdos, B. Hill, and J.-M. Tallon. Robust social decisions. *American Economic Review*, 106(9):2407–25, September 2016.

M. Dell, B. F. Jones, and B. A. Olken. Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3):66–95, 2012.

S. Dietz. High impact, low probability? An empirical analysis of risk in the economics of climate change. *Climatic Change*, 108(3):519–541, 2011.

S. Dietz and N. A. Matei. Spaces for agreement: A theory of time-stochastic dominance and an application to climate change. *Journal of the Association of Environmental and Resource Economists*, 3(1):85–130, 2016.

S. Dietz and N. Stern. Endogenous growth, convexity of damage and climate risk: How Nordhaus' framework supports deep cuts in carbon emissions. *The Economic Journal*, 125(583):574–620, 2015.

J. Etner, M. Jeleva, and J.-M. Tallon. Decision theory under ambiguity. *Journal of Economic Surveys*, 26(2):234–270, 2012.

P. C. Fishburn. *Utility theory for decision making.* Publications in Operations Research, No. 18. Wiley, 1970.

A. C. Fisher and P. V. Le. Climate policy: Science, economics, and extremes. *Review of Environmental Economics and Policy*, 8(2):307–327, 2014.

K. Fisher-Vanden. Introduction to the special issue on technological change and the environment. *Energy Economics*, 30(6):2731–2733, 2008.

T. Galaabaatar and E. Karni. Subjective expected utility with incomplete preferences. *Econometrica*, 81(1):255–284, 2013.

P. Ghirardato, F. Maccheroni, and M. Marinacci. Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, 118(2):133–173, 2004.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

I. Gilboa, A. Postlewaite, and D. Schmeidler. Is it always rational to satisfy Savage's axioms? *Economics and Philosophy*, 25:285–296, 2009.

I. Gilboa, F. Maccheroni, M. Marinacci, and D. Schmeidler. Objective and subjective rationality in a multiple prior model. *Econometrica*, 78(2):755–770, 2010.

K. Gillingham, W. D. Nordhaus, D. Anthoff, G. Blanford, V. Bosetti, P. Christensen, H. McJeon, J. Reilly, and P. Sztorc. Modeling uncertainty in climate change: A multi-model comparison. Working Paper 21637, National Bureau of Economic Research, October 2015.

M. Golosov, J. Hassler, P. Krusell, and A. Tsyvinski. Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, 82(1):41–88, 2014.

M. Hanemann. What is the economic cost of climate change? In S. H. Schneider, editor, *Climate Change Science and Policy*. Island Press, 2009.

J. Hansen and M. Sato. Paleoclimate implications for human-made climate change. In A. Berger, F. Mesinger, and D. Sijacki, editors, *Climate Change: Inferences from Paleoclimate and Regional Aspects*, pages 21–47. Springer Vienna, 2012.

L. P. Hansen and T. J. Sargent. Robust control and model uncertainty. *The American Economic Review*, 91(2):pp. 60–66, 2001.

J. Hassler and P. Krusell. Economics and climate change: Integrated assessment in a multi-region world. *Journal of the European Economic Association*, 10(5):974–1000, 2012.

J. Hassler, P. Krusell, and J. Nycander. Climate policy. *Economic Policy*, 31(87): 503–558, 2016.

G. Heal and A. Millner. Discounting under disagreement. Working Paper 18999, National Bureau of Economic Research, April 2013.

G. Heal and A. Millner. Reflections: Uncertainty and decision making in climate change economics. *Review of Environmental Economics and Policy*, 8(1):120–137, 2014a.

G. M. Heal and A. Millner. Agreeing to disagree on climate policy. *Proceedings of the National Academy of Sciences*, 2014b.

B. Hill. Confidence and decision. *Games and Economic Behavior*, 82(0):675–692, 2013.

C. Hope. The marginal impact of $CO_2$ from PAGE2002: An integrated assessment model incorporating the IPCC's five reasons for concern. *Integrated Assessment*, 6 (1):19–56, 2006.

C. Hope. Optimal carbon emissions and the social cost of carbon over time under uncertainty. *Integrated Assessment*, 8(1):107–122, 2008.

IPCC. *Working Group III Report: Mitigation of Climate Change, In Climate Change 2014, IPCC, Fifth Assessment Report (AR5).* Cambridge University Press, 2014.

IPCC. *Climate Change 2014 Synthesis Report: Longer Report.* Cambridge University Press, 2014.

E. Karni. Familiarity breeds completeness. *Economic Theory*, 56(1):109–124, 2014.

E. Karni, F. Maccheroni, and M. Marinacci. Chapter 17 - Ambiguity and Nonexpected Utility. In H. P. Young and S. Zamir, editors, *Handbook of Game Theory with Economic Applications*, volume 4 of *Handbook of Game Theory with Economic Applications*, pages 901–947. Elsevier, 2015.

P. Klibanoff, M. Marinacci, and S. Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, 2005.

C. Kolstad, K. Urama, J. Broome, A. Bruvoll, M. C. Olvera, D. Fullerton, C. Gollier, W. M. Hanemann, R. Hassan, M. R. K. F. Jotzo, L. Meyer, and L. Mundaca. Social and economic and ethical concepts and methods. In *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2014.

H. Kunreuther, G. Heal, M. Allen, O. Edenhofer, C. B. Field, and G. Yohe. Risk management and climate change. *Nature Clim. Change*, 3(5):447–450, May 2013.

D. Lemoine and H. C. McJeon. Trapped between two tails: trading off scientific uncertainties via climate targets. *Environmental Research Letters*, 8(3):034019, 2013.

R. J. Lempert. Climate targets: Values and uncertainty. *Nature Clim. Change*, 5(10): 914–915, Oct. 2015.

R. J. Lempert, D. G. Groves, S. W. Popper, and S. C. Bankes. A general, analytic method for generating robust strategies and narrative scenarios. *Management Science*, 52(4):514–528, 2006.

T. S. Lontzek, Y. Cai, K. L. Judd, and T. M. Lenton. Stochastic integrated assessment of climate tipping points indicates the need for strict climate policy. *Nature Clim. Change*, 5(5):441–444, May 2015.

F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.

M. J. Machina. “Expected utility” analysis without the independence axiom. *Econometrica*, 50(2):277–323, March 1982.

C. McGlade and P. Ekins. The geographical distribution of fossil fuels unused when limiting global warming to 2°C. *Nature*, 517(7533):187–190, Jan. 2015.

G. E. Metcalf and J. Stock. Integrated assessment models and the social cost of carbon: A review and assessment. Available at `http://works.bepress.com/gilbert_metcalf/114`, September 2015.

A. Millner. On welfare frameworks and catastrophic climate risks. *Journal of Environmental Economics and Management*, 65(2):310–325, 2013.

A. Millner and G. Heal. Resolving intertemporal conflicts: Economics vs politics. Working Paper 20705, National Bureau of Economic Research, November 2014.

A. Millner and T. K. McDermott. Model confirmation in climate economics. *Proceedings of the National Academy of Sciences*, 113(31):8675–8680, 2016.

A. Millner, S. Dietz, and G. Heal. Scientific ambiguity and climate policy. *Environmental and Resource Economics*, 55(1):21–46, 2013.

S. C. Newbold and A. L. Marten. The value of information for integrated assessment models of climate change. *Journal of Environmental Economics and Management*, 68 (1):111–123, 2014.

W. Nordhaus. Rolling the ‘DICE’: an optimal transition path for controlling greenhouse gases. *Resource and Energy Economics*, 15(1):27–50, March 1993.

W. Nordhaus. *A question of balance*. Yale University Press, 2008.

W. Nordhaus. Estimates of the social cost of carbon: Concepts and results from the DICE-2013R model and alternative approaches. *Journal of the Association of Environmental and Resource Economists*, 1(1/2):273–312, 2014.

W. Nordhaus and P. Sztorc. DICE 2013R: Introduction and User’s Manual. Available from `http://www.econ.yale.edu/~nordhaus/homepage/`, October 2013.

E. A. Ok, P. Ortoleva, and G. Riella. Incomplete preferences under uncertainty: Indecisiveness in beliefs versus tastes. *Econometrica*, 80(4):1791–1808, 2012.

A. Otto, B. J. Todd, N. Bowerman, D. J. Frame, and M. R. Allen. Climate system properties determining the social cost of carbon. *Environmental Research Letters*, 8 (2):024032 (6pp), 2013.

F. E. L. Otto, D. J. Frame, A. Otto, and M. R. Allen. Embracing uncertainty in climate change policy. *Nature Clim. Change*, 5(10):917–920, Oct. 2015.

R. S. Pindyck. Modelling the impact of warming in climate change economics. In *The Economics of Climate Change*, chapter 2, pages 47–71. University of Chicago Press, 2011a.

R. S. Pindyck. Fat tails, thin tails, and climate change policy. *Review of Environmental Economics and Policy*, 5(2):258–274, 2011b.

R. S. Pindyck. Uncertain outcomes and climate change policy. *Journal of Environmental Economics and Management*, 63(3):289–303, 2012.

R. S. Pindyck. The climate policy dilemma. *Review of Environmental Economics and Policy*, 7(2):219–237, 2013a.

R. S. Pindyck. Climate change policy: What do the models tell us? *Journal of Economic Literature*, 51(3):860–72, 2013b.

R. S. Pindyck. The use and misuse of models for climate policy. *Review of Environmental Economics and Policy*, 11(1):100–114, Jan. 2017.

S. Polasky, S. R. Carpenter, C. Folke, and B. Keeler. Decision-making under great uncertainty: environmental management in an era of global change. *Trends in Ecology & Evolution*, 26(8):398–404, 2011.

E. C. Prescott. Lawrence R. Klein lecture 1997: Needed: A theory of total factor productivity. *International Economic Review*, 39(3):525–551, 1998.

A. Rezai and F. Van Der Ploeg. Abandoning fossil fuel: How fast and how much. *The Manchester School*, 85:e16–e44, 2017.

K. Riahi, E. Kriegler, N. Johnson, C. Bertram, M. den Elzen, J. Eom, M. Schaeffer, J. Edmonds, M. Isaac, V. Krey, T. Longden, G. Luderer, A. Méjean, D. L. McCollum, S. Mima, H. Turton, D. P. van Vuuren, K. Wada, V. Bosetti, P. Capros, P. Criqui, M. Hamdi-Cherif, M. Kainuma, and O. Edenhofer. Locked into Copenhagen pledges? implications of short-term emission targets for the cost and feasibility of long-term climate goals. *Technological Forecasting and Social Change*, 90(Part A):8–23, 2015.

G. Riella. On the representation of incomplete preferences under uncertainty with indecisiveness in tastes and beliefs. *Economic Theory*, 58(3):571–600, 2015.

G. H. Roe and Y. Bauman. Climate sensitivity: should the climate tail wag the policy dog? *Climatic Change*, 117(4):647–662, 2013.

R. A. Rosen. IAMs and peer review. *Nature Climate Change*, 5(5):390–390, May 2015.

L. J. Savage. The theory of statistical decision. *Journal of the American Statistical Association*, 46(253):55–67, 1951.

L. J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

H. J. Schellnhuber, C. Reyer, B. Hare, K. Waha, I. M. Otto, O. Serdeczny, M. Schaeffer, C.-F. Schleussner, D. Reckien, R. Marcus, O. Kit, A. Eden, S. Adams, V. Aich, T. Albrecht, F. Baarsch, A. Boit, N. Canales Trujillo, M. Cartsburg, D. Coumou, M. Fader, H. Hoff, G. Jobbins, L. Jones, L. Krummenauer, F. Langerwisch, V. Le Masson, E. Ludi, M. Mengel, J. Möhring, B. Mosello, A. Norton, M. Perette, P. Pereznieto, A. Rammig, J. Reinhardt, A. Robinson, M. Rocha, B. Sakschewski, S. Schaphoff,

J. Schewe, J. Stagl, and K. Thonicke. Turn down the heat: confronting the new climate normal. Main report, volume 2, World Bank, 2014.

D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, May 1989.

T. Seidenfeld, M. J. Schervish, and J. B. Kadane. A representation of partially ordered preferences. *The Annals of Statistics*, 23(6):pp. 2168–2217, 1995.

N. Stern. Presidential address imperfections in the economics of public policy, imperfections in markets, and climate change. *Journal of the European Economic Association*, 8(2-3):253–288, 2010.

N. Stern. The structure of economic modeling of the potential impacts of climate change: Grafting gross underestimation of risk onto already narrow science models. *Journal of Economic Literature*, 51(3):838–59, 2013.

T. Sterner and U. M. Persson. An even sterner review: Introducing relative prices into the discounting debate. *Review of Environmental Economics and Policy*, 2(1):61–76, 2008.

J. Stoye. Axioms for minimax regret choice correspondences. *Journal of Economic Theory*, 146(6):2226–2251, 2011.

R. Tol. The economic effects of climate change. *Journal of Economic Perspectives*, 23 (2):29–51, 2009.

C. Traeger. Analytic integrated assessment and uncertainty. Mimeo., March 2015.

I. van den Bijgaart, R. Gerlagh, and M. Liski. A simple formula for the social cost of carbon. *Journal of Environmental Economics and Management*, 77:75–94, 2016.

C. P. Weaver, R. J. Lempert, C. Brown, J. A. Hall, D. Revell, and D. Sarewitz. Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks. *Wiley Interdisciplinary Reviews: Climate Change*, 4(1):39–60, 2013.

M. Weitzman. Prices vs. quantities. *The Review of Economic Studies*, 41(4):477–491, 1974.

M. Weitzman. A review of the Stern Review on the economics of climate change. *Journal of Economic Literature*, 45(3):703–724, 2007.

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

M. Weitzman. Some basic economics of extreme climate change. In J.-P. Touffut, editor, *Changing Climate, Changing Economy*, Cournot Centre for Economic Studies Series. Edward Elgar, 2009b.

M. Weitzman. What is the "damages function" for global warming – and what difference might it make? *Climate Change Economics*, 1:57–69, 2010.

M. Weitzman. Fat-tailed uncertainty in the economics of catastrophic climate change. *Review of Environmental Economics and Policy*, 5:275–292, 2011.

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

M. L. Weitzman. Gamma discounting. *The American Economic Review*, 91(1):260–271, 2001.

M. L. Weitzman. Additive damages, fat-tailed climate dynamics, and uncertain discounting. *Economics: The Open-Access, Open-Assessment E-Journal*, 3(2009-39): 1–22, 2009c.

M. L. Weitzman. Tail-hedge discounting and the social cost of carbon. *Journal of Economic Literature*, 51(3):873–82, 2013.

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

A. Xepapadeas. The cost of ambiguity and robustness in international pollution control. In R. W. Hahn and A. Ulph, editors, *Climate Change and Common Sense: Essays in Honour of Tom Schelling*, chapter 6. Oxford University Press, 2012.

# A Appendix: Proofs of Results in the Text

**Proposition 4.** *Make Assumptions 1-3. Suppose $f \in \mathcal{F}$, and there does not exist $g \in \mathcal{F}$ such that $\overline{g}_{u,\sigma} > \overline{f}_{u,\sigma}$ for all $(u, \sigma) \in \Psi$. Then there exists $(u_0, \sigma_0) \in \Psi$ such that $\overline{f}_{u_0,\sigma_0} \geq \overline{g}_{u_0,\sigma_0}$ for all $g' \in \mathcal{F}$.*

*Proof.* This result is immediate from generalisation of the "Abstract Pearce-Wald Lemma" recently provided by Battigalli et al. (2016, Lemma 1). □

**Proof of Proposition 1.** Suppose an act $f$ is justifiable, and so there exists $(u_0, \sigma_0) \in \Psi$ such that $\overline{f}_{u_0,\sigma_0} \geq \overline{g}_{u_0,\sigma_0}$ for all $g \in \mathcal{F}$. It is impossible that there exists $g$ as in (2). Conversely, suppose an act $f$ is not justifiable. Then, by Proposition 4, there does exist $g \in \mathcal{F}$ as in (2). □

**Proof of Proposition 3.** Suppose $f \in \mathcal{F}$ is not justifiable. By Proposition 1 there exists $g \in \mathcal{F}$ such that $\overline{g}_{u,\sigma} > \overline{f}_{u,\sigma}$ for all $(u, \sigma) \in \Psi$. Then, for any $H \in \mathcal{H}$, we know $H\left(\{\overline{f}_{u,\sigma}\}_{(u,\sigma)\in\Psi}\right) < H\left(\{\overline{g}_{u,\sigma}\}_{(u,\sigma)\in\Psi}\right)$, i.e. $f$ is not maximal with respect to $H$. So $f$ is not second-order justifiable.

Conversely, suppose $f \in \mathcal{F}$ is justifiable: there exists $u_0, \sigma_0 \in \Psi$, with $\overline{f}_{u_0,\sigma_0} \geq \overline{g}_{u_0,\sigma_0}$ for all $g \in \mathcal{F}$. Then $f$ is second-order justifiable via the SEU preference $H_{u_0,\sigma_0} \in \mathcal{H}$. □

**Proof of Lemma 1.** Write $\mathbf{x}(q, \phi) := f(q)(\phi)$ for the outcome from act $f(q)$ in $\{f(q) : q \in \mathcal{Q}\}$. Write $\widehat{\mathbf{x}}(\phi)$ for the baseline outcome (dependent on the state $\phi$) that would have resulted from act $f(q)$ if climate change were not damaging and if emission cuts to achieve $q$ were costless (which is thus *in*dependent of $q$). Write damages

$\mathbf{d}_t(q,\phi)$ and benefits $\mathbf{b}_t(q,\phi)$, in outcome terms, associated with $f(q)$ in state $\phi$: both are measured relative to $\widehat{\mathbf{x}}(q,\phi)$, so that in each period $\mathbf{x}(q,\phi)_t = f(q)(\phi)_t = \widehat{\mathbf{x}}_t(q,\phi) - \mathbf{d}_t(q,\phi) + \mathbf{b}_t(q,\phi)$. Then, taking the vectors across all time periods,

$$u(\mathbf{x}(q,\phi)) = u\big(\widehat{\mathbf{x}}(q,\phi)\big) - \big[u\big(\widehat{\mathbf{x}}(q,\phi)\big) - u\big(\widehat{\mathbf{x}}(q,\phi) - \mathbf{d}(q,\phi)\big)\big] \\ + \big[u\big(\mathbf{x}(q,\phi)\big) - u\big(\mathbf{x}(q,\phi) - \mathbf{b}(q,\phi)\big)\big]\,.$$

The first square-bracketed term on the right hand side is the difference in utility between the world as it would have been without a climate problem, and the damaged world. So this can be interpreted as the welfare damages from climate change, and written $D_u(q,\phi)$. The second term is the difference between the world as it is, and as it would have been if there had been no benefits from quantities of emissions; it may thus be interpreted as their private benefits, and may be written $B_u(q,\phi)$. □

**Proof of Proposition 2.** Under the assumptions of the proposition, $\bar{f}'_{u,\sigma}(\hat{q}_0) = 0$. Now, the second order Taylor series with integral remainders is:

$$\bar{f}_{u,\sigma}(q) = \bar{f}_{u,\sigma}(\hat{q}_0) + \bar{f}'_{u,\sigma}(\hat{q}_0)(q-\hat{q}_0) + \frac{1}{2}\bar{f}''_{u,\sigma}(\hat{q}_0)(q-\hat{q}_0)^2 + \frac{1}{6}\int_{\hat{q}_0}^{q} \bar{f}'''_{u,\sigma}(t)(q-t)^3 dt.$$

Following $(u,\sigma)$, we would choose $\hat{q}$ such that $\bar{f}'_{u,\sigma}(\hat{q}) = 0$. For such $\hat{q}$, then, by differentiating the expression above, we see:

$$\bar{f}'_{u,\sigma}(\hat{q}_0) + \bar{f}''_{u,\sigma}(\hat{q})(\hat{q}-\hat{q}_0) + \frac{1}{2}\int_{\hat{q}_0}^{\hat{q}} \bar{f}'''_{u,\sigma}(t)(\hat{q}-t)^2 dt = 0 \tag{2}$$

The disagreement regret $R^{\mathcal{Q}}(u,\sigma,u_0,\sigma_0)$ due to setting policy $\hat{q}_0$ instead is

$$\begin{aligned} R^{\mathcal{Q}}(u,\sigma,u_0,\sigma_0) &= \bar{f}_{u,\sigma}(\hat{q}) - \bar{f}_{u,\sigma}(\hat{q}_0) \\ &= \bar{f}'_{u,\sigma}(\hat{q}_0)(\hat{q}-\hat{q}_0) + \frac{1}{2}\bar{f}''_{u,\sigma}(\hat{q}_0)(\hat{q}-\hat{q}_0)^2 + \frac{1}{6}\int_{\hat{q}_0}^{\hat{q}} \bar{f}'''_{u,\sigma}(t)(\hat{q}-t)^3 dt. \end{aligned}$$

Use (2) to substitute for $\bar{f}''_{u,\sigma}(\hat{q})$, to see:

$$\begin{aligned} &R^{\mathcal{Q}}(u,\sigma,u_0,\sigma_0) \\ &\quad= \frac{1}{2}\bar{f}'_{u,\sigma}(\hat{q}_0)(\hat{q}-\hat{q}_0) + \int_{\hat{q}_0}^{\hat{q}} \bar{f}'''_{u,\sigma}(t)(\hat{q}-t)^2\left(\frac{1}{6}(\hat{q}-t) - \frac{1}{4}(\hat{q}-\hat{q}_0)\right) dt \\ &\quad= \frac{1}{2}\left(\bar{f}'_{u,\sigma}(\hat{q}_0) - \bar{f}'_{u_0,\sigma_0}(\hat{q}_0)\right)(\hat{q}-\hat{q}_0) - \int_{\hat{q}_0}^{\hat{q}} \bar{f}'''_{u,\sigma}(t)(\hat{q}-t)^2\left(\frac{1}{12}\hat{q} + \frac{1}{6}t - \frac{1}{4}\hat{q}_0\right) dt \end{aligned}$$

where we use the fact that $\bar{f}'_{u_0,\sigma_0}(\hat{q}_0) = 0$. This provides the exact result.

The result stated in the text applies when this quadratic approximation is sufficiently good that we can neglect the remainder term. □