

Knowledge and Reality A: Lecture Six

1. Formal Epistemology

Some choices are easy. If you like Pepsi and hate Coke, what should you decide about which to drink? PEPSI! If you want to go to the pub with your mates but your lover will make your life hell because of it, what do you do? Depends how bad a hell!

We do some *really basic* maths to show which option is better. Fix a 'utility' to picking Coke and Pepsi. If you *hate* Coke then you'll lose out. Let's say -10 Utility Points. If you *love* Pepsi then you'll gain. Let's say 10 Utility Points. The choice is clear – one option gives you 10, the other option loses you 10. Nor does it matter how difficult it is to determine these 'Utility Points'

Going to the pub is worth 20 UP. Say that if your missus/bloke hating you isn't that bad (say you hate them too, or they're too mild mannered to be really mean) then it's -5 UP. Say that if your missus/bloke would do terrible things to you then it's -25 UP. In the former you do it, in the latter you don't.

And other things can be factored in! If you'd feel *guilty* at doing it, maybe that's -20 UP and even in the first case you shouldn't do it. Things can get more complicated. Perhaps if your lover came home he/she would discover you gone and make your life hell to the tune of 25 UP. But you reckon there's only a 1 in 5 chance of them coming home. If you stay in there's a 100% chance of losing nothing and gaining nothing. So if you go to the pub there's a 100% chance of you gaining 20 UP and a 20% chance of you losing 25. The calculation is simple: your *expected* utility gain is:

$$20 - (0.2 \times 25) = 20 - 5 = 15$$

2. The Centipede Game

Two people, each taking a turn. There are 8 coins. You get to keep as many coins as you pick up. On your turn you can:

- (i) Pick up a coin and then let it be the other player's turn.
- (ii) Pick up two coins, and then the game ends. You keep what you've got.

The game ends when we run out of coins. What weird result apparently arises in this game?

3. Newcomb's Paradox: The One Box Solution

In this paradox the combination of two intuitively true principles lead us astray. The first is the principle of *maximum expected utility*.

You should do what is likely to bring about the maximum amount of utility.

If you have to gamble and betting it on Horse A is 80% likely to yield £30 and betting on Horse B is 20% likely to yield £200, what should you do?

Horse A expected yield: $0.8 \times 30 = £24$

Horse B expected yield: $0.2 \times 200 = £40$

Ceteris paribus bet on Horse B (and things might not be equal: maybe the Mafia will break your legs unless you get them £30. In which case the *utility* isn't in line with monetary gain. But for demonstration purposes we'll stick with simple numbers and money.)

Now turn to the paradox. Imagine there is a room with two boxes in it. One box is transparent: it has £1,000. Another box is opaque: but you know it either has nothing in it, or £1,000,000. You can take the contents of either one box, or both boxes. Here's the twist. There's some Derren Brown-esque character present. He has predicted, with 100% accuracy, which box you will pick. If you choose *two* boxes, there will be nothing in the opaque box. If you choose *one* box, there will be £1,000,000 in the opaque box. So, how many boxes do you pick?

If you pick one box, and pick the one you can see into, you will *definitely* get £1,000. If you pick one box, and pick the one you can't see into, what are the chances of you getting £1,000,000? Well, if Derren is a *perfect* predictor, the chance is 100%. So you'll definitely get £1,000,000.

If you pick two boxes, what's the chances of you getting what? As Derren is the perfect predictor, you have 100% chance of having nothing in the second box. So you will *only* get £1000. Clearly the Maximum Expected Utility principle says to pick one box – the one you can't see into.

4. Newcomb's Paradox: The Two-Box Solution

The other principle is the Dominance Principle.

If you have some choices, and one of them, choice X, has no downsides that the others don't have and there's at least a chance that doing X will be good for you then you should do it.

Example: Stay at home, or go to the pub. Imagine that staying at home has no upsides or downsides. Imagine that going to the pub has no downsides (I get free beer, I have no problem with walking to the pub etc.) but a high chance I'll have fun. I should go to the pub! Imagine there's a *small* chance I'll have fun. I should *still* go to the pub! Imagine there's a one in a trillion chance I'll have fun. I should *still* go to the pub! At least, if there's no difference whatsoever between being in the pub and being at home. One choice *dominates* the others. It is clearly better, because it is as good as the other choices *and* has the benefit of (possibly, or even definitely) being better.

How does this work in the Newcomb Paradox? Imagine your mother could see inside both boxes, and could advise you on how many to take. And let's imagine you trust your mother implicitly. If there were £1,000,000 in the opaque box and £1,000 in the transparent box, how many boxes would she tell you to pick? *Both of them!* You'd come away with £1,001,000 rather than £1,000,000. Clearly, then, you'd be well advised to take the advice of your mother in that case.

What if there were only £1,000 in the transparent box and nothing in the opaque box? Your mother would *still* tell you to pick both boxes as you'd walk away with £1,000 rather than nothing (as if you picked one box, you'd presumably pick the opaque, empty, box). *So in that case you also should follow her advice and take both boxes!*

Here's the rub. It doesn't matter whether your mother is there or not. You know that *either way* she'd tell you to take both boxes. So you know that *either way* it's in your interests to pick both boxes. That choice *dominates* the other choices. So, no matter what happens – you should pick both boxes. Think of it another way: it's too late for Derren to change the boxes now. *Whatever* you decide what's in the boxes is what's in the boxes. So, screw it, you may as well take as much as you can – and take *both* boxes.

5. Newcomb's Paradox: The Bit Where I Explain Why It's A Paradox

So NP is a paradox because one principle says to do one thing, and another principle says to do another.

MEU: Pick one box.

DP: Pick two boxes.

And *both* principles look to be good, rational principles. Something has to go! Something has to be revised! A principle concerning how choices are made has to be altered!

5. Skills

This topic really is about your own responses to the material. In the others, I've stressed that you could go off and read books on it. If you do read the material, you'll see that a lot of it is tricky. Lots of maths and tables and stuff. That can be okay. You need to master the art of learning what to read and what to ignore. Especially in a discipline like philosophy where you can be dipping in and out of many *other* disciplines (do applied ethics, read politics and economics; do metaphysics, read physics journals etc.) You have to try and get out of an article what you can. The complex bits may be so complex you won't get *anything* out of them. Just skip them. You can't argue *against* them, so you *may as well assume that it works*. If the author says it, then you may as well believe it! It's called 'accepting for the purpose of argument'