Ignore Risk; Maximize Expected Moral Value*

Michael Zhao

(forthcoming in *Noûs*; please cite published version)

Abstract

Many philosophers assume that, when making moral decisions under uncertainty, we should choose the option that has the greatest expected moral value, regardless of how risky it is. But their arguments for maximizing expected moral value do not support it over rival, risk-averse approaches. In this paper, I present a novel argument for maximizing expected value: when we think about larger series of decisions that each decision is a part of, all but the most risk-averse agents would prefer that we consistently choose the option with the highest expected value. To the extent that what we choose on a given occasion should be guided by the entire series of choices we prefer, then on each occasion, we should choose the option with the highest expected moral value.

Suppose that ten hikers have been trapped in an avalanche. You oversee a disaster response effort, and you're presented with two options. Option *A* is highly risky: it has a 50% chance of saying all ten people and a 50% chance of failing to save anyone. Option *B*, on the other hand, is safer: it will certainly save four people. Option *A* leads to more people saved on expectation, yet many of us, being risk-averse, would still prefer option *B*.

Now suppose you had to choose between options *A* and *B* as a *policy*, which would be followed a large number of times in similar situations. In that case, nearly all of us—all except the most risk-averse—would choose option *A*. After all, it's extremely unlikely that in the long run, option *B* would lead to more lives saved than option *A*. In fact, unless you are *extraordinarily* risk-averse, in some

*The ideas behind this paper originated over a decade ago; I am grateful to Peter Diao for early discussion. I am also grateful to Martín Abreu Zavaleta, Zach Barnett, Harjit Bhogal, and especially an anonymous referee for insightful comments on earlier drafts of this paper.

technical sense to be clarified later, whatever your level of risk aversion, there's some number of repetitions of A that you would prefer to that many repetitions of B.

Thoughts like this form the basis for an intuitive argument for choosing the option with the highest expected moral value when we face moral decisions under uncertainty, regardless of how risky that option is. In this paper, I'll advance a rigorous version of this argument. I'll begin, in §1, by motivating the question of how we should act morally under uncertainty. As I'll argue, although those in the discussion often assume that we should maximize expected moral value, it's not immediately obvious that we should. In §§2–3, I'll present the original argument for maximizing expected moral value. As I'll argue, even if we aren't risk-neutral, we should still typically act as if we were. §4 responds to objections.

1 Acting morally under uncertainty

How should we choose between different options when the outcomes of those options are uncertain? Let me make two simplifying restrictions. First, I'll restrict my attention to cases where none of the options involve violating any deontological constraints, including those about actively imposing the risk of harm on people. Second, although this is a paper on how to act morally under uncertainty, I won't delve into the question of *moral* uncertainty, the question of how to act morally if one is uncertain about which moral theory is correct.[1] Rather, I'm dealing with garden-variety *empirical* uncertainty, uncertainty about the consequences of different courses of action.[2]

Within the literature on moral decision under uncertainty, many have assumed that *expected-value maximization* (EVM) is the correct answer by default: that is, we should assess acts by averaging over the moral value of the possible outcomes of each act, weighted by the probability of those outcomes, and choose the act that has the highest expected moral value.[3] Formally, if an act has possible outcomes $A_1, \ldots, A_n$ with probabilities $p_1, \ldots, p_n$, the right act is the one

[1] There's now a burgeoning literature on the topic. See, for example, Sepielli (2009), MacAskill et al. (2020), Rosenthal (forthcoming).

[2] Some prefer to reserve "uncertainty" for cases in which the probabilities of the possible outcomes are unknown, and use "risk" for cases in which they are known. I'll use "uncertainty" in a broad sense, to cover any decision where the agent is uncertain of the outcome. The probabilities can be objective or subjective: they can be genuine chance-properties of events, or merely expressive of the agent's credences.

[3] See Parfit (1984, ch. 1), Jackson (1991), Hooker (2000, ch. 3), MacAskill et al. (2020).

that maximizes $\sum_{i=1}^{n} V(A_i)p_i$, where $V(A_i)$ is the moral value of outcome i.[4] Sometimes, EVM is even seen as a reformulation of consequentialism to account for action under uncertainty.

Note that EVM is, in a sense, risk-neutral: an agent who maximizes expected value does not care about how risky a particular option is, and would prefer an option that has a barely-higher-than-50% chance of saving ten lives to another that saves five lives with certainty, even though the first option is much riskier than the second. In contrast, other approaches typically embody various forms of risk aversion. For example, we can employ a *maximin* approach, according to which we compare the worst possible outcome that each act produces, and choose the act that produces the least bad of these outcomes. This would be admittedly an extreme form of risk aversion, since we would be willing to tolerate much worse *expected* outcomes, so long as the worst *possible* outcome is not as bad as for other acts. For this reason, I suspect that very few of us would adopt the maximin approach for reasons having to do with risk.[5] But there are other approaches that embody less extreme forms of risk aversion, like a rank-weighted view, where we give extra weight to the bad possible outcomes, or a view on which we maximize the expectation of some concave function (like the logarithm or square root) of the moral value of the outcome.

Now, although EVM is *one* principle for action under uncertainty, it is unclear what argument, if any, its proponents have against rival principles.[6] Frank

[4]Of course, this assumes that moral value can be represented by real numbers. This in turn assumes, among other things, that different types of moral value or disvalue are fully aggregable, and that we can aggregate moral value for particular individuals into impersonal moral value. These are controversial assumptions, but standard for those working on this issue.

[5]It's telling that those who accept a maximin principle (like Rawls 1971) typically do so not out of risk aversion, but for other reasons: according to Rawls, for example, reasonable agents behind the veil of ignorance choose the maximin because they are unwilling to impose possible sacrifices on others for their own possible gain, not because they are afraid of ending up in the worst possible outcome. Rawls (1971, 13f) goes so far as to resist using the term "maximin" for the difference principle, on the grounds that using it

> might wrongly suggest that the main argument for this principle from the original position derives from an assumption of very high risk aversion. There is indeed a relation between the difference principle and such an assumption, but extreme attitudes to risk are not postulated; and in any case, there are many considerations in favor of the difference principle in which the aversion to risk plays no role at all.

[6]Some take EVM to follow from the idea that rational agents maximize expected utility, which is part of orthodox decision theory (Jackson 1991, MacAskill et al. 2020). As I'll show

Jackson (1991, 463), for example, simply calls EVM "the obvious answer." But although it might seem like an obvious answer, given the plentitude of possible answers, it's not obvious that it is the correct one.

In fact, the claim that EVM is the correct way to act under uncertainty seems to conflict with our intuitions about certain cases. Consider the following case:

> (*Disease:*) Suppose that 1,000 people in a large city have been infected with a deadly disease. You're in charge of disease control for the city, and you're considering two responses. Response *A* is risky: it could be a complete success and save all 1,000 people, or it could be a complete failure and save no one; you estimate the chances of success at 50%. Response *B*, on the other hand, is guaranteed to save exactly 400 people.

I take it that, in such a case, many of us would prefer option *B* over option *A*.[7] This preference seems permissible: there doesn't seem anything obviously morally objectionable about playing it safe here, even though the riskier option would lead, on expectation, to more lives' being saved. Some might even think that it's morally *obligatory* to go with the safe option here, since, given the stakes, going with the risky one would be reckless.[8]

---

in the next section, this is badly mistaken, at least on the orthodox interpretation of what it is for rational agents to maximize expected utility. (For one thing, expected-utility theory can accommodate risk aversion, whereas EVM cannot.)

[7]This prediction is in line with ordinary people's reactions to similar vignettes, like the famous one Tversky and Kahneman (1981) devised to demonstrate loss aversion, in which a large majority of respondents chose a treatment that saves 200 people with certainty to one that saved 600 people with a 1/3 chance. Of course, Tversky and Kahneman also showed that, because what people take to be a loss depends on how the scenario is framed, a large majority chose the initially *dispreferred* treatment in a description of the scenario framed in terms of deaths rather than lives saved. Such a sensitivity to framing would undermine the evidential weight of people's intuitions in these cases. But it poses no problem for us, since our point in appealing to these intuitions is simply to support the claim that maximizing expected value often conflicts with ordinary intuitions.

[8]Rosenthal (forthcoming) argues for this point, although in the context of moral rather than empirical uncertainty; the considerations that she cites, however, extend to empirical uncertainty. As she writes,

> Consider what orthodox decision-theory [read: EVM] can say about a choice between the following two acts, when both acts' moral status is uncertain. The first act might be extremely morally choice-worthy or it might be morally horrific … A second act is much lower stakes, with an equal possibility that it's mildly good or mildly bad, but little chance of anything more significant … Depending upon how

Or consider the following case:

> (*Game show:*) Suppose you're playing for a charity on a game show where you answer questions correctly for money. Assume that each dollar won allows the charity to do the same amount of good as the previous dollar: say, every $5 won lets the charity buy an additional mosquito net that will protect someone against malaria. You've already won $50,000. You're faced with the choice between walking away with that amount of money, or taking a stab at the $1 million question, which you only have a 10% chance of answering correctly. If you answer that question incorrectly, you lose all of the money you've won.

Taking a stab at the $1 million question has higher expected moral value: on expectation, it does $.1 \times \$1,000,000 = \$100,000$ worth of good, whereas walking away does only $50,000 worth of good. But again, it seems morally permissible to opt for the safer choice, and walk away with the $50,000. (Note that the defender of EVM can't defend this decision by invoking the decreasing marginal value of money: we've stipulated that each additional dollar does as much good as the last dollar.)

Finally, consider the following:

> (*Wealthy donor:*) You head an effort to inoculate children in a very populous country against a deadly childhood disease. An extremely wealthy, and eccentric, philanthropist makes you the following offer. Either (*A*) he will flip a fair coin until it lands tails; if it lands tails on the first toss, he provides enough resources for one child to be inoculated; if on the second toss, for two children to be inoculated; if on the third toss, for four children; and so on; or (*B*) he will provide the resources for all ten of the children in a particular village to be inoculated.

---

we specify the particulars, an orthodox decision-theoretic approach to moral uncertainty may end up treating these two cases as equivalent or near equivalent, because the first act's extreme options should be able to balance each other and yield a fairly moderate expected moral rightness. But the stakes seem to make a moral difference. It will sometimes be *morally* reprehensible to risk moral catastrophe—in some cases this will be too reckless—while low-stakes risks may be acceptable, despite having the same expected moral rightness.

Of course, this is just a moral analog of the St. Petersburg paradox. The expected number of children inoculated on option A is the following infinite sum:

$$\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 4 + \ldots = \sum_{i=1}^{\infty} \frac{1}{2} = \infty.$$

Since this is a diverging series, for any finite number, option A will on expectation cause more than that many children to be inoculated—certainly more than ten.[9] So if EVM is correct, then you should choose option A over option B. But intuitively, you would be acting recklessly if you chose option A; it doesn't seem like we're permitted to sacrifice a small good thing for a very low probability of something much better. Some (Bostrom 2009, Monton 2019) have even argued that we should simply discount extremely low probabilities to zero in our decision-making, so that we should prefer the safer, lower-EV option in cases like this.[10]

Such examples may be extreme, but we face more mundane moral decisions under uncertainty all the time. We might face a decision between telling an impressionable student to follow his passion, and advising him to find a stable job; between donating to a charity with ambitious goals that are difficult to realize, and donating to one with more modest aims; between supporting a radical movement that might make society much better or much worse, and supporting the status quo. In each of these cases, it seems at least permissible to choose the safer option, even if it has lower expected moral value.

The fact that we apparently may prefer the less risky option to the riskier one, even though the riskier option has higher expected moral value, weakens

[9] One might object that there are only finitely many children to be inoculated, so the possible value of option A is bounded. Even so, the expected value of A is greater than that of B, assuming a large enough upper bound. For example, if there are ten million ($\approx 2^{23}$) children to be vaccinated, then the expected value of A is

$$\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 4 + \ldots + \frac{1}{2^{24}} \times 2^{23} + \frac{1}{2^{25}} \times 2^{23} + \frac{1}{2^{26}} \times 2^{23} + \ldots = \sum_{i=1}^{23} \frac{1}{2} + \sum_{i=1}^{\infty} \frac{1}{2^i} = 12.5,$$

which is still greater than the expected value of B.

[10] And again, unlike in the St. Petersburg paradox, which involves a lottery with infinite expected *monetary* value, one cannot defend EVM in this example on the basis that additional units of money generate less and less of whatever value we are trying to maximize, so that choosing option A does not necessarily maximize expected value. The tenth million dollars won create much less happiness than the first million, but the inoculation of the tenth million children does not have less moral value than that of the first million.

the claim that maximizing expected moral value is the *obviously* correct procedure for action under uncertainty. One might even doubt that there's anything like an obligatory choice in cases like this; one might think so long as one's risk attitude lies in some reasonable range, any choice supported by that attitude is permissible. And even if *we* think that choosing the higher-EV option in these cases is obligatory, it's easy to imagine someone who disagrees out of risk aversion. It might bother us that there's nothing we could say to a moderately risk-averse person that would convince him to maximize expected moral value.

How can a defender of EVM argue for the claim that his approach is obligatory in cases like these? One natural thought is to appeal to the *long run*, the consequences of many independent repetitions of the scenarios. If *everyone* in relevantly similar circumstances chose the riskier, but higher-EV option (option *A*), then the consequences would be better than those if everyone opted for the safer option (option *B*). After all, by the law of large numbers, if a large number, n, of people were placed in situations like *disease*, and the results of each situation were independent from the others, then everyone's choosing the riskier option would lead to $500n$ lives saved, whereas everyone's choosing the safer option would only lead to $400n$ lives saved.[11]

But this argument is flawed—or at least incomplete. First, we need a principle linking what agents ought to do *on a particular occasion* with what happens in the long run. Such a need is especially pressing given that many of these decisions are unlikely to be faced more than once: it's unlikely that someone will be in a situation exactly like *disease* or *game show* more than once. One might think that what happens in the long run is simply irrelevant for decisions that are unlikely to be repeated. Second, it's false that if everyone in a situation like *disease* were to choose option *A*, then around $500n$ people would *definitely* be saved. What the law of large numbers implies is that, as n increases, the average lives saved on each repetition *tends toward* 500 with probability 1. But it's still uncertain what outcome will actualize. With 100 repetitions, for example, we get the following distribution of possible outcomes and probabilities, centered around 50,000 lives saved: {100,000 saved, $(1/2)^{100}$; 99,000 saved, $100\times(1/2)^{100}$; 98,000 saved, $\frac{100\times 99}{2}\times (1/2)^{100}$; …; 1000 saved, $100\times(1/2)^{100}$; none saved, $(1/2)^{100}$}. Many of these outcomes are such that choosing option *B* would have led to more lives saved. So we can't say *with certainty* that the higher-EV option will lead to a better

[11] This argument emerges from time to time, although (as far as I know) there's no definitive statement of it in the literature. For a recent statement of the argument in the context of her REU theory, see Buchak (2013, 7.3).

outcome, even in the long run.

Of course, after a large number of independent repetitions, the probability that option *A* leads to a better outcome is very high; for example, there's only a 2.8% chance that, after 100 repetitions, option *A* leads to a worse outcome than option *B*. And one might think that this fact justifies choosing the EV-maximizing option. But this response runs the risk of begging the question: after all, to be risk-averse is just to give weight to negative outcomes disproportionate to their disvalue or their probability. So telling risk-averse agents that the probability of a worse outcome with the riskier option is very small in the long run might be unconvincing.

If appealing to the *finite* long run doesn't work for the EV-maximizer, perhaps he can appeal to the *infinite* long run. If people were either to choose option *A* or option *B* an infinite number of times, then the former would lead to a better outcome *with probability 1*. But a probability of 1 doesn't mean certainty.[12] After all, it's possible for the behavior of an actual series to depart from its limiting behavior: the ratio of heads in a series of tosses of a fair coin tends to 1/2 with probability 1, but the coin could keep turning up tails indefinitely. Similarly, even with an infinite number of repetitions of *disease*, option *A* could keep failing to save anyone. And, more importantly, it's unclear how an appeal to the consequences of actions when repeated an infinite number of times has any bearing on what an agent on a particular occasion should do, especially given the worries about repeatability mentioned earlier.

In the rest of this paper, I'll advance a refined version of the long-run argument that avoids the problems of this naive formulation. Such an argument aims to convince *all but extraordinarily risk-averse agents* that, when they face decisions under uncertainty, they morally ought to choose the option that maximizes expected moral value.

Let us take some preliminary steps before we begin the argument. First, we should replace talk of the consequences of a particular act, as if those consequences were known beforehand, with something that accommodates their uncertainty. So instead, we'll talk in terms of the *lotteries*, or sets of possible outcomes and their respective possibilities, yielded by the acts: so the lottery associated with the riskier response in *disease* is {1000 people saved, 1/2; no one saved, 1/2}; the lottery associated with going for the $1 million question is {win $1

[12] Those who endorse *the principle of regularity*, according to which every possible event has non-zero (possibly infinitesimal) probability, deny this. But if regularity is correct, then average value converges to expected value with some probability infinitesimally smaller than 1.

million, 1/10; win nothing, 9/10}. Second, since two rational agents might have different levels of risk aversion, hence have different preferences over the same set of lotteries, we won't say that an act yields a *better* lottery than another; instead, we'll say the lottery yielded by one is *preferable* to that yielded by another *for a particular rational agent*. (I'll use "rational" just to mean that the agent's preferences satisfy certain structural constraints—more on this in the next section.) I'll also sometimes talk about an agent's preferring one act or choice over another as shorthand for the agent's preferring the lottery yielded by the former to that yielded by the latter. Choosing the most preferable lottery becomes the probabilistic extension of maximizing value, and the generalization of maximizing expected value to cases where the agent isn't risk-neutral.

Next, we won't talk about the lotteries yielded by repeating a decision, since the decisions may not be likely to be repeated (that is, with the exact same possible lotteries). Instead, we'll talk in terms of *long series of decisions* that individual decisions are embedded in. The decision about which disease response to choose, for example, is part of the series of decisions under uncertainty that the agent faces across his lifetime, and it is also part of the set of decisions under uncertainty faced by agents collectively. So even though it might not make sense to talk about the long run in the sense of a large number of identical repetitions of some decision, we can still meaningfully talk about it in the sense of a long series of possibly non-identical decisions.

The structure of the next sections is as follows. In §2, I'll defend a claim about rational agents' preferences in the long run. When we think about decisions as one-offs, different agents have different preferences, even if they are all fully rational: in a one-off case, a risk-neutral agent would prefer option *A* in cases like *disease*, whereas a risk-averse one might prefer option *B*. Nonetheless, *when it comes to long enough series of decisions*, then given certain minimal conditions, the preferences of all but the most risk-averse agents converge: all agents who are not extraordinarily risk-averse will prefer the lottery yielded by our consistently choosing the option with the higher EV to that yielded by our consistently choosing the option with the lower EV, even if what those options are differs from occasion to occasion. In §3, I'll show how such a claim is relevant to what we ought do on a particular occasion by arguing that we should assess individual decisions indirectly, in terms of the larger series of decisions that they are part of.

2 Long-run preference

To begin the argument for maximizing expected moral value, I'll introduce and defend the following principle about rational agents' preferences over long series of decisions:

> **Long-run preference**: For any two series of lotteries $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ such that (1) the series are long enough, (2) the lotteries are all probabilistically independent of each other, (3) the series are well-behaved, and (4) and for all i, A_i has higher expected value than B_i, a rational agent prefers the compound lottery that results from accepting $A_1, \ldots, A_n$ to the compound lottery that results from accepting $B_1, \ldots, B_n$, so long as the agent is not extraordinarily risk-averse.

To motivate **long-run preference**, consider a special case: one in which a rational agent has a choice between a large number, n, of independent repetitions of some risky lottery A and that many independent repetitions of some safer, but lower-EV lottery B. (This is just the case where $A_1 = \ldots = A_n$ and $B_1 = \ldots = B_n$.) In this case, **long-run preference** says that so long as an agent is not extraordinarily risk-averse, then even if he prefers B to A as a one-off, he will prefer n repetitions of A to n repetitions of B. For example, although we might prefer the status quo over a one-time lottery in which we win \$200 with probability 1/2 and lose \$100 with probability 1/2, we would probably prefer a series of 100 such lotteries to the status quo. Although we might prefer to save four people with certainty over having a 50-50 chance of saving ten, we would probably prefer choosing the latter option 100 times over choosing the former option 100 times. In the long run, all but extraordinarily risk-averse agents will prefer the lottery with the higher expected value.[13]

Long-run preference generalizes this claim to the case where the lotteries on subsequent occasions are not necessarily identical. If a rational agent is presented with a long series of decisions, each of which is between a riskier but higher-EV

[13]The condition that the probabilities on subsequent repetitions be independent from one another is important. In the case where the lottery is {win \$200, 1/2; lose \$100, 1/2}, if a single coin toss at the beginning determines the result of all of the repetitions, then the only possible outcomes of the 100 repetitions are: {win \$20,000, 1/2; lose \$10,000, 1/2}. No one who would not accept the one-time lottery would accept this lottery either.

For this reason, the argument for maximizing expected moral value in the case of empirical uncertainty cannot be extended to the case of moral uncertainty: what happens on one occasion may be independent from what happens on another occasion, but what the true moral theory is on one occasion is not independent from what the true moral theory is on another occasion.

option and a safer but lower-EV option, then so long as the agent is not extraordinarily risk-averse and some other minimal conditions are satisfied, the agent will prefer the lottery that results from her consistently choosing the higher-EV option to the lottery that results from her consistently choosing the lower-EV one, even if what the higher- and lower-EV options are is different on each occasion.[14]

The basic idea behind the principle is simple. It's been understood for a long time that even if people are risk-averse, so that they would reject one-time lotteries with positive expected monetary value, because of how independent risks reduce when added, people may accept compound lotteries that are formed by adding many of the one-time lotteries. This is, after all, the basis for insurance. From the perspective of the policyholder, paying a small amount each month is better than standing to lose a large amount with a small probability, even if the expected monetary value of the policy is negative. From the perspective of the insurance company, even if the company is as risk-averse as its policyholders, it's rational to sell a large number of such policies (that is, accept a large number of lotteries that are risky but have positive expected monetary value). After all, if the risks are independent, the company is almost guaranteed to come out ahead.

In fact, a formal result similar to **long-run preference** has been proven in the framework of Von Neumann–Morgenstern expected-utility theory (von Neumann and Morgenstern 1953). In order to state the formal result, I'll first have to rehearse the expected-utility framework. In this framework, risk aversion (and risk attitudes in general) are captured through a device for representing agents' preferences known as a *utility function*, $U(\cdot)$, which associates each possible outcome with a real number representing its choiceworthiness for the agent. It's crucial to note that "utility" here doesn't mean what moral philosophers use it to mean, moral value. Rather, on the orthodox version of the expected-utility framework, the utility function of an agent is simply a theoretical construct de-

[14]Obviously, generalizing the result requires introducing additional requirements on the series, which I've lumped together under the vague requirement that they be "well-behaved." The requirement is meant to rule out improbable cases like those where one decision has stakes that swamp all others, like a decision between the lottery {one million people live, 1/2; no one lives, 1/2} and {400,000 people live, 1} in the context of the decisions that an ordinary person would be likely to face. After all, if one decision swamps all others, then the lottery yielded by the entire series of choices would be dominated by that one decision, so that someone who prefers the safer choice in that decision might prefer the entire safer series of choices. Similarly, it rules out outlandish cases like one where the difference in the expected value between the A_i's and the B_i's asymptotically approaches 0, but the A_i's do not become less risky, so that each additional A_i makes the entire series less choiceworthy relative to each additional B_i. But most of the decisions that we ordinarily face under uncertainty are not part of such pathological series.

signed to capture the agent's preferences over possible outcomes, on the assumption that those preferences satisfy some basic conditions of rationality.[15] (That the agent must satisfy these conditions on rationality secures the role of expected-utility theory as a *normative* theory.)

According to the framework, a utility function that represents the agent's preferences is such that *the agent always prefers the option that maximizes expected utility*. What this means is as follows. Suppose that L is a lottery with outcomes $A_1, \ldots, A_m$ with probabilities $p_1, \ldots, p_m$, and M is a lottery with outcomes $B_1, \ldots, B_n$ with probabilities $q_1, \ldots, q_n$. Then the expected value of the utility function, $EU(\cdot)$, is such that the agent prefers L to M if and only if

$$EU(L) = \sum_{i=1}^{m} p_i U(A_i) > EU(M) = \sum_{i=1}^{n} q_i U(B_i).$$

Colloquially, the expected utility of a lottery is the average of the utilities of its outcomes, weighted by their probabilities. A rational agent prefers one lottery to another if and only if the expected utility of the first is higher.

Here's an example that shows how expected-utility theory captures attitudes to risk. Consider choosing between the following two lotteries: in L, you win \$100 (for certain); in M, you have a 50% chance of winning \$200, and a 50% chance of winning nothing. Although they have the same expected monetary value, most of us, being risk-averse, would prefer L to M. Expected-utility theory captures this by assigning to us a *concave* utility function, that is, one that "opens downward"—mathematically, one whose second derivative is negative. What this means is that the utility of winning \$200 is less than twice the utility of winning \$100. For example, suppose the utility function $U(x) = \sqrt{x}$ represents our preferences. The expected utility of L is $1 \times \sqrt{100} = 10$, whereas the expected utility of M is $.5 \times \sqrt{200} + .5 \times \sqrt{0} = 7.1$. So even though both lotteries have the same expected monetary value, the non-risky lottery has a higher expected utility, which captures the fact that we prefer that lottery.

[15] See, for example, von Neumann and Morgenstern (1953), Arrow (1951), Luce and Raiffa (1957). von Neumann and Morgenstern (1953, 28) write: "We have practically defined numerical utility as being that thing for which the calculus of mathematical expectations is legitimate."

The conditions that an agent's preferences must satisfy are typically taken to be: (1) *completeness*, that for any pair of lotteries, the agent prefers one to the other or is indifferent between them; (2) *transitivity*, that the preference relation is transitive; and (3) *continuity*, that for any three lotteries, the agent is indifferent between some probabilistic combination of the best and worse lotteries and the third lottery.

Again, utilities in the decision-theoretic sense are theoretical constructs designed to represent the preferences of a rational agent, rather than anything of direct moral relevance. It's important not to confuse the idea that rational agents maximize expected utility, which is true tautologically (being equivalent to the claim that agents whose preferences satisfy certain conditions have preferences that satisfy those conditions), with the claim that we ought to maximize expected moral value, which is a substantive moral claim. It's conceptually possible for the former to be true and the latter false. For example, we might be morally required to maximize the logarithm of expected moral value, given some measure of moral value; in doing so, however, so long as our preferences are complete, transitive, and continuous, we can still be represented as maximizing expected utility.[16]

Having rehearsed expected-utility theory, we're in a position to discuss the formal result similar to **long-run preference**, which I'll call **long-run acceptance**. Colloquially, the result says that for a long enough series of lotteries with positive EVs, even if a rational agent rejects each lottery as a one-off because of its riskiness, unless the agent is extraordinarily risk-averse, she will accept the sum of all of those lotteries.[17]

More formally, suppose that there is some series of lotteries $X_1, X_2, X_3, \ldots$ such that (1) each lottery has positive EV, (2) the lotteries are independent of each other, and (3) the series is well-behaved; and that some agent has preferences representable by $U(\cdot)$. Let's call $X_1 + \ldots + X_i$ the lottery yielded by an agent's

[16]For this reason, some decision theorists write that it is misleading to say that decision theory enjoins us to maximize expected utility. Jamie Dreier (1996, 253), for example, writes,

> It is, I think, very misleading to think of decision theory as telling you to maximize your expected utility. If you don't obey its axioms, then there is no utility function constructable for you to maximize the expected value of. If you do obey the axioms, then your expected utility is always maximized, so the advice is unnecessary. The advice, 'Maximize Expected Utility' misleadingly suggests that there is some quantity, definable and discoverable independent of the formal construction of your utility function, that you are supposed to be maximizing.

[17]See Ross (1999). Ross actually proves a version of the result where acceptance is not necessarily monotonic: even if an agent accepts some long series of lotteries, he may not accept a continuation of that series with other positive-EV lotteries. He notes that whether an agent with a particular utility function will monotonically accept a particular series of lotteries depends on, in addition to that utility function, features of that series. I've added the "well-behaved" condition on the series to restrict the result to those for which acceptance is monotonic. Nielsen (1985), Lippman and Mamer (1988) proved a special case of this result, for independent repetitions of the same lottery, although theirs features monotonic acceptance.

accepting all of the lotteries $X_1,\ldots,X_i$. (So if each X_i is a lottery in which you win \$200 if a fair coin lands heads, and lose \$100 if it lands tails, then $X_1+\ldots+X_{100}$ is the lottery that results from flipping a fair coin 100 times, winning you \$200 for each heads and losing you \$100 for each tails.) **Long-run acceptance** says:

> There is a threshold number m such that for all $n > m$, $EU(X_1 + \ldots + X_n) > U(\varnothing)$, so long as $\lim_{x\to-\infty} U(x)e^{ax} = 0$ for all $a > 0$.

That is, the agent will always accept a sufficiently long series of lotteries with positive EVs, so long as some minimal conditions are met and the utility function representing her preferences doesn't grow exponentially or faster in the negative direction. This condition on the agent's utility function formalizes our talk of not being "extraordinarily risk-averse."[18]

The reader is advised to consult the papers cited for a formal proof of this result, but the intuition is as follows. Because each lottery has positive expected value, as the number of lotteries increases, the distribution of possible outcomes will be shifted in the positive direction without a proportional increase in its spread: in the special case of identical lotteries, for example, the mean outcome will grow proportionally to n, but the standard deviation only proportionally to $\sqrt{n}$. This means that the probability of getting an outcome worse than the status quo decreases as the number of lotteries increases. Unless the agent's utility function grows faster in the negative direction than the probability decreases, the expected utility of the series will increase as the number of lotteries increases.

Note that **long-run acceptance** concerns accepting a series of lotteries rather than rejecting it: that is, preferring that series to the status quo. **Long-run preference**, as I've defined it, concerns preferring a series to another series. But it's easy to extend the former to the latter. Intuitively, take two series of lotteries, $A_1,\ldots,A_n$ and $B_1,\ldots,B_n$, where A_i has greater expected value than B_i for all i. For each pair of A_i and B_i, consider a third lottery C_i, formed by receiving lottery A_i and *offering* lottery B_i to someone else (i.e., accepting a lottery whose outcomes are the negatives of the outcomes of B_i). Because A_i has greater expected value than B_i, this third lottery has positive expected value. As a result,

[18] Having an exponential utility function seems at least sufficient for being extraordinarily risk-averse in the ordinary sense. After all, if someone has an exponential utility function and is indifferent between losing 50 cents for certain or having a one-in-two chance of losing a dollar (so has preferences representable by $U(x) = -1/4^x$), then he will prefer to lose five dollars for certain rather than face a one-in-1,000 chance of losing ten dollars, and will prefer to lose 50 dollars for certain rather than face a one-in-10^{30} chance of losing 100 dollars. Very few, if any of us, exhibit such levels of risk aversion.

given **long-run acceptance**, if n is large enough, the agent will accept the series $C_1, \ldots, C_n$, assuming that he is not extarordinarily risk-averse. But accepting this series is equivalent to accepting $A_1, \ldots, A_n$ and offering someone else $B_1, \ldots, B_n$. If the agent accepts such a combination of lotteries, it seems clear that he prefers $A_1, \ldots, A_n$ to $B_1, \ldots, B_n$.[19]

While we used the framework of expected-utility theory to argue for **long-run preference**, I want to note that the plausibility of the principle is not tied to that theory. A rival approach to decision under uncertainty, like Lara Buchak (2013)'s risk-adjusted expected-utility (REU) theory, would do as well. I'm using the standard theory out of need for specificity, and for ease of exposition. In fact, theories like REU are more obviously accommodating of **long-run preference**, since they were *designed* in part to capture the intuition that people might rationally reject risky lotteries with positive expected values but accept compounds of those lotteries.

I also want to note that nothing in the argument relies on the idea that decisions need to be repeatable, where we individuate decisions based on the set of attainable lotteries. After all, **long-run preference** does not require that the *same* lotteries be repeated over and over; all it requires is that the choice between the two lotteries appear in a long enough series of decisions between higher-EV and lower-EV lotteries that satisfy some minimal conditions. As long as these requirements are satisfied, then all but the most risk-averse agents would prefer that the higher-EV option be chosen consistently. This is one advantage that the current argument has over the intuitive long-run argument for maximizing expected value. Again, recall that the intuitive argument went in terms of the consequences of repeating particular decisions a large number of times; it was vulnerable to the objection that many decisions are unlikely to be repeated, so it is unclear what the relevance of the long run to what an agent should do on a particular occasion is. The present argument does not assume repeatability; in its stead, all it assumes is that individual decisions under uncertainty are embedded

[19] More formally, suppose that the agent has already accepted some lottery B. Consider the lottery C, composed by accepting some lottery A (with higher expected value than B) and paying whatever payoff lottery B produces. Suppose that the agent now accepts C. Since he has already accepted B, he prefers to have A and B and to pay whatever B yields to having B. But having A and B and paying the payoff from B is equivalent to having A. So the agent prefers A to B. Let $A = \sum_{i=1}^{n} A_i$, $B = \sum_{i=1}^{n} B_i$, where the A_i's and B_i's are as above. Then $C = \sum_{i=1}^{n} C_i$, where each C_i consists of accepting A_i and paying the payoff from B_i, which the agent has already accepted. Since each C_i has positive expected value, **long-run acceptance** says that the agent will accept C, even if he has already accepted B. But this is equivalent to preferring A over B, which is **long-run preference**.

in larger series of decisions under uncertainty that agents face.

What does **long-run preference** say in the moral case? Suppose we have two options A and B, the former of which has higher expected moral value: choosing A leads, on expectation, to more lives saved, or welfare, or happiness, or preference-satisfaction. Even so, A might have a riskier profile, and we might morally prefer B over A on a one-off basis. Nonetheless, if such a decision is part of a larger series of decisions between riskier but higher-EV and safer but lower-EV options, all except the extraordinarily risk-averse among us would morally prefer that the higher-EV option be chosen consistently, so that A would be chosen over B.

As an example, consider *disease* again. In the one-off case, many of us would prefer the less risky response, although it leads to fewer lives saved on expectation. But imagine that such a decision is part of a larger series of decisions between riskier but higher-EV and safer but lower-EV outcomes: for example, imagine if decisions similar to this one were faced hundreds of times by disease control agencies around the world, even if the details varied on each occasion. Given this series of decisions, it is vanishingly unlikely that their always choosing the safer option would save more lives than their choosing the riskier option, so long as the risks are independent. Of course, there's a tiny chance that everyone's choosing the risky option would lead to more lives being lost. But unless an agent is extraordinarily risk-averse, **long-run preference** implies that the unchoiceworthiness of that unlikely bad outcome is outweighed by the choiceworthiness of the very likely outcome in which it would lead to more lives saved. So nearly all of us would prefer, on moral grounds, that everyone choose the riskier, higher-EV option over the safer, lower-EV one.[20]

[20]I want to note that, although we have been assuming for simplicity that future wellbeing is as valuable as current wellbeing, the argument does not need such an assumption. After all, the decisions we are concerned with are those between options that have their effects at the same time. So even if we applied a discount rate to the future, that will affect all options equally, and typically (depending on the agent's utility function) will not change an agent's preferences between them. For example, suppose that an agent has preferences representable by $U(x) = \sqrt{x}$, and that one unit of wellbeing ten years from now is worth half a unit of wellbeing right now. Then the decision between two lotteries A and B whose consequences occur in ten years is equivalent to that between lotteries A' and B' that take effect now, where the outcomes A' and B' are those of A and B divided by two. $EU(A') = EU(A)/\sqrt{2}$ and $EU(B') = EU(B)/\sqrt{2}$, so $A' \succ B'$ iff $A \succ B$.

3 Indirect evaluation

So far, I've argued that, in considering long enough series of decisions under uncertainty, consistently choosing the higher-EV option is preferable to choosing otherwise to all but the most risk-averse agents. In order for this to be relevant for what an agent should do on a particular occasion, it must be the case that we should evaluate individual choices in what I'll call an *indirect* way, by *first* evaluating larger sets or series of choices, and *then* evaluating the individual choices in terms of the larger set of choices they belong to.

Why evaluate choices indirectly? Well, no decision under uncertainty is an island; each decision we face is embedded in a larger series of decisions, like the set of decisions we face over an entire lifetime, or the set of decisions faced by agents altogether. As we'll see, even if each member of some set of choices is optimal as a one-off, the entire set of choices might be much worse than some other set of choices. So instead of viewing decisions in isolation from each other, we should look at some kind of larger series of decisions that each decision is embedded in, and choose in accordance with the *entire series of choices* that yields the most preferable lottery.

In this section, I'll defend two separate indirect ways of evaluating the decisions that an agent makes under uncertainty: first, in terms of the larger series of decisions that the agent faces across his life; second, in terms of the set of decisions that agents as a whole face.

3.1 Resolute choice

First, consider the series of decisions under uncertainty that an agent will face throughout his lifetime. Suppose that the agent prefers that, over the course of his lifetime, he choose some series of options over any other series; in our case, suppose that he prefers that he always choose the option that maximizes expected moral value. Now, suppose that the agent does not take this to imply that, on each particular occasion, he ought to choose in accordance with his long-term preference: on each decision, he ignores his long-term preference for his choosing the higher-EV option consistently, and chooses a safer, lower-EV option. There seems to be something irrational about such an agent, as if he cannot see how his individual decisions aggregate. After all, if on each particular occasion, he chooses contrary to his long-term preference, then his choices taken as a whole amount to his doing exactly what he *dis*prefers.

If we want to avoid such irrationality, then we cannot think of our decisions in isolation. Rather, we should think of ourselves as in a *dynamic choice* problem,

one in which an agent decides in light of the fact that his decision is embedded in an entire series of decisions that he faces over time. The principle of dynamic choice that I will endorse here is known as **resolute choice**, that an agent should first choose the *long-term plan* that yields the most preferable lottery, and then, on each occasion, choose as prescribed by the plan he has adopted, even though doing so might go against his one-time preferences (McClennen 1990). For example, suppose an agent prefers the lottery yielded by the plan to maximize EV whenever he faces a decision under uncertainty to any other achievable lottery. **Resolute choice** says that he ought to follow through with the plan, so that, on each occasion, he does the thing that maximizes EV.

It should be clear how someone who adopts **resolute choice** evaluates decisions indirectly. Instead of choosing the option that, considering the case as a one-off, yields the best lottery, the agent evaluates each option by whether it accords with the long-term plan that yields the best lottery. Although a detailed defense of **resolute choice** is beyond the scope of this paper,[21] one consideration in favor of it is that, at least in the kinds of cases we are considering, it avoids forms of irrationality caused by following two rival principles of dynamic choice, **myopic choice** and **sophisticated choice**. **Myopic choice** has an agent, in each decision, choose whichever long-term plan she prefers at that point and make individual choices accordingly. The problem is that the plan that she prefers may change: in the context of a series of decisions between a riskier, higher-EV option and a safer, lower-EV option, she starts off preferring the plan to choose the former consistently; but when she gets toward the end of the series, the remaining series is no longer long enough for **long-run preference** to hold, and she will prefer the plan to choose the latter consistently. She thus ends up being diachronically inconsistent, failing to adopt the same plan throughout.

Sophisticated choice has an agent think about the choices she is likely to make in each possible scenario, starting from those at the end of the series and moving backward in time, and make choices so that, *given those later choices*, the entire series of choices yields the best lottery. **Sophisticated choice** has an advantage over **resolute choice** in cases in which the agent is unlikely to follow through with the plan that yields the best lottery. Suppose that an old acquaintance from graduate school is giving a talk in my department. It might be best if I attend the talk (which I know will be bad) and ask a friendly question during the Q&A; but given that I won't be able to help myself and will make a scathing remark instead if I attend the talk, which would be worse than if I don't attend

[21] For defenses, see McClennen (1990, 1997), Gauthier (1997).

the talk in the first place, I should not attend the talk at all.

Note that the scenarios that we are considering, however, are not like this. I can easily stick to a plan to maximize expected moral value consistently, since there is nothing psychologically unfeasible about sticking to it.[22] And **sophisticated choice** has well-known problems, especially in cases like these. After all, consider a moderately risk-averse agent with the following preferences: (1) following **long-run preference**, he prefers that he choose the entire series of riskier but higher-EV options to the series of safer but lower-EV options; (2) in a one-off case, he prefers the safer option over the riskier one; and (3) in each decision, if he anticipates that he will consistently choose the safer option in all later decisions, he prefers the lottery yielded by those choices plus choosing the safer option in the present decision to that yielded by those choices plus choosing the riskier option. Now, suppose such an agent uses **sophisticated choice**. When the agent only has one decision remaining (say, toward the end of his life), he will prefer a lower-EV but safer option over a higher-EV but riskier one, and will end up choosing the former. When he has two decisions remaining, knowing that he will choose the safer option in the last decision, he chooses the safer option. Applying this reasoning recursively, at the outset, he knows that he will choose the safer option in all later decisions, and he chooses the safer option. The effect of his exercising **sophisticated choice** is that he will adopt a plan whose lottery he *dis*prefers to that yielded by another plan, one that is completely feasible. Such an agent is guilty of a kind of irrationality, that of performing a suboptimal series of acts when a clearly better, and completely feasible, series of acts is available to him.[23] This is not to say that there are not problems with **resolute choice**; but at least in the context of the cases that we are considering, **resolute choice** is preferable to **sophisticated choice**.[24]

[22]One might object that if **resolute choice** will force the agent to act against his preferences toward the end of the series, then the agent at that point may no longer have reason to stick to the plan; and if this is so, then it may be unfeasible for a rational agent to stick to the long-term plan. I address this objection in §4.

[23]Thoma (2019) offers a similar argument against sophisticated choice in this kind of case.

[24]More generally, McClennen (1990) argues that none of the forms of dynamic choice can satisfy all of three desiderata: (1) *diachronic consistency*, that the agent adopts the same plan throughout, (2) *normal form–extensive form convergence*, that the agent would choose the same plan if all of his choices were controlled by an initial single choice, and (3) *separability*, that in each decision, the agent is permitted to ignore historical background. **Myopic choice** violates 1, **sophisticated choice** violates 2, and **resolute choice** violates 3. McClennen argues that the violation of separability is the least bad option; in later work (McClennen 1997), he argues against taking separability to be a requirement on dynamic choice at all.

3.2 Rule consequentialism

Resolute choice requires the agent to think about his decisions more expansively: the agent must pay attention to the entire series of decisions he faces when making any particular choice. But it is possible to think about an even more expansive context of decision: we might think about the set of decisions faced by agents as a whole. What are the principles of choice that we should endorse in this context?

We can generalize, to some extent, the discussion of principles of dynamic choice in the last section to address this question. Just as an agent can evaluate decisions as one-offs, in isolation of any larger series of decisions he faces, he can also evaluate them in isolation of any decisions faced by other agents. Assuming that his moral preferences over different options are determined entirely by morally relevant features of the lotteries yielded, he might endorse a principle like the following:

> One should choose the option that, considered in isolation, yields the most preferable lottery.

We can label this principle **naive act consequentialism**, keeping in mind the differences from what usually goes by that label of "act consequentialism"; for one, we are restricting our attention here only to cases in which none of the options involves violating preexisting moral requirements. Nonetheless, just like traditional act consequentialism in normative ethics, **naive act consequentialism** assesses a choice solely in terms of the (probabilistic) consequences of that choice.

And, just as there are problems with viewing our decisions in isolation from our other decisions, so too are there problems with viewing them in isolation from the decisions of others. In the case of moderately risk-averse agents each facing a choice between a riskier, higher-EV option and a safer, lower-EV option, each might prefer the latter in a one-off decision; given **long-run preference**, however, each prefers that they *all* choose the riskier option. If they follow **naive act consequentialism**, then they will all choose the safer option, and the result is that their choices together yield a lottery that everyone strongly disprefers. While this might not constitute a form of collective *irrationality*, we still have strong reason to avoid this outcome, so we should not endorse **naive act consequentialism**.

One might think that we can avoid this problem by reformulating act consequentialism so that agents pay attention to how other agents are likely to act:

> One should choose the option that, in combination with those that other agents will choose, yields the most preferable lottery.

This principle is roughly analogous to **sophisticated choice**, so we might call it **sophisticated act consequentialism**. And again, there are problems analogous to those for **sophisticated choice**. Regardless of how he expects other agents to choose, a moderately risk-averse agent might prefer the combination of those choices and his choosing the safer option to the combination of those choices and his choosing the riskier option. If each agent follows **sophisticated act consequentialism**, then regardless of how he expects others to choose, he will choose the safer option, and the collective result is again some collective choice that all agents strongly disprefer.[25]

Finally, consider the following principle, analogous to **resolute choice**:

> One should choose the option prescribed by the rule whose being followed by everyone would yield the most preferable lottery.

Given that this principle assesses rules in terms of the lotteries they yield, and assesses individual choices based on whether or not they accord with the best rule, I will call it **rule consequentialism**. In fact, **rule consequentialism** is just a probabilistic generalization of what usually goes under that label, the claim that one should perform the act prescribed by the rule whose universal adoption would have the best consequences (Brandt 1959, Hooker 2000); we simply replace talk of the (deterministic) consequences of acts with talk of the (probabilistic) lotteries yielded by choices, and talk of the best consequences with talk of the most preferable lottery. **Rule consequentialism** avoids the problem of agents' collectively making a strongly dispreferred choice that **naive** and **sophisticated act consequentialism** faced: if each agent follows **rule consequentialism**, then each will choose in accordance with the rule to maximize EV (since, by **long-run preference**, that rule yields the most preferable lottery), so that they collectively choose the set of higher-EV options, just as each agent prefers. Just as with **resolute choice**, endorsement of **rule consequentialism** does not imply that there

[25] If we do not assume that risk-averse agents prefer the safer option *regardless* of how others choose (which is inconsistent with **long-run preference** on EU theory), then each agent prefers the riskier option if and only if enough other agents choose the riskier option. If each agent believes that not enough other agents will choose the riskier option, then everyone will choose the safer option, so we have the same problem as before. On the other hand, if each agent believes that enough other agents will choose the riskier option, then everyone will choose the riskier option, and **sophisticated act consequentialism** will coincide in its recommendations with **rule consequentialism**.

are no problems with the view; nonetheless, in the context of decision under uncertainty, it avoids a serious problem that affects other principles of decision.

As I've shown, then, **resolute choice** and **rule consequentialism** are both indirect ways to assess decisions under uncertainty: we first identify the best plan or policy for making choices, then choose accordingly in individual decisions. Together with **long-run preference**, both imply that, when an agent faces a decision under uncertainty, he morally should choose the option with the higher expected moral value.

4 Objections

To close, I will discuss and respond to some objections to the overall argument.

First, one might not be convinced by the justifications that I offered for the indirect modes of evaluation discussed in the last section, **resolute choice** and **rule consequentialism**. In the case of the justification for **rule consequentialism**, perhaps it is convincing when addressed to an entire set of agents: it shows that the set of agents have reason to follow **rule consequentialism**. But it doesn't follow from this that an individual agent has such reason. After all, a particular agent may not be in a position to influence what other agents do; and regardless of what they do, he prefers the combination of their choices and his making the safer choice. So why should the group-based justification for **rule consequentialism** move him? And in the case of the justification for **resolute choice**, we can imagine an agent who, previously having adopted **resolute choice**, is now nearing the end of her life. At this point, she knows that she will not face enough decisions for **long-run preference** to apply, so she prefers that she choose the safer, lower-EV option in all remaining decisions. It is true that choosing so would violate the plan that she adopted at the outset; but the self who adopted that plan is long gone, so why be bound by it?

To respond conclusively to this objection goes beyond the scope of this paper, but let me sketch a response. Let's first consider the response in the case of **rule consequentialism**. It may be that *my* deviating from the best rule yields a lottery that is preferable to that yielded by my following the rule. Nonetheless, I know that *everyone*'s deviating from that rule would yield a much worse lottery. I do not endorse this, so to justify my deviating from the rule, I need to find some reason why I alone should be excused from it. But there is no such reason: my position is identical to that of every other agent, who is aware of the same considerations that I am aware of. I am thus seeking an exception for myself from some rule that we would all endorse, even though there are no grounds

for that exception. In doing so, I am failing to see myself as an equal to other agents, which is *ipso facto* morally objectionable. If someone has a conception of morality as a set of rules that reasonable agents who regard each other as equals would accept, he will not be moved by the consideration that his violating one of these rules leads to a better result.[26]

What about the response in the case of **resolute choice**? Just as a willingness to break the best set of rules expresses a morally deficient conception of one's relation to others, one might think that a willingness to go against one's long-term preferences expresses a deficient conception of the relation between different temporal stages of the self. After all, it seems plausible that a requirement on thinking of oneself as an agent who exists across time is to form long-term preferences and to subordinate one's immediate preferences to them. If I'm unwilling to endure the momentary discomfort of exercise for the sake of my long-term goal to stay healthy, then in some sense I'm not thinking of the future self that will benefit from my current discomfort as *me*. Given that, from the perspective of the agent's entire life, the agent prefers the series of choices in which she chooses the higher-EV option consistently, those preferences must have some uptake in the agent's deliberations in each decision in order for her to think of herself as an agent who exists across time at all, rather than as a momentary self alienated from the past selves who made the earlier decisions. To the extent that the agent conceives of herself diachronically, she must take herself to be bound to the plans that express her long-term preferences, and that speaks in favor of **resolute choice**.[27]

Next, I want to discuss several related objections that concern the use of **resolute choice** in the overall argument. One might first object that they fail to show that agents should *always* maximize expected moral value. Consider someone toward the end of his life, who has only a few decisions under uncertainty remaining. Since the series of remaining decisions is too short for **long-run preference** to apply, it may very well be that the morally best plan for the agent to

[26] I am alluding here to a broadly *contractualist* justification of rule consequentialism, according to which moral principles are those that all reasonable agents would accept as those for a system of social interaction. Although the most familiar form of contractualism (Scanlon 1998) is quite different from rule consequentialism, Derek Parfit (2011) has (controversially) argued that rule consequentialism and the most plausible form of contractualism converge. And given persistent doubts about whether Scanlon's version of contractualism can restrict how much aggregation it allows (Norcross 2002, Kumar 2011), it may be that it collapses to rule consequentialism too.

[27] McClennen (1997) makes a similar point in greater detail, although in the language of coordination between different temporal stages rather than subordination to some temporally-extended self.

adopt is one on which he chooses the safer option in each decision. **Resolute choice** would then advise him, in each decision, to choose the safer option, even if it has lower expected moral value.

Second, continuing from this example, contrast this agent with someone toward the beginning of her life. Given that this younger agent likely has a large number of decisions under uncertainty remaining, **long-run preference** will apply to that series of decisions, so that the best plan for her to adopt is one on which she chooses the higher-EV option in each decision. Now, the two agents might find themselves in identical scenarios, but, given that the best plan for one is different from the best plan for the other, **resolute choice** says that what one ought to do differs from what the other ought to do: the younger agent should choose the riskier response in a case like *disease*, while the older agent should choose the safer one. But the mere fact that an agent has adopted some long-term plan does not seem like it should make a difference for what that agent should do in such an important decision.

My response to the first objection is simply to concede the point: the argument through **resolute choice** does not show that such an agent would have reason to maximize expected moral value. Indeed, the conditions on **long-run preference** mean that there are other cases in which an agent is not obligated to do so: for example, if there is one decision whose stakes swamp the combined stakes of all of the other decisions. Nonetheless, the argument still shows that, for almost all agents and in almost all decisions, the agent ought to choose the option that maximizes expected moral value. Given the fact that most people are risk-averse, such a conclusion still has highly revisionary consequences for how people should make moral decisions under uncertainty.

In response to the second objection, I want to note that the fact that **resolute choice** entails that what an agent should do in a particular decision depends not just on intrinsic features of the decision, but also on what the agent has decided in the past is a consequence of the well-known fact that **resolute choice** violates a principle known as *separability*: that, whenever the agent faces a decision, she is permitted to think about it in isolation from what she has already decided. Rejecting separability is not as counterintuitive as it seems. For one, there are cases in which we think our past commitments carry some normative weight: committing oneself to do something because one has promised to do so certainly gives us moral reason to do that thing. For another, as I've already mentioned, taking oneself to be bound by commitments formed in the past might be a prerequisite for thinking of oneself diachronically in important ways. So it is unclear why we have to take a temporally narrow view of the decision when we

deliberate morally.

Finally, while the argument might appear to have these counterintuitive consequences, this is a result of its conditional nature. Again, the argument aims to show that *even if moderate risk aversion is permissible*, agents should still typically maximize expected moral value. It is compatible with this conditional that no attitude besides perfect risk neutrality is permissible, from which it directly follows that one should always maximize expected moral value. To the extent that assuming the permissibility of risk aversion leads to counterintuitive consequences, such a consideration could form the basis of an argument against risk aversion itself.

References

Kenneth Arrow. "Alternative Approaches to the Theory of Choice in Risk-Taking Situations." *Econometrica*, 19:404–437, 1951.

Nick Bostrom. "Pascal's Mugging." *Analysis*, 69:443–445, 2009. doi: 10.1093/analys/anp062.

Richard Brandt. *Ethical Theory*. Prentice-Hall, 1959.

Lara Buchak. *Risk and Rationality*. Oxford University Press, 2013.

James Dreier. "Rational Preference: Decision Theory as a Theory of Practical Rationality." *Theory and Decision*, 40:249–276, 1996. doi: 10.1007/bf00134210.

David Gauthier. "Resolute Choice and Rational Deliberation." *Noûs*, 31:1–25, 1997. doi: 10.1111/0029-4624.00033.

Brad Hooker. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford University Press, 2000.

Frank Jackson. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics*, 101:461–482, 1991. doi: 10.1086/293312.

Rahul Kumar. "Contractualism on the Shoal of Aggregation." In R. Wallace, R. Kumar, and S. Freeman, editors, *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*. Oxford University Press, 2011.

Steven Lippman and John Mamer. "When Many Wrongs Make a Right." *Probability in the Engineering and Informational Sciences*, 2:115–127, 1988. doi: 10.1017/S0269964800000668.

R. Duncan Luce and Howard Raiffa. *Games and Decision*. Wiley, 1957.

William MacAskill, Krister Bykvist, and Toby Ord. *Moral Uncertainty*. Oxford University Press, 2020.

Edward McClennen. *Rationality and Dynamic Choice*. Cambridge University Press, 1990.

Edward McClennen. "Pragmatic Rationality and Rules." *Philosophy and Public Affairs*, 26, 1997. doi: 10.1111/j.1088-4963.1997.tb00054.x.

Bradley Monton. "How to Avoid Maximizing Expected Utility." *Philosophers' Imprint*, 19, 2019.

Lars Tyge Nielsen. "Attractive Compounds of Unattractive Investments and Gambles." *The Scandinavian Journal of Economics*, 87(3):463–473, 1985. doi: 10.2307/3439996.

Alastair Norcross. "Contractualism and Aggregation." *Social Theory and Practice*, 28:303–314, 2002. doi: 10.5840/soctheorpract200228213.

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

Derek Parfit. *On What Matters*. Oxford University Press, 2011.

John Rawls. *A Theory of Justice*. Harvard University Press, 1971.

Chelsea Rosenthal. "What Decision Theory Can't Tell Us About Moral Uncertainty." *Philosophical Studies*, forthcoming. doi: 10.1007/s11098-020-01571-3.

Stephen Ross. "Adding Risks: Samuelson's Fallacy of Large Numbers Revisited." *Journal of Financial and Quantitative Analysis*, 34:323–339, 1999. doi: 10.2307/2676262.

Thomas Scanlon. *What We Owe to Each Other*. Belknap Press of Harvard University Press, 1998.

Andrew Sepielli. "What to Do When You Don't Know What to Do." *Oxford Studies in Metaethics*, 4:5–28, 2009.

Johanna Thoma. "Risk Aversion and the Long Run." *Ethics*, 129:230–253, 2019. doi: 10.1086/699256.

Amos Tversky and Daniel Kahneman. "The Framing of decisions and the psychology of choice." *Science*, 211, 1981. doi: 10.1126/science.7455683.

John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 2nd edition edition, 1953.