Ethan Osterman
Longtermism
Dr. Knutzen

Structural Risk: Decision Theory in the Long Run

I

The evaluative standards of ethical longtermism are tightly bound up in the formalism of the expected value calculation. On such a view, in order to determine whether one ought to pursue a given strategy, they need only to take the weighted sum of the possible outcomes multiplied by their respective probabilities. This paper attempts to show the inadequacy of traditional expected value calculations for the longtermist's purposes, using environmental policy as a case study and an exemplar.

In section II, I outline two broad outlooks (loosely characterized as "accelerationist", A, and "decelerationist", D) on environmental policy, for which I provide rough-and-ready characterizations in terms of expected upside and downside, and the probability of realizing that prospective gain. In section III, I review three decision-theoretical strategies for evaluating investments: first, the naïve expected value calculation; second, the risk-weighted expected value calculation; and third, the Kelly-style calculation of expected growth rate. I develop these strategies progressively, showing how each supersedes its predecessor, at least with respect to investments in the long run. In section IV, I apply the final (Kelly-style) strategy to our prospective "investments" in A and D; I argue that, given some reasonable assumptions about their respective profiles, D emerges as the favored policy, contrary both to received wisdom and to the typical conclusions of longtermism. Finally, section V acts as a conclusion.

II

Longtermism is typically optimistic about our prospects in the face of environmental degradation. Particular focus is placed on decarbonization and other "innovation in clean energy" which furthers the longtermist project of value maximization (MacAskill 2022, 33). Although we do not yet see the benefits of these technologies, it is assumed that they will come into effect some time in the near (ish) future, thus solving the ecological problem without necessitating any major restructuring or sacrifice. Bostrom captures this sentiment nicely when he writes that

> w*hen [a] rocket is in midair, it is in an unsustainable, transitory state: Its engines are blazing and it will soon run out of fuel. Returning the rocket to a sustainable state is desirable, but this does not mean that any way to render its state more sustainable is desirable. For example, reducing its energy consumption so that it just barely manages to hold stationary might make its state more sustainable in the sense that it can remain in one place for longer; however, when its fuel runs out the rocket will crash to the ground. The best policy for a rocket in midair is, rather, to maintain enough thrust to escape Earth's gravitational field: a strategy that involves entering a less sustainable state (consuming fuel faster) in order to later achieve the most desirable sustainable state* (Bostrom 2013, 25-26).

In other words: the longtermists, traditionally understood, are bullish on renewable energy, and on the possibility of decoupling economic growth from environmental impact. And as a result, they are cautious about what they take to be *excessive* precaution. Such a view is mirrored in the climate change literature by authors who argue, for example, that "stakeholders care not only about the environmental costs of inaction, but also about the economic costs associated with (perhaps unnecessarily) aggressive actions on climate change" (Otto et al. 2015, 918).[1] Concerns of this kind—about a possible *overreaction*—ultimately collapse back into the idea of value-maximization: for it is thought that, by overreacting, we may preclude the opportunity for some highly-valuable end-state.

We will characterize this cluster of views as the *accelerationist* (or: maximizing) strategy: for the idea is, above all, to "keep things moving". Whatever problems there may be, the accelerationist supposes that they will be best solved by further technological and industrial developments; we should not be *too* precautious, since this would incur apparently unnecessary costs.

Views of this kind are, of course, not without opposition. One might wonder why, for example, in Bostrom's analogy, returning to earth is not an option. More importantly, we might argue that the accelerationist view misunderstands the nature of environmental problems, and that, as a result, it is insufficiently precautionary. Rupert Read, for example, emphasizes that "ruinous threats" are *unbounded*; there is "no known (or even knowable) upper bound" to the threat (Read 2017, 133). As a result, in such situations, a precautionary approach is necessitated. One interesting consequence of this view is that the absence of "full scientific confidence" implies that the potential downsides are unbounded, at the very least with respect to epistemic access.[2] But if they are unbounded, then we ought to adopt a precautionary approach, rather than trying to optimize against a potential "overreaction" whose costs, in the grand scheme of things, can only be so high; concerns about spending too much on mitigation, on Read's precautionary approach, simply misunderstand what mitigation is.

The connection (and opposition) to longtermism in particular is not hard to find: in a popular article written with Emille Torres, Read writes of MacAskill, for example, that he is "growth-manic at the very moment when the world is waking up to limits" (Read and Torres 2023). And in this way, the precautionary approach championed by Read is clearly connected to the burgeoning "degrowth" movement, whose central thesis is (quite aptly) that growth (understood as increasing material throughput) cannot be sustained indefinitely, that we have already passed some relevant boundaries, and that these facts combine to necessitate some form of contraction (see, e.g., Meadows et al. 1972; Kallis 2018; Schmelzer et al. 2022).

Contrary to the accelerationists, it should be obvious that the degrowth perspective is highly skeptical of the possibility of decoupling—and all the more so when we take the broader ecological situation into perspective. Rockström et al (2009), for example, discuss *nine* key "planetary boundaries" which we run the risk of overshooting: the "decoupling" which the accelerationist strategy depends upon would then need to be effective along *all nine* of these

[1] Two of the authors of the above paper—Frame and Allen 2008—also contributed to Bostrom's anthology on Global Catastrophic Risks, so the connection to longtermism is not entirely arbitrary.
[2] MacAskill actually agrees on this point. The key difference is not in his understanding of uncertainty, but in his estimation of the actual situation, and what sorts of solutions are feasible. See, e.g., (MacAskill 2022, 49).

dimensions, rather than just the dimension of carbon emissions. And this is clearly a more difficult challenge.

In light of this brief discussion, we will characterize this cluster of views as the *decelerationist* (or: precautionary) strategy: for here it is assumed that continued technological and industrial developments are closely linked to environmental degradation. As a result, what is needed is to decrease our material throughput, to attain a "steady-state economy", or something of that nature.

With the accelerationist and decelerationist strategies roughly articulated, we can give a rough characterization of their relative "profiles", in terms of relative upside, downside, and probability of success of any given action which comports with the respective strategy. First of all, the accelerationist strategy is clearly characterized by a larger upside (an all-things-considered more valuable final state), but courts greater downsides along the way; rather than mitigating risk as it arises, the accelerationist allows it to grow on the assumption that it will be able to fix it somewhere down the line.

The decelerationist, then, has a relatively smaller upside; they sacrifice the best possible outcome in order to mitigate risks as they arise. However, due to the relative modesty of their goals ("less material throughput"), we might expect decelerationism to have a higher probability of success for any given action. I.e., the accelerationist might invest in some technology which ends up not working, whereas all that the decelerationist has to do is to lower the material throughput. This is not to say that the decelerationist's task is *easy*, just that it engenders less uncertainty. And in any case this is only a provisional characterization; we will return to these nuances later. For now, the provisional relationships we've identified are summarized in Figure 1 below.

	Relative Upside	Relative Downside	Relative Probability
A	Higher	Higher	Lower
D	Lower	Lower	Higher

Figure 1: relative investment profiles of the accelerationist (A) and decelerationist (D) strategies.

Rather than attempting to argue for one or the other position on its own terms, our goal here is simply to understand how we might evaluate these strategies from a decision-theoretical perspective. Specifically, the aim will be to show how longtermist appeals to expected value are too coarse-grained to effectively evaluate the two strategies on offer, and that more robust methods are needed to capture the structure of the decision. It is to this task that we turn next.

III

As we have already indicated, the natural way of determining the relative desirability of the strategies above, at least in the longtermist literature, is to take recourse to an expected value calculation. The mathematics here is quite simple: we take the weighted sum of the n possible outcomes multiplied by their respective probabilities, as in:

$$E[X] = \sum_{1}^{n} x_n \cdot p(x_n) \tag{3.1}$$

In the table of strategies above (in section II) we have not given any specific numerical characterization of strategies A and D, and so at first glance expected value (expectation) leaves us more or less in the dark. This is not in-and-of-itself sufficient to make such an evaluative calculus inadequate, however. The longtermists traditionally assume that, given the apparently astronomical amount of unrealized value waiting for us in the future, the maximizing strategy—in this case, A—will dominate. In other words: given some apparently reasonable assumptions about the strategies under consideration—p(A) can only be so much lower than p(D), whereas A may be many orders of magnitude greater than D could ever be—we have the inequality

$$E[A] > E[D] \tag{3.2}$$

which does not depend on any *specific* values, but instead depends upon whether or not the inequality

$$\frac{A}{D} > \frac{p(D)}{p(A)} \tag{3.3}$$

Holds—which, by stipulation, it does.[3]

Here we ought to mention a brief caveat: for downsides A* and D* < 0, if the magnitude of A* were sufficiently large, and the magnitude of D* sufficiently small, (3.2) might not hold, even if (3.3) did. On the longtermist reading, however, this cannot be possible: for D, having abandoned the project of economic growth and development, results in what Bostrom calls *permanent stagnation*. Such a state represents "a prima facie enormous loss, because the capabilities of a technologically mature civilization"—which strategy D has preempted—"could be used to produce outcomes that would plausibly be of great value, such as astronomical numbers of extremely long and fulfilling lives" (Bostrom 2013, 20). It is not hard to see that these downsides would be incurred *even if* D itself were attained, and D (to say nothing of D*) is thus, for the longtermist, hardly meaningfully better than A*: a stunted, terrestrial humanity isn't much better, in terms of all-things-considered value, than complete extinction. To pass up the opportunity for the unrealized (but purportedly realizable) gains which strategy A represents is neither precautionary nor prudential but is instead an act of extreme profligacy.

[3] It is of course *mathematically* possible for p(A) to be many orders of magnitude smaller than p(D), such that neither (3.2) nor (3.3) holds. That this case does not apply is granted, again, by stipulation.

Something like this, then, is the canonical longtermist position, and the mathematics underlying it. But even putting aside any doubts we may have about the specifics of the argument—is it really the case, for example, that D and A* aren't much different? etc.—the expected value formalism is not without its problems. For consider the following problem: in a gambling game (the "St. Petersburg game"), you may flip a coin until it lands on tails; for each consecutive heads, you double your money (see, e.g., Jaynes 2003, 399). The expected value is represented by the following equation:

$$E = \sum_{n=1}^{\infty} \frac{1}{2^n} 2^n = \sum_{n=1}^{\infty} 1 = \infty \qquad (3.4)$$

Thus, according to the expected value formalism, a bet of this kind has an *infinite* expected value; there is no amount of money you should *not* be willing to pay to have an opportunity to play the game. And this in spite of the fact that, barring an extremely low cost of entry, such a bet is plainly not worth taking.

The case is instructive for two reasons. First, it bears some passing resemblance to the longtermist gamble outlined above: in both cases, the purportedly optimal result is produced by an *astronomical* upside which is capable of swamping even infinitesimally small probabilities. Second, it shows that the problem is not merely with some particular *case*—say, the longtermist's views on environmental policy—but a deeper problem with the equations being used: they are too coarse-grained to reflect many of our actual evaluations. This is not to say that A is necessarily the worse option, but that we ought not blindly trust an expected value calculation where astronomical upsides are involved; it may very well be missing something important.

For this reason, it will prove necessary to expand our evaluative calculus. And one intuitive place to begin would be by accounting for the *riskiness* of the bet (see Buchak 2013, Pettigrew 2024). The essential change here is that, rather than weighing the outcomes by the probabilities themselves, we modify the weights with some risk function, $R(p)$, which obeys the following constraints (Pettigrew 2024, 136):

$(i)\ R(0) = 0,\ R(1) = 1$
$(ii)\ R\ is\ strictly\ increasing\ such\ that\ p < q\ implies\ R(p) < R(q)$
$(iii)\ R\ is\ continuous$

By way of example, consider the risk function $R(p) = p^2$ at $p = 0.5$. It is plain to see that $R(0.5) = 0.5^2 = 0.25$, which is simply to say that, on the risk function thus defined, success will be evaluated as being *half as valuable* as it would have been on the standard expected value calculation. The calculation is not actually quite this simple, since the risk calculation *also* needs to account for "where [a given outcome] ranks in the ordering of [all] outcomes from best to worst" such that worst-case outcomes are given *greater* weight (Pettigrew 2024, 136). In order to do this, we take the initial expected value sum

$$E = p_1 v_1 + p_2 v_2 \qquad (3.5)$$[4]

[4] Note that this is just an instance of (3.1) with n = 2.

And rewrite it equivalently as follows:

$$E = ((p_1 + p_2) - p_2)v_1 + p_2 v_2 \tag{3.6}$$

Then, we apply R to each grouping of p-values in (3.6), yielding

$$RE = (R(p_1 + p_2) - R(p_2))v_1 + R(p_2)v_2 \tag{3.7}$$

It is important to note that for (3.7), unlike (3.5) and (3.6), v_1 and v_2 are not value agnostic; i.e., it matters *which* value is indexed as “1” and which as “2”. This is because of how the weights are constructed in Buchak’s formalism: rather than weighing v_1 by $R(p_1)$, we weigh it by the *difference* between by $R(p_2)$ and p = 1. In other words, v_1 is weighted with the remainder of all possible weight *not* applied to v_2 (see Figure 2). Thus, for any risk function R(p) for which the condition

$$\forall p \text{ in } [0, 1], R(p) \leq p \tag{3.8}$$

holds, v_1 is given *additional* weight, while v_2 is given less. The standard way of constructing (3.7), then, will be to assign the worst outcome to v_1, and the best to v_2 (or to v_n for any arbitrary number of possible outcomes). This produces a risk *averse* calculus, inasmuch as the worst-case scenario is given proportionally more weight in the calculation than the best case.[5]

Figure 2: a visual representation of the standard p-space versus Buchak’s risk-weighted p-space. v_2 is discounted by the risk function; as a result, the remainder of the space is given to v_1. For $n > 2$, the space allotted to v_2 above would be partitioned between v_2 and any additional elements; v_1 would be granted the remainder which is not allotted to *any* other v.

With such a calculus up and running, we can then apply it to various cases. For example, for $R(p) = p^2, v_1 = 0,\ p_1 = 0.5,\ v_2 = 2,\ p_2 = 0.5$ (i.e., a single double-or-nothing coin flip), we can plug the values into (3.7) to get

$$RE = ((0.5 + 0.5)^2 - 0.5^2) \cdot 0 + (0.5^2 \cdot 2) = 0 + (0.25 \ \cdot 2) = 0.5 \tag{3.9}$$

---

[5] Note that, in order to develop a risk-*seeking* calculus, it would suffice to violate condition (3.8). E.g., instead of $R(p) = p^2$, we might substitute $R(p) = p^{1/2}$. We could also characterize the distinction as having to do with whether the curve is *convex* (risk-seeking) or *concave* (risk-averse).

Which is only half of the result we would get from (3.5) for the same values. When applying the same principles to the St. Petersburg game (and having modified RE to account for the infinite number of possible outcomes), we have

$$RE = \sum_{n=1}^{\infty} 2^n \left(R\left(\sum_{m=n}^{\infty} \frac{1}{2^m} \right) - R\left(\sum_{m=n+1}^{\infty} \frac{1}{2^m} \right) \right) \tag{3.10}$$

Which, for any R(p) satisfying (3.8), will produce a risk-weighted expected value which is strictly speaking less than the unweighted expectation found in (3.4). For $R(p) = p^2$, for example, numerical methods[6] predict that (3.10) will converge to a (miniscule!) risk-weighted expected value of 3—a far cry from the supposedly *infinite* value produced by the naive calculation in (3.4)! This does not mean, however, that we have proven that the naïve expected value calculation is false, or that we have found the "right" calculus. After all, for $R(p) = p$, the equation simply reduces back to the naïve calculation of (3.4)[7], and if we abandon the constraint in (3.8), i.e., that "$\forall p$ in $[0, 1], R(p) \leq p$", we can produce infinitely many risk-*seeking* strategies for which the series does not converge at all. For the risk-seeking function $R(p) = p^{1/2}$, for example, numerical approximations over $[1, 200]$ already yield $RE = 2.668 \times 10^{30}$!

In other words: all we have shown is that there are (infinitely many!) mathematically precise risk functions R(p) such that, on the adoption of such a risk function, the St. Petersburg game is a poor investment; there are also infinitely many such strategies on which the game is a sound investment. We have not shown that one ought to adopt any such risk-averse function over and above the risk-neutral or risk-seeking alternatives: the equations themselves are silent on this matter. The calculus as it stands is therefore not *action-guiding*. It should go without saying that this is a problem. After all, we want to say more to the risk-seeking player than that we do not share his exceedingly high risk-tolerance: we want to say, over and above this, that the game itself is unsustainable *for anyone*.[8]

In order to make sense of the problem, it will be useful here to distinguish between what I will call *perspectival* risk on the one hand, and *structural* risk on the other. The former has, ultimately, to do with an individual's evaluation of whether a gamble is "worth it"—and this may vary wildly from person to person. Structural risk, then, is a deeper claim, not just about how individuals *feel* about a gamble, but about the internal dynamics of the gamble itself, and the ways in which the distribution of possible outcomes evolves over many rounds of play. Or, to put it slightly differently: whereas perspectival risk merely *deters* individuals, structural risk makes it such that,

---

[6] In this case "numerical methods" just means "plugging the equation into the right kind of calculator and substituting some arbitrarily large n_{max} for ∞", but (3.10) could easily be reformatted into an algorithm which computes the approximate solution step-wise—given, again, some arbitrarily large number of outcomes n_{max}.

[7] This is for the same reason that (3.5) is equivalent to (3.6): the introduction of duplicates which cancel one another out produces mathematically equivalent statements. The thing that distinguishes (3.7) is the introduction of a multiplicative factor which does not cancel out. Removing the multiplicative factor (setting $R(p) = p$) allows us to cancel out the duplicates once again, moving back from (3.6) to (3.5)

[8] Considerations of this kind appear to be at the heart of the debate between neoclassical economists—according to whom *everything* might ultimately be a matter of marginal utility, which is itself a matter of the individual's perspective—and the old-school Marxists—for whom irreducible *structural* considerations (e.g., socially necessary labor time) systematically constrain the possible outcomes. This ultimately has little bearing on our discussion, but is interesting enough in-and-of-itself to merit a brief aside.

across n instances of the same game, the outcomes converge towards a particular (low-value) attractor.[9]

We can see from this discussion that structural risk is intimately bound up with the evolutionary dynamics of a system. It should be no surprise, then, that the expected value formalism, both in its naïve and risk-weighted varieties, is poorly equipped to analyze structural risk: for in both instances, what we are looking at is not a system as it develops, but a single value which represents the expected final *state* of that system. For this reason, we might characterize the shortcomings of the expected value approach in terms of a preference for the long *term* over the long *run*.

We want to look at how the system is changing over time, and to find a strategy which is appropriate for the dynamics under consideration. And the way to do this is to find a strategy for which the *growth rate* is maximized, rather than the all-things-considered expected value (see Kelly 1956, Thorp 2006). Such a strategy ($\phi *$, the Kelly strategy) is said to be asymptotically optimal in the sense that 1) when compared with any other "essentially different strategy", ϕ, with respect to the value X_n at round n, $\lim_{n\to\infty} X_n(\phi *)/X_n(\phi)$ approaches infinity (so, the Kelly strategy diverges from the alternative) and 2) the Kelly strategy minimizes the time to reach any "fixed preassigned goal" (Thorp 2006, 5).[10]

Before discussing the mathematical formalism of the Kelly strategy, it may be worthwhile to try and give an intuitive motivation for the idea that growth rate is the essential thing to maximize. The essential point is that, given the uncertainty of the systems under consideration, the best way to maximize one's *final* standing is to ensure that one's *current* standing is optimal. Rather than gambling everything on the highest possible payout, we want to find an investment which does not compromise the integrity of our portfolio (by, say, losing so much that we can no longer make sensible investments). In this way, the growth-optimal strategy is robust against risk without being stagnant: what we are maximizing, for any given round of the game, is our *progress* towards the desired final state.

With that out of the way, we can quickly lay out the formalism for the Kelly strategy. To begin with, we define the growth rate as follows (Kelly 1956, 919):

$$g(f) = (1 + fW)^p \cdot (1 - fL)^{1-p} \tag{3.11}$$

With f representing the fraction of one's total wealth invested on a single bet, W representing the upside of the gamble, L representing the downside, and p representing the probability of attaining the upside. Note that, for W, L = 1, we have the standard double-or-nothing scenario which we have been working with above.

The question, then, is: for what value of f is (3.11) maximized? To perform such an optimization, we will want to take the derivative g'(f), setting the equation thus derived equal to zero. This is because $\frac{d}{df} = 0$ characterizes the local maxima and minima for any continuous curve;

---

[9] For the St. Petersburg game, a simple Monte Carlo simulation can be run over some arbitrarily large number of instances, say, n = 10,000. The average winnings of the games thus simulated vary, but typically come out around ~7 dollars, or whatever currency you'd like. In one instance (see Appendix A), the average came out to 112 (so, about six or seven consecutive successes) presumably due to some small number of highly successful games.

[10] Proofs for these results will not be presented here (although they can be found in Breiman 1961), but it is at least worth mentioning that the Kelly strategy is optimal in this sense.

since the function is concave with respect to f, we know that the solution here must be a maximum (See Figure 3).

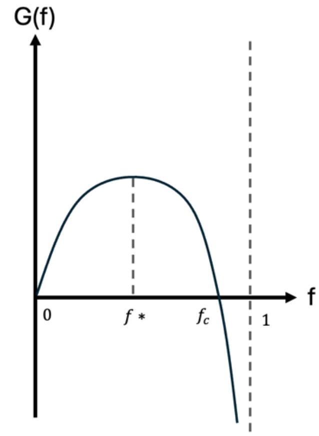

Figure 3: graph of the general growth-rate function (adapted from Thorp 2006).

In order to take the derivative, we begin by taking the log of the equation, thus eliminating the exponentials:

$$G = \log(g(f)) = p \cdot \log(1 + fW) + (1 - p) \cdot \log(1 - fL) \tag{3.12}$$

What we are really maximizing, then, is the *logarithmic* growth rate, listed here as G. And the derivative of G with respect to f is thus

$$\frac{dG}{df} = \frac{pW}{1 + fW} - \frac{(1 - p)L}{1 - fL} = 0 \tag{3.13}$$

When rearranging to solve for f, we find that

$$f = \frac{p}{L} - \frac{1 - p}{W} \qquad (3.14)^{11}$$

What this equation provides us with is a specification of the optimal fraction of our portfolio to invest in a given prospect, depending upon that prospect's risk profile, consisting, again, of the upside, W, the downside, L, and the probability, p, of attaining W. Thus, for a single round of the St. Petersburg Game, we have W = 1, L = 1, and p = 0.5. The Kelly strategy gives the following evaluation:

$$f = \frac{0.5}{1} - \frac{0.5}{1} = 0 \tag{3.15}$$

[11] For a long-form derivation of 3.14, see Appendix B

I.e., the optimal fraction of our portfolio to invest is 0%. And the reason why this is so should be obvious: the bet provides no competitive edge to the player, and so the two sides of the equation cancel out. Such an "investment" boils down to nothing more than blind luck. Suppose, however, that the coin had a 51% chance of landing on heads. Then we would have

$$f = \frac{0.51}{1} - \frac{0.49}{1} = 0.02 \tag{3.16}$$

In other words: to maximize one's growth rate in such a scenario, the optimal fraction of one's portfolio to invest is 2%. Now, strictly speaking one could invest more than this without risking financial ruin (so long as f does not exceed f_c, the rightmost zero of $g(f)$ in Figure 1), but the rate of growth will be slower than at the optimal point determined by the Kelly strategy. Beyond f_c, "even though [the player] may temporarily experience the pleasure of a faster win rate, eventual downward fluctuations will inexorably drive the value [of the player's portfolio] toward zero" (Thorp 2006, 7).

Now, for a slightly different example, suppose that we begin with a $1 portfolio. Suppose also that any money invested in the game cannot be withdrawn from the game until its conclusion (i.e., each player can only play once, and must withdraw everything in one go), and that any money "in the game" is forfeited on a loss. Using the optimal Kelly strategy from (3.16), we invest $0.02 to begin. If we win, our $0.02 investment doubles to $0.04. Our total portfolio consists of

$0.04 "in the game"
$0.98 "out of the game"

The question now is: should we play another round? To answer, note that the current fraction of our portfolio "in the game" is 0.04/0.98 = 0.0408 = 4.08%. This exceeds the optimal fractional investment of 2%, and so playing another round is suboptimal, even if it is not ruinous. This suboptimality can only increase with the number of wins, and so the Kelly strategy advises that we ought to quit while we're ahead, as it were.

The difference between expected value and growth-rate optimization can be brought into sharper relief with a simple analogy: the expected value formalism is a little bit like looking at a cliff face and imagining that it would be good to climb it simply because of how high it is, or how much more "efficient" it would be than walking up a switchback; the Kelly strategy, on the other hand, allows us to observe that the cliff face is, in fact, sheer, and slick with moss and lichen. We can get neither footholds nor handholds, and so another strategy (say, the switchback) is preferable. In fact, it is *faster* than trying to scale the cliff, even though the whole motivation for the cliff-scaling strategy was its ostensibly efficiency![12]

There is much more that could be said about the Kelly strategy, but for our purposes it will make sense to stop here. For we have found, at least in some nascent form, an evaluative calculus which, unlike expected value in its naïve and risk-weighted varieties, is capable of accounting for structural risk, and in this sense, it can be genuinely action-guiding: rather than bottoming out in considerations of perspectival risk tolerance, it is capable of identifying investments which are objectively structurally unsound. In what follows, I will attempt to show how such an evaluative calculus can help to shed light on matters of environmental policy.

---

[12] Thus mirroring the Kelly strategy's asymptotic optimality, defined above.

IV

We have seen that, rather than adopting a strategy which attempts to maximize our *final* standing, we ought to adopt a strategy which maximizes our *growth rate* at any given moment. Such a strategy is said to be robust against structural risk. But the model which we have developed here is admittedly crude, and could benefit from further elaboration. For this reason, before attempting to apply it to the analysis of the strategies laid out in Section II, we will begin by noting some shortcomings of the model as it is presently developed.

First, our reasoning must be broadly analogical when we apply the Kelly model to environmental strategy. What would it mean, for example, to say that we should invest 40% of our portfolio in accelerationism, and 40% in decelerationism? Aren't these strategies more or less diametrically opposed? And what would we do with the remaining 20% of our investment? But the fact of the matter is that the decision-space is highly constrained by external empirical factors. As such, the options laid out before us are by no means optimal. Our goal is not to find a strategy with *no* structural risk, with *no* chance of ruin, with *no* overinvestment, but to find the strategy which is *most robust* against these risks—since these will be incurred "come what may".

Nor is it true that, for any given round of the gamble, we *either* achieve the upsides, or otherwise incur the downsides. It will often be the case that both costs and benefits accrue to us; the question is just how *much* of a benefit is achieved by, say, investing in research into some renewable fuel-source, carbon sequestration technology, or what have you.

Furthermore, it will be exceedingly difficult to provide numerical approximations of the strategies of Section II: we could spend lifetimes trying to determine the exact upsides, the exact downsides, and the exact probabilities, and in the end we would probably not make much progress. For this reason, I will avoid numerical characterizations as much as possible in the discussion to follow. Instead, my aim is to determine the *structural* characteristics of each strategy, in order to get a *general* sense of how risky they might be.

To begin our analysis, it will be useful to develop some evaluative tools. As we saw above, the standard St. Petersburg game is inadvisable due to the relationship between W, L, and p producing an f-value of 0. At a point like this, even the slightest edge granted to the player will produce a sound investment (given a sufficiently small fraction of one's portfolio is invested), as we saw in (3.16). We can generalize this notion to develop a risk-satisficing frontier (RSF) which holds for *any* p-value. Such an equation would determine, for a given p, the ratio of W/L beyond which at least *some* investment is merited. This is accomplished simply by setting (3.14) = 0, and rearranging such that

$$\frac{W}{L} = \frac{1-p}{p} \tag{4.1}$$

This can then be represented graphically, as in Figure 4. Note here that any points *above* the curve ("beyond the frontier") represent an in-principle admissible investment; points on or below the curve are inadvisable due to their inherent structural risk.

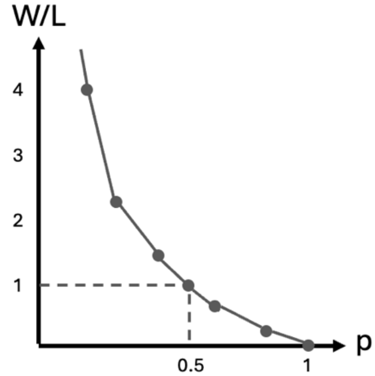

Figure 4: the risk-satisficing frontier (RSF) for all p-values. Points beneath the curve are inadvisable (structurally unsound) investments, whereas points above the curve are admissible. Note that at $p = 1$, $W/L = 0$. This ought to be interpreted as meaning that, when success is assured, it no longer makes sense to talk about a ratio of upsides to downsides.

The RSF clearly exhibits that, as the probability of success decreases, the necessary ratio of W to L (the "edge" afforded to the player) increases hyperbolically—i.e., at a rate which, at sufficiently low p-values, surpasses even exponential growth. In this sense, the Kelly strategy strongly disincentivizes strategies with low p-value: at $p = 0.2$, for example, upsides must exceed downsides by a factor of 4 for the investment to be structurally sound. At $p = 0.1$, this increases to a factor of 9. For $p < 0.1$, things only get worse.

At a first glance, such considerations may be thought to weigh against the accelerationist strategy. After all, as we outlined in our discussion in Section II above, this strategy is characterized by its comparatively low probability of success; and this is due to the argument that even if we invest into, say, renewables, it is not clear that they will actually suffice to reduce the strain placed on the earth-system by human economic activity. The same does not seem to be the case for the decelerationist strategy: for the most part, decreasing material throughput just means decreasing environmental impact; there's little in the way of uncertainty here. It may be that there is uncertainty in whether or not we will be able to *implement* the strategy (due to, say, entrenched political opposition, or structural inertia), but the payoff of the strategy *itself* is relatively clear.

We might formalize this intuition in the following manner: for strategies A with $0 < p < x$, the area *under* the RSF is greater than it would be for strategies D with $x < p < 1$. Given uncertainty about the exact parameters of the investments A and D, the probability q that a strategy falls *below* the frontier is accordingly higher for A than for D. The area under the curve from [0.1, 1], for example, is only about 1.4; the area under [0.001, 0.1] is about 4.5—i.e., about three times as large.[13]

This may provide some comparatively weak reason to prefer D over A, but we can in fact say more than this. Recall that our concern is with investments as they develop *over time*, and that the exact profile of a given investment can change over time. For this reason, it may be useful to

[13] These values are calculated just by taking the integral of function (4.1) over the range specified. Note that $p = 0$ is undefined, but that the function approaches infinity as p approaches 0. For this reason, the calculation above takes its lower limit to be 0.01 rather than 0; if we took the lower bound to be n as n approaches 0, the area of the leftmost region would trivially exceed the area of the rightmost region.

plot the value of both the upsides and downsides of a strategy as they develop. To begin, consider Figure 5 below, where W(t) and L(t) are constant with respect to time.

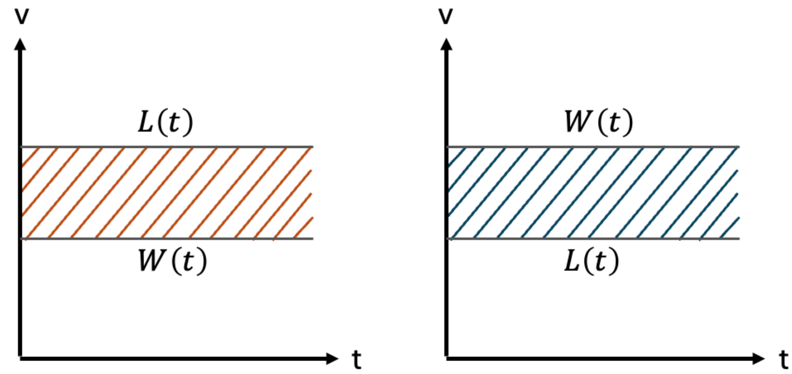

Figure 5: two simple representations of the risk-dynamics of a portfolio over time. The hatched region represents the *discrepancy* between the curves. Red hatching indicates a positive discrepancy (and thus the potential for structural risk), while blue hatching indicates a negative discrepancy, and thus robustness in the face of uncertainty.

Such a graph may be said to exhibit a discrepancy between upsides and downsides; this discrepancy can be mathematically specified by the following equation:

$$dis = \int_0^{t_{final}} L(t)dt - \int_0^{t_{final}} W(t)dt \tag{4.2}$$

Which is just the area *beneath* L(t) and *above* W(t). Thus, graphs where L(t) exceeds W(t) will have a positive discrepancy; graphs where W(t) exceeds L(t) will have a negative discrepancy. For any $p < 0.5$, the magnitude of a positive discrepancy serves as a useful index to the structural risk of the strategy, since we do not have a sufficiently high p-value to counteract the all-things-considered disadvantageous ratio of upsides to downsides. I.e., for $p = 0.5$, the RSF indicates that the ratio of upsides to downsides must exceed 1. Even for $p > 0.5$, there is a limit to how great the discrepancy can be: the RSF indicates that, for $p = 0.75$, for example, the ratio of upsides to downsides must exceed $\frac{1}{3}$. In this way, we have devised a general *graphical* method for identifying structural risk.

With this set of tools in mind, we can now give a graphical illustration of how the accelerationist and decelerationist strategies develop over time. Regarding the profile of the accelerationist strategy, recall the earlier suggestion from Bostrom that we might "enter … a less sustainable state (consuming fuel faster) in order to later achieve the most desirable sustainable state" (Bostrom 2013, 26). We can interpret this graphically as follows: $L_A(t)$ increases faster than $W_A(t)$ (downsides are accumulating in the here-and-now) but at some t_{final}, $W_A(t)$ meets or overtakes $L_A(t)$; from this point onward, $W_A(t)$is assumed to increase exponentially, while $L_A(t)$ is assumed to decay towards zero. We might imagine that t_{final} is something like "50 to 100 years from now", but the exact value is neither here nor there: what matters is that it takes some time for the dynamics to shift in favor of the payoffs.

The decelerationist strategy, on the other hand, aims to dampen the value of the downside $L_D(t)$, preventing short-term bad outcomes from accumulating. Furthermore, there are immediate upsides $W_D(t)$ associated with such a strategy, since the worst effects of environmental degradation are being actively mitigated, rather than pushed off to some imagined future state. We might provisionally assume that $W_D(t)$ and $L_D(t)$ follow more or less the same trajectory in the early stages of the dynamics, after which $L_D(t)$ decays towards zero, while $W_D(t)$ stagnates. We will assume the same t_{final} for ease of comparison. And with this in mind, both the accelerationist and the decelerationist ought to accept that $W_D(t_{final}) < W_A(t_{final})$—i.e., the final value of the upsides of the decelerationist strategy are fewer than the upsides of the accelerationist strategy. This is precisely the cost of taking a precautionary, rather than a maximizing, approach. These assumptions are represented graphically in Figure 6 below.

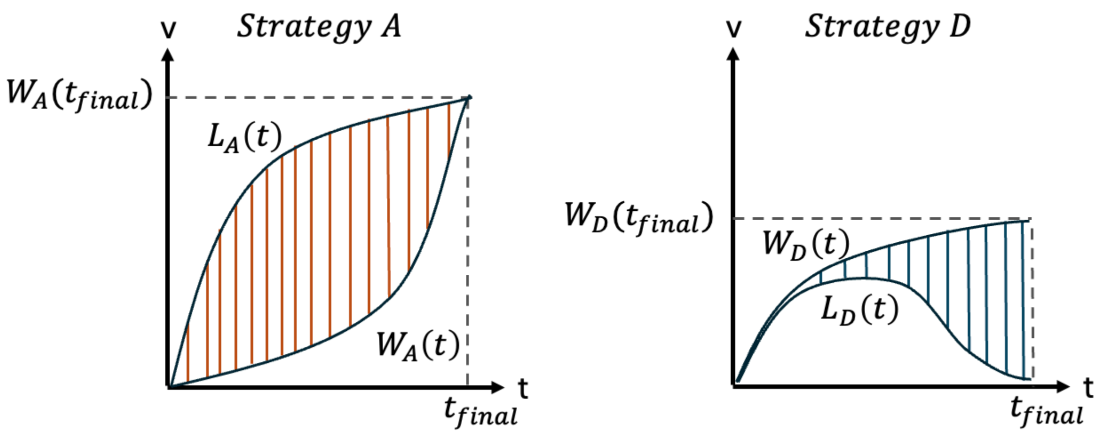

Figure 6: general long-run profiles of A and D. As in Figure 5, red hatching indicates a positive discrepancy, while blue hatching indicates a negative discrepancy. Strategy A postpones gains in favor of long-term payoffs but accumulates substantial structural risk along the way. Strategy D mitigates risks early, at the cost of a lower final value, but avoids structural fragility. Note that, beyond t_{final}, $W_A(t)$ is assumed to continue to increase, while $L_A(t)$ is assumed to decay towards zero.

Assuming that these characterizations are broadly accurate, then we have

$$\int_0^{t_{final}} L_A(t)dt - \int_0^{t_{final}} W_A(t)dt \ > \ 0 \ > \ \int_0^{t_{final}} L_D(t)dt - \int_0^{t_{final}} W_D(t)dt \qquad (4.3)$$

I.e., the discrepancy of the accelerationist strategy is positive, whereas the discrepancy of the decelerationist strategy is negative. If the evaluation condensed in Figure 6 and Equation (4.3) is an acceptable one, then it is damning for the accelerationist: for what it says, in effect, is that the accelerationist strategy is rife with structural risk, while the decelerationist strategy is robust against it. Even granted that the accelerationist is able to maximize their final value, the proposed strategy simply accepts too many structurally unsound gambles along the way to make the strategy advisable.

Note also that our analysis here relies on little to no exact numerical assumptions. All that is needed is the *general* character of the upside and downside curves for each strategy. It may be that, for $p > 0.5$, the discrepancy in A could be insufficient to create genuine structural risk—but given that p is generally assumed to be relatively low, this is an *edge case* rather than the default assumption.

V

In this paper, I have attempted to do two things. First, I have attempted to show why naïve expected value calculations are inadequate for thinking about decision theory in the long run. My hope is that I have done enough to substantiate that concern, and to describe the general shape of a more adequate formalism: one that emphasizes growth-rate over final value, structural risk over individual perception, the long run over the long term. Second, I have attempted to make some sense of our ways of thinking about environmental strategy, and to show how a more adequate formalism might push us to think differently about things. And although I do think that the conclusions I've reached are broadly correct, the discussion here is not meant to *prove* anything about environmental strategy—as I've said before, the models I've developed are crude—so much as it is to exhibit and *explore* some ways of thinking about risk under uncertainty, making *practical* use of the environmental question as a case study. One interesting upshot of the discussion is, perhaps ironically, that in order to think about the long *term*, we really need to think about the long *run*, which ultimately reduces to thinking about robustness against risk in the short term. For in the *absence* of this robustness, over the sufficiently long timeframes which the longtermists take to be of interest, any amount of structural risk would be sufficient to ruin us.

There is much that the longtermist might say in response to such an analysis—to the effect, for example, that we are discounting the value of future lives, that it would be *morally wrong* of us not to persist in the risk-courting accelerationist strategy, or, at the very least, that it's a shame that some astronomically-high-valued futures might ultimately be unattainable.

I do not have room here to respond at any great length to objections of this kind. Instead, I can say this: I have written previously of ethical language as emerging from the dynamics of social life. On such an account, ethical language plays the functional role of making these dynamics *explicit*, and identifying effective strategies for the *regulation* of individuals and of communities broadly construed. The important insight for our purposes is that ethical language is *derived*; it is a particular *tool* for navigating a life. It is dependent upon an understanding of the situations themselves, and this understanding generally *precedes* ethical talk; it provides ethical talk with *something* to evaluate. What I have attempted to produce here is an articulation of our situation: to produce a model, however crude, of the actual dynamics which underwrite our decisions. And if this is what I have done, then claims about the *ethical* inadequacy of my framework must be grounded in a kind of mistake.

Works Cited

Bostrom, existential risk prevention as a global priority (2013)

Breiman, optimal gambling systems for favorable games (1961)

Buchak, risk and rationality (2013)

Frame and Allen, climate change and global risk (2008)

Jaynes, probability theory: the logic of science (2003)

Kallis, degrowth (2018)

Kelly, a new interpretation of information rate (1956)

MacAskill, what we owe to the future (2022)

Meadows et al., limits to growth (1972)

Otto et al., embracing uncertainty in climate change policy (2015)

Pettigrew, should longtermists hasten extinction (2024)

Read, how to think about climate crisis via precautionary reasoning (2017)

Read and Torres, radical longtermism and the seduction of endless growth (2023)

Rockström et al., planetary boundaries (2009)

Schmelzer et al., the future is degrowth (2022)

Thorp, the kelly criterion in blackjack, sports betting, and the stock market (2006)

Appendix A: Monte Carlo Simulation of the St. Petersburg Game

```
import random

def St_Petersburg():
    value = 1
    loss = False
    while loss == False:
        if random.random() < 0.5:
            value *= 2
        else:
            loss = True
    return value

def Monte_Carlo(n = 10000):
    total = 0
    for i in range(0, n):
        total += St_Petersburg()
    return total / n

def Small_Sim(n = 10):
    total = 0
    for i in range(0, n):
        outcome = St_Petersburg()
        print(outcome)
        total += outcome
    return total / n
```

```
print(Monte_Carlo())
print(Monte_Carlo())
print(Monte_Carlo())
```

```
7.275
7.0115
112.1081
```

```
print(Small_Sim())
```

```
1
1
4
4
2
2
1
1
2
32
5.0
```

Appendix B: Derivation of Equation 3.14 from 3.13

$$\frac{dG}{df} = \frac{pW}{1 + fW} - \frac{(1 - p)L}{1 - fL} = 0 \tag{3.13}$$

$$\frac{pW}{1 + fW} = \frac{(1 - p)L}{1 - fL}$$

$$pW - fWLp = (1 - p)L + fWL(1 - p)$$

$$pW - (1 - p)L = f(WL)(1 - p + p)$$

$$pW - (1 - p)L = f(WL)$$

$$\frac{pW}{WL} - \frac{(1 - p)L}{WL} = f$$

$$f = \frac{p}{L} - \frac{1 - p}{W} \tag{3.14}$$