Moral Uncertainty and Climate Change

Some climate economists, most notably Martin Weitzman, have argued for strong action against climate change on the basis that unmitigated climate change brings a non-negligible probability of a catastrophic outcome. Those who ignore these catastrophic tail risks and focus instead on the costs of the most likely levels of warming will tend to advocate for more moderate action on climate change. Many of those especially worried about the tail risks of climate change make some variant of the following claims:

> *Empirical Claim*: The probability of extreme global warming sufficient to destroy, or otherwise undermine the long-run potential of, human civilisation is >~1%.
> *Moral Claim*: Destroying, or otherwise undermining the long-run potential of, human civilisation would be extremely, perhaps infinitely, bad; the costs would swamp many times over all levels of warming that would not destroy, or undermine the long-run future of, human civilisation.

There are, then, two fat tails here: the probability density across levels of warming has a fat tail; and utility as a function of warming also has a kind of fat tail such that the disutility of warming increases very rapidly at the upper end of warming. These two fat tails compound such that climate change has very large expected costs, and these expected costs are dominated by the tail risk. This argument has, as we know, been the subject of significant discussion among climate economists and ethicists.

Another of Weitzman's arguments is arguably equally, if not more, practically important, but has received very little attention in the literature. Here it is in full:

> "Suppose, for the sake of argument, that a policy maker believes the probability is 50 percent that my fat-tailed specification is correct and 50 percent that the thin-tailed specification of someone else is correct. Then, other things being equal, rational policy should lean more in the direction of my fat-tailed conclusions than in the direction of someone else's thin-tailed conclusions because of the highly asymmetric consequences of fat tails versus thin tails. In this sense, whether it is fair or unfair, the playing field is not level between me and someone else. To further illustrate this point, suppose one person advises you that a fire insurance policy protecting your house against extreme losses is unnecessary because so few houses of your kind burn to the ground, while another person advises you that a complete fire insurance policy is necessary in your case. Other things being equal, should you flip a coin to decide

> what to do just because both advisers seem to be giving equally credible guidance?" (Weitzman, 'Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change', 2011, p. 291).

As I understand him, the fat tails that Weitzman is referring to here are of the empirical *and* moral kind. Thus, Weitzman's argument is about how it is rational for policymakers to act in the face of empirical and moral uncertainty about the costs of climate change: it's about what it's rational to do when we have some non-negligible credence in the Empirical Claim and the Moral Claim. (It is not completely clear whether this is a correct interpretation, or whether Weitzman instead made a more narrow argument solely about empirical uncertainty. Noting this, I am more interested in the substance of the argument itself than whether Weitzman endorsed it.)

Call this argument the *Metanormative Climate Change Argument* (MCC). A sketch of the MCC is as follows. Note that this is focused on climate change policies rather than only on the expected social costs of climate change.

> 1. The Empirical Claim and the Moral Claim entail that strong action on climate change has extremely high, perhaps infinite, expected utility, whereas rival reasonable theories entail its expected utility is bounded and much lower.
> 2. Rational policymakers ought to have non-negligible credence in the Empirical Claim and the Moral Claim.
> 3. Rational choice is a function of the degree of credence one has in different normative theories, and the stakes of the choice according to the theories in which one has credence.
> 4. Therefore, rational climate change policy should be guided by the Empirical Claim and the Moral Claim, rather than by reasonable rival theories.

The question of how it is rational to act in the face of moral uncertainty has only recently received significant attention. My aim here is assess the soundness of the MCC. I argue that the MCC is sound on some assumptions and in some conditions. However, the truth value of a number of these assumptions and conditions is unclear.

1. How should we respond to epistemic peers and superiors?

It seems reasonable that rational policymakers ought to treat some people who endorse the Empirical Claim and the Moral Claim as epistemic peers or epistemic superiors. Although the

Empirical Claim is the subject of significant disagreement, some climate scientists cited by the IPCC endorse it, as does Weitzman, an eminent expert (IPCC 2013, chap. 10). Some climate economists and ethicists also endorse axiologies that imply the Moral Claim. Weitzman does so on the basis of a utility function with constant relative risk aversion, which entails that marginal utility is extremely high at low levels of consumption. Total Utilitarianism and generally any moral theory that is to a significant extent aggregative across persons and temporally neutral also imply the Moral Claim (Bostrom 2003). It seems plausible that some of these people should count as epistemic peers for policymakers.

Further, it is plausibly rational to split one's credence in some way across propositions endorsed by one's epistemic peers and superiors. I'll assume that rational policymakers should give non-negligible credence to the Empirical Claim and to the Moral Claim. The question then is: how should rational policymakers act given that they have these credences?

2. How should we respond to moral uncertainty?

There is a longstanding debate in moral philosophy between objectivists and subjectivists about rightness. Subjectivists hold that, necessarily, a person has the moral obligations she has at the time solely in virtue of facts about her evidential or doxastic situation at the time. Objectivists deny this. For example, subjectivist utilitarians will say that agents ought to maximise expected utility, whereas objectivists will say that agents ought to maximise utility. Suppose that the paracetamol a doctor is giving to her patient is in fact cyanide, but that the doctor does not and could not have known this. The subjectivist will say that the doctor ought to give the patient the pill, whereas the objectivist will say that the doctor ought not to do so.

Here, there is a disagreement about whether empirical uncertainty bears on what agents ought in some sense to do. There is an analogous debate about whether *moral* uncertainty bears on what agents ought, in some sense, to do. Call those who believe that moral uncertainty bears on what agents ought, in some sense, to do, *metanormativists*. Call those who deny this *normativists*.

Some metanormativists take the metanormative ought to be a moral ought, whereas others take it to be a rational ought. On my interpretation Weitzman endorses the latter view, and I find it more plausible, so I will deal with that version in what follows.

There is disagreement among metanormativists about how we rationally ought to respond to moral uncertainty. One might hold a My Favourite Theory approach, which says that one should follow the view that one thinks most likely to be true. But consider this example:

Chicken or Impossible Burger?

David is at a restaurant deciding whether to order chicken breast or the plant-based Impossible Burger. He has 49% credence in an animal welfare theory which says that ordering the Impossible Burger is obligatory and that ordering the chicken is very badly wrong; and he has 51% credence in an anthropocentric theory which says that ordering either is permissible and that there is no difference in expected value between the two.

On the My Favourite Theory approach, David is permitted to order chicken or the Impossible Burger. However, this seems counterintuitive because it is insensitive to what is at stake according to the theories in which David has credence. According to the animal welfare theory, a lot is at stake in this decision, whereas according to the anthropocentric theory, nothing is at stake. Ordering chicken seems like a big risk to take, given David's partial beliefs. Indeed, this idea of differential stakes is what Weitzman appeals to in the argument quoted previously.

A potential solution to this is offered by a theory I, following Brian Hedden (2016), call Maximise Intertheoretic Expectation (MITE). MITE extends expected utility theory to account for partial belief in moral propositions, as well as empirical ones. When we take the expected moral value of an action on each moral theory and sum them up, weighted by the probability of each theory, we get the 'intertheoretic expectation'. MITE is sensitive to what is at stake in the decision according to the theories in which one has credence.

Suppose that we can represent your doxastic, or belief-like, state with a probability function P and that T_i denotes moral theory i. Suppose that each theory in which one has credence can be represented in Expected Moral Value (*EMV*) terms. Thus, each theory gives a verdict about the *EMV* of an action A. Therefore:

Intertheoretic Expectation

$$\mathrm{IE}(A) = \sum_i P(T_i) EMV_i(A)$$

It is highly controversial whether intertheoretic comparisons between *any* rival moral theories are possible, and a number of philosophers deny that they are. It is even more controversial whether *all* moral theories in which agents should have credence can be represented in expected moral value terms (MacAskill 2014). For the purposes of this paper, I assume that

intertheoretic comparisons are possible at least between some theories, and will focus on these theories in what follows.

We can apply MITE to the Chicken example. Suppose that option A is to order the chicken, and option B is to order the Impossible Burger. David's decision can be represented in the following table, using numbers to represent how good the outcomes would be, according to the theories at stake:

	Animal Welfarist	Anthropocentric
A	-25	5
B	5	5

Using these numbers, we can deduce the intertheoretic expectation of actions A and B:

$$IE(A) = (.25*-50) + (.51*5) = -9.95$$
$$IE(B) = (.49*5) + (.51*5) = 5$$

Since IE(B) > IE(A), David ought to choose option B and order the Impossible Burger. Thus, on MITE, in this case the animal welfarist theory is David's *effective moral theory*: it is the moral theory that he should rationally use to guide his decisions. In this choice, to borrow Weitzman's phrase, "the playing field is not level" between the two theories: because the choice is high stakes according to one of the theories, that theory triumphs even though we have less credence in it than its rival.

3. Applying MITE to climate change

We can now sketch how MITE would apply to climate policy, and set out the assumptions which ensure that the MCC goes through. Assume in what follows that rational policymakers should have >>1% credence in the Empirical Claim. Uncertainty about the Empirical Claim is accounted for by the expected moral value calculations of different moral theories, which will be fed into the intertheoretic expectation calculation.

Suppose that we have a choice between two carbon prices, *High* ($100 per tonne) and *Low* ($20 per tonne). These are a proxy for the strength of the portfolio of policies we are willing to take against climate change that might also include other policy instruments, such as adaptation, Carbon Dioxide Removal, and solar geoengineering. From my own assessment

of expert estimates of the probability of greenhouse gas emissions pathways, and Weitzman's estimates of climate sensitivity, the unconditional probability of climate change-caused existential catastrophe is ~3%. Suppose that *High* eliminates this risk without introducing any other existential risks, and that *Low* reduces it to 2.5%, without introducing any other existential risks.

For the MCC to work, the choice between *High* and *Low* has to have differential stakes from the point of view of the theories in which rational policymakers should have credence. Moral theories which put extremely large weight on reducing existential risk would view the choice between *High* and *Low* to have very high stakes. Some theories without this feature would hold that the choice has much lower stakes.

There are a number of theories which put extremely high weight on reducing existential risk. Total Utilitarianism and generally any theory that is to a significant extent aggregative across persons and temporally neutral about welfare has this implication. Weitzman's moral theory also has this implication for somewhat different reasons. Here I will explore the implications of Total Utilitarianism, as this is in my view the most plausible of the theories which this feature. Similar implications follow from a number of other theories, *mutatis mutandis*.

The total utility of a possible population is the product of the average welfare of the population and its size. Because there are so many possible good lives in the future, on Total Utilitarianism the far future has astronomical value. Consequently, the expected value of increasing the probability that the far future is realised is astronomically high (Bostrom 2003). The expected value of even minute reductions in existential risk exceeds what can plausibly be achieved by benefiting the current generation.

Greaves and Ord have shown that, with some qualifications, on MITE, provided that one has nonzero credence in Total Utilitarianism, Total Utilitarianism dominates the intertheoretic expectation calculation (Greaves and Ord, n.d.). In brief, the explanation for this is as follows. As an option involves more and more people, Total Utilitarianism ascribes a choice a higher relative weight, eventually coming to dominate the intertheoretic expectation, regardless of one's credence in Total Utilitarianism. In the climate change case we are considering, the 'large population limit' is that in which the size of the possible future population (that could be brought about by the high carbon price) tends to infinity. As the possible future population tends to infinity, $EMV_{Total}(High)$ - $EMV_{Total}(Low)$ tends to positive infinity. On leading theories, T_j, that value *Low* over *High*, $EMV_{Tj}(Low)$ - $EMV_{Tj}(High)$ approaches a finite bound. Thus, the ratio of the value difference of *High* over *Low* according

to the Total View to the value difference according to leading rival theories approaches infinity. Consequently, with respect to *High* vs. *Low*, Total Utilitarianism swamps the intertheoretic expectation in the large population limit. In this case, it is the effective moral theory. In sum, on the outlined assumptions, the MCC succeeds.

The MCC is of interest chiefly because it bears on the practical question of whether rational policymakers should take strong against climate change. However, the MCC may not be necessary to get to the conclusion that strong action would be rational. Suppose that rational policymakers should believe that the probability distribution across levels of warming has a fat tail. If so, it is plausible that many reasonable axiologies that do not put a special weight on avoiding existential catastrophe entail that the expected costs of climate change are very high, much higher than assumed by some climate economists. Thus, one arguably need not endorse the Moral Claim to be committed to strong action on climate change.

4. Criticisms of the MCC

There are a number of possible objections the MCC. I discuss a selection in what follows. These objections can be grouped into four categories.

4.1. First-order criticisms

Many of the most prominent criticisms of Weitzman's argument for strong action on climate change have operated on the first order: they have tried to show that the Empirical Claim is false and that the Moral Claim is false. These criticisms might indeed make it very probable that the expected disutility of climate change is much lower than argued by Weitzman. However, this does not necessarily show that the intertheoretic expected costs of climate change are much lower than argued by Weitzman. To show that, one would have to demonstrate that rational policymakers should have zero small credence in the Empirical Claim and the Moral Claim. Proving that is a much bigger task than proving that the Empirical Claim and the Moral Claim are unlikely to be true. In this way, the MCC puts an insurmountable burden of proof on first-order arguments against tail risk-based defences of strong action on climate change. This is an important result.

4.2. Criticisms assuming that MITE is the correct metanormative theory

The defence of the MCC outlined in section 3 made a number of important empirical assumptions, which are open to question. If these assumptions turn out to be false, then climate change may not have differential stakes, as required by the MCC.

Firstly, I assumed that strong action on climate change does not introduce any new existential risks. This is not obvious. Here are some examples. First, nuclear power is a very attractive tool for cutting greenhouse gas emissions, as the only low carbon source of baseline power generation. However, domestic nuclear power could also increase the risk of nuclear proliferation, in turn increasing the risk of an existential catastrophe-level nuclear war. Second, Weitzman and others argue that the tail risks of climate change are so severe that research on solar geoengineering is justified. Weitzman's co-author on *Climate Shock*, Gernot Wagner, is now co-director of a new research programme on solar geoengineering at Harvard. However, solar geoengineering introduces a number of new risks, which could arguably exceed the existential risks posed by climate change. For example, it could increase knowledge about a doomsday weapon (Morton 2015, 342–43); increase the risk of catastrophic termination shock (Baum, Maher, and Haqq-Misra 2013); or increase the risk of political conflict (Nightingale and Cairns 2015).

Secondly, as a number of commentators, including Nordhaus (2011) and Pindyck (2011), have pointed out, spending on climate change may not be the most cost-effective way to reduce existential risk. There are a number of other serious existential risks aside from climate change. Weitzman argues that climate change is in fact the most serious existential risk. However, from my own personal conversations with researchers at the Future of Humanity Institute, there appears to be a consensus that climate change presents significantly less severe existential risks than artificial intelligence, nuclear war, and developments in biotechnology. Aggressive greenhouse gas mitigation would cost between 3% and 11% of global GDP by 2100 (IPCC 2014, chap. 6). These trillions of dollars could be spent on reducing the risks of AI, nuclear war, and biotechnology, and arguably reduce existential risk to a much greater extent. If, as I've argued, existential risk reduction swamps the intertheoretic expectation, then MITE requires that mitigation of these risks should be prioritised over climate change. The extent to which they should be prioritised is a very difficult question, which I cannot answer in full here.

4.3. Criticisms of MITE

Weitzman argues that, on some ethical assumptions he takes to be plausible, the expected disutility of the destruction of human civilisation is infinite. This creates problems for MITE

because all actions that reduce the probability of existential catastrophe would have positive infinite, and therefore equal, expected utility, even if they reduce the probability of existential catastrophe by different amounts. If so, MITE requires us to be indifferent between reducing existential risk by 1% and reducing it by 50%, which is counterintuitive. Thus, MITE would need to be replaced with a decision theory that accounts for infinities. In short, MITE inherits the problems that expected utility theory has with infinite ethics (Bostrom 2011).

4.4. Criticisms of metanormativism

Perhaps the leading alternative in the literature to MITE is normativism. Normativism holds that there is no sense in which we ought to respond to our moral uncertainty; what we ought to do is determined by the true moral theory, regardless of whether we believe it or have evidence for it. (Note that this is different to the My Favourite Theory approach. On normativism, the true moral theory, not necessarily the theory in which we have highest credence, determines what we ought to do.) Normativism entails the unsoundness of the MCC. Even if it is reasonable to have some credence in the Moral Claim, if the true moral theory entails that a low carbon price ought to be implemented, then that is what we ought to do.

Normativism is quite attractive, and I will now attempt briefly to rebut some criticisms of it. Perhaps the most common criticism is that it is not sufficiently action-guiding. Many philosophers endorse an access principle about moral obligations:

> *Access:* If one is rationally required to *x*, then one can know or have sufficient evidence for the fact that one is rationally required to *x*.

It is not obvious that *Access* is a genuine requirement on normativity, but assuming that it is, it is unclear whether normativism in fact violates it. As Brian Hedden (2016) has argued, it is plausible that all moral truths are accessible *a priori*: all the evidence for the true moral theory is in principle available. The fact that we are suboptimal reasoners who cannot eliminate our moral uncertainty does not change this fact. Thus, normativism arguably satisfies an evidentialist version of *Access*.

One possible response to this is to relativise rationality to the capacities of the reasoning agent. Moral philosophy is extremely difficult. For ordinary suboptimal reasoners, moral uncertainty is the appropriate doxastic state. If so, there is a sense in which normativism is not adequately action-guiding for suboptimal reasoners. One problem with

this is that suboptimal reasoners can come to have credence in abhorrent or crazy moral theories. However, it is counterintuitive that abhorrent or crazy moral theories can determine an agent's rational obligations.

Finally, it is unclear whether metanormativism is more action-guiding than normativism. It is extremely difficult not only to find out what the correct moral theory is, but also to find out what the correct metanormative theory is. MITE requires agents to maximise intertheoretic expectation even if they do not believe in MITE or even if, given their rational capacities, it would be irrational to give particularly high credence to MITE. Thus, the action-guidingness argument for metanormativism is potentially self-refuting.

Conclusion

The MCC is a neglected argument. If it is sound, it shows that some of the arguments made by opponents of tail risk-based arguments for strong action on climate change face an insurmountable burden of proof. However, it remains very unclear whether the MCC succeeds because many of its premises are questionable. It is unclear whether metanormativism is true, unclear whether MITE is true, and unclear whether the requisite empirical assumptions about climate policy hold.

References

Baum, Seth D., Timothy M. Maher, and Jacob Haqq-Misra. 2013. "Double Catastrophe: Intermittent Stratospheric Geoengineering Induced by Societal Collapse." *Environment Systems & Decisions* 33 (1): 168–80. doi:10.1007/s10669-012-9429-y.

Bostrom, Nick. 2003. "Astronomical Waste: The Opportunity Cost of Delayed Technological Development." *Utilitas* 15 (3): 308–314. doi:10.1017/S0953820800004076.

———. 2011. "Infinite Ethics." *Analysis and Metaphysics*, no. 10: 9–59.

Greaves, Hilary, and Toby Ord. n.d. "Moral Uncertainty about Population Axiology." http://users.ox.ac.uk/~mert2255/papers/mu-about-pe.pdf.

Hedden, Brian. 2016. "Does MITE Make Right? On Decision-Making Under Normative Uncertainty." In *Oxford Studies in Metaphysics*, edited by Russ Shafer-Landau. Vol. 11.

IPCC. 2013. *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by T. F. Stocker, D Qin, Gian-Kasper Plattner, M Tignor, S. K. Allen, J Boschung, A Nauels, Y Xia, V Bex, and P Midgley. Cambridge University Press.

———. 2014. *Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

MacAskill, William. 2014. "Normative Uncertainty."

Morton, Oliver. 2015. *The Planet Remade : How Geoengineering Could Change the World*. London: Granta.

Nightingale, Paul, and Rose Cairns. 2015. "The Security Implications of Geoengineering: Blame, Imposed Agreement and the Security of Critical Infrastructure." Climate Geoengineering Governance Working Paper Series. Tech. Rep., School of Business, Management and Economics, Univ. of Sussex.

Nordhaus, William. 2011. “The Economics of Tail Events with an Application to Climate Change.” *Review of Environmental Economics and Policy* 5 (2): 240–57. doi:10.1093/reep/rer004.
Pindyck, Robert S. 2011. “Fat Tails, Thin Tails, and Climate Change Policy.” *Review of Environmental Economics and Policy* 5 (2): 258–74. doi:10.1093/reep/rer005.