RATIONAL CHOICE LECTURE 2

1. It is important to note four things about the general theory that I introduced in the last lecture. First, the subjective probability function Pr really is *subjective*. There is no 'right' or 'wrong' probability function, or at least the theory does not have to assume this. And there is no assumption that anyone's subjective probabilities track the true chances of events. In this respect Savage's theory is an advance on that of von Neumann and Morgenstern 1944, on which all probabilities arise from the objective chances of outcomes of lotteries.

2. Second, the idea that people do in fact maximize expected utility seems to go wrong if we always assume, as I did in the example at no. 9, that U ($n) = n. That would imply that people are willing to risk $1000 on the toss of a fair coin that would pay out $1001 if it lands heads. In fact people generally are not. But this doesn't mean that (7.1) is wrong; it just means that we need to be more careful about the utility function. The St Petersburg paradox (Jeffrey 1983: 150-5) makes this especially vivid.

3. For a less vivid but more realistic example, suppose that in fact you have $1000 but would decline a bet, on a *fair* coin, that wins $1000 if heads comes up and loses $1000 if tails comes up. It follows from (7.1) that your utility function U satisfies:

   (3.1) U ($1000) > 0.5 U ($0) + 0.5 U ($2000)

   It follows that utility is not a linear function of money; instead it is *marginally declining* in the sense that every additional dollar is worth les to you than the previous one (mathematically, $dU(\$x)/dx > 0$ and $d^2U(\$x)/dx^2 < 0$). The hypothesis of declining marginal utility for money is perhaps the single most powerful thesis in economics. It explains why people take out insurance, why rich people take taxis and poor people do not, why trade unions exist, and many other social facts.

4. Third, we are not required to treat utility as a measure of some introspectively available psychological state, such as pleasure. This represents a break from the early utilitarian conception of utility as a hedonic sate (Bentham 1789 [1907] ch. 1), a break consequent upon the increasingly behaviouristic tenor of philosophy of mind in the early twentieth century. Instead, we should think of utility just as whatever it is that a rational person seeks to maximize the expectation of. It is thus a theoretical term, rather like mass and force in Newtonian mechanics.

5. Fourth, there is no assumption that rational agents are selfish, except in the truistic sense that every agent seeks to maximize the expectation of *his own* utility function. But that is compatible with any assignment of utilities to outcomes that are other-regarding: for instance, giving money to

charity might be utility-maximizing and hence rational if you care about the welfare of certain others, so that your utility function assigns a high value to states that improve the welfare of others (see further my 2013). These two last points might seem to make the theory a tautology. But they do not: the fact that it makes definite predictions about what we observe follows from the fact that some observations seem to contradict the theory: this is in fact one of the objections to rational choice theory to which I now turn.

6. One very simple objection is that the theory is psychologically unrealistic: it implies that agents calculate weighted averages of utilities when making decisions between options in everyday situations. But agents obviously do no such thing: even in the very simple cases discussed in the exercises in Lectures 1 and 2, it is implausible that anyone *calculates* whether or not to take an umbrella before going out.

7. M. Friedman (1953) replies that economic theory in general (and rational choice theory in particular) should be assessed by its observable implications for the subject matter of interest: in this case the subject matter is human *behaviour*, not whatever psychological processes precede it. The theory is adequate if people do in fact act as if they are more-or-less rational calculators, whether or not any human being ever is or could be such a thing. An analogy (Friedman 1951: 21): suppose we are trying to predict the behaviour of an expert billiards player. Then a theory that derives his actions from the assumption that he calculates angles of incidence, momenta etc. using Newtonian mechanics might not be psychologically accurate but it is likely to be more or less behaviourally adequate—otherwise he wouldn't be an *expert*. Friedman's position is a form of instrumentalism about theories: for discussion and criticism see Hausman 1992.

8. Perhaps a more serious objection is that the theory is not even *behaviourally* accurate. To get a vivid sense of this, consider two games of Russian Roulette. Case 1: a six-shooter contains four bullets, and you're asked how much you'll pay to remove one of them. Case 2: a six-shooter contains two bullets, and you're asked how much you'll pay to remove both of them. Question: if money is of no value to you if you are dead, would you pay more in Case 1 or in Case 2? Almost everybody would pay more in Case 2 but rational choice theory says that this is wrong.

9. But perhaps the most influential counterexamples to rational choice theory are the 'paradoxes' of Allais and Ellsberg. Here is the Allais paradox (Allais 1953). There are three states, S1, S2 and S3 with Pr (S1) = 0.33, Pr (S2) = 0.66 and Pr (S3) = 0.01. Consider these two gambles:

A has a prize of $250 if S1, $240 if S2 and $0 if S3

B has a prize of $240 for sure

Which do you prefer?

Now consider the following two gambles:

C has a prize of $250 if S1, $0 if S2 *or* S3
D has a prize of $240 if S1 *or* S3, $0 if S2

Which do you prefer?

Experimental evidence (Kahneman and Tversky 1979) shows clearly that most people prefer B to A and C to D; again this is inconsistent with rational choice theory. That is: there is *no* utility function and subjective probability assignment to S1, S2 and S3 on which $B \succ A$ and $C \succ D$ are jointly consistent with equation (7.1) from Lecture 1.

10. On the other hand there is also evidence that with time and practice people do conform more closely to the rational choice ideal (Binmore 2009: 23f.). So on this empirical question perhaps Ramsey got it right when he wrote of the theory that it 'cannot be made adequate to all the facts; but it seems to me a useful approximation to the truth particularly in the case of our self-conscious or professional life, and it is presupposed in a great deal of our thought' (1926 [1990]: 69).

Exercises

1. The example in no. 8 is known as *Zeckhauser's paradox*. Show that rational choice theory forbids paying more in Case 2 of Zeckhauser's paradox.
2. Prove the statement at the end of no. 9.

References

Ahmed, A. 2013. From Game Theoretical Accounts of Cooperation to Meta-Ethical Choices. *Studies in Christian Ethics* 26: 176-83.

Allais, M. 1953. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine. *Econometrica* 21: 503-546.

Bentham, J. 1789 [1907]. *Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press.

Binmore, K. 2009. *Rational Decisions*. Princeton: Princeton, UP.

Friedman, M. 1953. The methodology of positive economics. In his *Essays in Positive Economics*. Chicago: Chicago UP: 3-43.

Jeffrey, R. 1983. *The Logic of Decision*. 2d ed. Chicago: Chicago UP.

Ramsey, F. P. 1926 [1990]. Truth and probability. In his *Philosophical Papers*, ed. D. H. Mellor. Cambridge: CUP 1990: 52-94.

Von Neumann, J. and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton: Princeton UP 1944.