On welfare frameworks and catastrophic climate risks

By Antony Millner[*]

Draft: March 22, 2012

Recent theoretical work in the economics of climate change has suggested that climate policy is highly sensitive to 'fat-tailed' risks of catastrophic outcomes (Weitzman, 2009b). Such risks are suggested to be an inevitable consequence of scientific uncertainty about the effects of increased greenhouse gas concentrations on climate. Criticisms of this controversial result fall into three categories: The first suggests it may be irrelevant to cost benefit analysis of climate policy, the second challenges the fat-tails assumption, and the third questions the behaviour of the utility function assumed in the result. This paper analyses these critiques, and suggests that those in the first two categories have formal validity, but that they apply only to the restricted setup of the original result, which may be extended to address their concerns. They are thus ultimately unconvincing. Critiques in the third category are shown to be robust, however they open up new ethical and empirical challenges for climate economics that have thus far been neglected – how should we 'value' catastrophes as a society? I demonstrate that applying results from social choice to this problem can lead to counterintuitive results, in which society values catastrophes as infinitely bad, even though each individual's utility function is bounded. Finally, I suggest that the welfare functions traditionally used in climate economics are ill-equipped to deal with climate catastrophes in which population size changes. Drawing on recent work in population ethics I propose an alternative welfare framework with normatively desirable properties, which has the effect of dampening the contribution of catastrophes to welfare.
JEL: D61, D63, D81, Q54
Keywords: Climate Change, Catastrophes, Welfare, Uncertainty, Social Choice, Population

Since the publication of the Stern Review on the Economics of Climate Change (Stern, 2007), economists have been heatedly debating the appropriate welfare

framework for evaluating climate change policy (Weitzman, 2007; Nordhaus, 2007; Dasgupta, 2008; Sterner and Persson, 2008; Dietz and Stern, 2008; Heal, 2009). Choices of key parameters in the standard discounted utilitarian model– the pure rate of time preference (PRTP), and the elasticity of the marginal utility of consumption – were identified as key determinants of the radical difference between the policy prescriptions offered by Stern, and those of his most prominent North American counterpart (Nordhaus, 2008). In the thick of this debate, Weitzman introduced his 'dismal theorem' in an influential and much debated paper (Weitzman, 2009*b*). In informal terms, the dismal theorem states that the evaluation of climate change policy is highly sensitive to catastrophic outcomes, even if they occur with vanishingly small, but 'fat-tailed'[1], probability. The dismal theorem suggests that such fat-tailed risks are an inescapable consequence of Bayesian statistics, and our 'structural' uncertainty about the response of the climate system to increases in greenhouse gas concentrations. It argues that such risks can swamp all other effects on welfare, including those that arise from conventional discounting, and make willingness to pay to avoid them formally infinite. Weitzman argues that this has radical implications for the way we evaluate policies that aim to mitigate climate change, and suggests that existing analyses based on integrated assessment models do not take sufficient account of such catastrophic risks, and thus recommend mitigation policies that are not ambitious enough.

This paper provides a critical analysis of Weitzman's dismal theorem, and the by now sizeable literature that comments on its content and interpretation. While the mathematical result Weitzman obtains (described in detail in section 2 below) is beyond reproach, the assumptions upon which it is based are controversial, and have been heavily criticized. In section 3, I divide these critiques into three categories – criticisms of the relevance of the dismal theorem for cost-benefit analysis, criticisms of fat-tailed probability distributions, and criticisms of the behaviour of the utility function it assumes. Critiques of the dismal theorem in the first two categories show that the framework Weitzman employed is not robust to plausible changes in the assumptions of the modeling. However, in both these cases, I demonstrate that the intellectual content of the dismal theorem can be reinstated by appropriate extensions and modifications of Weitzman's results, and by assessing the plausibility of the assumptions made in the criticisms themselves. Thus, I find that criticisms based on these two strains of inquiry are ultimately unconvincing. Criticisms of the behaviour of the utility function are however robust – the dismal theorem may be 'resolved' by accepting that the traditional tool of constant relative risk aversion (CRRA) utility functions is not up to the task of evaluating fat-tailed risks. I demonstrate that a generalization of the CRRA-like utility functions assumed by the theorem – Harmonic Absolute Risk Aversion utility functions – provides finite welfare measures even when risks are fat-tailed, and moreover makes policy evaluation relatively insensitive to the

---

[1]We will refer to any probability distribution which decays polynomially in the tails (upper or lower, as appropriate) as 'fat-tailed'.

tails of the consumption distribution for a plausible range of parameters. While such utility functions provide a technical resolution to the dismal theorem, they highlight the fact that the theorem's message is not so much a technical point about specific utility functions, but an ethical and empirical point about what the appropriate utility function for analysis of catastrophic risks should be. The key issue is: How should we 'value' catastrophic outcomes as a society?

If we attempt to address this question via Harsanyi-style aggregation over the preferences of diverse individuals, some counterintuitive results are possible. While individuals are unlikely to be willing to pay large sums to avoid very low probability catastrophic risks, implying that CRRA utility functions are indeed an inappropriate representation of their preferences, social willingness to pay need not be similarly constrained. I demonstrate that even if every individual's utility function is bounded, aggregate social welfare may still be unbounded below at positive consumption levels. This suggests that in some cases social evaluation of catastrophic risks is highly sensitive to the preferences of the most risk-averse members of society, even if they make up an exponentially vanishing fraction of the population. Thus it is not a simple matter to reason from willingness to pay to avoid catastrophes as elicited from a sample of the population to social preferences over catastrophic risks.

In section 4 I suggest that the ethical issues the dismal theorem raises run even deeper than the choice of an appropriate social utility function. Any welfare analysis which admits catastrophic impacts on consumption must also of necessity deal with population change. The role of population change is not made explicit in the analyses of Weitzman and his critics – they implicitly assume an average utilitarian population principle. Yet average utilitarianism has been heavily criticized by philosophers and social choice theorists. I discuss an attractive alternative to average utilitarianism – critical level utilitarianism – which provides a more defensible approach to population ethics, and dampens the effect of catastrophic risks on welfare. Section 5 concludes by suggesting that the dismal theorem shows that the welfare framework that has thus far been standard in integrated assessment modeling is an inadequate normative basis for climate policy analyses which account for possible catastrophes. Nevertheless, alternatives with desirable properties are available, and should be implemented in the next round of economic analyses of mitigation policies. These however require us to face up to ethical and empirical issues that have thus far been neglected.

I. Weitzman's Dismal Theorem

Weitzman (2009b) phrases his analysis in terms of a two period model. Current consumption is normalized to 1. Future consumption is given by the random variable c. Consider the following welfare function:

$$(1) \qquad\qquad W = U(1) + \beta \mathsf{E} U(c)$$

where U is the social utility function (which we will assume is concave and twice differentiable), E is the expectation operator, and $\beta \in [0,1]$ is a discount factor on future utility. Suppose now that we are able to transfer an infinitesimal amount δ of consumption into the future, *with certainty*. What amount ϵ of current consumption should we be willing to sacrifice in order to make this transfer? In other words, what is the relationship between the values of ϵ and δ that will leave total welfare unchanged? Formally, this question is answered by setting

(2) $$U(1-\epsilon) + \beta\mathsf{E}U(c+\delta) = U(1) + \beta\mathsf{E}U(c).$$

In the limit as $\epsilon, \delta \to 0$, one finds

(3) $$\epsilon = \left(\beta \frac{\mathsf{E}U'(c)}{U'(1)} \right) \delta.$$

This expression tells us the marginal willingness to pay for a transfer of a certain infinitesimal unit δ of consumption into the future. The stochastic discount factor M is defined as

(4) $$M := \beta \frac{\mathsf{E}U'(c)}{U'(1)}.$$

Weitzman proves the following theorem:

THEOREM 1: *Suppose that*

- *The coefficient of relative risk aversion $\eta(c) = -c\frac{U''(c)}{U'(c)}$ is strictly greater than zero as $c \to 0$.*

- *The consumption growth rate $y := \log c$ is distributed according to a probability density function $h(y|s) = \frac{1}{s}f(\frac{y-\mu}{s})$, where f is a normalizable function, μ is a known location parameter, and s is an uncertain scale parameter. Weitzman interprets s as a 'generalized climate sensitivity' parameter – this will be discussed below.*

- *The prior on s is of the form $p(s) \propto s^{-k}$, where $k > 0$, and you are given a finite number n of independent observations $\mathbf{y}_n$ of the random variable y.*

Then we can conclude that,

1) *The posterior distribution for y, $q(y|\mathbf{y}_n) \propto \int_0^\infty h(y|s) \prod_n h(y_n|s)p(s)ds$ scales like $|y|^{-(n+k)}$ as $y \to -\infty$.*

2) *The stochastic discount factor $M \to \infty$.*

The first part of the conclusion of the theorem is a result due to Schwarz (1999), while the second part had been pointed out by Geweke (2001) in the special case

of some fat tailed parametric distributions (e.g. t-distribution, inverted-gamma distribution) for the posterior $q(y|\mathbf{y}_n)$. Weitzman's contribution was to combine these two results, and give an interpretation of their consequences for climate change economics. He suggests that the correct way to interpret this result is that when there is what he terms 'structural uncertainty' (i.e. uncertainty about the parameters of the distribution for future consumption), policy choice is heavily dependent on the catastrophic tails of the distribution of future consumption, which in general are 'fat', i.e. decay to zero only polynomially in consumption. In fact, he suggests, the effect of these catastrophic fat tails swamps all other effects on the discount factor, including that of the PRTP (given by $-\ln\beta$ in this model) which had hitherto been the chief source of contention amongst economists.

II. Criticisms of the Dismal Theorem

A. Marginal vs. total willingness to pay

Weitzman chose to frame his results in terms of the stochastic discount factor, which as demonstrated in equation (3), determines willingness to pay for an infinitesimal certain transfer of consumption to the future. The fact that his result is framed in terms of marginal, and not total, willingness to pay, has been criticized by several authors, most notably Horowitz and Lange (2009) and Karp (2009). They point out that the fact that marginal willingness to pay may be infinite is of very limited relevance for cost-benefit analysis – all it implies is that one would certainly want to transfer a non-infinitesimal amount of consumption to the future. Nordhaus (2011) also emphasizes the absence of a policy variable from Weitzman's analysis, and suggests that the dismal theorem has no special consequences for policy choice. I reprise and generalize these authors' arguments below.

Suppose that we wish to determine the optimal transfer a from current consumption into future consumption. Then we need to pick a to maximize

$$(5) \qquad U(1-a) + \beta\mathsf{E}U(c+a).$$

The first order condition is

$$(6) \qquad U'(1-a) = \beta\mathsf{E}U'(c+a).$$

Now ask yourself whether this equation has a solution at $a = 0$, which corresponds to the case of an infinitesimally small transfer. The answer is clearly no, if the conditions of Weitzman's theorem on U and the probability distribution for c are satisfied. However, this in itself does not mean that there is *no* solution other than the vacuous $a = 1$. In fact, when the conditions of Weitzman's theorem are satisfied, a solution $a^* \in (0,1)$ to the first order condition (6) is guaranteed to exist under very mild continuity conditions:

PROPOSITION 1:   *Assume that $U'(0) = \infty$, $U'(x)$ is finite for all $x > 0$, and that $U'(x)$ is a continuous function. In addition, assume $\mathsf{E}U'(c) = \infty$. Then the first order condition (6) has a unique solution $a^* \in (0,1)$.*

PROOF:

Let $F(a) := \beta\mathsf{E}U'(c+a) - U'(1-a)$. By assumption, $F(0) = \infty, F(1) = -\infty$. The continuity of U' implies that $F(a)$ is continuous on $[0,1]$. Therefore the intermediate value theorem tells us that there exists $a^* \in (0,1)$ such that $F(a^*) = 0$. The fact that $F(a)$ is decreasing (since U is concave) means that a^* is unique.

For Horowitz and Lange (2009) and Karp (2009), this fact renders the dismal theorem uninteresting. Karp (2009) in particular shows that the optimal transfer a^* is about as sensitive to the PRTP, as it is to the probability of catastrophe in a simple binary lottery. Using this result, he suggests that even a moderate interpretation of the dismal theorem which suggests that catastrophic risks can swamp the effect of the PRTP, is flawed. I am however hesitant to dismiss the dismal theorem on these grounds. To see why, notice that this argument is predicated on the assumption that it is possible to transfer a finite amount of consumption a into the future *with certainty*. This seems a very strong assumption, especially so in the case of climate change, where catastrophic outcomes are at least conceivable for a wide range of abatement policies, owing to the long residence time of atmospheric CO_2 (Solomon et al., 2009).

Consider a generalization of the decision problem in (5), in which this time it is no longer certain that a sacrifice of an amount a of current consumption gives rise to a certain payoff in the future. We will model the effect of a sacrifice of size a on the distribution of future consumption through the distribution function $q(c;a)$. The decision problem is now to choose a to maximize

$$(7) \qquad\qquad U(1-a) + \beta \int_0^\infty U(c)q(c;a)dc.$$

The first order condition for this problem is:

$$(8) \qquad\qquad U'(1-a) = \beta \int_0^\infty U(c)\frac{\partial q(c;a)}{\partial a}dc.$$

Define

$$(9) \qquad\qquad I(a) := \int_0^\infty U(c)\frac{\partial q(c;a)}{\partial a}dc.$$

We identify three cases of interest:

1) $I(a)$ is finite for all values of a

2) $I(a)$ diverges for $a \in [0,d]$, but is finite for $a \in (d,1]$, with $d \in (0,1)$.

3) $I(a)$ diverges for all $a \in [0, 1]$.

In case 1, both marginal and total willingness to pay are finite, and an optimal value of a exists. In the second case, marginal willingness to pay is infinite, but total willingness to pay is finite and greater than or equal to d – this is analogous to the case examined by Karp (2009). In this case the optimal value of a must be greater than d. In the third case, both marginal *and* total willingness to pay are infinite, and there is no $a < 1$ which satisfies the first order condition.

To make these three cases concrete, and tie them to some physical and economic parameters, consider the following set of assumptions: Suppose that future consumption depends on the increase in global average temperature above preindustrial levels (T) through a multiplicative damage function, i.e.:

$$(10) \qquad c \propto \frac{1}{1 + D(T)}$$

Assume, as is standard in the integrated assessment modeling literature (e.g. Nordhaus, 2008; Hope, 2006; Weitzman, 2010) that $D(T)$ grows at most polynomially in temperature T, with leading order exponent γ, so that for large T,

$$(11) \qquad c \sim T^{-\gamma}.$$

Now pick a CRRA utility function,

$$(12) \qquad U(c) = \frac{c^{1-\eta}}{1 - \eta}$$

and assume that the temperature change T is distributed like

$$(13) \qquad T \sim T^{-g(a)}, \text{ for large } T,$$

where $g(a)$ is a nondecreasing function of a, and we require $g(a) > 1$ for all a in order for the temperature distribution to be normalizable. In this model, a may be interpreted as the level of abatement spending. More abatement is assumed to flatten out the tail of the distribution of future temperatures[2]. With these assumptions, the contribution to the integral $I(a)$ from temperatures larger than

---

[2]It is of course possible that the tail becomes 'thin' at some finite value of a, in which case we may say that $g(a) \to \infty$.

$\overline{T} \gg 1$ is:

$$I(a; \overline{T}) \propto \int_{\overline{T}}^{\infty} \frac{(T^{-\gamma})^{1-\eta}}{1-\eta} \frac{\partial}{\partial a} \left[T^{-g(a)} \right] dT$$

$$\propto \int_{\overline{T}}^{\infty} (T^{-\gamma})^{1-\eta} \left[-(T^{-g(a)} \ln T) g'(a) \right] dT$$

$$\propto \int_{\overline{T}}^{\infty} T^{\gamma(\eta-1)-g(a)} \ln T dT.$$

Now for large T, we know that $\ln T$ grows more slowly than T^{θ} for any $\theta > 0$, so we can conclude that for large $\overline{T}$,

$$(14) \qquad\qquad I(a; \overline{T}) \propto \int_{\overline{T}}^{\infty} T^{\gamma(\eta-1)-g(a)} dT.$$

We now make use of the following simple fact of integration:

$$(15) \qquad\qquad \int_{\overline{T}}^{\infty} T^{-p} dT \text{ converges if and only if } p > 1.$$

Using this, the integral $I(a)$ converges if and only if,

$$(16) \qquad\qquad g(a) > 1 + \gamma(\eta - 1).$$

Our three cases correspond to three types of behaviour for the function $g(a)$:

1) $I(a)$ converges for all $a \rightarrow g(a) > 1 + \gamma(\eta - 1)$ for all $a \in [0, 1]$.

2) $I(a)$ converges only for $a \in [d, 1] \rightarrow g(a) > 1 + \gamma(\eta - 1)$ for $a \in [d, 1]$.

3) $I(a)$ diverges for all $a \in [0, 1] \rightarrow g(a) < 1 + \gamma(\eta - 1)$ for all $a \in [0, 1]$.

Notice how the critical value of $g(a)$ at which the convergence properties of $I(a)$ changes depends on the parameter γ, which measures the steepness of the climate damages function, and η which measures the curvature of the utility function. Increases in either of these parameters decrease the range of values of a for which the integral converges. In general, increases in either of these parameters increase the importance of the tails of the distribution for T for welfare calculations.

To get a sense of the implications of (16), pick $\gamma = 2$, the value used in the DICE integrated assessment model (Nordhaus, 2008). Next, consider the theoretical distribution for climate sensitivity[3] derived in Roe and Baker (2007). This distribution falls off as T^{-2} for large T. Let a_2 be the value of abatement that

---

[3]Climate sensitivity is the amount of surface warming expected from a doubling of CO_2 concentrations, in equilibrium.

stabilizes CO_2 concentrations at twice their preindustrial level. Then, as we approach equilibrium, we have $g(a_2) \approx 2$. The inequality (16) therefore implies that $I(a_2)$ converges only if $\eta < 3/2$. Thus, under these assumptions, we know for certain that if $\eta > 3/2$, the optimal level of abatement spending must be larger than a_2, i.e. it is optimal to keep CO_2 concentrations below twice their preindustrial level, given the assumptions of this calculation. Moreover, if $g(a) = 2$ for all abatement levels[4], $\eta > 3/2$ implies that welfare diverges for all values of a.

It seems to me that the correct value for γ, and in particular the behaviour of the function $g(a)$, are not very well tied down in the existing literature. I would expect a $\gamma > 2$ – the damage function used by Nordhaus (2008) has been argued to be implausibly flat (Ackerman, Stanton and Bueno, 2010), with $5°C$ of warming giving rise to only a 6.5% drop in economic output[5]. Weitzman (2010), in contrast, estimates $\gamma \approx 6.7$. This value of γ makes the dismal theorem even more difficult to get rid of than in the example above, requiring $\eta < 1.15$ for convergence. As for $g(a)$, I do not think we can rule out the possibility of it being bounded above by $1 + \gamma(\eta - 1)$, in which case the dismal theorem holds for all values of a, and not just for marginal willingness to pay[6].

B. Where do fat tailed probability distributions come from?

The analysis above shows that we cannot plausibly rule out a divergence in social welfare at all abatement levels when consumption risks are fat tailed, and the utility function has the CRRA form. Yet where do such fat-tailed risks come from? Two arguments have been proposed – an abstract Bayesian argument, and an argument based on stylized models of the temperature response to increases in greenhouse gas concentrations, which centers around the concept of climate sensitivity.

THE BAYESIAN ARGUMENT. — The analysis in Weitzman (2009b) proceeds in the Bayesian paradigm, in which beliefs about the uncertain value of future consumption are based on a set of observations (he uses this as a loose proxy for scientific data or discoveries), a likelihood function that determines how observations are treated as evidence, and a prior distribution over the parameters of the likelihood function. This Bayesian framework is used to represent what Weitzman

---

[4]If we believe the model of Roe and Baker (2007), this seems to be an inevitable conclusion, unless radiative forcing is reduced to exactly zero at some abatement level, a virtual impossibility on any reasonable time horizon due to the slow decay of atmospheric CO_2. This is so since the quadratic tail in the temperature distribution arises only from uncertainty in feedback processes, and the central limit theorem – more on this later.

[5]For comparison, earth's average temperature was approximately $5°C$ less than today at the time of the last glacial maximum, when two thirds of the planet was covered in ice. It is not unreasonable to expect changes of similar magnitude for a warming of $5°C$.

[6]Nordhaus (2011) claims that *'when we introduce policies, the analysis underlying the dismal theorem no longer applies directly'*. The analysis in this section demonstrates that, under the standard assumptions, Weitzman's results are easily extended to policy choice as well, and that one needs only moderate risk aversion for dismal theorem like results to apply.

calls 'structural uncertainty' in our knowledge. What he means by this is that the parameters of the distribution from which consumption values are drawn are themselves uncertain. Thus, in arriving at posterior beliefs, we need to account for the evidence our observations provide, given particular values of parameters of the future consumption distribution, but then average over all possible values of these parameters with an appropriate prior. For reasons to be discussed below, Weitzman thinks the appropriate prior for the scale parameter of the consumption distribution is itself fat tailed.

Formally, by Bayes' theorem, the posterior $q(y|\mathbf{y}_n)$ for the growth in consumption y is related to the likelihood function $h(y|s)$ and the prior $p(s)$ for the scale parameter via:

$$(17) \qquad q(y|\mathbf{y}_n) \propto \int_0^\infty h(y|s) \left[\prod_n h(y_n|s) \right] p(s)ds.$$

Schwarz (1999) showed that, when $p(s) \propto s^{-k}$ (the Jeffreys prior), then $q(y|\mathbf{y}_n)$ is fat tailed, i.e. decays only polynomially, as $y \to -\infty$.

The crucial ingredient that leads to this result is the choice of prior, $p(s)$. It is the fact that this prior is itself fat-tailed that ultimately leads to a fat-tailed posterior as well. So, where does this prior come from? The prior that Weitzman deploys is an example of an 'uninformative' prior. That is, it is a prior that is supposed to represent a complete absence of prior knowledge. The argument that an uninformative prior for scale parameters in a distribution is best chosen to be of the form $p(s) \propto s^{-k}$ is due to Jeffreys (1946), and was elaborated on by Jaynes (1965, 1968). The treatment in Jaynes (1965) is particularly compelling. His argument is that, in a sense to be made precise below, the choice of prior over s is completely determined by its being a scale parameter. He begins by assuming that the observations $\mathbf{y}_n$ are samples from a two parameter probability density $h(y; \mu, s)$. If one is truly uninformed about the values of the parameters μ, s, he argues, this set of information should be indistinguishable from a transformed system, in which

$$(18) \qquad \mu' = \mu + b$$
$$(19) \qquad s' = \alpha s$$
$$(20) \qquad y' - \mu' = \alpha(y - \mu).$$

This is so, since the simple act of transforming the system (i.e. blowing up or shrinking some of its variables by a factor, and changing their points of origin) should not reveal anything new to us. Thus, he requires our knowledge to be *invariant* under the transformation group defined by (18–20). He shows that, in order for invariance to hold, the probability density must take the form $h(y; \mu, s) =$

$\frac{1}{s} f\left(\frac{y-\mu}{s}\right)$ for some function f, and the prior on s must be $p(s) = 1/s$[7]. In this framework, the transformation law (19) for s *defines* it as a scale parameter, i.e. it is a parameter that is only meaningful up to scale transformations. Similarly, the transformation law (18) for μ defines it as a location parameter. It is the specification of a transformation group for the system that pins down the prior, and gives s its interpretation.

Supposing one accepts this Bayesian approach, and Jaynes' argument that lack of knowledge should be represented by system invariance with respect to a specified transformation group, the only escape from a fat tailed prior for s can be either to argue that an uninformative prior is somehow inappropriate for the application in question, or that we have focussed our attentions on the wrong transformation group. The first option is the one preferred by Geweke (2001). I am wary of it in the case of the dismal theorem. In this case, the scale parameter s is argued by Weitzman to be a 'generalized climate sensitivity' parameter – I will discuss this in depth in the next section. For the moment, I simply point out that it is logically impossible to define an informative prior for climate sensitivity that is independent of knowledge of climate observations. Any informative prior, for example based on expert opinion (Frame et al., 2005), will of necessity be informed by such knowledge. This should rightfully be represented in our choice of likelihood function $h(\mathbf{y}_n|s)$ (i.e. a choice of climate model and noise distribution over temperature observations), and *not* in the prior. If we truly wish the prior to be independent of the observations (as it must be if we are to use the observations to infer a posterior – we can't use them twice), it must represent *a priori* (in the Kantian sense), rather than *a postiori* knowledge. Hence the appeal of abstract mathematical techniques for specifying an uninformative prior, which are divorced from any detailed knowledge of the data.

I believe that the second option – that we may have chosen the wrong transformation group – carries more weight. In standard statistical applications the choice of a transformation group may be pinned down by thinking about the symmetries of the underlying system that generates our data $\mathbf{y}_n$. For example, if the data are measurements of the locations of a set of events, then one might reasonably expect our information about the system to remain the same if we make measurements in meters or feet (scale transformations), and if we make them in Paris or London (location transformations), if we are truly uninformed *a priori*. What is much less clear is what the appropriate transformation group should be for the growth in consumption, the variable in question in the dismal theorem. More specifically, it is unclear why one would expect Weitzman's 'generalized climate sensitivity' to act as a scale parameter for the distribution of consumption growth, or indeed whether any such scale parameter exists. I can think of no good argument that

---

[7]Weitzman's more general prior $p(s) = s^{-k}$ does *not* follow from this transformation group. Although Weitzman's prior is itself scale invariant, the full system $(\mathbf{y}_n, \mu, s)$ is no longer invariant under the transformation group (18–20) for this choice. In fact, Jaynes (1965) specifically argues against this more general prior. This point has little consequence for the dismal theorem however, as the improper prior $p(s) = 1/s$ is the *most* fat tailed of all the priors in the class Weitzman considers!

pins down the transformation group for consumption growth, other than epistemically arbitrary considerations of simplicity and familiarity. At a broader level, the transformation group method is not universally accepted as a satisfactory solution to the problem of choosing uninformative priors. A common alternative is to use the principle of maximum entropy (Jaynes, 1968). Remarkably, this principle recommends an exponential distribution as the prior for a continuous positive parameter such as Weitzman's generalized climate sensitivity. With such a prior, the posterior (17) will be thin-tailed. Thus, even taking the Bayesian paradigm as read, Weitzman's framing of the problem is controversial, and relies on implicit assumptions which are not justified.

As a final remark on Weitzman's chosen framework for representing a state of ignorance, it is vital to recognize that the Bayesian approach, and the expected utility welfare framework it feeds into, is by no means the only game in town in situations of informational paucity. It has been widely argued (Ellsberg, 1961; Gilboa, 2009; Gilboa, Postlewaite and Schmeidler, 2009) that the normative application of expected utility theory is only appropriate in situations of risk (where probabilities are known), rather than uncertainty (where probabilities are unknown), a distinction going back at least as far as Knight (1921). Strong Bayesians contest this, saying that Savage's subjective expected utility axioms define rationality in the face of uncertainty. If one holds this position, one is forced to describe beliefs with a subjective probability distribution, even when there is no information upon which to base such a choice. Hence the need to go through the mathematical gymnastics associated with choosing an uninformative prior. If instead one is willing to accept that the domain of application of expected utility theory is limited, then one must search for an alternative evaluation criterion that is appropriate in situations of ambiguity or ignorance. There are several such criteria available (Savage, 1954; Arrow and Hurwicz, 1977; Gilboa and Schmeidler, 1989; Klibanoff, Marinacci and Mukerji, 2005), which go variously under the names of ambiguity averse, or ignorant, decision rules. Weitzman (2009b) argues that one does not need to account for ambiguity aversion in order to represent aversion to unknown probabilities, since 'structural uncertainty' leads to fat-tailed risks via the Bayesian non-informative prior discussed above, and hence, via the dismal theorem, to policies that exhibit a kind of precautionary principle. As discussed above, this is a highly model-dependent interpretation, relying on a very specific Bayesian setup of questionable relevance to the problem at hand, and more importantly, as we will see below, heavily sensitive to the choice of utility function. Moreover, it erroneously presupposes that ambiguity averse decision rules must recommend highly precautionary mitigation policies. In Appendix A, I demonstrate that applying the representation of ambiguity averse preferences obtained in Klibanoff, Marinacci and Mukerji (2005) to the setup of the dismal theorem can in fact render the stochastic discount factor (4) finite, even when consumption distributions are fat-tailed. I also derive a general condition on the expected utilities which enter the stochastic discount factor calculation

which must be satisfied for this to be possible. These results show that ambiguity aversion can lead to policies that are *less* precautionary than fat-tailed expected utility theory with a CRRA utility function, but that in general, the direction of the effect depends on the empirical details. The more detailed analysis of climate policy under ambiguity in Millner, Dietz and Heal (2010), finds that ambiguity aversion has a substantial positive effect on the value of abatement policies when the climate damages function is steep at high temperatures. It is obviously important to ascribe this effect to the correct preferences – aversion to fat-tailed, but known, risks is fundamentally different to aversion to unknown probabilities.

THE CLIMATE SENSITIVITY ARGUMENT. — I will now set the abstract Bayesian approach aside, and ask what climate science tells us about the relevance of the dismal theorem. As the inequality (16) demonstrates, if the probability distribution over future warming at any finite time in the future is fat tailed enough (no matter how stringent our abatement policy), then a version of the dismal theorem applies. This holds for any such fat-tailed distribution, whether it is justified on abstract Bayesian grounds, or derived from scientific and impacts models[8].

A commonly used summary statistic for the effect of increases in CO_2 concentrations on temperature is the climate sensitivity. The climate sensitivity (S) is defined as the increase in global mean surface temperatures that results from a doubling of CO_2 concentrations, *in equilibrium*. Weitzman (2009b) motivates his argument for fat tails being relevant to climate policy by appealing to scientific work on climate sensitivity estimation, which suggests that the distribution for S may itself be fat tailed. In order to investigate this argument, I will work with a conceptual model of climate sensitivity proposed by Roe and Baker (2007).

Much of the uncertainty about the long term effects of increases in CO_2 concentrations derives from uncertainty about *feedbacks* in the climate system. An easily visualized example of a feedback is the following: Imagine that CO_2 concentrations are increased. This leads to a rise in temperatures due to the increase in radiative forcing that results from more out-going long-wave radiation being trapped by CO_2 molecules. Now an increase in temperature may melt ice and snow. Since ice is a highly reflective surface, whereas water is not, if some ice is lost when temperatures rise, the amount of incoming radiation that is reflected back into space decreases. This gives rise to even more warming than would have occurred purely due to the radiative forcing effect of CO_2 molecules. That is, there is a positive feedback in the system – increases in temperature give rise to changes in the physical characteristics of the planet, which increases the amount of warming even further. There are many feedback processes in the climate system, some positive, and some negative. The net effect of all these feedbacks is widely believed to be positive.

Consider a very simple model of the climate sensitivity's dependence on the

---

[8]In reality, these two approaches are of course not independent. See e.g. Frame et al. (2005).

feedbacks in the climate system. Formally, define

$$S = \frac{\Delta T_0}{1 - f}. \tag{21}$$

where ΔT_0 is the 'grey-body' climate sensitivity, and f measures the strength of feedbacks. The grey-body sensitivity is the amount of warming that would occur if there were no feedbacks in the climate system, and is easily calculated based on very simple energy balance considerations[9]. Values quoted in the literature for ΔT_0 are in the range 1.1–1.3°C (Hansen et al., 1984; Schwartz, 2010). Roe and Baker (2007) observe that what we actually measure from the instrumental record is the size of the feedbacks and *not* the climate sensitivity itself. Assuming that the total feedback f is the sum of a large number of independent feedback processes, the uncertainty in f is well approximated by a normal distribution. Using the formula (21) to transform the probability density for f into a probability density for S, they find that the distribution for S must be skew, and has a quadratic fat tail, i.e. $p(S) \sim S^{-2}$ as $S \to \infty$.

Weitzman (2009b) uses arguments similar to these to argue that temperature distributions are fat-tailed. There are however several criticisms of the manner in which Weitzman applies these arguments to the economics of climate change, based on our scientific understanding of how increases in CO_2 concentrations affect temperature distributions (Nordhaus, 2011; Costello et al., 2010; Newbold and Daigneault, 2010; Roe and Bauman, 2010). I will explore only one version of these criticisms, which I believe is illustrative of the wider concerns on this front.

The key point to absorb about the climate sensitivity, is that even if we accept that its probability distribution is fat-tailed, the fact that it is an equilibrium quantity means that it tells us only about the amount of warming in the infinite future. Thus the focus on the climate sensitivity as a legitimator of the fat tails argument is probably misguided, as any blow-up in the expected marginal utility will be suppressed by an exponentially declining discount factor $\lim_{t \to \infty} e^{-\rho t} = 0$. The possibility remains however that temperatures may be fat tailed at finite time. However, there are some reasons for believing that this is not the case. In order to understand this, one must first appreciate a second crucial fact about the influence of climate sensitivity on temperature trajectories. It has been argued on very general grounds based on dimensional analysis (Hansen et al., 1985), and confirmed in climate models of all stripes (Baker and Roe, 2009; Held et al., 2010), that the atmosphere's *adjustment time* is proportional to S^2. The adjustment time is a measure of the time it takes for the system to reach equilibrium.

In order to understand the consequences of this fact, consider the following highly stylized reduced form model[10] of the temperature trajectory as a function

---

[9]All one needs to know for this calculation is the Stefan-Boltzmann law, the solar constant, the values for the effective emissivity and albedo of the earth, and the radiative forcing that results from CO_2 doubling – all very well established physics. See e.g. Hartmann (1994).

[10]I am grateful to Gerard Roe for suggesting this model to me. It can be derived from the following

of climate sensitivity S and time t:

$$(23) \qquad T(t) = S\left[1 - \exp\left(-\frac{K\sqrt{t}}{S}\right)\right],$$

where K is a constant. Note that this trajectory is the warming pathway that would arise if the concentration of CO_2 were instantaneously doubled from its preindustrial value – it thus corresponds to a very specific emissions pathway. This model incorporates three desirable properties – first, the adjustment time is proportional to S^2, second, $\frac{\partial T}{\partial S} > 0$, i.e. temperatures are increasing in sensitivity for all time, and finally, $\lim_{t\to\infty} T(t) = S$.

Notice the following fact about this model:

$$(24) \qquad \text{For any finite time } t, \ \lim_{S\to\infty} T(t) = K\sqrt{t}.$$

Thus in this stylized model, the amount of warming possible at any finite time t is *bounded*, even if the distribution for S itself is unbounded. In general, it seems reasonable to believe that any reduced form model of the dependence of temperature trajectories on the climate sensitivity that accounts for its effect on adjustment times will have this property. The important message of this stylized analysis is that there may be physical constraints which bound temperature change at finite time. It has been argued (Costello et al., 2010; Newbold and Daigneault, 2010) that bounded temperature distributions can dispense with the dismal theorem. The model in (23) has the benefit of proposing an (admittedly stylized) physical mechanism that bounds temperature distributions, rather than assuming an *ad hoc* bound as Costello et al. (2010) and Newbold and Daigneault (2010) do.

While I agree that it may be possible to rule out fat tailed temperature distributions on physical grounds (at least for some well defined emissions pathways), I do not agree that this rules out fat tailed consumption distributions, or indeed consumption distributions that put non-zero weight on zero consumption (Dietz, 2010). I believe the objections expressed in Costello et al. (2010); Newbold and Daigneault (2010); Roe and Bauman (2010) are an artifact of the unrealistic behaviour of their damage functions for large temperatures. Following the standard

---

differential equation for temperature change:

$$(22) \qquad c_0\sqrt{t}\frac{dT}{dt} = R(t) - \frac{T}{\lambda}.$$

Here the factor $c_0\sqrt{t}$ represents the effective heat capacity of the oceans, which is known to grow with the square root of time as heat fluxes at the earth's surface diffuse into the deep oceans (Hansen et al., 1985). $R(t)$ is radiative forcing, which is equal to S/λ in the case of CO_2 doubling, and λ is proportional to S. Setting the initial condition $T(0) = 0$, the solution of this equation is (23). Weitzman (2009a) considers a similar model, except that he neglects the $\sqrt{t}$ factor, so the temperature trajectories he arrives at do not have adjustment times with the correct quadratic dependence on S. See Baker and Roe (2009) for a more realistic, yet still analytically tractable, model.

assumptions in the integrated assessment literature (Nordhaus, 2008; Weitzman, 2010; Hope, 2006), their analyses assume multiplicative damage functions that imply that consumption is non-zero for any finite temperature. This seems a highly unrealistic choice – for example, it implies that consumption will still be above zero when the average temperature reaches the boiling point of blood! If instead one accepts that the damage multiplier hits zero at some finite temperature T_H, the dismal theorem is back in play. This is so because, as one can see from (24), the maximum possible value of warming grows unboundedly as $t \to \infty$, even for the relatively moderate case of a doubling of CO_2 concentrations. Thus there will eventually be a time at which there is a non-zero probability that $T > T_H$, provided the distribution over S has support on $[0, \infty)$. Of course, for utility functions with coefficient of relative risk aversion bounded above zero, this will ensure that expected marginal utility diverges at this time, and so the dismal theorem reappears. Dietz (2010) demonstrated that the PAGE integrated assessment model (Hope, 2006) is able to produce scenarios in which consumption is driven to zero at finite time for climate sensitivity values as low at $3°C$, adding to the plausibility of this story.

Weitzman ($2009a$) seems to have taken the criticisms of his result based on neglecting the dynamics of temperature trajectories on board. He goes to considerable lengths to show that a combination of factors – additive damage functions and uncertainty about the PRTP[11] – can still give rise to the dismal theorem even if temperature distributions remain bounded for all finite time, and the distribution over discount rates places an infinitesimally small weight on a zero discount rate. This is an interesting mathematical result about the limiting behaviour of certain functions, but it is perhaps overly sophisticated for the application at hand. Provided we accept that there is some temperature T_H at which consumption hits zero, and there is positive weight on $T > T_H$ at finite time, a dismal theorem like result it guaranteed. Weitzman's work shows that this is not a necessary condition for a divergence in the welfare function – one can send $T_H \to \infty$ and have probability distributions that place zero weight on the worst events (and the lowest discount rates) and still get a divergence. This is a valuable result, but to obtain the essential economic insights of the dismal theorem, one needs far less mathematical machinery than he employs: If there is any positive probability of consumption hitting zero, no matter how small, then welfare diverges for utility functions with coefficient of relative risk aversion bounded above one.

The upshot of this analysis – of both the Bayesian argument, and that based on toy scientific models – is that if we interpret the dismal theorem in the narrow sense in which it is presented in Weitzman ($2009b$), then its conclusions are not robust. Changes to the Bayesian framework used to justify it can cause it to disappear, as indeed can properly accounting for the physics of the climate system.

---

[11]Dasgupta (2001) is critical of the manner in which Weitzman uses uncertainty about the PRTP to deduce declining discount rates (e.g. Weitzman, 2001). Heal (2009) also emphasizes that in general this approach leads to dynamically inconsistent plans. Thus it is not clear that this is a legitimate approach to normative welfare analysis.

However, such a narrow interpretation does not do the dismal theorem justice. Weitzman (2009b) is careful to describe his modeling work as an abstract parable, suggesting we should focus on the broad-brush message, rather than the technical content, of the theorem. I believe this is the correct way to interpret his result. As we have seen, if we allow for consumption approaching zero at finite (but very large) temperature, as must be the case, and are not able to make certain transfers to the future (as seems very likely), then the criticisms of the dismal theorem that have been discussed thus far fail to persuade.

C.   The role of the utility function

The previous section highlighted the arguments for and against fat tailed probability distributions in the economic analysis of climate change. Much of the debate about the applicability of Weitzman's results has focussed on these issues. In the current section, I will argue that a more pertinent line of inquiry may be to interrogate the role that the utility function plays in the dismal theorem. Indeed, taking the conventional expected utility welfare framework as read at this point, this is the main content of Weitzman's result – it shows that welfare calculations are highly sensitive to the behaviour of the utility function at low consumption levels, and that the wide-spread CRRA utility function is perhaps ill suited to applications with fat-tailed risks.

The fact that the expected utility model of choice under uncertainty cannot be trusted to provide meaningful results for certain utility functions has been appreciated since Menger (1934) demonstrated that one requires utility functions to be *bounded* in order to be able to rank arbitrary lotteries. In fact, Menger (1934) showed that for any unbounded utility function, a variant of the St. Petersburg paradox exists. If we restrict attention to lotteries with finite expected payoffs, then the constraints on the utility function may be weakened. Arrow (1974) showed that if the utility function is concave, monotonically increasing, and *bounded below*, then expected utility exists for all lotteries with finite expected payoff. Note that this result does not require marginal utility to be finite at the origin – Arrow (1974) shows explicitly that this condition is not needed. Thus utility functions which are finite at the origin, but have infinite marginal utility at the origin, are still able to rank all lotteries of interest. In the case of CRRA utility functions (12), these conditions are fulfilled only when $\eta < 1$. In order to extend this result to an existence result for expected marginal utility (and thereby the stochastic discount factor), define $V(c) = -U'(c)$, and assume that $U''' > 0$, i.e. the decision maker is *prudent* (see e.g. Gollier, 2001, p. 237). Then $V(c)$ is increasing and concave, and provided $V(0) = -U'(0)$ is finite, expected marginal utility will exist for all lotteries with finite expected payoff. In the case of a CRRA utility function (for which $U''' > 0$ is always satisfied), $U'(c) = c^{-\eta}$, and thus $V(0)$ is not finite for *any* value of η. With these results in mind, the Dismal theorem is unsurprising – we are simply pushing a particular choice of utility function beyond its region of usefulness. Arrow (2009) and Nordhaus (2009) come

to much the same conclusion.

HARA UTILITY FUNCTIONS. — There are alternative choices of utility function that are almost as easy to handle analytically as the CRRA utility function, and do not suffer from this difficulty. Weitzman (2009*b*) himself discusses a 'value of statistical life' cut-off parameter D such that $U(c) = 0$ for all $c < D$, and Pindyck (2011) performs a set of illustrative welfare calculations with a similar model that places a cut-off on marginal utility, and demonstrates (unsurprisingly), that depending on the position of the cut-off, fat-tailed distributions may in fact be preferred to thin-tailed ones. Both of these approaches provide a technical resolution to the dismal theorem, but suffer from some drawbacks, as they require assuming that the utility function is not everywhere differentiable. As we shall see below, this is by no means a necessary assumption.

Consider the smooth Harmonic Absolute Risk Aversion (HARA) utility function, which is widely used in the finance literature:

$$(25) \qquad\qquad U(c) = \zeta \left(\nu + \frac{c}{\eta} \right)^{1-\eta}, \ \text{ where } \zeta \left(1 - \frac{1}{\eta} \right) > 0.$$

$$(26) \qquad \Rightarrow -c \frac{U''(c)}{U'(c)} = c \left(\nu + \frac{c}{\eta} \right)^{-1}.$$

Clearly, this utility function has a coefficient of absolute risk aversion that is decreasing in consumption, and a coefficient of relative risk aversion that is increasing in consumption. In the limit $\nu \to 0$, the HARA utility function approaches a CRRA utility function (12). Notice however that for all $\nu > 0$, the coefficient of relative risk aversion for the HARA function is equal to zero when $c = 0$. Moreover, it is easy to see that $U^{(n)}(0)$, the n-th derivative of U at the origin, exists for all n. Thus both expected utility and expected marginal utility exist for this utility function for arbitrary finite mean probability distributions for consumption.

To illustrate how the stochastic discount factor M depends on the parameters of the HARA utility function, suppose that $y = \log c$ is a Cauchy distributed random variable[12] – a prototypical example of a fat-tailed distribution, considered by Weitzman (2009*b*) and Costello et al. (2010). In Figure 1, we plot the logarithm of the stochastic discount factor (4) as a function of the parameter ν in (25). Of course, $\lim_{\nu \to 0} M = \infty$ – this is obviously not representable on the curve in Figure 1. What the figure shows is that M varies by a factor of about 10^8 between $\nu = 10^{-3}$ and $\nu = 1$. Thus, under this choice of utility function, social evaluation of fat-tailed consumption risks will be *extremely* sensitive to the choice of ν for low values of ν. However for ν greater than about 1, evaluation of fat-tailed risks is relatively *insensitive* to the value of ν – this suggests that the HARA utility

---

[12]The Cauchy distribution is the same as a Student's t-distribution with 1 degree of freedom.
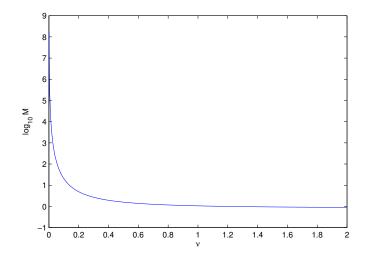
FIGURE 1. LOGARITHM OF THE STOCHASTIC DISCOUNT FACTOR M, AS A FUNCTION OF THE PARAMETER ν OF THE HARA UTILITY FUNCTION (25). $\eta = 4, \zeta = 1, \beta = 1$ IN THIS FIGURE, AND CONSUMPTION GROWTH IS ASSUMED TO BE CAUCHY DISTRIBUTED.

function with $\nu > 1$ might provide a good alternative to CRRA utility functions when risks are fat-tailed. Note that the HARA utility function provides finite answers even when there is a non-zero chance that consumption is identically zero at finite time, e.g. if damage multipliers hit zero at finite temperature.

While the analysis above shows that welfare *evaluations* (i.e. the computation of a welfare function, or the stochastic discount factor, for a given policy) may still be very sensitive to the tails of a distribution when the utility function is bounded (i.e. for a small value of ν in (25)), this does not mean that welfare *comparisons* (i.e. policy choices) are also heavily dependent on the tails. In fact the following result, due to Chichilnisky (2000), implies just the opposite:

THEOREM 2: *Let $x(t), y(t)$ be lotteries over a random variable $t \in \mathbb{R}$ with probability measure $\mu(t)$. Then if preferences over lotteries can be represented by expected utility functionals with a bounded utility function,*

$$x \succ y \iff \exists \epsilon > 0 : x' \succ y' \text{ where } x = x' \text{ and } y = y' \text{ a.e. on any } \Sigma \subset \mathbb{R} : \mu(\Sigma^c) < \epsilon.$$

In the words of Chichilnisky (2000), *"expected utility is insensitive to small probability events"*, provided of course that the utility function is bounded. Thus, even if the ν parameter of the HARA utility function is arbitrarily close to zero, implying that welfare measures are very large negative for fat-tailed risks, the ranking of alternative policies is *independent* of these very low probability outcomes. Thus boundedness of the utility function, even at a very large negative

value (i.e. $\nu \ll 1$), dispatches with the dismal theorem entirely for the purposes of policy choice.

For Weitzman, the approach to resolving the dismal theorem via bounding the utility function is unsatisfactory. In his words, *"It is easy to put arbitrary bounds on utility functions...that arbitrarily cut off or otherwise severely dampen...low values of C. Introducing...these changes formally closes the model in the sense of replacing the symbol $+\infty$ with an arbitrarily large but finite number."* As Figure 1 shows, this is not the case if one allows for utility functions with coefficient of relative risk aversion that tends to zero for low c. In this case one can choose perfectly reasonable looking parameter values that guarantee that the stochastic discount factor is not only not 'arbitrarily large', but of order 1 for fat-tailed risks. Moreover, policy choice is independent of low probability events in this case. This modeling choice is only 'arbitrary', if one is wedded to the CRRA utility function as the normative baseline of welfare calculations. The widespread focus on the CRRA utility function is perhaps an instance of the 'tools-to-theories' heuristic (Gigerenzer, 1991), in which an analytical tool – the CRRA utility function – is elevated to the status of a theory, in this case of normative welfare analysis. Indeed, Weitzman sometimes writes as if expected utility with a CRRA-like utility function is synonymous with expected utility theory itself. This is of course not the case. If we believe the axioms of expected utility theory, as elucidated by von Neumann and Morgenstern (1944), and especially Savage (1954), we should evaluate uncertain prospects with expected utility functionals. However these axioms are silent as to which utility function, or indeed which probability distribution, is appropriate for a given application. It is up to us to furnish expected utility theory with a utility function that is true to our tastes, and a probability distribution that is true to our beliefs. Thus, if we feel that the CRRA utility function does not represent our preferences in the case of catastrophic climate change, we should not hesitate to replace it with another utility function that does. Nordhaus (2009) performs some illustrative calculations that suggest that our willingness to pay to avoid catastrophes is not very large (he considers the case of payments to avoid killer astroids), suggesting that the CRRA utility function may be too cautious to reflect individual preferences.

AGGREGATION OF HETEROGENEOUS PREFERENCES. — Although HARA utility functions may provide a more acceptable representation of individual preferences, it is social preferences that should enter welfare calculations. Indeed it is unclear in the analysis presented thus far exactly where the utility function comes from, and what it represents. Does it include prioritarian ethical judgements about welfare inequality, or measure only the empirical risk preferences of a representative agent (Kaplow and Weisbach, 2011), and more generally, how does it relate to the preferences of the individuals that make up society? Clearly, in addressing this question, we should draw on the resources provided by social choice theory.

It is important to stress that even though individuals' utility functions may

all be bounded at $c = 0$, this does not imply that society's welfare is similarly constrained. The analysis of HARA utility functions presented above assumed a homogenous population, each of whom has the same attitude to risk, or equivalently, that a single representative agent may be identified whose risk preferences are given by the HARA utility functions examined above. This assumption masks some of the complexity of welfare analysis under catastrophes. Assuming an anonymous social ordering, and that the Harsanyi (1955, 1978) social aggregation axioms hold, we should aggregate over the expected utility obtained by all individuals in order to arrive at social welfare under the catastrophic risk. Treating this aggregation exercise as a *fait accompli* can be misleading. In fact, one can show that even if every individual's utility function is bounded for all c, aggregate social welfare may still diverge (or be very sensitive to the tails of the distribution over c). This suggests that it is not a straightforward matter to simply presuppose a HARA social utility function because our individual willingness to pay to avoid catastrophe, or that of a sample of the population, is small.

A simple example will suffice to make this point. Suppose that each agent in a heterogenous population consumes the same amount in all states of the world, but each has a different utility function[13]. Suppose that agents' utility functions are:

(27) $$U(c; \phi) = -(v + c)^{-\phi}.$$

where $v, \phi > 0$. Clearly, $U' > 0, U'' < 0$, and $U^{(n)}(0)$ is bounded for all agents. Assume an anonymous social ordering, that each agent has the same value of v, and that the density of individuals with preferences ϕ is given by $w(\phi)$. Then a social planner who evaluates consumption risks would need to compute the following aggregated expected utility integral[14]:

(28) $$V = \int \left[\int U(c, \phi) p(c) dc \right] w(\phi) d\phi$$

(29) $$= \int \left[\int U(c, \phi) w(\phi) d\phi \right] p(c) dc$$

where $p(c)$ is the consumption risk (we normalize the total population size to 1). Suppose now that $w(\phi)$ is a gamma distribution, with parameters (θ, k), and

---

[13]Note that Harsanyi's axioms do not require these utility functions to be interpersonally comparable.
[14]Fubini's theorem guarantees that if the integral converges, we can change the order of integration. A simple *reductio ad absurdum* argument makes the result that follows water-tight.

consider the inner integral in (29). This evaluates to

$$(30) \qquad \int U(c,\phi)w(\phi)d\phi = -\frac{1}{\theta^k \Gamma(k)} \int_0^\infty (v+c)^{-\phi} \phi^{k-1} e^{-\phi/\theta} d\phi$$

$$(31) \qquad \qquad \propto -\int_0^\infty \phi^{k-1} \exp\left(-\phi \left[\ln(v+c) + \frac{1}{\theta}\right]\right) d\phi$$

Now in order for this integral to be finite, we require $c > c^*$, where

$$(32) \qquad \qquad c^* := e^{-\frac{1}{\theta}} - v$$

Thus if $1/\theta < \ln(1/v)$, the social welfare function diverges at $c^* > 0$, provided $p(c)$ puts positive weight on $c < c^*$, even though each individual's utility function is bounded for all c. Since $1/\theta = \frac{\text{mean}(\phi)}{\text{var}(\phi)}$ for the gamma distribution, the divergence occurs if $v < 1$, and the mean of $w(\phi)$ is smaller than $\ln(1/v)$ times its variance.

Note that this result is in stark contrast to the well known Arrow-Lind theorem on public risk bearing (Arrow and Lind, 1970) – the reason being that in this case the risks to individuals are correlated (the relevant case for applications to climate change), whereas the Arrow-Lind theorem assumes uncorrelated, and therefore diversifiable risks. The result has a similar flavour to the dismal theorem, in that it relies on the extremes of the distribution of ϕ to generate a divergence in expected social utility[15]. It is thus subject to some of the same criticisms – e.g. perhaps it is unreasonable to assume that the distribution of ϕ has full support on $[0, \infty)$, or that $v < 1$, or indeed that the Harsanyi aggregation theorem is a reasonable basis for social choice[16] (Weymark, 1991). Excepting these objections, its conceptual message is clear – social evaluation of correlated risks can be highly sensitive to those individuals who are highly risk averse, even if they make up an infinitesimal fraction of the population. This adds a complication to the already difficult issue of social aggregation when catastrophic risks are involved. It is not a simple matter to reason (as Nordhaus (2009, 2011) does) that since spending on, for example, asteroid defense systems, is small, social willingness to pay is similarly constrained. Such spending decisions are made by small samples of the population, which have a very low chance of including the most risk-averse members of society.

---

[15]In fact, some might say that this is a stronger result than the dismal theorem. It shows that social welfare may diverge *at a positive value of* c, even for *exponentially* decaying distributions over ϕ. No fat tails in sight here.

[16]Note however that a similar result would obtain if instead of adopting Harsanyi's aggregation theorem, we took a Rawlsian approach to distributive justice, and maximized the utility of the least well off individual in society. Under catastrophic risks, this would be the individual with $\phi \to \infty$, for whom $U(c;\phi) \to -\infty$ for $c < 1 - v$, $U(c;\phi) \to 0$ for $c > 1 - v$. Thus social welfare would still diverge if $p(c)$ puts any weight on $c < 1 - v$.

III.   The elephant in the room: Population change

The HARA utility functions discussed in the previous section provide a technical resolution to the dismal theorem, but they do not resolve the ethical problems it raises. Correctly interpreted, the dismal theorem says that social evaluation of fat-tailed risks depends sensitively on how we choose to treat catastrophic outcomes. If we believe such outcomes to be infinitely bad for society, then they can dominate all other ethical choices in a conventional welfare analysis based on expected utility theory. If they are not infinitely bad, we still face the difficult task of deciding how to 'value' a catastrophe that wipes out the whole human race (and other species as well). For example, if one adopts the HARA utility function, one still needs to justify a particular choice of ν. As the previous section demonstrated, this is a by no means trivial exercise in social choice.

In this section, I want to suggest that the ethical problems that the dismal theorem raises run deeper than simply deciding on the behaviour of the social utility function for low values of consumption. There is an implicit assumption in Weitzman's welfare framework (and that employed by his critics), as represented by equation (1), which is by no means innocuous. This assumption relates to how the welfare function (1) accounts for population change.

Many applications of welfare economics need not concern themselves with population change, as they are concerned with marginal projects which are unlikely to affect which 'potential persons' come into being. However if there is any situation in which concerns about population change are likely to be relevant, then it has to be catastrophic climate change. Indeed, almost by definition, a climate catastrophe would entail a drastic, perhaps total, collapse in the earth's population. The idea that we can consider states of the world in which consumption is drawn down to zero without accounting for a change in population is manifestly absurd.

The only way in which it is possible to interpret Weitzman's welfare function (1) as accounting for population change, is if we interpret it as adopting an *average utilitarian* approach to population ethics. That is, an uncertain prospect is evaluated by computing the average utility of all people who happen to be alive in that prospect, and welfare is just the expectation of these average utilities. Then the function $U(c)$ that appears in (1) is just the average utility of a set of identical agents when their consumption is c. But is average utilitarianism a defensible approach to welfare evaluations when population varies?

A.   *Population principles and the dismal theorem*

Since the publication of Derek Parfit's 'Reasons and Persons' (Parfit, 1984), formal approaches to population issues in social choice theory have been gaining

momentum[17]. Two recently published books (Broome, 2004; Blackorby, Bossert and Donaldson, 2005) provide good overviews of the welfarist[18] approach to the problem. Blackorby, Bossert and Donaldson (2005) in particular set out a list of axioms that one might reasonably want variable population social welfare functions, or population principles, to satisfy, and provide representation results for various subsets of these axioms. They also prove several impossibility theorems which show that some combinations of these axioms cannot provide consistent and complete social orderings. Dasgupta (2001) provides an alternative perspective on population ethics.

A key new concept that is introduced when considering welfare functions that account for population change is the *critical level*. Suppose that social welfare at a point in time may be represented by a function $P(\bar{U}, N)$, where $\bar{U}$ is average utility and N is population size. This is the case for all population principles of interest[19] (Blackorby, Bossert and Donaldson, 2005). Now imagine adding a single individual to a population, holding the utility of all other members of the population constant. The critical level is defined as the utility level of the additional individual which leaves total welfare unchanged, i.e. his existence is socially neutral if his utility is at the critical level. Formally, the critical level of a population principle $P(\bar{U}, N)$, evaluated at a given level of average utility and population size, is defined as the utility level $C(\bar{U}, N)$ which satisfies:

$$(33) \qquad P(\bar{U}, N) = P\left(\frac{N\bar{U} + C(\bar{U}, N)}{N + 1}, N + 1\right)$$

As an example, consider average utilitarianism, for which $P(\bar{U}, N) = \bar{U}$. In this case $C(\bar{U}, N) = \bar{U}$, i.e. the critical level is equal to the average utility of those who live, and is independent of population size. From this we see that average utilitarianism entails a large amount of dependence between the value of additional population, and the welfare of those who live. This conflicts with two key independence axioms: *utility independence*, and *existence independence*. Utility independence requires rankings of alternatives to be independent of the utility of individuals whose utilities are the same in both alternatives. Existence independence is a stronger requirement, which insists that rankings are also independent of the existence of individuals with the same utility in both alternatives. To see that average utilitarianism violates both these axioms, consider two different util-

---

[17]Of course, population issues have always been central to economics (Malthus, 1798; Sidgwick, 1907), and had been treated formally before Parfit (e.g. Dasgupta, 1969). However Parfit's book provided the impetus for an axiomatic approach to the problem that is rooted in social choice theory.

[18]I use the term 'welfarist' to refer to the consequentialist welfare function type approach to social evaluation, which both these books largely follow. Strictly speaking, neither book is restricted to a purely welfarist approach, as both consider cases in which non-welfare information, e.g. birth date, affects social choice. For a deontological approach to population issues, see Kamm (1993, 1996).

[19]We do not consider generalized utilitarian rules. Such rules depend on the nonlinear average $\bar{U}_g = g^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} g(U_i)\right)$, where the curvature of g measures aversion to welfare inequality.

ity profiles: $U_1 = (a, b)$, $U_2 = (a, c, d)$. Clearly $U_1 \succ U_2$ iff $\frac{a+b}{2} > \frac{a+c+d}{3}$ under average utilitarianism. This ranking depends on the value of a, and therefore violates utility independence. Moreover, if a did not exist, $U_1 \succ U_2$ iff $b > \frac{c+d}{2}$. It is easy to see that both these inequalities cannot hold for all values of (a, b, c, d), and so existence independence is also violated. Finally, the fact that the critical level for average utilitarianism depends on $\bar{U}$ implies that it violates a further desirable axiom – the *negative expansion principle*. Suppose, without loss of generality, that we define $U = 0$ as the value of utility at which an individual's life is just worth living[20]. Then the negative expansion principle says that if we add an individual with $U < 0$ to the population, keeping everyone else's condition unchanged, then social welfare should decrease. Since the critical level for average utilitarianism is just average utility, which may be negative, it is possible to add an individual with negative utility to the population and increase social welfare, thus violating the negative expansion principle. In general, Broome (2004, chapter 13) argues that population principles with critical levels that depend on average utility are undesirable.

The most commonly deployed alternative to average utilitarianism is classical utilitarianism, whose population principle is $P(\bar{U}, N) = N\bar{U}$. This is the welfare function employed by the much used DICE integrated assessment model (Nordhaus, 2008). The critical level for classical utilitarianism is zero, and it satisfies existence, and therefore utility, independence. However it exhibits Parfit's *repugnant conclusion*. For every value of N, $\bar{U} > 0$ and $0 < \epsilon < \bar{U}$, the repugnant conclusion implies that it is possible to find an $M > 0$ such that $P(\bar{U}, N) < P(\bar{U} - \epsilon, N + M)$. In words, the repugnant conclusion says that a sufficiently large increase in population size can substitute for any decrease in average utility (provided average utility stays positive). It is clear simply by looking at the product representation of the classical utilitarian welfare function, that the repugnant conclusion is implied. Population principles that do not avoid the repugnant conclusion thus place a high value on population.

Neither of these two commonly used population principles seems to provide a satisfactory approach to population ethics. A simple alternative, critical level utilitarianism (CLU), avoids the pitfalls of both these principles, and is advocated by Broome (2004). Blackorby, Bossert and Donaldson (2005, pp. 147-151) give an exhaustive discussion of CLU's desirable properties, relative to other population principles. CLU's population principle is:

$$(34) \qquad P_{CL}(\bar{U}, N) = N(\bar{U} - \alpha)$$

where $\alpha > 0$ is a constant, the critical level. CLU satisfies utility and existence independence, avoidance of the repugnant conclusion, and the negative expansion principle. It does not, however, satisfy a further requirement known as *prior-*

---

[20]This requires us to assume that utilities are ratio-scale measurable, rather than cardinally measurable, a standard assumption in population ethics.

ity for lives worth living. Population principles which satisfy this requirement have the property that all alternatives in which every individual's utility is positive are ranked higher than all alternatives in which every individual's utility is negative. That CLU violates priority for lives worth living seems an acceptable compromise, and more desirable than dispensing with either utility or existence independence, avoidance of the repugnant conclusion, or the negative expansion principle. Indeed, priority for lives worth living seems the least inevitable of the axioms discussed thus far. It implies a kind of reverse repugnant conclusion, in that it insists that a population consisting of an arbitrary number of individuals with utility $-\epsilon$ is ranked as worse than an alternative consisting of a single individual with utility ϵ, for any $\epsilon > 0$. Thus, it does not accept that population may be an end in itself, for example because there is some existence value to the human species, and it may thus be worth maintaining some small population with reproductive capacity, even if its constituents live extremely unpleasant lives. Thus principles that satisfy priority for lives worth living, such as average utilitarianism, place a low value on population.

How would the dismal theorem be affected if we were to employ a CLU function to evaluate climate policy? Let $p(c, N)$ be the joint distribution over (c, N), then the CLU analogue of (1) is

$$(35) \qquad W_{CL} = \bar{U}(1) - \alpha + \beta \int N \left(\bar{U}(c) - \alpha\right) p(c, N) dc \, dN$$

$$(36) \qquad = \bar{U}(1) - \alpha + \beta \int N \left(\bar{U}(c) - \alpha\right) p(N|c)p(c) dc \, dN$$

$$(37) \qquad = \bar{U}(1) - \alpha + \beta \int \bar{N}(c) \left(\bar{U}(c) - \alpha\right) p(c) \, dc$$

where $\bar{N}(c) = \int N p(N|c) dN$ is the mean of the distribution for N conditional on c, and $p(c)$ is the marginal distribution for c. Biological necessity requires that as $c \to 0$, $\bar{N}$ must approach zero. Thus catastrophic events are given less weight in welfare assessments with CLU welfare functions than they are by average utilitarianism, since they correspond to low values of $\bar{N}$. This does not however mean that catastrophic risks cannot dominate welfare. If $\bar{U}$ is unbounded below, fat-tailed consumption risks may still theoretically lead to a divergence in W_{CL}. However, this seems a much bigger stretch for the current model than it does for the original welfare function (1), as it requires the product $\bar{N}(c)\bar{U}(c)$ to be sufficiently steeply curved for low c. If $\bar{U}(c) \sim c^{1-\eta}$ and $\bar{N}(c) \sim c^{\theta}$ as $c \to 0$, a sufficient condition for W_{CL} to be finite is that $\theta > \eta - 1$. If population declines exponentially as $c \to 0$, as seems very likely based on conventional population models, this inequality is always satisfied. Nevertheless, we cannot unequivocally rule out a dismal-theorem like divergence. The only sure way to guarantee that welfare converges for all possible risks is to require that the utility function $\bar{U}$ be bounded.

While the CLU welfare function provides a more acceptable framework for the analysis of policies which may affect the size of future populations, it also requires new inputs, i.e. values for the critical level α and the function $\bar{N}(c)$. $\bar{N}(c)$ is a purely empirical input, and may be estimated (at least in principle), by examining the dependence of fertility and mortality rates on nutrition levels (Fogel, 1997). α however is an ethical judgement about how good an individual's life must be for her existence to be socially desirable. Its value should not be too large (for obvious reasons), nor too small, as small values of α approximate the repugnant conclusion. Much like the PRTP, its value should be determined through public reason and debate. Choices for such ethical parameters *"should reflect conscious political decisions with respect to what are, after all, political questions."* (Dasgupta, Sen and Marglin, 1972, p. 120).

IV. Conclusions

Weitzman's dismal theorem, and his subsequent work on the climate damages function (Weitzman, 2010), has done much to show that existing analyses of climate mitigation policy based on integrated assessment models are heavily dependent on poorly substantiated model assumptions, and may be insufficiently sensitive to low-probability, high-impact, outcomes. These points are all well made, however this paper has argued that perhaps the true message of the dismal theorem is to call into question the appropriateness of standard welfare frameworks for the analysis of climate change policy. Such concerns have been raised before (Heal, 2009) – the dismal theorem throws them into the spotlight. Climate change is a mega-problem – it requires us to evaluate and compare market and non-market impacts, come to terms with deep uncertainty, make inter-temporal trade-offs far into the next century, divide mitigation responsibilities equitably, and contemplate potential catastrophes. It is unsurprising that the old workhorse of a constant relative risk aversion expected utility functional is not up to the task of providing normatively acceptable assessments of climate policy. The dismal theorem makes this plain. While efforts to improve the empirical adequacy of integrated assessment models are underway, it is vital that a parallel program interrogates the ethical and formal foundations of such models. Accounting for uncertainty or ambiguity, social aggregation problems, population change, and catastrophes are certain to be necessary components of a more satisfactory approach. Any given treatment of these issues is likely to be controversial, yet not making them explicit is more problematic, as implicit, unjustified assumptions then determine policy prescriptions.

REFERENCES

Ackerman, F., E.A. Stanton, and R. Bueno. 2010. "Fat tails, exponents, and extreme uncertainty: Simulating catastrophe in DICE." *Ecological Economics*, 69(8): 1657–1665.

Arrow, Kenneth J. 1974. "The use of unbounded utility functions in expected-utility maximization: Response." *Quarterly Journal of Economics*, 88(1): 136–138.

Arrow, Kenneth J. 2009. "A note on uncertainty and discounting in models of economic growth." *Journal of Risk and Uncertainty*, 38(2): 87–94.

Arrow, Kenneth J., and Robert C. Lind. 1970. "Uncertainty and the Evaluation of Public Investment Decisions." *The American Economic Review*, 60(3): 364–378.

Arrow, K.J., and L. Hurwicz. 1977. "An optimality criterion for decision-making under ignorance." In *Studies in resource allocation processes*. 482. Cambridge, UK:Cambridge University Press.

Baker, Marcia B., and Gerard H. Roe. 2009. "The Shape of Things to Come: Why Is Climate Change So Predictable?" *Journal of Climate*, 22(17): 4574–4589.

Blackorby, Charles, Walter Bossert, and David J. Donaldson. 2005. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics. Econometric Society Monographs*, Cambridge University Press.

Broome, John. 2004. *Weighing Lives.* Oxford University Press, USA.

Chichilnisky, Graciela. 2000. "An axiomatic approach to choice under uncertainty with catastrophic risks." *Resource and Energy Economics*, 22(3): 221–231.

Costello, Christopher J., Michael G. Neubert, Stephen A. Polasky, and Andrew R. Solow. 2010. "Bounded uncertainty and climate change economics." *Proceedings of the National Academy of Sciences*, 107(18): 8108–8110.

Dasgupta, P. 1969. "On the Concept of Optimum Population." *The Review of Economic Studies*, 36(3): 295–318.

Dasgupta, Partha. 2001. *Human well-being and the natural environment.* Oxford:Oxford University Press.

Dasgupta, Partha. 2008. "Discounting climate change." *Journal of Risk and Uncertainty*, 37(2): 141–169.

Dasgupta, Partha, Amartya Sen, and Stephen A Marglin. 1972. *Guidelines for project evaluation.* New York:United Nations.

Dietz, Simon. 2010. "High impact, low probability? An empirical analysis of risk in the economics of climate change." *Climatic Change, to appear.*

Dietz, Simon, and Nicholas Stern. 2008. "Why Economic Analysis Supports Strong Action on Climate Change: A Response to the Stern Review's Critics." *Review of Environmental Economics and Policy*, 2(1): 94 –113.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *The Quarterly Journal of Economics*, 75(4): 643–669.

Fogel, Robert William. 1997. "New findings on secular trends in nutrition and mortality: Some implications for population theory." In *Handbook of population and family economics.* Vol. 1, Part 1, , ed. Mark R. Rosenzweig and Oded Stark, 433–481. Elsevier.

Frame, D. J., B. B. B. Booth, J. A. Kettleborough, D. A. Stainforth, J. M. Gregory, M. Collins, and M. R. Allen. 2005. "Constraining climate forecasts: The role of prior assumptions." *Geophysical Research Letters*, 32.

Geweke, John. 2001. "A note on some limitations of CRRA utility." *Economics letters*, 71: 341–345.

Gigerenzer, Gerd. 1991. "From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology." *Psychological Review*, 98(2): 254–267.

Gilboa, Itzhak. 2009. *Theory of Decision under Uncertainty.* . 1 ed., Cambridge University Press.

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

Gilboa, Itzhak, Andrew Postlewaite, and David Schmeidler. 2009. "Is It Always Rational to Satisfy Savage's Axioms?" *Economics and Philosophy*, 25(3): 285–296.

Gollier, Christian. 2001. *The economics of risk and time.* Cambridge, Mass.; London:MIT Press.

Gollier, Christian, and Johannes Gierlinger. 2008. "Socially efficient discounting under ambiguity aversion, http://idei.fr/display.php?a=9848." *IDEI Working Paper*, , (561).

Hansen, J., A. Lacis, D. Rind, and G. Russell. 1984. "Climate sensitivity: Analysis of feedback mechanisms." In *Climate processes and Climate sensitivity. Geophysical Monograph*, , ed. Maurice Ewing. American Geophysical Union.

Hansen, J., G. Russell, A. Lacis, I. Fung, D. Rind, and P. Stone. 1985. "Climate Response Times: Dependence on Climate Sensitivity and Ocean Mixing." *Science*, 229(4716): 857–859.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *The Journal of Political Economy*, 63(4): 309–321.

Harsanyi, John C. 1978. "Bayesian Decision Theory and Utilitarian Ethics." *American Economic Review*, 68(2): 223–228.

Hartmann, Dennis L. 1994. *Global Physical Climatology.* . 1 ed., Academic Press.

Heal, Geoffrey. 2009. "Climate Economics: A Meta-Review and Some Suggestions for Future Research." *Review of Environmental Economics and Policy*, 3(1): 4–21.

Held, Isaac M., Michael Winton, Ken Takahashi, Thomas Delworth, Fanrong Zeng, and Geoffrey K. Vallis. 2010. "Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing." *Journal of Climate*, 23(9): 2418–2427.

Hope, Chris. 2006. "The marginal impact of CO2 from PAGE2002: An integrated assessment model incorporating the IPCC's five reasons for concern." *Integrated Assessment*, 6(1).

Horowitz, John, and Andreas Lange. 2009. "What's wrong with infinity - a note on Weitzman's dismal theorem." `http://faculty.arec.umd.edu/jhorowitz/weitzman_final.pdf`.

Jaynes, E.T. 1965. "Prior probabilities and transformation groups." Unpublished, available at: `http://bayes.wustl.edu/etj/articles/groups.pdf`.

Jaynes, E.T. 1968. "Prior probabilities." *IEEE Transactions on system science and cybernetics*, 4(3): 227–241.

Jeffreys, Harold. 1946. "An Invariant Form for the Prior Probability in Estimation Problems." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007): 453–461.

Ju, Nengjiu, and Jianjun Miao. 2009. "Ambiguity, learning, and asset returns." *Econometrica (forthcoming)*, , (`http://people.bu.edu/miaoj/jm17.pdf`).

Kamm, F. M. 1993. *Morality, Mortality: Volume I: Death and Whom to Save from It.* Oxford University Press, USA.

Kamm, F. M. 1996. *Morality, Mortality: Volume II: Rights, Duties, and Status.* Oxford University Press, USA.

Kaplow, Louis, and David Weisbach. 2011. "Discount rates, social judgments, individuals' risk preferences, and uncertainty." *Journal of Risk and Uncertainty*, , (forthcoming).

Karp, Larry. 2009. "Sacrifice, discounting, and climate policy: five questions." `http://are.berkeley.edu/~karp/FiveQuestionsSubmit.pdf`.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. "A Smooth Model of Decision Making under Ambiguity." *Econometrica*, 73(6): 1849–1892.

Knight, Frank. 1921. *Risk, uncertainty, and profit.* New York:Houghton Mifflin.

Malthus, Thomas Robert. 1798. *An essay on the princple of population.* J. Johnson (London).

Menger, Karl. 1934. "Das Unsicherheitsmoment in der Wertlehre. Betrachtungen im Anschluss an das sogennante Petersburger Spiel." *Zeitschrift für Nationalökonomie*, 5: 459–485.

Millner, Antony, Simon Dietz, and Geoffrey Heal. 2010. "Ambiguity and climate policy." *NBER working paper no. 16050*, , (`http://www.nber.org/papers/w16050.pdf`).

Newbold, Stephen C., and Adam Daigneault. 2010. "Climate response uncertainty and the benefits of greenhouse gas emission reductions." *Environmental and Resource Economics*, 44(3): 351–377.

Nordhaus, William D. 2007. "A Review of the Stern Review on the Economics of Climate Change." *Journal of Economic Literature*, 45(3): 686–702.

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

Nordhaus, William D. 2009. "An analysis of Weitzman's dismal theorem." `http://nordhaus.econ.yale.edu/documents/weitz_011609.pdf`.

Nordhaus, William D. 2011. "The Economics of Tail Events with an Application to Climate Change." *Review of Environmental Economics and Policy*, 5(2): 240 –257.

Parfit, Derek. 1984. *Reasons and Persons.* Oxford University Press, USA.

Pindyck, Robert S. 2011. "Fat Tails, Thin Tails, and Climate Change Policy." *Review of Environmental Economics and Policy*, 5(2): 258–274.

Roe, Gerard, and Yoram Bauman. 2010. "Should the climate tail wag the policy dog?" In preparation, available at: `http://earthweb.ess.washington.edu/roe/GerardWeb/Publications_files/martydom_draft.pdf`.

Roe, Gerard H., and Marcia B. Baker. 2007. "Why Is Climate Sensitivity So Unpredictable?" *Science*, 318(5850): 629–632.

Savage, Leonard J. 1954. *The Foundations of Statistics.* Wiley and Sons.

Schwartz, Stephen E. 2010. "Feedback and sensitivity in an electrical circuit: An analog for climate models." *Climatic Change, to appear.*

Schwarz, Michael. 1999. "Decision making under extreme uncertainty." PhD diss. Stanford University Graduate School of Business.

Sidgwick, Henry. 1907. *The method of ethics. .* 7 ed., Macmillan and Co.

Solomon, Susan, Gian-Kasper Plattner, Reto Knutti, and Pierre Friedlingstein. 2009. "Irreversible climate change due to carbon dioxide emissions." *Proceedings of the National Academy of Sciences*, 106(6): 1704–1709.

Sterner, Thomas, and U. Martin Persson. 2008. "An Even Sterner Review: Introducing Relative Prices into the Discounting Debate." *Review of Environmental Economics and Policy*, 2(1): 61 –76.

Stern, N. H. 2007. *The economics of climate change : the Stern review.* Cambridge:Cambridge University Press.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behaviour.* Princeton University Press.

Weitzman, Martin L. 2001. "Gamma Discounting." *The American Economic Review*, 91(1): 260–271.

Weitzman, Martin L. 2007. "A Review of The Stern Review on the Economics of Climate Change." *Journal of Economic Literature*, 45(3): 703–724.

Weitzman, Martin L. 2009*a*. "Additive damages, fat-tailed climate dynamics and uncertain discounting." *Economics - The Open-Access, Open-Assessment E-Journal*, 3(2009-39).

Weitzman, Martin L. 2009*b*. "On Modeling and Interpreting the Economics of Catastrophic Climate Change." *Review of Economics and Statistics*, 91(1): 1–19.

Weitzman, Martin L. 2010. "GHG targets as insurance against catastrophic climate damages." National Bureau of Economic Research.

Weymark, J.A. 1991. "A reconsideration of the Harsanyi-Sen debate on Utilitarianism." In *Interpersonal comparisons of well-being. ,* ed. J. Elster and J. Roemer. Cambridge University Press.

AMBIGUITY AVERSION AND THE DISMAL THEOREM

Formally, the scale parameter s in the expression (17) for posterior beliefs is an uncertain model parameter which specifies a distribution for consumption at each of its values. How would Weitzman's calculation of the stochastic discount factor change if we account for the fact that the decision maker may be averse to the ambiguity over consumption distributions implied by uncertainty in s?

We work in the framework of Klibanoff, Marinacci and Mukerji (2005). The analogue of (1) which accounts for ambiguity aversion is:

$$(\text{A1}) \qquad\qquad W' = U(1) + \beta\Phi^{-1}\left(\mathsf{E}_s\Phi(\mathsf{E}_{c|s}U(c))\right)$$

where Φ encodes ambiguity averse preferences, and thus satisfies $\Phi' > 0, \Phi'' < 0$. Using the fact that $\frac{d}{dx}f^{-1}(x) = [f'(f^{-1}(x))]^{-1}$, we can go through the standard manipulations to show that the appropriate discount factor for this model is:

$$(\text{A2}) \qquad\qquad M' = \beta\frac{\mathsf{E}_s\left[\Phi'(\mathsf{E}_{c|s}U(c))\mathsf{E}_{c|s}U'(c)/U'(1)\right]}{\Phi'\left(\Phi^{-1}(\mathsf{E}_s\Phi(\mathsf{E}_{c|s}U(c)))\right)}.$$

Notice that when Φ is linear (i.e. ambiguity neutrality) this expression reduces to (4). We specialize to a simple extension of the example Weitzman presents in his paper. We assume, with him, that the distribution of $y = \log c$ given s is $h(y|s) \sim \mathcal{N}(\mu, s^2)$, and choose $U(c) = \frac{c^{1-\eta}}{1-\eta}$. We will assume that $\eta < 1$ in what follows, so that utilities are always positive. This allows us to take $\Phi(u) = \frac{u^{1-\xi}}{1-\xi}$, which implies constant relative ambiguity aversion.

Now some preliminary calculations:

$$\mathsf{E}_{c|s}U'(c) = \mathsf{E}_{y|s}e^{-\eta y} = e^{-\eta\mu + \frac{1}{2}\eta^2 s^2}$$

$$(\text{A3}) \qquad \mathsf{E}_{c|s}U(c) = \frac{1}{1-\eta}\mathsf{E}_{y|s}e^{(1-\eta)y} = \frac{1}{1-\eta}e^{(1-\eta)\mu + \frac{1}{2}(1-\eta)^2 s^2}$$

Using the fact that $\Phi'(u) = u^{-\xi}$, $\Phi^{-1}(x) = [(1-\xi)x]^{\frac{1}{1-\xi}}$, the above formulas, and quite a bit of algebraic manipulation, we find that:

$$(\text{A4}) \qquad\qquad M' = \beta e^{-\eta\mu}\frac{\mathsf{E}_s\exp\left(-\frac{s^2}{2}[\xi(1-\eta)^2 - \eta^2]\right)}{\left[\mathsf{E}_s\exp\left(-\frac{s^2}{2}(\xi - 1)(1-\eta)^2\right)\right]^{-\frac{\xi}{1-\xi}}}.$$

Now in order to compute these expectations, we need to know the probability distribution for s. We use the same distribution as Weitzman (2009b), which is just the Bayesian update of the scale-invariant prior based on the data $\mathbf{y}_n$:

$$(\text{A5}) \qquad\qquad \begin{aligned} p(s|\mathbf{y}_n) &\propto \prod_{j=1}^n h(y_j|s)p_0(s) \\ &\propto \frac{1}{s^{k+n}}\exp\left(-\frac{1}{2s^2}z\right) \end{aligned}$$

where we define $z := \sum_{j=1}^{n}(y_j - \mu)^2$, which is clearly positive.

So, in order to compute M', and see if it diverges, we need to examine integrals of the form:

$$(A6) \qquad I = \int_0^\infty s^{-(k+n)} \exp\left(-\frac{1}{2}\left[\frac{z}{s^2} + \Gamma s^2 \right] \right) ds$$

where Γ is a constant that represents the coefficient of $-\frac{1}{2}s^2$ in the exponent of either the numerator or the denominator of (A4). The integral I converges (and is positive) whenever both z and Γ are greater than zero. Since $z > 0$ is automatic, we need $\Gamma > 0$ for both the numerator and denominator of (A4). Thus we require

$$(A7) \qquad \xi > \max\left\{ \left(\frac{\eta}{1-\eta} \right)^2, 1 \right\}.$$

When this condition is satisfied, M' is finite.

Some comments about this result. Notice that it is obviously not satisfied for $\xi = 0$, the case Weitzman examines. It is easy to see by inspection that (A4) does indeed diverge in this case. Also, it is interesting that the condition is of this form – it suggests that if the social planner is averse enough to ambiguity, relative to his degree of risk aversion, the dismal theorem does not apply. To get a feel for the numbers involved here, assume that $\xi = 10$, a roughly plausible estimate (Ju and Miao, 2009; Gollier and Gierlinger, 2008). This implies that we must have $\eta < \frac{\sqrt{10}}{1+\sqrt{10}} \approx 0.76$ in order to ensure convergence. This is an implausibly low value of η, so although this result is theoretically interesting, it is unlikely to have much practical significance.

Now consider the case $\eta > 1$. The appropriate constant relative ambiguity aversion function when utilities are negative (i.e. when $\eta > 1$) is:

$$(A8) \qquad \Phi(u) = -\frac{(-u)^{1+\xi}}{1+\xi}.$$

If we take this choice, and go through the same calculations as above, we can show that:

$$(A9) \qquad M' = \beta e^{-\eta\mu} \frac{\mathsf{E}_s \exp\left(\frac{s^2}{2}[\xi(1-\eta)^2 + \eta^2] \right)}{\left[\mathsf{E}_s \exp\left(\frac{s^2}{2}(1+\xi)(1-\eta)^2 \right) \right]^{\frac{\xi}{1+\xi}}}.$$

Thus Γ is negative for both the numerator and the denominator in this case, and they both diverge. We might hope to get some kind of interesting limiting behaviour when considering the ratio of these two divergent quantities, however

this can be ruled out. Notice that the coefficient of $\frac{1}{2}s^2$ in the numerator is larger than that in the denominator when $\eta^2 > (1-\eta)^2$, which is always satisfied when $\eta > 1$. Also, $\xi/(1+\xi)$, the exponent of the denominator, is less than one. So, in general the numerator goes to infinity faster than the denominator, implying that $M' \to \infty$.

Why is it that ambiguity aversion can render the stochastic discount factor finite when $\eta < 1$, but not when $\eta > 1$? Consider just the numerator of (A2), which may be written as

$$(A10) \qquad \mathsf{E}_s\left(\Phi'(\mathsf{E}_{c|s}U(c))\right)\mathsf{E}_s\left(\mathsf{E}_{c|s}U'(c)\right) + \mathrm{cov}_s\left\{\Phi'(\mathsf{E}_{c|s}U(c)), \mathsf{E}_{c|s}U'(c)\right\}$$

The first term in (A10) diverges to $+\infty$, since the factor $\mathsf{E}_s\left(\mathsf{E}_{c|s}U'(c)\right)$ is just the term that appears in the ambiguity neutral stochastic discount factor (4). Therefore the only hope we can have of rendering the expression (A10) finite is if the second covariance term is negative, and also diverges. Now since Φ' is decreasing, the covariance term is negative if and only if

$$(A11) \qquad \qquad \mathrm{cov}_s\left\{\mathsf{E}_{c|s}U(c), \mathsf{E}_{c|s}U'(c)\right\} > 0.$$

Note that this condition is necessary for ambiguity to have an effect on the dismal theorem for any utility function and probability distributions. Whether it is satisfied is in general an empirical question, which depends on the shape of the utility function, as well as the dependence of the distribution of consumption on s. For the functional forms in our model above, an examination of (A3) makes it is clear that:

$$(A12) \qquad \qquad \mathrm{cov}_s\left\{\mathsf{E}_{c|s}U(c), \mathsf{E}_{c|s}U'(c)\right\} \begin{cases} > 0 & \text{For } \eta < 1. \\ < 0 & \text{For } \eta > 1. \end{cases}$$

Thus we can only hope for ambiguity aversion to resolve the divergence in the stochastic discount factor when $\eta < 1$ in the case examined above.