8 Risk

8.1 Why maximize expected utility?

So far, we have largely taken for granted that rational agents maximize expected utility. It is time to put this assumption under scrutiny.

In chapter 1, I gave a simple initial argument for the MEU Principle. An adequate decision rule, I said, should consider all the outcomes an act might bring about – not just the best, the worst, or the most likely – and that it should weigh outcomes in proportion to their probability, so that more likely outcomes are given proportionally greater weight.

In chapter 5, we looked at the internal structure of utility. I didn't mention it at the time, but the account we developed can be used to support the MEU Principle. Consider a schematic decision problem with two acts and two states.

	S_1	S_2
A	O_1	O_2
B	O_3	O_4

Let's assume the four outcomes are logically incompatible with each other. (We can always make them incompatible by describing the outcomes in more detail.) By Jeffrey's axiom,

$$\mathrm{U}(O_1 \vee O_2) = \mathrm{U}(O_1) \cdot \mathrm{Cr}(O_1 \,/\, O_1 \vee O_2) + \mathrm{U}(O_2) \cdot \mathrm{Cr}(O_2 \,/\, O_1 \vee O_2).$$

Since choosing A effectively means choosing $O_1 \vee O_2$, we have

$$\mathrm{Cr}((O_1 \vee O_2) \leftrightarrow A) = 1.$$

Moreover, on the supposition that $O_1 \vee O_2$ is true, it is certain that O_1 comes about just in case state S_1 obtains. That is,

$$\mathrm{Cr}(O_1 \,/\, O_1 \vee O_2) = \mathrm{Cr}(S_1 \,/\, O_1 \vee O_2).$$

Together, the previous two observations entail that

$$\mathrm{Cr}(O_1 \,/\, O_1 \vee O_2) = \mathrm{Cr}(S_1 / A).$$

In a well-defined decision matrix, the states must be independent of the acts. This suggests that $\mathrm{Cr}(S_1 / A) = \mathrm{Cr}(S_1)$. We get

$$\mathrm{Cr}(O_1 \,/\, O_1 \vee O_2) = \mathrm{Cr}(S_1).$$

By the same reasoning, $\mathrm{Cr}(O_2 \,/\, O_1 \vee O_2) = \mathrm{Cr}(S_2)$. The above instance of Jeffrey's axiom can therefore be rewritten as

$$\mathrm{U}(A) = \mathrm{U}(O_1) \cdot \mathrm{Cr}(S_1) + \mathrm{U}(O_2) \cdot \mathrm{Cr}(S_2).$$

This says that the *utility* of the act A equals the *expected utility* of A!

Now, utility is a measure of (all-things-considered) desirability. The MEU principle is therefore equivalent to the seemingly innocuous claim that rational agents choose an act that they desire at least as strongly as any alternative. (We are going to challenge this seemingly innocuous claim, and the present argument, in chapter 9.)

In chapter 6, we met yet another argument for the MEU Principle. The argument began with an idea about how to measure (or define) an agent's intrinsic utility function. The idea was to look at the agent's preferences between outcomes and lotteries. Assuming that the agent always chooses a most preferred option, von Neumann's construction of utility entails that an agent obeys the MEU Principle (in choices between lotteries) iff their preferences satisfy the "axioms" of Completeness, Transitivity, Continuity, Independence, and Reduction.

To finish this argument for the MEU Principle (for choices between lotteries), we would need to explain why the five axioms should be considered requirements of rationality.

An influential argument in support of the axioms attempts suggests that if an

agent's preferences violate the axioms, then the agent is disposed to make patently bad choices in certain multi-stage decision problems.

To illustrate, suppose an agent violates the Transitivity axiom. The agent prefers A to B, B to C, but C to A. These preferences form a cycle. Whichever of A, B or C the agent has, she would prefer to have one of the others. If she is willing to pay a small amount to get the preferred option, we could exploit her in a kind of multi-stage Dutch Book.

Concretely, let's assume the agent starts with C. Since she prefers B to C, she should be willing to pay an insignificant amount (say, 1p) if we let her swap C for B. Once she has B, we let her swap B for A in exchange for another penny. She is happy to do that, since she prefers A to B. Finally, we let her swap A for C, again in exchange for 1p. The agent should accept, since she prefers C to A. The agent is back where she started, with C, and we have gained three pence. We could start over, letting the agent swap C for B for A for C until we have emptied her wallet.

This kind of argument is called a **money-pump argument** (for obvious reasons). It's worth spelling out in more detail. In its present form, the argument has a serious flaw.

8.2 Money pumps and sequential choice

We are dealing with an agent with cyclical preferences:

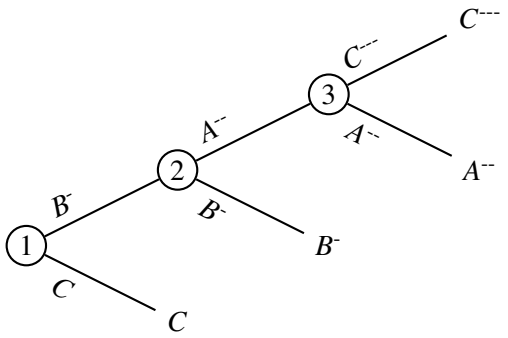
$$A \succ B$$
$$B \succ C$$
$$C \succ A$$

We imagine presenting this agent with a sequence of choices. A decision problem with more than one choice point is called a **sequential decision problem**. The branch of decision theory that studies sequential decision problems is called **sequential decision theory** or **dynamic decision theory**. Our money-pump argument invites us to take a brief look into this area.

We have assumed that the agent starts with C. At the first choice point in our money-pump scenario, the agent can either keep C or exchange it for B, at a small cost. Let B^- express B with the added small cost: $B^- = B \wedge$-1p. So the first choice

is between C and B^-. If the agent chooses B^-, she is given the option to pay another penny to swap B for A. If she goes in for the trade, she is left with $A^{--} = A\wedge$-2p. She is then offered a third choice, in which she can stick with A^{--} or get $C^{---} = C\wedge$-3p.

We can picture the sequential decision problem in a tree diagram, called an **extensive form representation**.



The circled nodes are choice points. Now, what path through this tree will the agent take?

Earlier, we have assumed that the agent will choose B^- at node 1, because she prefers B to C, and we take for granted that the preference is strong enough that she also prefers B^- to C. Similarly, we have assumed that the agent would choose A^{--} at node 2 (because she prefers A to B), and C^{---} at node 3 (because she prefers C to A). She ends up with $C^{---} = C\wedge$-3p even though she could have gotten C at no cost by "turning right" at the first node.

But would the agent really make these choices?

Look again at node 1. Superficially, the agent is here offered a choice between C and B^-. But if she "chooses B^-" she isn't actually getting B^- unless she "turns right" at node 2. If she turns left at node 2 and again at node 3, as we assumed she will, then "choosing B^-" at node 1 actually means getting C^{---}. And C^{---} is worse than C. If the agent can foresee that she will turn left at nodes 2 and 3, then she will *not* turn left at node 1.

The flaw in our argument is that we have ignored any information the agent might have about her predicament and about what she would do at later stages in the scenario. We have adopted what is called a **myopic** approach to sequential choice. The myopic approach treats each choice point as if it were the only decision the agent ever faces, ignoring any downstream consequences. This won't do. An adequate

evaluation of the agent's options should take into account what the agent is likely to do later. This approach to sequential choice is called **sophisticated**.

To investigate the above decision problem from a sophisticated perspective, we need to say what the agent knows about her situation. Let's assume that she is fully informed about the sequential decision problem that she is facing. Let's also assume that she has perfect knowledge of her preferences, so that she can figure out what she will do at any future choice point.

What the agent should do at node 1 now depends on what she will do at node 2. What she should do at node 2 similarly depends on what she will do at node 3. But if there are no relevant choices after node 3 then we can figure out what the agent will do here. The choice at node 3 then really is between A^{--} and C^{---}. Since the agent prefers C to A, it is plausible that she will choose C^{---}.

With this information in hand, we can return to node 2. Her choice at node 2 is effectively between C^{---} (*via* node 3) and B^{-}. The agent prefers B to C. So we can expect her to choose B^{-} at node 2.

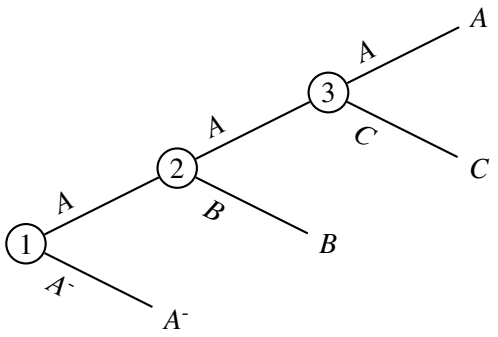
Now return to node 1. Given what we have just figured out, the choice at node 1 is effectively between C and B^{-}. The agent prefers B (and B^{-}) to C. We may therefore expect her to choose B^{-} at node 1. She will "turn left" at node 1 and right at node 2.

This kind of reasoning is called **backward induction**. We'll meet it again in section 10.5.

Exercise 8.1 †††

Draw a decision matrix (without utility values, but with credences) for the agent's choice at node 1.

Our attempted money pump doesn't work. At least not if the agent knows about our plan. But this can be fixed. In the following sequential decision problem, our agent would trade A for A^{-} at node 1, assuming again that she is fully informed about the scenario and her preferences. She would make a guaranteed and avoidable loss of 1 penny.



Exercise 8.2 ††

Explain by backward induction why the agent would choose A^- at node 1.

Exercise 8.3 †

Where would the agent end up if her preferences were transitive, so that $A \succ B$, $B \succ C$, and $A \succ C$?

The real point is, of course, not about money. The point is that cyclical preferences effectively lead to the choice of a dominated strategy. The agent could have gotten A, by "turning left" at each node. Due to her cyclical preferences, she ends up with a strictly worse outcome A^-.

We have assumed that the agent prefers A to B, B to C, and C to A. Not all violations of Transitivity involve cycles of this kind. Instead of preferring C to A, the agent could be indifferent between C and A. Or she could have no attitude at all about the comparison between A and C, violating both Transitivity and Completeness. These preferences, too, can be shown to support the choice of a dominated strategy. The same is true, more generally, for almost all preferences that violate the von Neumann and Morgenstern axioms.

8.3 The long run

Let's look at one last argument for the MEU Principle. This one turns on a connection between probability and relative frequency.

Suppose you repeatedly toss a fair coin, keeping track of the number of heads and tails. You will find that over time, the proportion of heads approaches its objective probability, 1/2. After one toss, you will have 100% heads or 100% tails. After ten tosses, it's very unlikely that you'll still have 100% heads or 100% tails. 60% heads and 40% tails wouldn't be unusual. The (objective) probability of getting 40% tails or less in 10 independent tosses of a coin is 0.377. For 100 tosses, it is 0.028; for 1000, it is less than 0.000001. After 1000 tosses, the probability that the proportion of tails lies between 45% and 55% is 0.999.

In general, the rules of probability entail that if there is a sequence of "trials" $T_1, T_2, T_3 \ldots$ in which the same outcomes (like heads and tails) can occur with the same probabilities, then the probability that the *proportion* of any outcome in the sequence differs from its *probability* by more than an arbitrarily small amount ϵ converges to 0 as the number of trials gets larger and larger. This is known as the **(weak) law of large numbers**. Loosely speaking: in the long run, probabilities turn into proportions.

How is this relevant to the MEU Principle? Consider a bet on a fair coin flip: if the coin lands heads, you get £1, otherwise you get £0. The bet costs £0.40. If you are offered this deal again and again, the law of large numbers entails that the percentage of heads will (with high probability) converge to 50%. If you buy the bet each time, you can be confident that you will loose £0.40 in about half the trials and win £0.60 in the other half. The £0.10 *expected payoff* turns into an *average payoff*. In this kind of scenario, the MEU Principle effectively says that you should prefer acts with greater average utility (and therefore greater total utility) over acts with lower average (and total) utility. If you face the same decision problem over and over, then you are almost certain to achieve greater total utility if you follow the MEU Principle than if you follow any other rule.

In reality, of course, there are limits to how often one can encounter the very same decision problem. "In the long run, we are all dead", as John Maynard Keynes quipped. Fortunately, we saw in the coin flip example that the convergence of proportions to probabilities tends to be quick. It does not take millions of tosses until the percentage of heads is almost certain to exceed 40%.

As it stands, the long-run argument still assumes that the same decision problem is faced over and over. But we can weaken this assumption. Suppose you face a sequence of decision problems that may involve different outcomes, different states, and different probabilities. One can show that if the states in these problems are

probabilistically independent, and the relevant probabilities and utilities are not too extreme, then over time, maximizing expected utility is likely to maximize average (and total) utility.

From all this, you might expect that professional gamblers and investors generally put their money on the options with greatest expected payoff, since this would give them the greatest overall profit in the long run. But they do not. (Those who do don't remain professional gamblers or investors for long.) To see why, imagine you are offered an investment in a startup that tries to find a cure for snoring. If the startup succeeds, your investment will pay back tenfold. If the startup fails, the investment is lost. The chance of success is 20%, so the expected return is $0.2 \cdot 1000\% + 0.8 \cdot 0\% = 200\%$. Even if this exceeds the expected return of all other investment possibilities, you would be mad to put all your money into this gamble. If you repeatedly face this kind of decision and go all-in each time, then after ten rounds you are bankrupt with a probability of $1 - 0.2^{10} = 0.9999998976$.

This does not contradict the law of large numbers. In the startup example, you are not facing the same decision problem again and again. If you lose all your money in the first round, you don't have anything left to invest in later rounds. Still, the example illustrates that by maximizing expected utility you don't always make it likely that you will maximize average or total utility in the long run. More importantly, the example suggests that there is something wrong with the MEU Principle. Sensible investors balance expected returns and risks. A safe investment with lower expected returns is often preferred to a risky investment with greater expected returns. Shouldn't we adjust the MEU Principle, so that agents can factor in the riskiness of their options?

Exercise 8.4 ††

Every year, an investor is given £100,000, which she can either invest in a risky startup of the kind described (a different one each year), or put in a bank account at 0% interest. If she always chooses the second option, she will have £1,000,000 after ten years.

(a) What are the chances that she would do at least as well (after ten years) if she always chooses the first option, without reinvesting previous profits? (Hint: Compute the chance that she would do worse.)

(b) How does the answer to (a) mesh with my claim in the text that an investor who always goes with the risky option is virtually guaranteed to go bankrupt?

8.4 Risk aversion

Many people are risk averse, at least for certain kinds of choices. They prefer situations with a predictable outcome over highly unpredictable situations. This does not seem irrational. Does it pose a threat to the MEU Principle?

A standard way to measure risk aversion involves lotteries. Consider a lottery with an 80% chance of £0 and a 20% chance of £1000. The expected payoff is £200. Given a choice between the lottery and £100 for sure, a risk averse agent might prefer the £100. Can we account for these preferences?

We can. We could, for example, assume that the difference in utility between £1000 and £100 is, for this agent, less than five times the difference in utility between £100 and £0. For example, if U(£0) = 0, U(£100) = 1, and U(£1000) = 4, then the lottery has expected utility $0.8 \cdot 0 + 0.2 \cdot 4 = 0.8$, which is less than the guaranteed utility of the £100.

This is how economists model risk aversion. They assume that for risk averse agents, utility is a "concave function of money", meaning that the amount of utility that an extra £100 would add to an outcome of £1000 is less than the amount of utility the same £100 would add to a lesser outcome of, say, £100. We have already encountered this phenomenon in chapter 5, where we saw that money has declining marginal utility: the more you have, the less utility you get from an extra £100. According to standard economics, risk aversion is the flip side of declining marginal utility.

This should seem strange. Intuitively, the fact that the same amount of money becomes less valuable the more money you already have has nothing to do with risk. Money could have declining marginal utility even for an agent who loves the thrill of risky options. Conversely, an agent might value every penny as much as the previous one, but shy away from risks.

No doubt some actions that appear to display risk aversion (say, among professional gamblers) are really explained by the declining marginal utility of money.

But many people prefer predictable situations in a way that can't be explained along these lines. The following example is due to Maurice Allais,

Example 8.1 (Allais's Paradox)

A ball is drawn from an urn containing 80 red balls, 19 green balls, and 1 blue ball. Consider first a choice between the following two lotteries. Which do you prefer?

	Red (0.8)	Green (0.19)	Blue (0.01)
A	£0	£1000	£1000
B	£0	£1200	£0

Next, consider the alternative lotteries *C* and *D*, based on the same draw from the urn. Which of these do you prefer?

	Red (0.8)	Green (0.19)	Blue (0.01)
C	£1000	£1000	£1000
D	£1000	£1200	£0

If you choose *C* in the second choice, you get £1000 for sure. If you choose *D*, you get either £1000 (most likely) or £0 (least likely) or £1200. If you're risk averse, it makes sense to take the sure £1000.

In the first choice, the most likely outcome is £0 no matter what you do. It may seem reasonable to take the 19% chance of getting £1200 (by choosing *B*) rather than the 20% chance of getting £1000 (by choosing *A*).

Many people, when confronted with Allais's puzzle, seem to reason in this way. They prefer *C* to *D* and *B* to *A*. These preferences can't be explained by the declining marginal utility of money. Indeed, there is no way of assigning utilities to monetary payoffs that makes a preference of *C* over *D* and *B* over *A* conform to the MEU Principle. If you have the risk-averse preferences, you appear to violate the MEU Principle.

Exercise 8.5 †††

The preference for C over D and B over A appears to violate the Independence axiom of von Neumann and Morgenstern. Explain. (The axiom states that, for any A, B, C, if $A \succsim B$, and L_1 is a lottery that leads to A with some probability x and otherwise to C, and L_2 is a lottery that leads to B with probability x and otherwise to C, then $L_1 \succsim L_2$.)

Some say that the kind of risk aversion that is manifested by a preference of B over A and C over D is irrational. Rational agents, they say, can't prefer predictable situations over unpredictable situations. This might be OK if our topic were "economic rationality". But it's not OK if we're interested in a general model of how coherent beliefs and desires relate to choice. There is nothing incoherent about a desire for predictability.

The following scenario, presented as a counterexample to the MEU Principle by Mark J. Machina, reinforces this verdict.

Example 8.2

A mother has a treat that she can give either to her daughter Abbie or to her son Ben. She considers three options: giving the treat to Abbie, giving it to Ben, and tossing a fair coin, so that Abbie gets the treat on heads and Ben on tails. Her decision problem might be summarized by the following matrix (assuming for simplicity that if the mother decides to give the treat directly to one of her children, she nonetheless tosses the coin, just for fun).

	Heads	Tails
Give treat to Abbie (A)	Abbie gets treat	Abbie gets treat
Give treat to Ben (B)	Ben gets treat	Ben gets treat
Let the coin decide (C)	Abbie gets treat	Ben gets treat

The mother's preferences are $C \succ A$, $C \succ B$, $B \succ A$.

As in Allais's Paradox, there is no way of assigning utilities to the outcomes in the mother's decision matrix that makes her preferences conform to the MEU Principle.

Yet these preferences are surely not irrational. The mother prefers C because it is the most fair of the three options. It would be absurd to claim that rational agents cannot value fairness.

8.5 Redescribing the outcomes

When confronted with an apparent counterexample to the MEU Principle, the first thing to check is always whether the decision matrix has been set up correctly. In particular, we need to check if the outcomes in the matrix specify everything that matters to the agent.

My matrices in example 8.4 (Allais's Paradox) specify how much money you get depending on your choice and the draw. But if you're genuinely risk averse, then you don't just care about how much money you will have. You also care about risk. We need to add more information to the outcomes.

There are two ways of doing this. The first adds to the monetary payoffs further things that will happen as a result of the relevant choice (and draw).

Consider the bottom right cell of the second matrix in example 8.4. What will happen if you choose D and the blue ball is drawn? You get £0. But you might also feel frustrated about your bad luck: there was a 99% chance of getting at least £1000, and you got nothing! You might also feel regret about your choice: if only you had chosen the safe alternative C, you'd now have £1000. You probably don't like feelings of frustration and regret. If so, these feelings should be added to the outcome. The outcome in the bottom right cell of the second matrix might now say something like '£0 and considerable frustration/regret'.

By contrast, consider the bottom right cell of the first matrix. If you choose B and the blue ball is drawn, you get £0. The chance of getting £0 was 81%, so you'll be much less frustrated about your bad luck. The outcome in that cell might say something like '£0 and a little frustration/regret'. With these changes, the preference for B over A and C over D is easily reconciled with the MEU Principle.

Exercise 8.6 †

Assign utilities to the outcomes in the two matrices, with the changes just described, so that $\mathrm{EU}(B) > \mathrm{EU}(A)$ and $\mathrm{EU}(C) > \mathrm{EU}(D)$.

A problem for this first type of response is that it doesn't always work. Suppose you face Allais's Paradox towards the end of your life. The ball will only be drawn after your death, and the money will go to your children. You will not be around to experience frustration or regret, nor might your children, if the whole process is kept secret from them. But if you're risk averse, you might still prefer B to A and C to D.

The second strategy for redescribing outcomes gets around this problem. As before, we want to distinguish the outcomes in the bottom right cell of the two decision matrices. Let's ask again what will happen if you choose D and the blue ball is drawn. One thing that will happen is that you get £0. You may or may not experience frustration and regret. But here's another thing that is guaranteed to happen: you *will have chosen a risky option instead of a safe alternative*. If you are risk averse, then plausibly (indeed, obviously!) you care about whether your choices are risky. So we should put that into the outcome.

The outcome in the bottom right cell of the first matrix does not have this feature – that you will have chosen a risky option instead of a safe alternative. There is no safe alternative. We can once again distinguish the two outcomes, and reconcile your preferences with the MEU Principle.

Exercise 8.7 †

We should also distinguish the outcomes in the top right cell of the two matrices. Can you explain how?

In general, the first strategy assumes that the "attributes" that make up an outcome are events that occur as a causal consequence of the relevant choice. Your frustration or regret might be such events. That you have chosen a risky option is not. This is not a separate event, caused by your choice of a risky option, as you can see from the fact that the "occurrence" of this event after your choice does not depend on the causal structure of the world.

Let's say that an outcome is **individuated locally** if it only comprises causal consequences of the relevant choice. A locally individuated outcome entails nothing about what happened at or before the time of choice.

Genuine risk aversion arguably calls for a non-local individuation of outcomes. Just as (genuine) risk aversion is not the same as declining marginal utility of money,

it is not the same as fear of regret or frustration. If you are risk averse, then one of the things you care about is predictability. Since the outcomes should specify everything you care about, they should specify whether an outcome was brought about by a risky gamble or whether it was predictable.

Exercise 8.8 †

Redescribe the outcomes in example 8.4 so that the mother's preferences conform to the MEU Principle.

Exercise 8.9 ††

(a) In your solution to exercise 8.5, did you individuate the outcomes locally or non-locally?

(b) Either way, can you find another answer to the exercise that individuates outcomes the other way?

Many decision theorists, especially outside philosophy, assume that outcomes must be individuated locally. The assumption is so common that it doesn't even have a name. Let's call it **localism**. According to localism, genuine risk aversion (as manifested, for example, by the preference for C over D and B over A in Allais's Paradox) is incompatible with the MEU Principle.

There are several reasons for the prevalence of localism. Some are historical. As we saw in chapter 5, 'utility' was originally used to denote something like pleasure or wealth or welfare, and it is still often used in that sense. On this usage, the utility of an outcome is clearly not affected by how the outcome was brought about. Once it is settled how much pleasure or wealth or welfare the agent has, we know how much utility they have, no matter whether the outcome was brought about in a risky manner.

Even authors who don't directly identify utility with pleasure or welfare or wealth often assume that utility is a measure of *something like* pleasure of welfare or wealth. It is assumed to measure the extent to which the agent desires the outcome "in itself", irrespective of its origin.

On either interpretation of 'utility', an agent's utility function may not capture all their basic desires. A basic desire to act fairly or avoid risks, for example, may not

show up in the agent's utility function. The MEU Principle then effectively tells the agent to disregard these desires. This is obviously problematic. If we don't want to declare the relevant desires irrational, we need to revise the MEU Principle.

Many localists have therefore put forward alternatives to the MEU Principle that are meant to take some of these other desires into account, while still assuming that utility functions are only sensitive to local features of outcomes. In Lara Buchak's "Risk-Weighted Expected Utility Theory", for example, rational choice is determined by three parameters: an agent's credences, their (local) utilities, and a third parameter that captures the agent's attitude towards risk.

We have taken a different approach – the standard approach at least in some quarters of theoretical philosophy. We have assumed that an agent's utility function reflects all their basic desires. We have put no official constraints on what sorts of things an agent might desire.

To some extent, this is just a difference in bookkeeping. But it has some important ramifications. On our approach, the MEU Principle never tells an agent to disregard some of their basic desires. We can easily accommodate risk aversion. More generally, we don't need to find a new modification of the MEU Principle for every desire that doesn't pertain to local features of outcomes.

Exercise 8.10 ††

Risk and fairness are two non-local attributes that many people care about. Can you think of another such attribute?

That basic desires can pertain to non-local features of outcomes has consequences for preference-based approaches to utility. In particular, it seems to break von Neumann's methods for determining an agent's intrinsic utility function from their preferences.

Suppose, for example, that we want to determine the utility function for the mother in example 8.4. Let a and b be the outcomes of directly giving the treat to Abby or Ben, respectively. If the mother cares about fairness, then one relevant (non-local) aspect of a and b is that who gets the treat is not decided by a chance process. By von Neumann's method, we should now ask whether the mother prefers some other outcome c to a *lottery L between a and b*. This lottery would be a chance process that leads to outcomes which don't come about through a chance process. That's

logically impossible. L entails that either a or b comes about, and it also entails that neither of them come about. We can hardly assume that the mother has interesting views about how L compares to a and b.

If many of the lotteries in the von Neumann construction are logically impossible, then either the Completeness axiom or most of the other axioms become highly implausible. We loose a popular approach to defining utility, and a popular argument for the MEU Principle.

Exercise 8.11 ††

Explain how a non-local individuation of outcomes can undermine the money-pump argument for Transitivity from section 8.2.

This doesn't mean that we have to give up the whole preference-based approach to utility. Ramsey's account might still work, but it depends on how some details are filled in. A clear example of an account that is compatible with arbitrary basic desires was developed by Ethan Bolker and Richard Jeffrey in the 1960s.

Instead of lotteries or gambles, Bolker and Jeffrey use unspecific propositions. If a and b are two outcomes or concerns, then their disjunction $a \vee b$ can play the role of a gamble. As long as a and b are consistent, $a \vee b$ is guaranteed to be consistent as well. In general, Bolker and Jeffrey assume that an agent's preferences simply relate propositions. The **Bolker-Jeffrey representation theorem** shows that if these preferences satisfy certain formal conditions, then they are represented by a probabilistic credence function Cr and a utility function U relative to which the agent evaluates arbitrary propositions in line with Jeffrey's axiom. Utility is still derived from preference – although the relevant preferences, relating arbitrary propositions, are even further removed from choice dispositions as they are in von Neumann's or Ramsey's construction.

Essay Question 8.1

The money-pump argument from section 8.2 relies on non-trivial assumptions about the agent's basic desires. Can you find a way to tweak the argument to show that cyclical preferences are *always* problematic, even for agents who don't have the relevant basic desires? (You might, for example, try to make

some of the moves I made in section 3.5.)

Sources and Further Reading

A useful survey of money-pump arguments for the von Neumann and Morgenstern axioms is Johan E. Gustafsson, *Money-Pump Arguments* (2022). Katie Steele, "Dynamic Decision Theory" (2018) briefly summarizes some of the philosophical controversy over these arguments.

I don't know any good literature on the long-run argument. I describe some moves towards generalising the argument beyond cases where the agent faces the same decision problem over and over at www.umsu.de/wo/2018/678.

For an intro to Allais's Paradox, see Philippe Mongin, "The Allais paradox: What it became, what it really was, what it now suggests to us" (2019). The example of the mother and the treat is from Mark J. Machina, "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty" (1989).

That risk aversion should be handled by including risk as an "attribute" of outcomes is defended, for example, in Paul Weirich, "Expected Utility and Risk" (1986). For arguments against a non-local individuation of outcomes, see Jean Baccelli and Philippe Mongin, "Can redescriptions of outcomes salvage the axioms of decision theory?" (2021) and chapter 4 of Lara Buchak, *Risk and Rationality* (2013). This book defend Buchak's risk-weighted expected utility theory.

The Jeffrey-Bolker construction is described in chapter 9 of Richard Jeffrey, *The Logic of Decision* (1965/83). Unless the agent's utilities are unbounded, Jeffrey and Bolker actually don't manage to secure a unique representation. On this issue, see James Joyce, "Why we still need the logic of decision" (2000).