Sen on Social Welfare

Brian Weatherson, March 15, 2018

Interpersonal Utility Comparisons

The big question is can we ever make sense of questions like these:

- Is Brian better off than Josh in this situation?
- Brian is better off in situation A than in situation B, and Josh is better off in B than in A. But is the difference between A and B bigger for Brian or Josh?

The **Impossibility of Interpersonal Utility Comparisons** thesis (IIUC) is that these questions have no meaningful answer. This is extremely unintuitive. We can easily imagine cases where there is a very intuitive answer to them. But there are a couple of arguments to the effect that the comparison is impossible.

First, there is an *epistemological argument*, as follows.

1. We can only tell what people choose, not absolute questions about how absolutely well off they are, or how their situation compares with others.
2. What's scientifically true is a function of what we can tell.
3. So there are no facts about how absolutely well off people are, or how their situation compares with others.

Premise 2 of this argument is a statement of (a fairly crude form of) logical positivism. If IIUC rests on this crude a philosophy of science, then it isn't really worth taking seriously. But maybe there are more careful arguments in the vicinity. One of these is a *metaphysical argument* concerning choice.

1. How well off someone is just is a matter of what preferences they have satisfied.
2. But preferences are always *comparative* and *intra-personal*.
3. So there is no fact of the matter about someone's *absolute* welfare, because preferences are comparative, or *inter-personal* comparative welfare, because choices are intra-personal.

The assumption here is that the **preference satisfaction theory of welfare** is true. This is the theory that, as premise 1 says, how well off you are is a matter of what preferences get satisfied. This is the stock-standard view of welfare in economics, and it is a popular view in philosophy as well. And this argument does, I think, go through if premise 1 is true.

Sen alludes to another argument, also a broadly metaphysical argument, of the following form.

1. How well off someone is just is a matter of how they feel.
2. There is no way to compare feelings across people, or to measure absolutely how someone feels.
3. So there is no absolute or inter-personal, measure of how well off people are.

This is a much less interesting argument, because both premises seem very dubious. It is hard to measure precisely how different people are feeling, but only a sceptic should say it is literally impossible. We know what happy and sad people are basically like. So premise 2 is pretty implausible. And premise 1 is also hard to defend. Just which feelings people value is different. I *hate* some of the feelings that go along with strenuous exercise, but some people seem to positively value them. (Some people exercise for instrumental reasons, but some really seem to genuinely like it.) On the theory of welfare being supposed here, there is a fact about whether having those feelings makes you better or worse off, even though I hate them and others love them. It seems better to say that welfare is getting what you want, and sometimes we want certain kinds of phenomenal states, and sometimes we want to rule the world, and sometimes we want to be remembered well, and so on.

Two Kinds of Comparisons

The IIUC, as I'm interpreting it, really has two parts. And the two parts are at least logically independent.

1. There are no facts of the matter about how well off anyone is on any kind of absolute scale.
2. There are no facts about how one person's improvement in welfare between two states compares to someone else's improvement in welfare.

If you want less abstract versions of those, see the two questions at the very top. At least in principle, you could reject one but not the other.

What we might call a **special point** theory of welfare says that there is at least one special point in the welfare scale, and everyone's welfare can be compared to that point, and everyone 'above' the point is better off than everyone 'below' the point. A really absolutist theory of welfare, something that said welfare was like mass or height, would say that all points are special points. But we might just believe that there is one special point. (List calls this the 0 point in some cases.)

A theory that said this and nothing more would reject 1 - it says that sometimes we can compare people, at least if one of them is above the special point, and one is below. But it would accept 2 - there are no comparisons in how much better off someone is than someone else.

It's a little more of a challenge, but we can come up with theories that are at least consistent, and reject 1 but not 2. (For the record, allow arbitrary additive transformations of any one person's utility function.) But such theories are less interesting than special point theories, because they don't have an interesting philosophical interpretation.

But special point theories do. And they do because we have some sense of a division between **needs** and **wants**. We might think the special point is the point where someone has everything they need, and nothing more. So here is our toy theory of welfare.

- There are two great classes of people, those who have everything they need, and those who do not.
- Everyone in the first class is better off than everyone in the second class.
- Within each of the two classes, the IIUC holds. (Perhaps people are ordered by preference satisfaction.)

This theory of welfare does allow a basic kind of social welfare metric. We can ask, for different societies, what percentage of people are in the good place, and order societies by this percentage. It's a pretty crude social welfare function, but it's not useless. And we really do see some useful measures of society by this kind of metric - e.g., measuring what percentage of different societies are above the poverty line.

Transformations

One way of putting the point about IIUC is that utility functions are only defined up to **positive affine transformations**. That's to say, if U_1 is an accurate representation of how well off I am in some situations, then so is U_2, provided the following equation holds for some $a > 0$ and b.

$$U_2(X) = aU_1(x) + b$$

One consequence of this is that if we transform one person's utility measure, but not another's, we can be left with a very odd account of how well off society is. For example, on this table, I'm showing the utility that each of the two people in the little society (Brian and Josh) get from two possible situations, A and B.

Utility	Brian	Josh
A	11	-4
B	4	4

It looks like B is the better option according to each of the following three rules:

Utilitarianism The social welfare is the sum of each person's welfare.
Maximin The social welfare is the welfare of the worst off.
Above-Zero The social welfare is the percentage of people with welfare above 0.

But now consider the following table

Utility	Brian	Josh
A	6	9
B	-1	11

All three rules now say that A is better. But all I've done is apply the following two transformations. I used the $y = x - 5$ transformation on Brian's utility, and the $y = x/4 + 10$ transformation on Josh's. According to IIUP, these transformations are meaning preserving - these two tables say *exactly the same thing*. But that means that all three of these rules are sensitive to transformations that IIUP says are meaning preserving. If we are going to use any of these rules, then we need to find some way to resist IIUP.

Why Even Utility

You might note that affine transformations are still more informative than a purely ordinal ranking. Why not just any positive monotonic transformation/A transformation f is positive monotonic if $f(x) > f(y)$ iff $x > y$. Well, that's because people can make choices between bets. So if we know that I'm indifferent between B, on the one hand, and a 50/50 chance of getting A and getting C, on the other, then we know that $U(B)$ is half-way between $U(A)$ and $U(C)$. And only transformations that preserve that feature of my utility function can be meaning preserving. That means, in effect, the positive affine transformations. But this still leaves a huge range.

Arrow

As I said in class, it's really natural to think of Arrow as putting constraints on **voting systems**. After all, he starts with preference orderings, and those are really the inputs to voting systems, not to social welfare functions in general. But as Sen makes clear, when we look at Arrow's work in its proper historical context, it's clear he's aiming for something much stronger than this. It's a really pessimistic result about the very possibility of a social welfare function. And remember, without a social welfare function, we can't answer questions like

- Was this society, or that society, better off?
- Did this event make society better off, or worse off?
- Would this charitable plan make society better off, or that charitable plan?

This is a really dramatically sceptical result. And we get there in two steps.

1. IIUP is true, so any social welfare function can only use intra-personal comparisons (i.e., preference orderings.)
2. There is no good way to combine intra-personal comparisons into a social ordering (thanks to Arrow's theorem).

Looked at this way, Arrow's result is not just an annoyance for people designing voting systems, it's a challenge for anyone who wants to say anything systematic about social welfare in general.

Sen

Sen's work has involved pushing back on this scepticism on basically every possible front.

He noted that the impossibility theorem isn't as much a dead end as we thought. Ideally, what we'd have is a set of axiomatic constraints on a rule for combining preferences that are collectively satisfied by precisely one combination rule. Arrow shows that we can build some plausible rules that are satisfied by precisely zero combination rules. Not great, but one is really close to zero. Maybe we're close! That's why thinking about things like dropping the idea that the output of the combination rule is connected matter. Maybe a very slight weakening of the Arrow conditions can get us from zero back to one. To be fair, while this is a good idea in theory, it hasn't really worked in practice. But the other parts of his theory have been more productive.

He argues that we have been too quick to equate welfare with preference satisfaction. The problem here concerns **adaptive preferences**. People adjust their aims to what is available. And while this might be good for mental health on the whole, it makes using preference satisfaction as a measure of welfare problematic. People who have low expectations because of oppressive situations aren't well off when those low expectations are met.

But he also argues that we've been too slow to acknowledge the more physical, and at least roughly measurable, ways in which things can be better or worse for people. I mentioned a rough and ready version of this above, the distinction between needs and wants. But we can be more subtle than that.

It would take us way far afield to go into the details of this, but one important idea is to identify how well off someone is with their **capacities**. (Or, less contentiously, to say that improving capacities is a way to increase welfare.) Most fundamentally, someone who has their needs met has the capacity to stay alive. But other capacities might increase our welfare beyond that.

The capacities approach is controversial. Imagine that I have the capacity to wiggle my ears. But I don't like wiggling my ears, so I never do this. Then one day, as a side-effect of a virus, I lose this capacity. Does this really make my welfare go down? It isn't obvious that it does.

But the bigger picture is that we shouldn't expect questions about how to measure **social** welfare to be distinct from questions about how to measure **individual** welfare. Even once we have said something about individual welfare, there are further interesting social questions. (For example, is it a bad thing that the utilitarian rule neglects distributional questions? Sen thinks it is.) But the two are going to be tied together.