6 Preference

6.1 The ordinalist challenge

If the utility of an outcome for an agent is not measured by the amount of money the agent gains or loses, how is it measured? How can we find out whether an outcome has utility 5 or 500 or -27? What does it even mean to say that an outcome has utility 5?

At the beginning of the 20th century, doubts arose about the coherence of numerical utilities. **Ordinalists** like Vilfredo Pareto argued that the only secure foundation for utility judgements are people's choices. If you are given a choice between tea and coffee, and you choose tea, we can conclude that tea has greater utility for you than coffee. We may similarly find that you prefer coffee to milk, etc., but how could we find that your utility for tea is twice your utility for coffee – let alone that it has the exact value 5? The ordinalists argued that we should give up the conception of utility as a numerical magnitude.

Ordinalism posed a serious threat to the idea of expected utility maximization. If there is no numerical quantity of utility, we can't demand that rational agents maximize the probability-weighted average of that quantity, as the MEU Principle requires.

In 1926, Frank Ramsey pointed out that if we look at the choices an agent makes in a state of uncertainty then we can find out more about the agent's utility function than how it orders the relevant outcomes – enough to vindicate the MEU Principle. Ramsey's idea was rediscovered by John von Neumann, who published a simpler version of it in the 1944 monograph *Game Theory and Economic Behaviour*, co-authored with Oskar Morgenstern. This work is widely taken to provide the foundations of modern expected utility theory.

Before we have a closer look at von Neumann's approach, let's think a little more about the ordinalist challenge.

Ordinalism was inspired by a wider "positivist" movement in science and philosophy. The aim of the positivists was to cleanse scientific reasoning of obscure and untestable doctrines. Every meaningful statement was to have clear conditions of verification or falsification. A hypothesis whose truth or falsity is impossible to establish by either proof or observation was to be rejected as meaningless. In psychology, this movement gave rise to **behaviourism**, the view that statements about emotions, desires, and other psychological states should be defined in terms of observable behaviour.

Today, behaviourism, and positivism more generally, have been almost entirely abandoned. In part, this is because people came to appreciate the holistic nature of scientific confirmation. Statements in successful scientific theories often have observable consequences only in conjunction with other theoretical assumptions. More practically, the behaviourist paradigm was simply found to stand in the way of scientific progress. It is hard to explain even the behaviour of simple animals without appealing to inner representational states like goals or perceptions as causes of the behaviour.

On the basis of these historical developments, it may be tempting to dismiss the ordinalist challenge as outdated and misguided. But even if their general view of science was mistaken, the ordinalists raised an important issue.

In chapter 3, I emphasized that we should not think of an agent's credences as little numbers written in the head. If your credence in rain is $1/2$, then this must be grounded in other, more basic facts about you – facts that do not involve the number $1/2$. Even if we accept your state of belief as a genuine internal state, a cause of your behaviour, we need to explain why we represent the state with the number $1/2$ rather than $3/4$ or $12/5$.

There's nothing special here about credence. Numerical representations in scientific models are always based on non-numerical facts about the represented objects. For the numerical representations to have meaning, we need to specify what underlying non-numerical facts the different numbers are meant to represent.

The same is true for utility. The utility of a proposition for an agent is supposed to represents the extent to which the agent, on balance, wants the proposition to be true. But what non-numerical fact about an agent makes it correct to say that their utility for a certain proposition is 5? This question still needs an answer. And there is something to be said for the idea that the answer should involve the agent's choices.

The main reason to think that an agent has such-and-such goals or desires is that this would explain their behaviour. The point is even more obvious for the relative strength of goals or desires. I got out of bed because my sense of duty was stronger than my desire to stay in bed. Absent further explanation, the claim that my desire to stay in bed was stronger, even though I got up, is unintelligible. If we seek a standard to measure the comparative strength of different motives or goals, a natural idea is thus to look at what the agent is prepared to do.

6.2 Scales

Utility, like credence, mass, or length, is a numerical representation of an essentially non-numerical phenomenon. All such representations are to some extent conventional. We can represent the length of my pencil as 19 centimetres or as 7.48 inches. It's the same length either way. We must take care to distinguish real features of the represented properties from arbitrary consequences of a particular representation. For example, it is nonsense to ask whether the length of my pencil – the length itself, not the length in any particular system of representation – is a whole number. By contrast, it is not meaningless to ask whether the length of my pencil is greater than the length of my hand.

In the case of length, the conventionality of measurement essentially boils down to the choice of a **unit**. You can introduce a new measure of length simply by picking out a particular object (say, your left foot) and declare that its length is 1, with the understanding that if an object is n times as long as the chosen object then its length in your new system is n. (You could fix the unit by assigning any number greater than zero to your left foot; it doesn't have to be the number 1.)

Quantities like mass and length, for which only the unit of measurement is conventional, are said to have a **ratio scale** because even though the particular numbers are conventional, ratios between them are not. If the length of my arm is four times the length of my pencil in centimetres, then that is also true in inches, feet, light years, and any other sensible system of representation. That my arm is four times as long as my pencil is an objective, representation-independent fact.

Temperature is different. (Or has appeared to be different until the 19th century.) People have long known that metals like mercury expand as the temperature goes up. This can be used to define a numerical representation. Imagine we put a certain amount of mercury in a narrow glass tube. The higher the temperature, the more of the glass tube is filled with the expanding mercury. To get a numerical measure of temperature, we now need to mark *two* points on the tube, a unit and a **zero**. We could, for example, mark the point at which water freezes as 0 and the point at which it boils as 100. We can then say that if the mercury has expanded to x% of the distance between 0 and 100, then the temperature is x.

The Celsius scale for temperature and the Fahrenheit scale have different units and zeroes. As a result, 10 degrees Celsius is 50 degrees Fahrenheit, and 20 degrees Celsius is 68 degrees Fahrenheit. The ratio between the two temperatures is not preserved, so these scales are not ratio scales. Scales in which both the zero and the unit are a matter of convention are called **interval scales**.

Exercise 6.1 ††

Someone might suggest that we only need to mark a unit on the glass tube, since we are effectively measuring the volume of the mercury in the tube, and volume has a ratio scale: zero simply means that the mercury fills up none of the tube. Does this show that temperature has a (natural) ratio scale?

Ratio scales and interval scales are both called **cardinal** scales, in contrast to **ordinal scales**, in which the only thing that is not conventional is which of two objects is assigned a greater number.

The ordinalists held that utility has only an ordinal scale (hence the name of the movement). All we have to go by in order to measure utilities, the ordinalists assumed, are the agent's choices. If you choose tea over coffee and coffee over milk, we may infer that your utility for tea is greater than your utility for coffee, which in turn is greater than your utility for milk. But any assignment of numbers that respects this ordering is as correct as any other. We could say that for you, tea has utility 3, coffee 2, and milk 1, but we could equally say that tea has utility 100, coffee 0, and milk -8.

If the ordinalists were right, then whether an act in a decision problem maximizes expected utility would often depend on arbitrary choices in the measurement of utility. The MEU Principle would be indefensible. If, on the other hand, utility has an interval scale, then different measures of utility never disagree on the ranking of acts in a decision problem. A ratio scale is not required.

Exercise 6.2 †

In the Mushroom Problem as described by the matrix on page 12 (section 1.3), not eating the mushroom has greater expected utility than eating the mushroom. Describe a different assignment of utilities to the four outcomes which preserves their ordering but gives eating the mushroom greater expected utility than not eating.

Exercise 6.3 ††

Suppose two utility functions U and U' differ merely by their choice of unit and zero. It follows that there are numbers $x > 0$ and y such that, for any A, $U(A) = x \cdot U'(A) + y$. Suppose some act A in some decision problem has greater expected utility than some act B if the utility of the outcomes is measured by U. Show that A also has greater expected utility than B if the utility of the outcomes is measured by U'. (You can assume for simplicity that the outcome of either act depends only on whether some state S obtains; so the states are S and $\neg S$.)

If we want to rescue the MEU Principle from ordinalist skepticism, we therefore don't need to explain what makes it the case that your utility for tea is 3 rather than 100. We can accept that the exact numbers are a conventional matter of representation. Nor do we need to explain what makes your utility for tea twice your utility for coffee; such ratios also need not track anything real. But we do have to explain why, if we arbitrarily mark your utility for tea as (say) 1 and your utility for coffee as 0, then your utility for milk is fixed at a particular value: why it has to be -1 (say), rather than -7, even though both hypotheses appear to be in line with your choices.

6.3 Utility from preference

I am now going to describe John von Neumann's method for determining an agent's utility function from their preferences or choice dispositions. More precisely, what we are going to determine is the utility of the agent's *concerns*. Recall from the previous chapter that a concern settles everything that matters to the agent, leaving open only questions towards which the agent is indifferent. Once we know the utility an agent assigns to her concerns, we can fill in the rest of their utility function with the help of Jeffrey's Axiom. (Assuming we know the agent's credences.)

To make the following discussion a little more concrete (and to bypass some problems that will occupy us later), let's imagine an agent who is only ultimately interested in getting certain "rewards", which may be lumps of money or commodity bundles or pleasant sensations. I will use lower-case letters $a, b, c, \ldots$ for rewards. Our goal is to find the agent's numerical utility for $a, b, c, \ldots$.

We will determine the agent's utilities from their **preferences**, which we assume to represent their choice dispositions. For example, if the agent would choose reward a when given a choice between a and b, we say that the agent prefers a to b. The ordinalists did not challenge the assumption that people have preferences.

Let's introduce some shorthand notation:

$$A \succ B \Leftrightarrow \text{The agent prefers } A \text{ to } B.$$
$$A \sim B \Leftrightarrow \text{The agent is indifferent between } A \text{ and } B.$$
$$A \succsim B \Leftrightarrow A \succ B \text{ or } A \succsim B.$$

(Note that '$\succ$', '$\sim$', and '$\succsim$' had a different meaning in section 3.6. You always have to look at the context to figure out what these symbols mean.)

Our aim is to use facts about the agent's preferences to construct a utility function U (an assignment of numbers to rewards) that **represents** the agent's preferences, in the sense that for all rewards a and b, $\mathrm{U}(a) > \mathrm{U}(b)$ iff $a \succ b$, and $\mathrm{U}(a) = \mathrm{U}(b)$ iff $a \sim b$.

Let's begin. We accept that the choice of unit and zero is a matter of convention, so we take arbitrary rewards a and b such that $b \succ a$ and set $\mathrm{U}(a) = 0$ and $\mathrm{U}(b) = 1$. This resembles the conventional choice of using 0 for the temperature at which water freezes and 100 for the temperature at which it boils.

Exercise 6.4 ††

If our agent is indifferent between all rewards, then the procedure stalls at this step. Nonetheless, we can easily find a utility function for such an agent. What does it look like?

Having fixed the utility of two rewards a and b, we can now determine the utility of any other reward c. We distinguish three cases, depending on how the agent ranks c relative to a and b.

Suppose first that c "lies between" a and b in the sense that $b \succ c$ and $c \succ a$. To find the utility of c, we look at the agent's preferences between c and a **lottery** between a and b. By a 'lottery between a and b', I mean an event that leads to a with some objective probability x and otherwise to b. For example, suppose we offer our agent a choice between c for sure and the following gamble L: we'll toss a fair coin; on heads the agent gets a, on tails b. By the Probability Coordination Principle, the expected utility of L is

$$\mathrm{EU}(L) = 1/2 \cdot \mathrm{U}(a) + 1/2 \cdot \mathrm{U}(b) = 1/2 \cdot 0 + 1/2 \cdot 1 = 1/2.$$

If the agent obeys the Probability Coordination Principle and the MEU Principle, and she is indifferent between L and c, we can infer that c has utility $1/2$.

Exercise 6.5 †

Suppose $\mathrm{U}(a) = 0, \mathrm{U}(b) = 1$, and $\mathrm{U}(c) = 1/2$. Draw a decision matrix representing a choice between c and L, and verify that the two options have equal expected utility.

Exercise 6.6 ††

Why do we need to assume that the agent obeys the Probability Coordination Principle?

If the agent isn't indifferent between L and c, we try other lotteries, until we find one the agent regards as equally good as c. For example, suppose the agent is indifferent between c and a lottery L' that gives them a with probability $4/5$ and b with probability $1/5$. Since the expected utility of this lottery is $1/5$, we could infer that the agent's utility for c is $1/5$.

We have assumed that c lies between a and b. What if the agent prefers c to both a and b? In this case, we look for a lottery between a and c such that the agent is indifferent between b and the lottery. For example, if the agent is indifferent between b for sure and a lottery L'' that gives them either a or c with equal probability, then

c must have utility 2. That's because the expected utility of L'' is

$$\mathrm{EU}(L'') = {}^1\!/_2 \cdot \mathrm{U}(a) + {}^1\!/_2 \cdot \mathrm{U}(c) = {}^1\!/_2 \cdot 0 + {}^1\!/_2 \cdot \mathrm{U}(c) = {}^1\!/_2 \cdot \mathrm{U}(c).$$

Since the agent is indifferent between L'' and b, which has a guaranteed utility of 1, the lottery must have expected utility 1. So $1 = {}^1\!/_2 \cdot U(c)$. And so $\mathrm{U}(c) = 2$. In general, if the agent is indifferent between b and a lottery that leads to c with probability x and a with probability $1 - x$, then $\mathrm{U}(c) = {}^1\!/_x$.

Exercise 6.7 ††

Can you complete the argument for the case where the agent prefers both a and b to c?

In this manner, we can determine the agent's utility for all rewards from their preferences between rewards and lotteries. The resulting utility function has an arbitrary unit and zero, but once these are fixed, the other utilities are no longer an arbitrary matter of convention. We have a cardinal utility scale, and thus an answer the ordinalist challenge. Or so it seems.

6.4 The von Neumann and Morgenstern axioms

The method described in the previous section assumes that the agent obeys the MEU Principle. This may seem strange. The ordinalists argued that the MEU Principle makes no sense. How can we respond to them by *assuming* the principle? Besides, doesn't application of the MEU Principle presuppose that we already know the agent's utilities?

The trick is that we are applying the principle backwards. Normally, when we apply the MEU Principle, we start with an agent's beliefs and desires and try to find out the optimal choices. Now we start with the agent's choices and try to find out the agent's desires, relying on the Probability Coordination Principle to fix the relevant beliefs.

There is nothing dodgy about this. Whenever we want to measure a quantity whose value can't be directly observed, we have to rely on assumptions about how the quantity relates to other things that we can observe. Together with the Probability

Coordination Probability, the MEU Principle tells us what lotteries an agent should be disposed to accept if she has a given utility function. If she doesn't accept these lotteries, we can infer that she doesn't have the utility function.

You may wonder, though, what happened to the normativity of the MEU Principle. If we follow von Neumann's method to define an agent's utility function, won't the agent automatically come out as obeying the MEU Principle?

Not quite. It's true that *if the method works*, then the agent will evaluate lotteries by their expected utility, relative to the utility function identified by the method. But the method is not guaranteed to work. Nor does it settle how the agent evaluates the options in decision problems in which the relevant objective probabilities are unknown.

Here is one way in which the method might fail to work. We have assumed that if an agent ranks some reward c as between a and b, then the agent is indifferent between c and some lottery between a and b. This is not a logical truth. An agent could in principle prefer c to any lottery between a and b, yet still prefer c to a and b to c. Von Neumann's method does not identify a utility function for such an agent.

Von Neumann and Morgenstern investigated just what conditions an agent's preferences must satisfy in order for the method to work. To state these conditions, we assume that '$\succ$', '$\sim$', and '$\succsim$' are defined not just for basic rewards but also for lotteries between rewards as well as "compound lotteries" whose payoff is another lottery. For example, if I toss a fair coin and offer you lottery L on heads and L' on tails, that would be a compound lottery.

Here are the conditions we need. 'A', 'B', 'C' range over arbitrary lotteries or rewards.

Completeness

For any A and B, exactly one of $A \succ B$, $B \succ A$, or $A \sim B$ is the case.

Transitivity

If $A \succ B$ and $B \succ C$ then $A \succ C$; if $A \sim B$ and $B \sim C$ then $A \sim C$.

Continuity

If $A \succ B$ and $B \succ C$ then there are lotteries L_1 and L_2 between A and C such that $A \succ L_1 \succ B$ and $B \succ L_2 \succ C$.

Independence (of Irrelevant Alternatives)

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 $L_1 \succsim L_2$.

Reduction (of Compound Lotteries)

If a L_1 and L_2 are two (possibly compound) lotteries that lead to the same rewards with the same objective probabilities, then $L_1 \sim L_2$.

Von Neumann and Morgenstern proved that if (and only if) an agent's preferences satisfy all these conditions, then there is a utility function U, determined by the method from the previous section, that represents the agent's preferences (in the sense that $\mathrm{U}(A) > \mathrm{U}(B)$ iff $A \succ B$, and $\mathrm{U}(A) = \mathrm{U}(B)$ iff $A \sim B$). Von Neumann and Morgenstern also proved that the function U is unique except for the choice of unit and zero: any two functions U and U′ that represent the agent's preferences differ at most in the choice of unit and zero. These two results are known as the **von Neumann-Morgenstern Representation Theorem**.

If we adopt von Neumann's method for measuring an agent's utilities in terms of their choice dispositions, then the MEU Principle for choices involving lotteries is automatically satisfied by any agent whose preferences satisfy the above conditions – Completeness, Transitivity, etc. The normative claim that an agent ought to evaluate lotteries by their expected utility reduces to the claim that their preferences ought to satisfy the conditions. For this reason, the conditions are often called the **axioms of expected utility theory**.

Von Neumann therefore discovered not only a response to the ordinalist challenge (at least for agents who satisfy the axioms). He also discovered a powerful argument for the MEU Principle. The argument could be spelled out as follows.

1. The preferences of a rational agent satisfy Completeness, Transitivity, Continuity, Independence, and Reduction.
2. If an agent's preferences satisfy these conditions, then (by the Representation Theorem) they are represented by a utility function U relative to which the agent ranks lotteries by their expected utility.
3. That utility function U is the agent's true utility function.
4. Therefore: A rational agent ranks lotteries by their expected utility.

Exercise 6.8 †

Maurice would go to Rome if he were offered a choice between Rome and going to the mountains, because the mountains frighten him. Offered a choice between staying at home and going to Rome, he would prefer to stay at home, because he finds sightseeing boring. If he were offered a choice between going to the mountains and staying at home, he would choose the mountains because it would be cowardly, he believes, to stay at home. Which of the axioms does Maurice appear to violate?

6.5 Utility and credence from preference

In chapter 3, we asked how an agent's credences could be measured or defined. The betting interpretation gave a simple answer, but we found that it relies on implausible assumptions about the agent's utility function. In the meantime, we have learned from von Neumann how we might derive an agent's true utilities from their choice dispositions. With that information in hand, we might try again to determine the agent's credence function by offering them suitable bets.

Frank Ramsey, way ahead of his time in 1926, showed how the two tasks can be combined. He described a method for determining both an agent's credence function and their utility function from their preferences.

Instead of lotteries, Ramsey uses deals whose outcome depends on a proposition the agent doesn't intrinsically care about. Suppose N is a proposition whose truth-value you don't care about (say, that the number of stars is even), and suppose your credence in N is $1/2$. Instead of offering you a lottery that yields outcome a or outcome b with equal chance, we can offer you a deal that leads to a if N and to b if $\neg N$.

I will refer to conditional deals of the form 'A if X, B if $\neg X$' as **gambles**. Notice that every act in every decision problem corresponds to a (possibly nested) gamble. In the mushroom problem from chapter 1, for example, eating the mushroom amounts to choosing the gamble '*Dead* if *Poisonous*, *Satisfied* if not *Poisonous*'; not eating the mushroom amounts to choosing '*Hungry* if *Poisonous*, *Hungry* if not *Poisonous*'.

To determine an agent's credences and utilities, Ramsey begins by identifying a suitable proposition N with credence $1/2$. Recall that at this stage, we have nothing but the agent's preferences to go by.

Let's say that a proposition A is *neutral* for an agent if, for any conjunction of rewards R, the agent is indifferent between $R \wedge A$ and $R \wedge \neg A$. Intuitively, a neutral proposition is one the agent doesn't care about. Now let a and b be two rewards such that $a \succ b$. Suppose we find a neutral proposition N such that the agent is indifferent between the gambles 'a if N, b if $\neg N$' and 'b if N, a if $\neg N$'. Assuming that our agent ranks gambles by their expected utility, we can infer that the two gambles have equal expected utility:

$$\mathrm{Cr}(N) \cdot \mathrm{U}(a) + \mathrm{Cr}(\neg N) \cdot \mathrm{U}(b) = \mathrm{Cr}(N) \cdot \mathrm{U}(b) + \mathrm{Cr}(\neg N) \cdot \mathrm{U}(a).$$

Assuming further that the agent's credences are probabilistic, so that $\mathrm{Cr}(\neg N) = 1 - \mathrm{Cr}(N)$, it follows that $\mathrm{Cr}(N) = 1/2$. (As you may check.)

We now use this proposition N to determine the agent's utility function.

As before, we fix the unit and zero by taking arbitrary rewards with $a \succ b$ and set $\mathrm{U}(a) = 0$ and $\mathrm{U}(b) = 1$. Then we go through the rewards until we find one for which the agent is indifferent between c and the gamble 'a if N, b if $\neg N$'. Since this gamble has expected utility $1/2$, and we assume that our agent ranks gambles by their expected utility, we can infer that c has utility $1/2$.

In the next step, we can use gambles involving a, b, and c to determine the utility of further rewards. For example, if the agent is indifferent between a reward d and the gamble 'a if N, c if $\neg N$', then d must have utility $1/4$. And so on.

We can also determine the utility of rewards that don't lie between a and b. Suppose, for example, that a reward e is preferred to b, and the agent is indifferent between the gambles 'a if N, e if $\neg N$' and 'c if N, e if $\neg N$', where c is the earlier reward whose utility we've determined to be 1/2. Then the utility of e must be 1.5.

Exercise 6.9 †

Explain this last claim. That is, show that if $\mathrm{U}(a) = 0, \mathrm{U}(b) = 1, \mathrm{U}(c) = 1/2$, and the agent evaluates gambles by their expected utility, then they are indifferent between 'a if N, e if $\neg N$' and 'c if N, b if $\neg N$' only if $\mathrm{U}(e) = 1.5$.

If all went well, we now know the utility the agent assigns to all rewards. We still need to determine the agent's credence in propositions other than N.

Let X be some proposition. We need to find rewards a, b, and c such that the agent is indifferent between a and the gamble 'if X then b, if $\neg X$ then c'. The gamble's expected utility is $\mathrm{Cr}(X) \cdot \mathrm{U}(b) + \mathrm{Cr}(\neg X) \cdot \mathrm{U}(c)$. Since the agent is indifferent between the gamble and a, we can infer that

$$\mathrm{U}(a) = \mathrm{Cr}(X) \cdot \mathrm{U}(b) + (1 - \mathrm{Cr}(X)) \cdot \mathrm{U}(c).$$

Solving for $\mathrm{Cr}(X)$ yields

$$\mathrm{Cr}(X) = \frac{\mathrm{U}(a) - \mathrm{U}(c)}{\mathrm{U}(b) - \mathrm{U}(c)}.$$

All quantities on the right-hand side are known. We have determined $\mathrm{Cr}(X)$.

Like von Neumann's method, Ramsey's method only works if the agent's preferences satisfy certain formal conditions or "axioms". Ramsey lists eight axioms, the details of which won't be important for us.

Ramsey's Representation Theorem states that if (but not only if) an agent's preferences satisfy his eight conditions, then there is a utility function U and a probability function Cr that together represent the agent's preferences in the sense that (i) $A \succ B$ iff the expected utility of A, relative to Cr and U, is greater than that of B, and (ii) $A \sim B$ iff A and B have equal expected utility. The theorem also says that Cr is unique and U is unique expect for the choice of zero and unit.

What can this do for us? Ramsey's idea is that we may *define* an agent's credence and utility functions as whatever functions Cr and U "make sense of their preferences", meaning that the agent prefers an option A to an option B iff A has greater expected utility than B, where the expected utilities are computed with Cr and U. We also assume that in order for Cr to "makes sense" of the agent's preferences it must conform to the rules of probability. Ramsey's Representation Theorem assures us that if the agent's preferences satisfy his axioms, then *there are* functions Cr and U that make sense of the agent's preferences. Moreover, while there are different such pairs of functions Cr and U, they all involve the exact same function Cr, and the different U functions differ only in their choice of unit and zero. For agents who satisfy the axioms, our definition is therefore guaranteed to identify a unique credence function and a utility function that is determinate enough to vindicate the MEU Principle.

If we could convince ourselves that Ramsey's axioms are requirements of rationality, Ramsey's approach would also deliver a more comprehensive argument for the MEU Principle than what we got from von Neumann and Morgenstern. Their argument only showed that agents should rank *lotteries* by their expected utility. But not all choices involve lotteries. In real life, people often face options for which they don't know the objective probability of the outcomes. Why should they rank such options by their expected utility? On Ramsey's approach, the only way they could fail to rank options by their expected utility is that they violate one of the axioms.

Ramsey's approach also suggests a new argument for probabilism, the claim that rational degrees of belief conform to the rules of probability. Again, the requirement reduces to the preference axioms. On the proposed definition of credence, any agent who obeys the axioms automatically has probabilistic credences. If you don't have probabilistic credences, you violate the axioms.

Exercise 6.10 ††

Can you spell out the argument for probabilism I just outlined in more detail, in parallel to the argument for the MEU Principle from the end of section 6.4?

Unfortunately, Ramsey's axioms can hardly be considered requirements of rationality. For example, his method doesn't work unless there is a neutral proposition N with credence $1/2$, or unless there is a proposition c for which the agent is indifferent between c and the gamble 'a if N, b if $\neg N$'. Ramsey's axioms 1 and 6 ensure that these conditions are met, but it is hard to see why that should be a requirement of rationality.

Later authors have improved upon Ramsey in this respect, coming up with different (and generally more complicated) methods for determining credences and utilities from preferences. The best known of these proposals is due to Leonard Savage, published in his *Foundations of Statistics* (1954) – the second-most influential book in the history of decision theory, after *Game Theory and Economic Behaviour*. I won't go through Savage's method and axioms. Suffice it to say that his axioms still include conditions that nobody can seriously regard as requirements of rationality, let alone as requirements that anyone must meet in order to have credences and utilities.

That's bad news. Things get worse if we take a closer look at the connection between preference and choice behaviour.

6.6 Preference from choice?

Von Neumann and Ramsey both take as their starting point an agent's preferences, represented by the relations $\succ$, $\sim$, and $\succsim$. I suggested that we might read '$A \succ B$' as saying that the agent would choose A if given a choice between A and B. On this interpretation, von Neumann and Ramsey showed how we might determine an agent's utilities (and credences, in Ramsey's case) from their choice dispositions, assuming that these dispositions satisfy certain conditions ("axioms").

Let's be clear why I talk about dispositions. An agent's *dispositions* reflect what the agent *would* do if such-and-such circumstances were to arise. There is little hope of determining an agent's utilities or credences from their actual choices alone. Von Neumann and Ramsey certainly appeal to all sorts of choices most real agents never face.

Exercise 6.11 ††

Suppose we define '$A \sim B$' as 'the agent has faced a choice between A and B and expressed indifference', and '$A \succ B$' as 'the agent has faced a choice between A and B and expressed a preference for A. Which of the von Neumann and Morgenstern axioms then become highly implausible (no matter what exactly we mean by "expressing" indifference or preference)?

Now one of the problems for the betting interpretation, from section 3.4, returns with a vengeance. If an agent is not facing a choice between two options A and B, then offering her the choice would change their beliefs. Among other things, she would come to believe that they face that choice. From the fact that the agent *would* choose (say) A if she *were* offered the choice, we can't infer that the agent's *actual* expected utility of A is greater than that of B, even if we assume that the agent obeys the MEU Principle. Expected utilities depend on credences, and perhaps A only has greater expected utility after the agent's credences are updated by the information that she can choose between A and B.

The problem gets worse if we drop the simplifying assumptions that agents only care about lumps of money, commodity bundles, or pleasant sensations. Suppose one thing you desire (one "reward") is peace in Syria, another is being able to play the piano. Von Neumann's definition then determines your utilities in part by your preferences between peace in Syria and a lottery that leads to peace in Syria with objective probability $1/4$ and to an ability to play the piano with probability $3/4$. Ramsey's method might similarly look at your preferences between peace in Syria and gambles like 'peace in Syria if the number of stars is even, being able to play the piano if the number is odd'. If you thought you'd face this bizarre choice, your beliefs would surely be quite different from your actual beliefs. (Indeed, merely from being offered the choice, you could figure out that either there is peace in Syria or you can play the piano.)

Even in the rare case where an agent actually faces a relevant choice between A and B, we arguably can't infer that whichever option they choose (say, A) has greater expected utility.

For one thing, the agent might be indifferent between A and B and have chosen A at random. Choice dispositions arguably can't tell apart $A \succ B$ and $A \sim B$. The agent might also be mistaken about their options. If I offer you a choice between an apple and a banana, and you falsely believe that the banana is a wax banana, your choice of the apple doesn't show that you prefer an apple over a (real) banana. You might be similarly mistaken about which gambles or lotteries are on offer.

The upshot is that we need to distinguish (at least) two notions of preference. One represents the agent's choice dispositions: whether they would choose A over B in a hypothetical situation in which they face this choice. The other represents the agent's current ranking of rewards and gambles or lotteries: whether by the lights of the agent's current beliefs and desires, A is better than B. Von Neumann and Ramsey have at best shown how to derive utilities and credences from preferences in the second sense.

This could still be valuable. We might still get an interesting argument for probabilism and the MEU Principle. Moreover, there is plausibly *some* connection between preference in the second sense and choices dispositions. We haven't fully solved the measurement problem for credences and utilities. But one might hope that we are at least a few steps closer.

Essay Question 6.1

An agent's choice dispositions provide information about their beliefs and desires, but perhaps it is a mistake to think that one can determine the agent's beliefs and desires by looking at nothing but their choice dispositions. What other facts about the agent might one take into account? Evaluate the prospects of measuring an agent's utilities and/or credences based on these other facts, perhaps in combination with the agent's choice dispositions.

Sources and Further Reading

The 1926 draft in which Ramsey shows how one might derive utilities and credences from preferences is called "Truth and Probability". Richard Bradley, "Ramsey's Representation Theorem" (2004) provides some useful guidance to Ramsey's paper.

For a good discussion of Savage's approach, see chapter 3 of James Joyce, *The Foundations of Causal Decision Theory*. A useful, but mathematically heavy, survey of representation theorems in the tradition of Ramsey, Savage, and von Neumann and Morgensterm is Peter Fishburn, "Utility and Subjective Probability" (1994).

Preference-based approaches to utility are standard in economics, but fairly unpopular in philosophy. Christopher J.G. Meacham and Jonathan Weisberg, "Representation theorems and the foundations of decision theory" (2011) lists some common philosophical misgivings.

You may have noticed that both von Neumann's and Ramsey's method assume that we already know the agent's concerns. We will turn to this issue, and some related problems, in chapter 8.

On the connection between preference and choice behaviour, see Johanna Thoma, "In defence of revealed preference theory" (2021).

The Maurice exercise is from John Broome, *Weighing Goods* (1991, p.101).