Anton's Game: Deontological Decision Theory for an Iterated Decision Problem

SETH LAZAR
Australian National University

How should deontologists approach decision-making under uncertainty, for an iterated decision problem? In this article I explore the shortcomings of a simple expected value approach, using a novel example to raise questions about attitudes to risk, the moral significance of tiny probabilities, the independent moral reasons against imposing risks, the morality of sunk costs, and the role of agent-relativity in iterated decision problems.

I. ANTON'S GAME

In *No Country for Old Men*, Anton Chigurh is a hit man who toys with his victims.[1] He gives them a last chance at survival; tossing a coin, he tells them to 'call it'. If they call correctly, he lets them live. My story is somewhat different. In my story, Chigurh invites you to toss a coin. If you refuse, he will kill one innocent person – Victim – and the game ends. If you flip and the coin lands heads, he lets Victim go and the game ends. But if you flip and get tails, then he kills a different innocent person, Sacrifice-1, and offers you the same choice again. If you refuse to flip this time, he kills Victim. If you flip and it lands heads, he lets Victim go and the game ends. But if you get tails, then he kills Sacrifice-2 and offers you the same choice again. Assume that: all of the Sacrifices and Victim have just as much to live for as each other; you have no personal connection to any of them; you are certain that Chigurh will fulfil his threats; the coin is fair; and Chigurh has an infinite supply of Sacrifices, while you can make infinite choices in

[1] C. McCarthy, *No Country for Old Men* (London, 2005).

© Cambridge University Press 2016
Utilitas Vol. 29, No. 1, March 2017
doi:10.1017/S0953820816000236

CrossMark
https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

a finite period of time (the game is a supertask).[2] Ought you to flip or refuse?[3]

Suppose that Chigurh had offered you a simpler choice: either he kills n Victims or, if you flip the coin, he kills Sacrifice-1. All risk is removed from the picture. Give n whatever value you think warrants indifference between flipping and not flipping. For simplicity, I will assume that $n = 1$, but the ensuing discussion would work equally well for any value of n.[4]

Now, add risk. Each course of action has some probability of leading to a given number of deaths. You could evaluate your options by considering the number of deaths in each possible outcome, multiplying each by its probability of coming about, summing those products to calculate the option's mathematical expectation, and choosing the option with the best expectation. Call this the expected-deaths approach.

If you refuse to flip at T_1, then Victim certainly dies. So refusal leads to one expected death. If you flip, there is a 1/2 chance that nobody dies and a 1/2 chance that Sacrifice-1 takes a bullet. Should you flip? Considered in isolation, your choice at T_1 leads to 0.5 expected deaths. But of course the story does not end there. If you flip once, but refuse to flip at T_2, then Victim is killed anyway. So the expectation of flipping once and then stopping is again 1 expected death (1/2 x nobody dies, 1/2 x Sacrifice-1 and Victim both die). And if you flip at T_2 as well, there is a 1/2 probability that the coin will land heads and end the game, but a

[2] Anton's game has some superficial affinities with infinite decision problems in decision theory, such as the St Petersburg Game, and Satan's Apple. However, besides the fact that those are puzzles for rationality not morality, the St Petersburg Game involves a divergent infinite series (the gamble has infinite expected utility), whereas Anton's game involves a convergent series (one expected death rather than an infinity of them); and Satan's Apple uses infinity to reveal a tension between choice-by-choice rationality, and the rationality of those choices' cumulative effects. See e.g. H. Nover and A. Hájek, 'Vexing Expectations', *Mind* 113 (2004), pp. 237–49; F. Arntzenius, A. Elga and J. Hawthorne, 'Bayesianism, Infinite Decisions, and Binding', *Mind* 113 (2004), pp. 251–83. For an intriguing take on ethical versions of these and other problems of infinite decision theory, see N. Bostrom, 'Infinite Ethics', *Analysis and Metaphysics* 10 (2011), pp. 9–59. Notice that, since Anton's game is a supertask, we need not address the otherwise very interesting question of time discounting.

[3] I mean 'ought' in the belief- or perhaps evidence-relative sense; not, in any case, the fact-relative one.

[4] Anton's game has something in common with trolley problems. We might be inclined, therefore, to think that $n = 5$, as people typically use that ratio when working out when it is permissible to turn a trolley. I am not sure if the analogy is apt. Although Anton has *said* that he will kill Victim if you refuse to flip, by flipping you are not really redirecting an existing threat. Anton can change his mind; he has not yet decided whom to kill. I think your causal involvement in either death is about the same. But, as already noted, nothing depends on this view: if you think the analogy to trolley cases is right, then just assume throughout that there are five Victims.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

further 1/2 probability that Sacrifice-2 will be killed and you will face the same choice again.

You cannot evaluate flipping at T_1 in isolation from your subsequent choices (I discuss this in more detail below). This is an iterated decision problem: your decision now must factor in the decisions you will make when the problem is repeated. You can get off the bus at any time; but if you do, Victim will die. So you need to consider at the outset, when deciding whether to set the sequence in motion, how you will decide at later times.

We can quickly see that the expected deaths, if you flip every time or if you flip multiple times and then stop, are the same as if you refuse. Whichever course of action you take, there will be one expected death.

If heads comes up on the first flip, nobody dies. If it comes up on the second, only Sacrifice-1 dies; on the third, Sacrifice-1 and Sacrifice-2 die, and so on. If you keep flipping until heads comes up, and heads comes up on the n^{th} flip, there will be n-1 deaths in total. To calculate the expected deaths if you flip until you get heads, we must multiply these numbers by their probabilities of coming about, and then sum the products. The probability of getting heads on the n^{th} flip is $1/2^n$. So the expectation is the sum of $1/2^n(n\text{-}1)$ for all n. That is

$$\frac{0}{2}+\frac{1}{4}+\frac{2}{8}+\frac{3}{16}+\frac{4}{32}+\frac{5}{64}\cdots$$

If the series is infinite, the sum of these products is, at the limit, 1. To get an intuitive grasp of this point, notice that the numerator increases additively, by only one each time, while the denominator increases geometrically, by a factor of two each time. So each new term adds exponentially less to the sum than the last.

If we drop the assumption that Chigurh has an infinite number of Sacrifices, and he kills Victim after shooting the last Sacrifice, or if you simply choose to flip a finite number of times and then stop, letting him kill Victim, the expectation is the same: one expected death (see the appendix for the maths).

If the appropriate principle for decision-making under uncertainty in iterated decision problems where lives are at stake is simply to minimize expected deaths, then all of your options are equally good (or bad). In particular, it is permissible to flip the coin, believing that you will flip it every time you are offered the choice.

In other words, you may rationally entertain the possibility of Chigurh killing an infinite number of Sacrifices, in order to prevent him from killing Victim. More proximately, you can run a 1/2 risk of

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

killing at least one person, 1/4 of killing at least two, 1/8 of killing at least three, and so on.[5]

I find this result surprising. Some other ethicists think it outrageous.[6] In the remaining sections of this article, I will try to work out why. I consider a series of questions that must be answered by any attempt to develop a *deontological decision theory* – an attempt to apply a broadly deontological ethical theory to probabilistic moral problems, in defiance of the accusation that deontologists are hopeless at dealing with uncertainty.[7]

I am using 'deontological' here more in a sociological sense than a philosophical one. My primary aim is to explore how philosophers who self-identify as deontologists might build a decision theory that works for them. Anton's game helps us to think through the problems of the most natural approach, which is simply to discount each of our reasons by their probability of being actual, such that our reason to avert one death for certain is precisely the same as our reason to avert a one in a million probability of a million deaths occurring. The game forces us to challenge each element of that approach: I consider its risk-neutrality; its hyper-sensitivity to tiny variations in the stakes; its indifference to the independent moral significance of risk; its disregard for sunk costs; and its assumption of agent-neutrality (that it makes no difference whether it is you who flips the coin at every decision-point, or whether after you flip it at T_1, the decision is passed on to someone else). Undoubtedly other challenges could also be raised. And of course we should not assume that only deontologists should be sceptical of the expected-deaths approach. Consequentialists too at least need an argument for their risk-neutrality, if not for an alternative to it; the same goes for the other relevant considerations.

One might question why it is worth considering all of these topics in a single article. Could we not develop simpler cases to isolate each of these issues? Certainly we could – I do so elsewhere. The advantage of Anton's Game is that it serves as a tasting platter for the many problems to do with decision-making under uncertainty that deontologists have yet to address adequately. It admits of seemingly endless permutations, each

[5] It is important to remember here that the number of victims should correspond to whatever number you thought would warrant indifference between flipping and not flipping in the certain version of the game described above.

[6] See Fabre's and Moellendorf's arguments against totally discounting sunk costs in the ethics of war, in C. Fabre, 'War Exit', *Ethics* 125 (2015), pp. 631–52; D. Moellendorf, 'Two Doctrines of Jus Ex Bello', *Ethics* 125 (2015), pp. 653–73.

[7] For the accusation, see, for example, B. H. Fried, 'What Does Matter? The Case for Killing the Trolley Problem (or Letting It Die)', *Philosophical Quarterly* 62 (2012), pp. 505–29.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

of which can bring into focus a different question. I consider only some of those below – you will no doubt come up with others.

One might also question the unrealistic nature of the game. Such fantastical puzzles are common in decision theory, but perhaps ethicists should be more grounded. I'm not so sure. I am interested in Anton's game out of sheer intellectual curiosity, and because of how it can shed light on deontological decision theory. But this does not mean it is practically irrelevant. Anton's game is an idealized version of a common kind of iterated decision problem: we wish to pursue a good, despite the fact that some bads will unavoidably come about; we know that there is some chance that our efforts will fail, and that we will have an opportunity for a second attempt, and a third, and so on. What we learn from Anton, we can apply to real practical problems in the ethics of war, public health, bushfire management, and indeed in many of our interpersonal decisions. The central practical insight of the case, as I will argue in section V, is that in iterated decision problems the costs can mount up to the point that, even if our forward-looking prospects are no worse than they were at the start, at some point we should cease our efforts to achieve our objective. In Anton's game, you should gamble a finite number of times and then stop. The same may be true in war, and other areas of politics and public policy.

II. ATTITUDES TO RISK

The expected-deaths approach is risk-neutral. This is one ground for surprise at its conclusions: if you are risk-averse, you will choose not to flip rather than run the risk of so many more deaths; if you are risk-seeking, then you will focus on the upside, deeming refusal impermissible since it precludes the chance of nobody dying. Informally, I think people's attitudes to the case are often determined by their attitudes to risk.

Theories of prudential rationality can be prescriptive about attitudes to risk, but they can also just model whatever attitudes you happen to have, as they often do for your utility and probability functions.[8] But for objectivists about moral reasons, ethics should be prescriptive. For objectivists, your beliefs about morality cannot affect what it is morally permissible for you to do – otherwise you could avoid obligations if you could change your beliefs, and people with more permissive beliefs would be held to weaker moral standards.[9] One's attitudes to risk are

[8] Lara Buchak, for example, aims to show that a range of attitudes to risk can be rationally permissible, not that any of them are rationally required. See L. Buchak, *Risk and Rationality* (Oxford, 2013).

[9] Even if one cannot simply choose to believe p, there are surely things one can do to bring it about that one believes p.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

equivalent to one's moral beliefs in this respect. Objectivists cannot allow moral permissibility to hang on something so idiosyncratic, and so susceptible to change.

With the exception of some unpublished work by Lara Buchak, and a rarely cited paper by J. E. J. Altham, I do not think anyone has considered the question of which attitude to risk one would be required to take, in moral decision-making under uncertainty.[10] There is an incredibly rich seam to be mined here. For present purposes, I can only scratch the surface, and ask how we can support or gainsay the risk-neutrality of the expected-deaths approach. It has in its favour not only simplicity, but also a sensible argument against double-counting. It calculates the expectation of each option by simply aggregating the expected values of the outcomes. Each possible outcome contributes to the total only the probability-discounted weight of its intrinsic value. Endorsing risk-aversion or risk-seeking means giving additional weight to the worst or best outcomes, just in virtue of their position within the option's overall outcome profile. This looks like double-counting.

To defend risk-aversion or risk-seeking, we would have to rebut this objection, and show that an option's outcome profile can be valuable or disvaluable over and above the aggregated values of the outcomes themselves. As Buchak has observed, this generates an interesting analogy between defending non-neutral attitudes to risk and arguing for a principle of distributive justice that is not reducible to how a distribution benefits those within it.[11] She notes that risk-aversion could be motivated by a kind of relative prioritarianism, according to which we should give additional weight to the interests of those who end up worst off, in the worst outcomes (prioritarianism in general states that we should give extra weight to the interests of those who are absolutely badly off). This would support refusing to flip in Anton's game.

The analogy with distributive justice is intriguing, but relative prioritarianism does not seem a promising avenue here. It gives the interests of those who suffer in the worst outcome more weight than those (often the same people) who suffer in some other outcome that follows the same option. These outcomes are different possible worlds that might result from that option. Why should we care about relative levels of well-being across possible worlds?[12] Suppose Z is the worst

[10] J. E. J. Altham, 'Ethics of Risk', *Proceedings of the Aristotelian Society* 84 (1983), pp. 15–29.

[11] Buchak, *Risk and Rationality*, p. 233.

[12] For a similar point, criticising *ex post* forms of contractualism, see J. Frick, 'Contractualism and Social Risk', *Philosophy and Public Affairs* 43 (2015), pp. 175–223.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

possible outcome and A the best. Z and A are mutually exclusive: if Z comes about, A does not. Why, then, should we care about how people in Z fare, relative to the people in A? We are animated by relative levels of well-being, in general, because we believe that fairness demands a certain kind of distributive equality between moral equals. But fairness does not plausibly demand distributive equality across possible worlds, between possible versions of the same people.

On the other side, we might think that risk-seeking is warranted because deaths have diminishing marginal moral disvalue. If each additional Sacrifice's death adds somewhat less to the total than did the last, then flipping the coin would be morally required. I find this idea troubling, however. Why should Sacrifice-100's death count for less than Sacrifice-1's? People often quote the hackneyed saw, 'one death is a tragedy, 1,000,000 deaths is a statistic', but we generally recognize this as a failure of our moral imagination, rather than a principled standpoint to be reflectively defended (the quote is often attributed to Stalin). Each individual's death matters morally to the same degree, and should contribute the same strength of reason against flipping as do any others, albeit subject to the discount for how likely it is to occur.

However, one factor that is definitely in favour of risk-seeking is that it offers the chance that everybody might get the same result (survival). If we care about equality, and perhaps in this instance a kind of solidarity, this should tell in that outcome's favour.[13] Nevertheless, this is ultimately an additional reason that tells in favour of flipping the coin, rather than a justification for adopting a particular attitude to risk. Pending further argument, risk-neutrality still seems the most sensible option: I do not think we can plausibly assign moral value to outcome profiles that are independent of the expected values of the outcomes by which they are comprised. In this respect at least, the expected-deaths approach seems vindicated.

III. EQUAL EXPECTATIONS

The expected-deaths approach also has the surprising feature that, in a case as finely balanced as this, *all* your options have the same expectation, and so all are morally permissible. As a result, highly improbable possibilities could make the difference between permissibility and impermissibility.

Also relevant are M. Otsuka and A. Voorhoeve, 'Why It Matters That Some Are Worse Off Than Others: An Argument against the Priority View', *Philosophy and Public Affairs* 37 (2009), pp. 171–99.

[13] For the idea that we have independent reasons to care about equality, over and above how it contributes to the well-being of those affected, see L. S. Temkin, *Inequality* (Oxford, 1993).

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

On the first point, we can show that that the expectation if you flip the coin any finite number of times and then stop is also 1. If you toss the coin k times then stop, the sequence will be the same as above until the k^{th} iteration; then if you get tails on the k^{th} toss Anton will not only kill Sacrifice-k, but also kill the Victim. So, suppose $k = 4$. Your decision table looks like this:

	Heads at T_1	Heads at T_2	Heads at T_3	Heads at T_4	Tails at t_4
Flip four times then stop	$1/2 \times 0$ $= \frac{0}{2}$	$1/4 \times 1$ $= \frac{1}{4}$	$1/8 \times 2$ $= \frac{2}{8}$	$1/16 \times 3$ $= \frac{3}{16}$	$1/16 \times (4+1)$ $= \frac{5}{16}$

This gives the following sequence:

$$\frac{0}{2} + \frac{1}{4} + \frac{2}{8} + \frac{3}{16} + \frac{5}{16} = 1$$

Up to the k^{th} term this is the same as before; but then there is an additional term that reflects the cost of giving up tossing the coin, which takes the total expected deaths up to 1, just the same as refusal and infinite flipping. In general, the expectation of tossing the coin any finite number of times, k, and then stopping, is

$$\sum_{n=1}^{k} \frac{n-1}{2^n} + \frac{k+1}{2^k} = 1$$

So, the expected deaths approach implies that all your options are equally good. So all options are open. This might well give us pause; we might think that *surely* more can be said. Given this infinite array of possible strategies, can each really be as good as all the others?

More surprising still, the fine balance in this case means that even vanishingly improbable outcomes can tip the balance, to render flipping the coin at T_1 impermissible. Suppose then that instead of an infinite supply, Anton has only 1,001 potential Sacrifices. However, if you get tails 1,000 times in a row, he will kill Sacrifice-1,001, as well as Sacrifice 1,000, and the Victim. There is as much moral reason against causing Sacrifice-1,001's death as against causing Sacrifice-1,000's. The expectation of flipping the coin 1,000 times then stopping (setting sacrifice-1,001 to one side) is 1. Adding in the $1/2^{1,000}$ probability of an additional 1 death tips the balance against flipping every time, rendering its expectation worse than the alternatives. This despite how extraordinarily unlikely it is that it should ever come about.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

Of course, in that case the better option would be to flip the coin 999 times then stop. But suppose your choice at T_1 were irreversible: flip every time, or not at all (instead of tossing the coin yourself, you set in train a machine, which will run until the game ends). In this version, the $1/2^{1,000}$ risk of killing Sacrifice-1,001 would make it impermissible to proceed. This is very counterintuitive. But what might vindicate that sceptical intuition?

There are two possible approaches. The first is to argue that a $1/2^{1,000}$ risk of one more death is, in non-comparative terms, too small to be morally relevant. The second is to argue that it is too small to be relevant to comparisons, or clashes between values. Although the first approach is tempting, it cannot be right. When the *only* difference between two options is the presence in one of the $1/2^{1,000}$ risk of a death, then of course you ought to avoid the additional risk, no matter how tiny it is.

We must turn, then, to arguments that show this tiny risk is irrelevant to cases where the possibilities are different in other respects as well. There are, again, two promising possibilities. On the first, it would be disrespectful to appeal to such a slight reason in order to justify failing to save Victim. On the second, our moral values are somewhat indeterminate, such that flipping the coin and refusing to flip are only *roughly* equal: adding such a small increment to either cannot make the difference between choosing one or the other option.

The first possibility is that the risk to Sacrifice-1,001 is slight enough to be what F. M. Kamm calls an 'irrelevant utility', and cannot appropriately be appealed to in a conflict between moral reasons. Suppose that if the coin lands tails, Anton will kill Sacrifice-n for each n^{th} toss, and will also crush a flower. And suppose that there is some weak reason against crushing flowers. It would seem disrespectful to Victim to allow such a slight moral reason to make the difference between his being saved and being left to die. Perhaps the same is true where the reason invoked is a strong one, but its probability of coming into play is very low.

Second, ethicists typically agree that our choices do not always stand in a simple ordering relation to one another, such that we can always say of two options either that one is better than the other, or that they are precisely equal.[14] There is a relation of rough equality, such that two options are roughly equal when slight improvements to one do not render it preferable to the other.[15] I think that our moral reasons are themselves vague, with rough edges, so assigning a precise value

[14] R. Chang, *Incommensurability, Incomparability, and Practical Reason* (London, 1997).

[15] R. Chang, 'The Possibility of Parity', *Ethics* 112 (2002), pp. 659–88.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

to them is always somewhat elliptical – instead they range over a space of possible sharpenings.[16] In Anton's game, we are calculating in terms of expected deaths, which gives us the illusion of precision. In fact what matters, for each death, is the reason against letting it come about. And those reasons are at best epistemically and perhaps ontically vague.[17] The resulting expectations – of flipping and not flipping – are sufficiently indeterminate that small improvements to either cannot plausibly make the difference between permissibility and impermissibility. For moral decision-making under uncertainty, numbers are just a way of modelling our reasons, and we should not take them too literally. Some implications of this model, in particular its precision, are mere artefacts that should not be imputed to the properties being represented.[18]

Although both of these explanations can account for why it is acceptable to disregard Sacrifice-1,001's tiny claim, I think the appeal to vagueness is more theoretically robust. The 'irrelevant utility' idea has little in its favour besides intuitions about cases, and it is hard to see why merely reducing the probability that Sacrifice-1,001 will be affected should make what is clearly a *relevant* utility – his survival – into an irrelevant one.

Conversely, it is already independently plausible that our moral reasons are at least epistemically vague – it would surely be crazy to believe that we could assign precise numerical values (down to infinite decimal places) to each of our reasons.

Additionally, the irrelevant utilities approach has trouble with aggregation. Suppose that several people, hundreds perhaps, each faces the risk of being killed alongside Sacrifice-1,001 if the coin lands tails for the thousandth time. If a $1/2^{1,000}$ risk of one death is an irrelevant utility, then any number of such risks should be irrelevant too. And yet if there are enough potential victims, then even if the probability that they will be harmed is vanishingly small, it seems they should be able

[16] J. Broome, 'Is Incommensurability Vagueness?', *Incommensurability, Incomparability, and Practical Reason*, ed. Ruth Chang (Oxford, 1997), pp. 67–89. Chang rejects the view that what she calls parity can be understood as vagueness.

[17] In support of the idea that moral vagueness is ontic vagueness, see M. Schoenfield, 'Moral Vagueness Is Ontic Vagueness', *Ethics* 126 (2015), pp. 257–82. On vague value generally: T. Dougherty, 'Vague Value', *Philosophy and Phenomenological Research* 89 (2013), pp. 352–72.

[18] For a similar point about formal semantics and vagueness in language, see R. T. Cook, 'Vagueness and Mathematical Precision', *Mind* 111 (2002), pp. 225–47. What he says about degree-theoretic semantics and vague language is apt here: 'Degree-theoretic semantics provides a good model of how vague language behaves. On this view, the formalism is not a description of what is really occurring but is instead a fruitful way to represent the phenomenon, that is, it is merely one tool among many that can further our understanding of the dissatisfied in question. In particular, not every aspect of the model need correspond to actual aspects of the phenomenon being modeled' (p. 234).

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

to affect whether it is permissible to proceed.[19] This is no problem for the vagueness approach, since if you add enough people not flipping will eventually become determinately better than flipping.

The previous section concluded that the risk-neutrality of the expected-deaths approach is defensible, at least for now. This section suggests that we do need to modify it somewhat, albeit in a relatively non-invasive manner: the appearance of precision that the expected deaths approach betrays should be explicitly recognized as an artefact of the mathematical model, rather than as a realistic representation of our moral reasons. This concession does not change our deontic verdict, however; it still seems that all your original options in Anton's game are permissible.

IV. RISKY HARM AND HARMLESS RISKS

In the expected-deaths approach to Anton's Game, probabilities have no other role than to discount the objective weights of our reasons to avoid deaths. I think this is a mistake. Setting aside our attitudes to risk, the risks we run contribute to our objective moral reasons in two ways. First, exposing the Sacrifices to risks of harm wrongs them, even if the threatened harms do not result. Indeed, there is a *prima facie* objection to exposing an infinite array of Sacrifices to some risk of being killed, in order to save Victim's life, just because the net expected deaths are the same. Second, if the coin does land tails and a Sacrifice dies, the probabilistic way in which the death was brought about diminishes its moral weight. The first consideration tells against flipping the coin; the second in its favour. I elaborate on each in turn.

We have a claim not to be exposed to significant risks of wrongful harm, which is independent from our claim not to suffer harm *simpliciter*.[20] When you flip the coin for the n^{th} time, you subject Sacrifice-n to a 1/2 risk of being killed, which you have reason not

[19] The right response here might be to develop the 'irrelevant utility' idea to accommodate these kinds of cases, and argue that what matters here is not the probability, for each person, that she will suffer this harm, but the probability that *someone* will suffer this harm. Thanks to a referee for pointing this out. For more on this idea, see in particular M. Otsuka, 'Risking Life and Limb: How to Discount Harms by Their Probability', *Identified Versus Statistical Lives: An Interdisciplinary Perspective*, ed. Nir Eyal, I. Glenn Cohen and Norman Daniels (Oxford, 2015), pp. 77–93.

[20] This is one view among many others about the moral significance of risk, and since I cannot hope to survey and address them all, I focus on summarizing the heart of the argument I find most plausible. But for further discussion, see J. Oberdiek, 'The Moral Significance of Risking', *Legal Theory* 18 (2012), pp. 339–56; S. Perry, 'Risk, Harm, Interests, and Rights', *Risk: Philosophical Perspectives*, ed. Tim Lewens (New York, 2007), pp. 190–210; M. Hayenhjelm and J. Wolff, 'The Moral Problem of Risk Impositions: A Survey', *European Journal of Philosophy* 20 (2012), pp. 26–51; M. J. Zimmerman, 'Risk, Rights, and Restitution', *Philosophical Studies* 128 (2006), pp. 285–311. The most

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

to do, independent of your reason not to contribute to his death. This reason applies regardless of how the coin lands.

We could substantiate this thesis in three ways: the claim not to be exposed to risks could be grounded in one's broader claims not to suffer harms; we might posit an intrinsic interest in avoiding risks, independent of one's interest in avoiding the resulting harms; or we might argue that wrongful risk-imposition has nothing to do with harm at all – it is a harmless wrong. The differences between these possibilities are immaterial here. I mean simply to observe that when deciding what to do at T_1, you must acknowledge that flipping the coin and getting heads would not leave you with a clean slate. Even if you get lucky, you have subjected Sacrifice-1 to a 1/2 risk of being wrongfully killed. This is a serious risk! Sacrifice-1 has a complaint against you, just on these grounds, however the coin lands. Similarly if the coin lands heads at T_2, Sacrifice-2 has the same complaint, and so on each time. These are additional reasons that you must consider when deciding what to do – each of them must be assigned an appropriate weight and discounted for its probability of applying.

The practice of diverting costs from one person to others, just in case the expected harms are the same, is in tension with familiar deontological restrictions on marginal interpersonal trade-offs – in which we harm one person in order to provide another with some marginally greater benefit. Anton's Game makes this especially vivid, by imposing some risk of harm (however infinitesimal) on infinite Sacrifices, in order to save Victim from the same degree of expected harm as the aggregate of that imposed on all those Sacrifices. Recognizing that the Sacrifices have a claim not to be subjected to risks, independently of whether those risks come about, allows us to buttress their interests in a way that rules out this kind of immoderate aggregation.

But how precisely does the reason not to endanger the Sacrifices fit into your decision? There are two possibilities. First: when you flip the coin at T_1, you presumptively wrong Sacrifice-1 just by subjecting him to a 1/2 risk of being killed, independently of the wrong he suffers if the coin lands tails and Chigurh kills him. Likewise if you flip at T_2, you will wrong Sacrifice-2 in the same way, and so on.

But there is another possibility. When you flip at T_1, you wrong Sacrifice-1 in virtue of subjecting him to a 1/2 risk of being killed, and you also wrong subsequent Sacrifices by subjecting *them* to similar risks. After all, assuming that you will continue to flip the coin, you

sustained individual treatment of the ethics of risk is S. O. Hansson, *The Ethics of Risk* (London, 2013).

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

subject Sacrifice-2 to a 1/4 risk of being killed, Sacrifice-3 to a 1/8 risk, and so on.

I think the first approach is the right one. The risk that matters is the risk at the last point when it is in your voluntary control. After all, you can decide to stop flipping the coin at any point up until then, so it is not strictly true that you have really *imposed* a risk on Sacrifice-*n* until you flip the coin for the n^{th} time, even if there is, at that stage, a positive probability that he will be harmed.

In practice, this means that there is an additional downside omitted from the expected-deaths framework. Moreover, it seems weightier than the equality-based reasons considered above: we should care more about the badness of exposing Sacrifice-1 to a 1/2 risk of being killed than about the goodness of everyone getting the same result. So the scales tip against flipping.

While the argument from harmless risks tells against flipping, I think probabilities also matter in a way that diminishes the reasons against flipping, in the following way.[21] Sacrifice-*n*'s complaint against you when the coin lands tails is somewhat weaker than it would have been if Chigurh had announced an intention to kill Sacrifice-*n* however the coin landed. Very roughly, I think that innocent people have weaker complaints against being killed when the probability that they would be killed is lower, at least when the agent is aiming not at their deaths but at achieving some good.

Suppose that Chigurh breaks out a roulette wheel and offers you the following bet. If the ball lands on red, he pays you £100; if it lands on black 2, it will detonate an explosive that will kill me. Now imagine you got the same pay-out for red, but if the ball lands on any black slot the explosive will detonate, and I will die. Suppose that you bet, and in each case the ball lands on black 2, and I die. Your contribution to my death is morally worse in the second variation, I think, because you have given my life even less weight in your deliberations than in the first. In the first, you proved yourself ready to overlook a 1/37 probability of my death for the sake of a trivial gain; in the second, you overlook an 18/37 probability of my death for the same gain. Each shows reprehensible disregard for my safety; but clearly the second is worse.

This relationship also holds when you have good reasons for risking others' lives, as in the Chigurh case. The higher the probability that your action would kill an innocent person, the greater the wrong suffered by him when you proceed, because the more you are treating him as a resource that can be sacrificed for the greater good. The higher the probability that the Sacrifices will die, the readier you show

[21] S. Lazar, 'Risky Killing and the Ethics of War', *Ethics* 126 (2015), pp. 91–117.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

yourself, by proceeding, to treat their deaths as a way to give the Victim a chance at survival. Killing someone in this way is, other things equal, a morally worse way to treat that person.[22]

Actions which lead to others' deaths wrong their victims more gravely when the probability that they would have this outcome was greater. Conversely, they are less wrongful when the probability was lower. So the cardinal weight assigned to the Sacrifices' deaths should reflect the fact that your decision to proceed, even if it causes their deaths, does so with only 1/2 probability.

Do these considerations balance out? For every possible outcome (you get heads at flip 1, 2, 3, etc.), the second argument discounts the moral significance of each Sacrifice's death in that outcome. The first argument adds only one additional complaint to each outcome: that of the Sacrifice who luckily escapes death when the coin lands heads. This implies that the additional reason in favour of flipping outweighs that against. In this section, then, we have seen that the expected-deaths approach omits crucial morally relevant information – the independent moral significance of risks. Moreover, this makes a difference to our deontic verdict. The additional reasons in favour of flipping outweigh those against, so it seems that you ought to flip, rather than refuse.

V. SUNK COSTS

The expected-deaths solution to Anton's Game permits you to flip the coin infinite times, with an infinite number of deaths resulting if the coin never lands heads. If the coin repeatedly comes up tails, then nothing changes: your decision problem is no different at T_1 than at T_{100}. If you (amazingly) reached the 100^{th} coin flip, the 99 Sacrifices who have already died would be irrelevant to your decision to continue. They are 'sunk costs'. Since you can do nothing about them, you can ignore them. You know all this at T_1.

I find this viscerally unsettling.[23] But the worry about sunk costs is double-edged: on the one hand, we might think it is wrong to keep flipping the coin, no matter how many people have died in the attempt to save Victim; but on the other, we might think that those very Sacrifices give reason to keep flipping, since otherwise they will have died in

[22] This argument has some affinities with Philip Pettit's account of the robustly demanding good of respect, in P. Pettit, *The Robust Demands of the Good* (Oxford, 2015).

[23] Others would agree. See Fabre, 'War Exit', p. 637; Moellendorf, 'Two Doctrines', p. 666.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

vain.[24] I will call these respectively the Accumulation and Died in Vain variants of the sunk-costs worry.

I reject the Died in Vain variant. If your reasons to keep flipping the coin were an increasing function of the number of Sacrifices who have died in the attempt to save Victim, then, when enough Sacrifices have died, you would be required to keep flipping even if Anton adds another Victim. Similarly, that additional reason could justify flipping even if Anton were gradually to weight the coin in favour of tails. This is an unacceptable conclusion: ensuring that the dead have not suffered in vain cannot justify inflicting a greater risk of death on the living. And knowing at the outset that Anton is going to up the stakes should make you less inclined to flip at T_1, not more. Additionally, even if continuing to flip reduces the chance that the dead Sacrifices will have died in vain, it also increases the chance that they will have died as part of an even more costly effort to save Victim.

Perhaps if each of the Sacrifices had cared about Victim's survival, then continuing to flip would satisfy an interest of theirs, which might have some value. But the wishes of the dead must surely be of little weight, other things equal, otherwise the interests of the 107 billion of the world's human inhabitants who have died would dominate the moral lives of the 7 billion people now alive. And anyway, we can stipulate that none of the Sacrifices especially cares about Victim's survival.

The Accumulation variant of the sunk-costs worry is much more plausible. The basic idea is this. The more people die in the attempt to save Victim, the weaker grows Victim's claim to be saved. The sacrifices already made on her behalf are deficits in her moral ledger. Saving the Victim is always worth *something* – she is, after all, an innocent person whose life has value.[25] But the more people die along the way, the weaker her claim that others risk their lives for her sake.

To support this argument, consider some variations on Anton's game. First, if the coin lands tails, Chigurh kills Sacrifice-1, but sets Victim-1 free, replacing her with Victim-2. The game continues, and each time the coin lands tails the Victim is replaced. All the Victims are drawn

[24] Jeff McMahan tentatively defends a version of the Died in Vain variant (he calls it the Redemption Thesis). In general, though, he is in favour of an absolute discount for sunk costs: permissibility in these cases is determined only by forward-looking considerations. David Rodin has a subtly different but deontically equivalent view. See J. McMahan, 'Proportionality and Time', *Ethics* 125 (2015), pp. 696–719; D. Rodin, 'The War Trap: Dilemmas of Jus Terminatio', *Ethics* 125 (2015), pp. 674–95.

[25] This is contrary to the views of Fabre, 'War Exit' and Moellendorf, 'Two Doctrines', who think that we have a 'proportionality budget' which is set at T_1, such that once it is expended, no further sacrifices are permissible. In our case, that would mean if you get tails on the first flip, you must not flip again. This is surely too severe.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

from a different group from the Sacrifices. In this variant, I think you have the same reason to save each Victim as you had to save Victim-1. None of them uses up her claim to be saved, so that moral reason has the same weight at each time.

Second, suppose Anton has two games in progress; you can play only one of them. The Victim in Game A has already had three Sacrifices die on her behalf, while Game B's Victim is at the first iteration, with no bodies on her account. I think you have reason to choose Game B over Game A. The A-Victim has already had a fair go; her claim to aid is weaker than that of the B-Victim.

Third, suppose there is only one Sacrifice and only one Victim. Chigurh offers you this deal. Refuse to flip, he inflicts excruciating but temporary and instantly forgotten pain on Victim, and the game ends.[26] Flip, and if you get heads the game ends; if you get tails, he inflicts an identical pain on Sacrifice, and you choose again with the same options. What ought you to do?

I think that clearly you ought not to flip *ad infinitum*. After a point Sacrifice has borne enough pain on Victim's behalf that we cannot justify inflicting any extra risk on her. There is a limit to what Victim can expect other people to bear for her sake.

This last observation underpins the doctrine, alluded to above, that prohibits marginal interpersonal trade-offs. Some philosophers reject *all* trade-offs, but at least most deontologists think that trade-offs are justified only when the benefit is significantly greater than the harm, precisely because people are not merely sites for the realization of value: they are beings with a kind of moral status that we have to respect, even when that means realizing less value overall.[27]

Although Anton's original game is different from the pain-based version – the costs are imposed on different Sacrifices, instead of the same individual bearing all the cost – the basic point is the same. There is a limit to the costs that Victim can expect others to bear on her behalf. Suppose that Chigurh offered Victim the choice of whether to flip the coin; she flips and gets a sequence of tails. Surely at some point she must stop flipping. Externalizing the risks Chigurh has imposed on her can be permissible only up to a point.

One might also think that continuing to shift the risk away from Victim towards the Sacrifices is *unfair*. To see this, suppose Chigurh

[26] Thanks here to Garrett Cullity.

[27] This kind of moderate deontology has been the norm at least since it was first characterized by S. Kagan, *The Limits of Morality* (Oxford, 1989). See W. S. Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing', *Philosophical Review* 89 (1989), pp. 287–312; J. McMahan, *The Ethics of Killing: Problems at the Margins of Life* (New York, 2002); H. Frowe, *Defensive Killing* (Oxford, 2014).

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

offers you a different choice, with benefits instead of harms. If you refuse to flip, then a randomly chosen member of the group of twenty people on his left will get 1 util (unit of well-being). If you flip and get tails, a randomly chosen member of the group of twenty on his right will get 1 util, and you will get the opportunity to flip again. If it lands heads, nobody gets anything and the game ends. Since the expectations of refusing to flip and flipping each time are identical, you flip; it lands tails, giving one of the Righties a 1 util benefit. Other things being equal, I think you now have somewhat more reason to refuse to flip. The Righties have each been given a 1/20 chance of getting the util, but the Lefties have not. Fairness dictates that, at some point, you should give the Lefties a chance, even if that means forgoing the possibility of your run of tails continuing.

The same is true when we are distributing burdens, instead of benefits. Considerations of fairness can favour imposing a certain cost on the Victim, instead of a 1/2 probability of death on another Sacrifice. At each prior iteration, you gave a member of the group of Sacrifices a 1/2 chance of being killed, in order to give Victim a 1/2 chance of being set free. By T_5, say, you have made that trade-off four times. Moreover, each time a Sacrifice has been killed. Fairness must surely dictate that you reconsider the merits of saving Victim. Note that this consideration would apply even if Chigurh swapped the Victims out at each iteration.

One might object that, if you flip the coin at T_1, you will ensure that nobody faces as great a risk to their life as Victim faced. Victim was certain of being killed, but Sacrifice-1 faces only a 1/2 probability of death, Sacrifice-2 only 1/4, and so on. Might we not think, then, that fairness requires you to flip the coin, since no individual then faces the same magnitude of risk as did Victim?[28]

I reject this position. Flipping at T_1, with the expectation that you will continue to flip *ad infinitum*, does not merely subject Sacrifice-1 and the others to risk at T_1, it also includes the expectation that you will subject them to risk at T_2, T_3, and so on, until you get heads. It means diverting risk away from Victim every time the coin is flipped. Even if doing so means that nobody else is subjected to as great a risk as Victim faces if you do not flip, this does not seem sufficient reason to keep distributing risk away from Victim. After all, this line of reasoning would justify flipping even if the coin was weighted so that its probability of landing tails was 99/100 – still each Sacrifice would face a smaller risk than would Victim if you refuse to flip. Perhaps we sometimes have reason to spread risks around, rather than allow them

[28] Thanks to Johann Frick for this suggestion.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

to concentrate.[29] But in this case the risks to each individual are great enough, and the prospect of the choice situation repeating high enough, that those considerations are, if not silenced, then at least drowned out.

We can strengthen this argument from fairness by considering a further modification. Suppose that when the game begins, Chigurh has not yet identified his Victim or his first Sacrifice. There is one group of candidates for both roles. If you refuse to flip, he will kill one randomly chosen person. If you flip, he will choose Victim and Sacrifice-1 at random. Victim will remain the same until the game ends.

In this case there is a clear difference between flipping at T_1, when *anyone* could be Victim, and flipping at T_2, when Victim is set. Since everyone is equally likely to be the Victim at T_1, flipping the coin makes nobody better off at the expense of anyone else. Everyone is antecedently just as well off if you flip as if you do not. But at T_2, things have changed. Now Victim is fixed, as is the group of potential Sacrifices. From now on, flipping distributes risk away from Victim towards the Sacrifices, and considerations of fairness are raised (as well as of the relative diminution of Victim's claim to aid).

In sum: when we must distribute expected harms and benefits in conditions of uncertainty, fairness requires that we attend to the distribution of both risks and harms. As the Sacrifices bear more and more cost on Victim's behalf, it becomes harder to justify diverting risks away from Victim towards them.

The expected-deaths approach takes no account of sunk costs; in that it makes a mistake. If the foregoing reasoning is right, then your options have narrowed still further. You know at T_1 that flipping infinite times is impermissible. Instead, you ought to flip some finite number of times and then stop. So flipping at T_1 might still be permissible, but only if you think you will stick to the optimal strategy. If you are likely to flip too much or too little – perhaps in a desperate attempt to vindicate your decision or out of shock at the early deaths – then you had better not flip at all.

VI. AGENT-RELATIVITY

This raises a further puzzling feature of the case. I said at the outset that you cannot simply consider your choice at T_1 in isolation from what you will do at subsequent iterations. One might go further, and argue that the proper object of choice in an iterated decision problem is not the action of flipping at T_1, but a whole profile of options at

[29] See N. Daniels, 'Can There Be Moral Force to Favoring an Identified over a Statistical Life?', *Identified Versus Statistical Lives: An Interdisciplinary Perspective*, ed. I. Glenn Cohen, Norman Daniels and Nir Eyal (Oxford, 2015), pp. 110–23.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

every time point.[30] Certainly the sunk-costs reasoning suggests that line. Although I do not rule it out, I am ambivalent. You cannot, by selecting a profile of options, cause yourself to stick to it. You can make it more likely that you will do so, but you cannot necessitate your future actions. Still, you clearly have to consider, when deciding what to do at T_1, what you are likely to do if the game repeats itself. You have to factor your future voluntary choices into your decision now. Does this simply mean treating your future self as though it were another person? Can we plausibly take this kind of external attitude to our own future choices?[31] The expected-deaths approach is silent on these questions, and totally indifferent to who makes the choices at each decision point.

Suppose, then, that you are offered only the choice at T_1. At each subsequent iteration, Chigurh offers it to a different agent. How should you factor *their* choices into your calculations? If all strategies are equally acceptable, as they are on the simple expected-deaths framework, then it does not matter. But if the foregoing arguments are right and flipping is better than refusing only if the number of flips is finite, then you should clearly try to coordinate with the other agents if you can, and predict how they are likely to decide, factoring those probabilities into your expectations. I think this is also what you should do in the single-agent version of the case. Moreover, you *should* be better able to coordinate with yourself than you are with the other possible agents. One way to do this is to give yourself additional reasons by resolving to stick to the strategy, so that if you fail to adhere to your strategy you will let yourself down.

But I think there is one crucial difference between the multi-agent variation of the case and the single-agent version. Moral reasons, especially those governing permissible harm, often include an agent-relative component. I have a special responsibility for making sure that *I* do not violate rights, for example, rather than simply an agent-neutral responsibility to minimize rights-violations.

The reason to save Victim is probably agent-neutral. Nobody (I think) believes that I ought to save one stranger even if my doing so prevents

[30] For discussion along these lines, see B. Hedden, 'Options and Diachronic Tragedy', *Philosophy and Phenomenological Research* 87 (2013), pp. 1–35; S. Tenenbaum and D. Raffman, 'Vague Projects and the Puzzle of the Self-Torturer', *Ethics* 123 (2012), pp. 86–112. For discussion among decision theorists about whether rationality can require you to bind yourself to a sequence of actions, see, for example, Arntzenius et al., 'Infinite Decisions'; A. Elga, 'Subjective Probabilities Should Be Sharp', *Philosophers' Imprint* 10 (2010), pp. 1–11.

[31] These questions raise a nest of problems to do with the actualism/possibilism debate in ethical theory, to which I cannot do justice here. For some highlights of that debate, see F. Jackson and R. Pargetter, 'Oughts, Options, and Actualism', *Philosophical Review* 95 (1986), pp. 233–55; J. Ross, 'Actualism, Possibilism, and Beyond', *Oxford Studies in Normative Ethics* 2 (2012), pp. 74–96.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

you from saving two other strangers. I also think the sunk-costs reasons are agent-neutral. They derive from the dissipation of the reason to save Victim and the unfairness of repeatedly putting all the risk on the Sacrifices. Neither of these depends on who is flipping the coin.

The reasons not to contribute to the harm to the Sacrifices, however, are plausibly agent-relative. It seems right that you should give greater negative weight to outcomes in which you are proximately involved in their deaths than you do when other agents make those decisions. Although you of course have a responsibility to minimize all wrongful deaths, you have a somewhat greater responsibility to minimize wrongful deaths to which you have proximately contributed. Obviously, Chigurh is the one actually doing the killing, but your decision to flip involves you in the Sacrifices' deaths in a way that plausibly generates agent-relative reasons. In effect, this means that in the multi-agent version of the case, you can tolerate more expected deviation from the optimal strategy than you can in the single-agent version. That it is other people deviating means that you are less responsible for it than you would otherwise be, so it can weigh less heavily against your decision to flip.

Might one go further, and use reasoning like this to argue that by flipping the coin at T_1 you insert yourself into the causal chain that would lead to Victim's death if you subsequently refuse? Might this give you agent-relative reasons to avoid Victim's death, which weigh against the diminishment of her claim to be saved, advocated in section III? I think not, on both counts. Victim would clearly want you to flip the coin at T_1, whatever you decide at subsequent iterations. That gives her at least a 1/2 chance of survival, against certainty of being killed. Any additional agent-relative responsibilities that might otherwise be acquired by inserting oneself into the causal chain in this way are clearly annulled by the fact that doing so is so obviously in Victim's interests.

In the end it seems that although the agent-relative dimension of the case gives you some additional reason not to flip, we should really have taken that into account when setting the number of victims. Agent-relativity really has most bearing on which strategy you can adopt, assuming that there is some optimal strategy, between refusing at T_1, and flipping every time.

VII. CONCLUSION

Anton's game forces us to think about how to apply a broadly deontological approach to ethics to complicated problems of decision-making under uncertainty. It raises a host of issues, to which I have undoubtedly not done justice here. Any serious deontological decision

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

theory would have to develop a comprehensive and compelling account of what our attitude to risk should be, whether our reasons are imprecise and what we can learn from quasi-mathematical modelling; whether we have rights against being exposed to risks; how we should treat moral sunk costs; and how we should regard our own future choices when making decisions now that will affect which choices we get to make. In this article, I have only scratched the surface of that account. But I have, at least, arrived at a tentative answer to the puzzle with which we began. In Anton's game, you should flip the coin, but only if you are confident that you will stop flipping at the right time. When that will be depends on how steep is the discount applied to Victim's interests and how strong the objection against the unfairness of repeatedly flipping.

Anton's Game is pretty outlandish. And yet it has lessons for more mundane, realistic cases. Moral decision-making under risk, when lives are at stake, cannot plausibly be collapsed into a simple expected-deaths framework. Other reasons must be attended to – the value of an equal outcome; harmless risks and risky harms; the diminution of the Victim's claim to be saved; considerations of fairness; and the need to cause yourself to stick to the optimal plan. These reasons will arise in iterated decision problems that are much more quotidian: whenever we risk inflicting harms on innocent people in the pursuit of some goal and, if we fail at the first pass, face the prospect of having another go. These problems arise in war and counterinsurgency, the imposition of sanctions, individual medical practice and large-scale policy decisions (e.g. on vaccination), and in bush firefighting. In these cases, we cannot simply consider the outcomes that follow if we take a chance at T_1, but must also consider the subsequent decisions we will make. And we must acknowledge from the outset that, if things repeatedly go wrong, there comes a point when we simply ought to stop.[32]

sethlazar@gmail.com

[32] For helpful discussion of the ideas that this article is based on, I thank Rachael Briggs, David Grewal, Alan Hájek and Brendan O'Grady. I presented earlier versions of this article, and another related to it, at the Australasian Association of Philosophy, ANU's CRNAP workshop, and in departmental seminars at Roskilde, Oxford, St Andrews, Wellington, Stirling, Auckland and Otago. Thanks to the conveners of those seminars, and the participants, for their helpful comments and discussion. Thanks for comments on the draft to Ed Elliott, Cécile Fabre, Johann Frick, Al Hájek, Theron Pummer, Victor Tadros, Frej Klem Thomsen, David Wiens and Robbie Williams. Lastly, thanks to a reviewer for this journal.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press

APPENDIX

Here's one simple proof that being ready to flip infinite times results in one expected death (thanks to Santhosh Karnik):

Let S be the sum of the series:

(1) $S = \frac{0}{2} + \frac{1}{4} + \frac{2}{8} + \frac{3}{16} + \frac{4}{32} + \ldots \frac{n-1}{2^n}$

Then, double both sides of the equation, ignoring the 0 term to get:

(2) $2S = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{5}{32} + \ldots \frac{n}{2^n}$

Subtract (1) from (2) term by term to get

(3) $S = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \ldots \frac{1}{2^n}$

This is a Geometric Series, whose sum is

(4) $S = \frac{\frac{1}{2}}{1-\frac{1}{2}} = 1$

And to show that flipping a finite number of times and then stopping has the same sum (thanks again to Santhosh Karnik):

$$\text{Define } f(k) = \frac{k+1}{2^k} + \sum_{n=1}^{k} \frac{n-1}{2^n}.$$

$$\text{Then, } f(k+1) = \frac{k+2}{2^{k+1}} + \sum_{n=1}^{k+1} \frac{n-1}{2^n}$$

So the difference quotient of *f(k)* is

$$f(k+1) - f(k) = \frac{k+2}{2^{k+1}} + \frac{k}{2^{k+1}} - \frac{k+1}{2^k} = \frac{2k+2}{2 \cdot 2^k} - -\frac{k+1}{2^k}$$

$$= \frac{k+1}{2^k} - \frac{k+1}{2^k} = 0$$

Since *f(k+1) – f(k)* = 0 for all non-negative integers k, we know *f(k)* is constant.
Plug in $k = 0$ to see that $f(0) = 0 + \frac{0+1}{2^0} = 1$
Therefore *f(k)*=1 for all non-negative integers k.

https://doi.org/10.1017/S0953820816000236 Published online by Cambridge University Press