Chapter 14 Subjective Utility

14.1 Preference Based Theories

So far we've looked at two big theories of the nature of preferences. Both of them have thought that in some sense people don't get a say in what's good for them. There is an impersonal fact about what is best for a person, and that is good for you whether you like it or not. The experience theory says that it is the having of good experiences, and the objective list theory says that it includes a larger number of features. Preference-based, or 'subjective' theories of welfare start with the idea that what's good for different people might be radically different. It also takes the idea that people often are the best judge of what's best for them very seriously.

What we end up with is the theory that A is better for an agent than B if and only if the agent prefers A to B. We'll look at some complications to this, but for now we'll work with the simple picture that welfare is a matter of preference satisfaction. This theory has a number of advantages over the more objective theories.

First, it easily deals with the idea that different things might be good for different people. That's accommodated by the simple fact that people have very different desires, so different things increase their welfare.

Second, it also deals easily with the issues about comparing bundles of goods, either bundles of different goods, or bundles of goods at different times. An agent need not only have preferences about whether they, for instance, prefer time with their family to material possessions. They also have more fine-grained preferences about various trade offs between different goods, and trade offs about sequences of goods across time. So if one person has a strong preference for getting goods now, and another person is prepared to wait for greater goods later, the theory can accommodate that difference. Or if one person is prepared to put up with unpleasant events in order to have greater goods at other times, the theory can accommodate that, as well as the person who prefers a more steady life. If they are both doing what they want, then even though they are doing different things, they are both maximising their welfare.

But there are several serious problems concerning this approach to welfare. We'll start with the intuitive idea that people sometimes don't know what is good for them.

We probably all can think about things in everyday life where we, or a friend of ours, has done things that quite clearly are not in their own best interests. In many such cases, it won't be that the person is doing what they don't want to do. Indeed, part of the reason that people acting against their own best interests is such a problem is that the actions in question are ones they very much want to perform. Or so we might think antecedently. If a person's interests are just measured by their desires, then it is impossible to want what's bad for you. That seems very odd.

It is particularly odd when you think about the effect of advertising and other forms of persuasion. The point of advertising is to change your preferences, and presumably it works frequently enough to be worth spending a lot of money on. But it is hard to believe that the effect of advertising is to change how good for you various products are. Yet if your welfare is measured by how many of your desires are satisfied, then anything that changes your desires changes what is good for you.

Note that sometimes we even have internalised the fact that we desire the wrong things. Sometimes we desire something, while desiring that we don't desire it. So we can say things like "I wish I didn't want to smoke so much". In that case it seems that what would, on a strict subjective standpoint, have our best outcome be smoking and wanting not to smoke, since then both our 'first-order' desire to smoke and our 'second-order' desire not to want to smoke would be satisfied. But that sounds crazy.

Perhaps the best thing to do here would be to modify the subjective theory of welfare. Perhaps we could say that our welfare is maximised by the satisfaction of those desires we wish we had. Or perhaps we could say that it is maximised by the satisfaction of our 'undefeated' desires, i.e. desires that we don't wish we didn't have. There are various options here for keeping the spirit of a subjective approach to welfare, while allowing that people sometimes desire the bad.

14.2 Interpersonal Comparisons

I mentioned above that the subjective approach does better than the other approaches at converting the welfare someone gets from the different parts of their life into a coherent whole. That's because agent's don't only have preferences over how the parts of their lives go, they also have preferences over different distributions of welfare over the different parts of their lives, and preferences over bundles of goods they may receive. The downside of this is that a kind of comparison that the objective theory might do well at, interpersonal comparisons, are very hard for the subjective theorist to make.

Intuitively there are cases where the welfare of a group is improved or decreased by a change in events. But this is hard, in general, to capture on a subjective theory of welfare. There is one kind of group comparison that we can make. If some individuals prefer A to B, and none prefer B to A, then A is said to be a Pareto-improvement over B. (The name comes from the Italian economist Wilfredo Pareto.) An outcome is Pareto-optimal if no outcome is a Pareto-improvement over it.

But Pareto-improvements, and even Pareto-inefficiency, are rare. If I'm trying to decide who to give $1000 to, then pretty much whatever choice I make will be Pareto-optimal. Assume I give the money to x. Then any other choice will involve x not getting $1000, and hence not preferring that outcome. So not everyone will prefer the alternative.

But intuitively, there are cases which are not Pareto-improvements which make a group better off. Consider again the fact that the marginal utility of money is declining. That suggests that if we took $1,000,000 from Bill Gates, and gave $10,000 each to 100 people on the borderline of losing their houses, then we'd have increased the net welfare. It might not be just to simply take money from Gates in this way, so many people will think it would be wrong to do even if it wouldn't increase welfare. But it would be odd to say that this didn't increase welfare. It might be odder still to say, as the subjective theory seems forced to say, that there's no way to tell whether it increased welfare, or perhaps that there is no fact of the matter about whether it increased net welfare, because welfare comparisons only make sense for something that has desires, e.g. an agent, not something that does not, e.g. a group.

There have been various attempts to get around this problem. Most of them start with the idea that we can put everyone's preferences on a scale with some fixed points. Perhaps for each person we can say that utility of 0 is where they have none of their desires satisfied, and utility of 1 is where they have all of their desires satisfied. The difficulty with this approach is that it suggests that one way to become very very well off is to have few desires. The easily satisfied do just as well as the super wealthy on such a model. So this doesn't look like a promising way forward.

Since we're only looking at decisions made by a single individual here, the difficulties that subjective theories of

welfare have with interpersonal comparisons might not be the biggest concern in the world. But it is an issue that comes up whenever we try to apply subjective theories broadly.

14.3 Which Desires Count

There is another technical problem about using preferences as a foundation for utilities. Sometimes I'll choose A over B, not because A really will produce more welfare for me than B, but because I think that A will produce more utility. In particular, if A is a gamble, then I might take the gamble even though the actual result of A will be be worse, by anyone's lights, including my own, than B.

Now the subjectivist about welfare does want to use preferences over gambles in the theory. In particular, it is important for figuring out how much an agent prefers A to B to look at the agent's preferences over gambles. In particular, if the agent thinks that one gamble has a 50% chance of generating A, and a 50% chance of generating C, and the agent is indifferent between that gamble and B, then the utility of B is exactly half-way between A's utility and C's utility. That's a very useful thing to be able to say. But it doesn't help with the original problem - how much do we value actual outcomes, not gambles over outcomes.

What we want is a way of separating *instrumental* from *non-instrumental* desires. Most of our desires are, at least to some extent, instrumental. But that's a problem for using them in generating welfare functions. If I have an instrumental desire for A, that means I regard A as a gamble that will, under conditions I give a high probability of obtaining, lead to some result C that I want. What we really want to do is to specify these non-instrumental desires.

A tempting thing to say here is to look at our desires under conditions of full knowledge. If I know that the train and the car will take equally long to get to a destination I desire, and I still want to take the train, that's a sign that I have a genuine preference for catching the train. In normal circumstances, I might catch the train rather than take the car not because I have such a preference, but because I could be stuck in arbitrarily long traffic jams when driving, and I'd rather not take that risk.

But focussing on conditions of full knowledge won't get us quite the results that we want. For one thing, there are many things where full knowledge changes the relevant preferences. Right now I might like to watch a football game, even though this is something of a gamble. I'd rather do other things conditional on my team losing, but I'd rather watch conditional on them winning. But if I knew the result of the game, I wouldn't watch - it's a little boring to watch games where you know the result. The same goes of course for books, movies etc. And if I had full knowledge I wouldn't want to learn so much, but I do prefer learning to not learning.

A better option is to look at desires over fully specific options. A fully specific option is an option where, no matter how the further details are filled out, it doesn't change how much you'd prefer it. So if we were making choices over complete possible worlds, we'd be making choices over fully specific options. But even less detailed options might be fully specific in this sense. Whether it rains in an uninhabited planet on the other side of the universe on a given day doesn't affect how much I like the world, for instance.

The nice thing about fully specific options is that preferences for one rather than the other can't be just instrumental. In the fully specific options, all the possible consequences are played out, so preferences for one rather than another must be non-instrumental. The problem is that this is psychologically very unrealistic. We simply don't have that fine-grained a preference set. In some cases we have sufficient dispositions to say that we do prefer one fully specific option to another, even if we hadn't thought of them under those descriptions. But it isn't clear that this will always be the case.

To the extent that the subjective theory of welfare requires us to have preferences over options that are more complex than we have the capacity to consider, it is something of an idealisation. It isn't clear that this is necessarily a bad thing, but it is worth noting that the theory is in this sense a little unrealistic.

14.4 Money and Utility

In simple puzzles involving money, it is easy to think of the dollar amounts involved as being proxy for the utility of each outcome. In a lot of cases, that's a very misleading way of thinking about things though. In general, a certain amount of money will be less useful to you if you have more money. So $1000 will be more useful to a person who earns $20,000 per year than a person who earns $100,000 per year. And $1,000,000 will be more useful to either of them than it will be to, say, Bill Gates.

This matters for decision making. It matters because it implies that in an important sense, $\$2x$ is generally not twice as valuable to you as $\$x$. That's because $\$2x$ is like getting $\$x$, and then getting $\$x$ again. (A lot like it really!) And when we're thinking about the utility of the second $\$x$, we have to think about its utility not to you, but to the person you'll be once you've already got the first $\$x$. And that person might not value the second $\$x$ that much.

To put this in perspective, consider having a choice between $1,000,000 for certain, and a 50% chance at $2,000,000. Almost everyone would take the sure million. And that would be rational, because it has a higher utility. It's a tricky question to think about just what is the smallest x for which you'd prefer a 50% chance at $\$x$ to $1,000,000. It might be many many times more than a million.

The way economists' put this is that money (like most goods) has a *declining marginal utility*. The marginal utility of a good is, roughly, the utility of an extra unit of the good. For a good like money that comes in (more or less) continuous quantities, the marginal utility is the slope of the utility graph, as below.

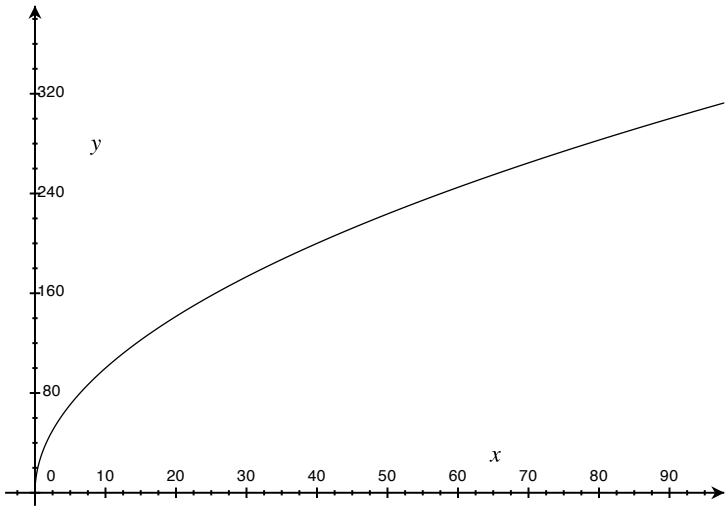

Figure 14.1: Declining Marginal Utility of Money

You should read the x-axis there are measuring possible incomes in thousands of dollars per year, and the y-axis as measuring utility. The curve there is $y = x^{\frac{1}{2}}$. That isn't necessarily a plausible account of how much utility each income might give you, but it's close enough for our purposes. Note that although more income gives you more utility, the

amount of extra utility you get from each extra bit of income goes down as you get more income. More precisely, the slope of the income-utility graph keeps getting shallower and shallower as your income/utility rises. (More precisely yet, a little calculus shows that the slope of the graph at any point is $\frac{1}{2y}$, which is obviously always positive, but gets less and less as your income/utility gets higher and higher.)

The fact that there is a declining marginal utility of money explains certain features of economic life. Imagine the utility an agent gets from an income of x dollars is $x^{\frac{1}{2}}$. And imagine that right now their income is \$90,000. But there is a 5% chance that something catastrophic will happen, and their income will be just \$14,400. So their expected income is $0.95 \times 90,000 + 0.05 \times 14,400 = 86220$. But their expected utility is just $0.95 \times 300 + 0.05 \times 120 = 291$, or the utility they would have with an income of \$84,861.

Now imagine this person is offered insurance against the catastrophic scenario. They can pay, say, \$4,736, and the insurance company will restore the \$75,600 that they will lose if the catastrophic event takes place. Their income is now sure to be \$85,264 (after the insurance is taken out), so they have a utility of 292. That's higher than what their utility was, so this is a good deal for them.

But note that it might also be a good deal for the insurance company. They receive in premiums \$4,736. And they have a 5% chance of paying out \$75,600. So the expected outlay, in dollars, for them, is \$3,780. So they turn an expected profit on the deal. If they repeat this deal often enough, the probability that they will make a profit goes very close to 1.

The point of the example is that people are trying to maximise expected utility, while insurance companies are trying to maximise expected profits. Since there are cases where lowering your expected income can raise your expected utility, there is a chance for a win-win trade. And this possibility, that expected income can go down while expected utility can go up, is explained in virtue of the fact that there is a declining marginal utility of money.