In a 2009 review of 22 peer-reviewed climate change studies, Martin Weitzman warned of a small but significant chance of ten or even twenty degrees of global warming resulting in “disastrous collapse of planetary welfare.” Such an outcome could have essentially infinite disutility—we might be prepared to give up anything to avoid it. If so, even a small chance of catastrophe could count for more in a cost-benefit analysis than the more likely scenarios.

Weitzman suggested that this reasoning, expressed formally in a “dismal theorem,” could yield “a valid economic-statistical precautionary principle which, at least theoretically, might dominate decision-making.”¹

Weitzman’s paper prompted an exchange with William Nordhaus in which the latter argued that for such a catastrophe to materialize, the climate would have to prove improbably sensitive to greenhouse gas emissions, the damages from climate change would have to be improbably high, and scientists and governments would have to remain improbably complacent in the face of danger signs. Weitzman’s guesstimates were thus exaggerated.²

The latter’s reply was noteworthy. Suppose, he wrote, that a policymaker believed that Weitzman’s probability distribution and another economist’s less alarming ones were equally likely to be right. The policy maker should give Weitzman’s conclusions greater weight:

[S]uppose 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?³

Whether or not he realized it, Weitzman was advocating Pascalian wagering. In the seventeenth century, Blaise Pascal argued that believing in God had infinite value, since if God existed, the reward would be eternal bliss. Any costs associated with believing, on the other hand—such as getting up early on Sunday mornings—were minor in comparison. If in doubt, then, in our own self-interest, we should try to believe. If at first we were unable to do so, we could at least improve our chances of believing by observing religious rituals. Similarly, Weitzman held that an infinitely bad disaster might justifiably dominate decision making even if its probability was small. Confronted, moreover, with a competing theory according to which the probability was negligible, it could be rational to act as if the more pessimistic account were right, even if one considered both equally plausible. As Pascal recognized, the value of acting on a theory depends not only on the likelihood that it is true, but the benefits if it turns out to be right, and the costs if it turns out to be wrong.⁴

Pascalian wagering also faces well-known objections. Critics argue that its implications are indeterminate on the one hand, or paralyzing on the other.⁵ Analogous criticisms are levied against the precautionary principle.⁶ Nordhaus warned in his critique of Weitzman that ‘if we were to accept the dismal theorem, we would likely drown in a sea of anxiety at the

⁴ S. F. Haller, Apocalypse Soon? Wagering on Warnings of Global Catastrophe.
⁵ See A Hájek, ‘Pascal’s Wager’ for a review.
prospect of the infinity of infinitely bad outcomes’. Yet in contrast to Pascal’s Wager, the loss from the devastation of the planet or human extinction would be enormous—on a scale far beyond all disasters in recorded history—but finite. That makes precautionary reasoning possible, and provides a new justification for Hans Jonas’s imperative of responsibility.

Infinite gain, infinite loss

A seeming advantage of Pascalian reasoning is that we need not know probabilities, merely possibilities. The expected value of a choice is the product of a cost or benefit and its probability of materializing. Infinity times any number is still infinity. Pascal’s God need not have a large chance of existing to make it worthwhile to try to believe in Him; indeed, provided there is any chance that He exists, the expected value of seeking to believe is infinite. Unfortunately for Pascal, that is just as true of anything else we might do. ‘Suppose that you choose to ignore the Wager’, observes Alan Hájek, ‘and to go and have a hamburger instead. Still, you may well assign positive and finite probability to your winding up wagering for God nonetheless; and this probability multiplied by infinity again gives infinity’. Even if you decide to disbelieve in God, there is some chance that the effort will lead you to end up believing anyway. However small that chance, trying to disbelieve also has infinite expected value.\footnote{A. Duff, ‘Pascal’s Wager and Infinite Utilities’; A. Hájek, ‘Pascal’s Wager’}

A similar problem arises if one seeks to rule out the worst-case scenario. For Pascal, this is the risk of eternal damnation. But a refusal to consider probabilities renders the wager vulnerable to the “many gods” objection. It seems reasonable to assign some probability to the possibility that that the true God is Odin, or Cthulhu, or even Barney the pink purple dinosaur. It certainly seems unlikely, but not inconceivable. Some of these gods might be jealous and vindictive, and reserve a special section in hell for believers in Jehovah. If the

\footnote{W. D. Nordhaus, ‘The Economics of Tail Events with an Application to Climate Change’, 254.}
value of heaven is infinitely positive, the value of hell may well be infinitely negative, and so wagering on Jehovah will have both infinite positive and negative value.\(^9\)

Some writers have interpreted Weitzman’s claim that runaway climate change might entail “unlimited downside exposure” as a claim that its costs could be infinite.\(^10\) Just as with the Wager, if the cost of the devastation of the planet or human extinction were literally infinite, the dismal theorem would offer no practical guidance. Suppose, for example, that the United States government decides to tear up its commitments to emissions cuts, slash support for alternative energy and burn as many fossil fuels as possible. This might set off runaway climate change. It might also prompt a political reaction leading to the government’s fall and the election of the first Green government in the nation’s history. Or it might be that benevolent space aliens, who had hitherto refrained from intervening to correct humanity’s folly, would be provoked into stepping in and setting things to right. Neither of the latter scenarios is probable, but both are possible. By the same token, while maintaining high emissions may bring about catastrophe, so could emissions cuts. To take Neil Manson’s example, they could conceivably cause a worldwide depression, the rise of militarist regimes and global thermonuclear war.\(^11\) That may be unlikely, but the expected value of an outcome is its probability multiplied by its value. Either maintaining emissions or cutting them would have infinite expected utility. They would also have infinite expected disutility.\(^12\)

In fact, Weitzman does not seem to have meant that the costs of a catastrophe could be literally infinite. Rather, he was claiming that it could be so large that low-probability risks could justifiably dominate our decision-making process.\(^13\) This is a thoroughly plausible

---

\(^9\) Hájek, ‘Pascal’s Wager’.


\(^12\) This problem is familiar to Pascal scholars as the ‘many gods’ problem; see Hájek, ‘Pascal’s Wager’.

claim. While we can permissibly discount costs and benefits that would leave future people better off than we are, we should at least give full weight to those that would leave them worse off. Either runaway climate change or thermonuclear war could consign a vast number of future people to miserable conditions. Even a small chance of permanent catastrophe thus has enormous expected disvalue. Yet provided the amount of utility at stake is not infinite, probabilities still matter. Either cutting emissions or increasing them might bring about permanent impoverishment, but the latter is much more likely to do so. The dismal theorem will encourage strong action to prevent it.

Betting on theories

Policy debates often implicitly follow the ‘maxiprobability method’ or the ‘my-favorite-theory approach’. The assumption is that we should try to figure out which theory is most likely to be true, and pursue the policies that would be optimal if it is right. This may seem like common sense, but it can be seriously misleading. Imagine that you are a U.S. senator in the run-up to the war of 2003. Assume that your sole concerns are that Saddam will use WMD and the cost in blood and treasure of overthrowing him, and that your goal is to adopt the best policy under the circumstances. One of your advisors argues that Saddam has WMD, and he is likely to use them; a second that he has WMD, but that he can reliably be deterred; a third, that he does not have WMD, and that he can be prevented from acquiring them. You find the first advisor’s argument most persuasive. If you were to say how likely she was to be right, you would assign a subjective probability of 0.4 to the truth of her claim, and of 0.3 to each of the others. If you act on what you consider the most plausible claim, then you will vote for war, because if Saddam is armed and dangerous, that will be optimal. This would be the wrong conclusion. You should vote against war, since you believe that

---

14 See my paper ‘Discounting and the Fallacy of Division’.
16 S. O. Hansson and M. Johanneson, ‘Decision-Theoretic Approaches to Global Climate Change’ 164; T. Lockhart, Moral Uncertainty and Its Consequences, 5-6, 42.
there is a 60% chance that one of the other claims is correct, and in that case it will be best to remain at peace. Sometimes when more than one less plausible theories point to the same alternative, we will maximize our chance of doing the best thing by choosing it instead.\(^\text{17}\)

In other cases, it is unwise to try at all to bring about the best outcome. Suppose a doctor, Jane, can treat a patient suffering moderate acne with one of two drugs.\(^\text{18}\) Jane is confident that A will relieve, but not cure the acne; she believes that drug B has a 90% chance of clearing it up, but a 10% chance of killing the patient. Only drug B has any chance of bringing about the best outcome, and if she prescribes it, it probably will bring about the best outcome. Nevertheless, Jane should not do so. The reason, of course, is that prescribing drug B also entails the chance of a disastrous error.\(^\text{19}\) Despite Jane’s belief that drug B is probably the best choice, she appropriately chooses drug A.

We should not ignore probabilities. Still, Pascal’s insight was profound that decision-making is less about determining what is most likely to be true than what we have the most reason to do. It is easy to confuse the two in the case of the Wager, since the action Pascal recommends is to believe in God. Yet even here the two could come apart. If the logical case and empirical evidence for atheism is much stronger than that for theism, then we might have overriding epistemic reason to disbelieve. Yet in light of what is at stake, we could nevertheless have overriding practical reason to believe in God. Even if the evidence suggested Christianity was probably not true, it could be rational all things considered to try to believe in it.\(^\text{20}\) If, for example, by taking a drug that impaired our reasoning, we could cause ourselves to believe, we would no longer believe what the balance of the evidence supported. Nevertheless, it could be quite rational to take the drug.\(^\text{21}\)

\(^{17}\) Lockhart, *Moral Uncertainty*, p. 43, 144.

\(^{18}\) Here I combine and adapt two examples from Frank Jackson’s ‘Decision-Theoretic Consequentialism and the Nearest and Dearest Objection’.


\(^{20}\) G. Mougin and E. Sober, ‘Betting Against Pascal’s Wager’.

In most cases, what we ought to believe and what we ought to do are more easily distinguished. We all recognize this when making decisions under risk. Nobody believes her house will burn down next year. Yet most of us consider it sensible to take out fire insurance. Though we justifiably believe that our money will be wasted, we have good practical reason to hedge against our belief being wrong. When we do so, it is typically for fear of being mistaken about our specific circumstances (e.g., that our house’s wiring is reliable, or that the matches are safely hidden from our toddler). What is less commonly recognized is the risk of error in our beliefs about the world (for example, that water is a good way to douse burning cooking oil). Even when the balance of available evidence supports them, they may still be wrong. While we speak of ourselves as believing in theories in the natural and social sciences, we seldom can be absolutely confident they are right. It would be more accurate to speak of ‘degrees of belief’ or ‘degrees of acceptance’.

The expected value of acting on a hypothesis depends not only on the payoff if it is correct, but also on the costs if it proves to be wrong. If those costs are high enough, it can be rational to act as if one believed a different theory to be true, even if one suspects that it is not. Instead of maximizing the probability of the best outcome, we ought to choose policies with the highest expected value. This is the product of the value of the possible outcomes and the probability that each will ensue. Such probabilities can be hard to assess, but a rational agent must work with the best estimates she has. (In some cases seeking further information will be the action with the greatest expected value.) We then assign subjective probabilities to the likelihood that competing theories are true, and then use these probabilities to weight the expected value of the actions’ outcomes.

---

23 J. Ross, ‘Rejecting Ethical Deflationism’, p. 743.
This has become a popular approach among moral philosophers examining how agents should decide when they are unsure of which moral theory to follow. The core insight is we should not always do what is prescribed by the theory we think most likely to be right. We must also consider how good or bad the action would be if competing theories, including those we find less credible, should be correct instead. Our goal, as one writer puts it, should be to “maximize expected moral rightness.” While this concerns moral uncertainty, it applies whenever we are unsure whether relevant facts or theories are true.

Suppose in Weitzman’s example the first advisor tells you your house has only one chance in a thousand of burning down, whereas the second advisor advises you that it is one in 1,000. Were the house to burn down, you would value the extra compensation provided by the generous policy at $500,000, and the cost of the catastrophe insurance is $200/year. Suppose that you estimate a 60% chance that the first advisor is right, and a 40% chance that the second advisor is. Incorporating these probabilities, the expected value of insuring your house is $229.91. You should insure. We should also take out insurance against extreme climate change.

The imperative of responsibility

The foregoing arguments provide a new justification for Hans Jonas’s imperative of responsibility. Technological development, Jonas noted, has brought mankind to the point where we can inflict apocalyptic levels of damage on ourselves and on the future. When we contemplate gambles that could lead to humanity’s destruction the loss is potentially infinite, whereas the potential winnings are generally finite. In such cases, we should give ‘priority to well-grounded possibilities of disaster (different from mere fearful fantasies) over hopes.

26 Lockhart, Moral Uncertainty, p. 144.
27 Ross, ‘Rejecting Ethical Deflationism’, 744, 755-58. Indeed, the case for doing so seems more compelling when applied to positive theories. Critics charge that we cannot compare degrees of rightness and wrongness across different moral theories, but such objections do not seem applicable when the uncertainty is about which positive theories about the world are correct. For an overview of the debate, see T. Żuradzki, ‘Meta-Reasoning in Making Moral Decisions under Normative Uncertainty’.
even if no less well grounded.’ Faced with apocalyptic threats, even if the reasons for hope are better than those for fear, we ought to be guided the latter in our decision making, because the consequences will be dire and irreversible should they materialize. Just as with Pascal, Jonas’s claim is not about what we have epistemic reason to believe, but what we have practical reason to do, and he too is recommending a bet.29

The imperative of responsibility is usually understood as a rival to utilitarian cost-benefit analysis.30 Yet it can be justified as a consequentialist decision-making procedure.31 Any consequentialist theory that does not discount harms that could permanently impoverish the future will accord great weight to averting such threats. If the survival of humanity matters, then this gives us still greater reason for precaution.32 Cost-benefit analysis that gave proper weight to these considerations would militate against running apocalyptic risks. In fact, it would be neither practical nor necessary to perform a cost-benefit calculation in each case.33 To prevent such gambles from being taken lightly, they should simply be ruled out.34 A familiar theme in consequentialist ethics is that we will often achieve the best results not by directly seeking to maximize expected value, but rather by following decision rules that do so.35 The imperative of responsibility should be regarded as such a rule. Jonas’s own language suggests as much: ‘The magnitude of these stakes, taken together with the insufficiency of our predictive knowledge, leads to the pragmatic rule to give the prophecy of doom priority over the prophecy of bliss’.36

29 H. Jonas, The Imperative of Responsibility: In Search of an Ethics for the Technological Age, x, 31-38; quoted passage at p. 32.
33 Werner, ‘Risiko und Rationalität’, 286.
34 Cf. D. Birnbacher, Verantwortung für zukünftige Generationen, 204.
35 R. E. Bales, ‘Act-Utilitarianism: Account of Right-making Characteristics or Decision-making Procedure?’
Jonas, however, feared not only the physical destruction of humanity but also losing ‘the essence of man’. His imperative also forbids taking any risks with the latter. Depending on how we interpret ‘the essence of man’, this might give it a much broader scope. Indeed, Jonas maintained that there were many “apocalyptic” threats, whether nuclear weapons or nuclear reactors. Interpreted that way, the imperative may seem to threaten the same paralysis as does Pascal’s Wager. Nordhaus warns that if we attached infinite disutility to existential risks, it could be rational to fight preventive wars to prevent barely discerned power shifts in the far future. If the imperative of responsibility is to avoid paralysis and paranoia, its scope must be restricted to a small number of truly apocalyptic risks.

While it is true that nearly anything we do could in principle prove catastrophic, in most cases we have no more reason to believe that acting is riskier than abstaining. It is often more dangerous, for example, to launch preventive wars than to forego them. With climate change, on the other hand, we have a well-theorized causal mechanism, with good theoretical and empirical reasons for believing that higher greenhouse gas concentrations are more dangerous than lower ones. There are a small number of other apocalyptic threats that fall into this category—notably thermonuclear war—but for now there are not scores or hundreds.

Still, there might be many actions for which we can construct a theoretically possible, albeit improbable, story of how they could lead to doom, and that seem slightly more likely than not to bring on the apocalypse. To adapt an example of Jon Elster’s, it is conceivable that commercial TV transmissions will attract visitors from outer space. While such

37 Jonas, Imperative of Responsibility, 37.
38 D. Bourg and J.-L. Schlegel, Purar aux risques de demain: le principe de précaution, 168.
40 Godard, ‘Précaution’, 102.
41 Nordhaus, ‘Economics of Tail Events’, 254.
extraterrestrials might either save the earth from some other threat or destroy it, it is unlikely that these probabilities exactly cancel each other out. In the absence of reliable information, it is rational to make use of whatever information we do have—even if it is merely a flimsy analogy or a vague hunch. Suppose we judge—very tentatively—that the chance that alien visitors will destroy the earth is slightly greater than that they will destroy it. Should this be a decisive consideration in whether or not to broadcast *The Simpsons*?\(^{45}\)

Intuitively, the question seems absurd, but we should not reject it out of hand. In other contexts when policies appear to carry a small chance of apocalyptic destruction they should be taken seriously, and sometimes are. It was reasonable for Compton to take into account the chance that the atomic bomb would ignite the earth’s atmosphere even if it was only on the order of three in a million. It is reasonable today to consider the possibility that particle accelerators might destroy the earth in deciding whether to run them.\(^{46}\) The TV transmission question may seem absurd because we are by now accustomed to television broadcasts, and they have become an integral part of modern societies. That does not show that they are safe. It may instead indicate that the industrial revolution has led to the proliferation of existential risks. The *Stern Review* estimated a ten percent chance of human extinction per century. Surely no British government before the twentieth century, had it considered the matter at all, would have assessed the risk as nearly this high.\(^{47}\)

Even if the number of apocalyptic threats to which we must respond is presently circumscribed, with technological development it is likely to grow. In that case the imperative of responsibility could become far more confining.\(^{48}\) Yet in a nightmare world where existential risks had become common, what could we reasonably do but exercise

---


\(^{47}\) For discussion, see D. Jamieson, *Reason in a Dark Time: Why the Struggle Against Climate Change Failed--and What It Means for Our Future*, 122.

\(^{48}\) Cf. Godard, ‘Précaution’, 92.
extreme caution? Perhaps the problem is not that the imperative of responsibility is too restrictive, but that technological development is too dangerous. We should not take for granted that industrialization’s story has a happy ending.\textsuperscript{49} If century after century we continue to run even small existential risks, sooner or later, we are bound to lose the gamble.\textsuperscript{50}

REFERENCES


Gronke, H. ‘Epoché der Utopie. Verteidigung des ”Prinzips Verantwortung” gegen seine

\textsuperscript{49} See D. Lyons, ‘Are Luddites Confused?’

\textsuperscript{50} Bostrom, ‘Existential Risk Prevention as Global Priority’, 25.


Nordhaus, W., A Question of Balance: Weighing the Options on Global Warming Policies


