34. Social Welfare

The one sub–topic which I will cover from this short chapter is the concept of a *(Bergson–Samuelson) Social Welfare Function*, which is covered in subsections 34.2 – 34.4 of the text.

The whole point of Pareto efficiency — and the whole problem with Pareto efficiency — is Pareto's unwillingness to make interpersonal comparisons of utility.

That is, suppose we have two different allocations, both of which are on the contract curve (in a two–person exchange economy, as in chapter 32). We cannot rank these allocations using the concept of Pareto efficiency. All that concept tells us is that there is no other feasible allocation which everyone likes better.

A *Social Welfare Function* is a way of talking about the ranking of different Pareto efficient allocations. Now it's not really a way of ranking the allocations. It's just a way of talking about how someone else might rank the allocations.

The starting point for a social welfare function is individual utility functions for the people in the economy. Varian starts out (in subsection 34.2) with a fairly general formulation of individual's utility functions. I am going to be more narrow. My utility function for person 1 is going to be a function

$$u^1(x_1^1, x_2^1, \ldots, x_N^1)$$

if there are N consumption goods.

So notice that my utility function for person depends — only — on the bundle of goods and quantities which she herself consumes. It's exactly the utility function used in chapters 4 and 5 to explain consumer choice and consumer demand (in econ 2300).

This restriction, that person 1's utility function is the same utility function she uses to figure what to consume in a market economy, is what makes the social welfare function a "Bergson–Samuelson" social welfare function. Varian starts out more generally in section 34.2. In particular, he considers the possibility that I get utility not only from the bundle of goods and services which I am consuming, but also from the bundles of goods and services that other people are consuming. That's an interesting generalization. It's also very relevant to the theory of externalities, which Varian discusses in chapter 35 — and which I will not discuss in econ 2350.

Now I will remind you of a few features of the utility functions we used in econ 2300. A utility person is a way of measuring how much a person likes different consumption bundles. It can be thought of as a way of keeping score : if $\mathbf{x} \equiv (x_1^1, x_2^1, \ldots, x_N^1)$ and $\mathbf{z} \equiv (z_1^1, z_2^1, \ldots, z_N^1)$ are two different consumption bundles for person 1, saying that "$u^1(\mathbf{x}) > u^1(\mathbf{z})$" is another way of saying that person #1 prefers the bundle $\mathbf{x}$ to the bundle $\mathbf{z}$, or that the bundle $\mathbf{x}$ is on a higher indifference curve for person #1 than is the bundle $\mathbf{z}$.

Now in econ 2300, we emphasized that these utility functions were just a way of keeping score. In subsection 4.1, Varian explained that he thinks it would be very silly to regard the actual value of $u^1(x_1^1, x_2^1, \ldots, x_N^1)$ as having any significance.

In particular, in econ 2300, we could change the utility function, by a **monotonic transformation**. That is, in econ 2300, the utility functions

$$u^1(x_1^1, x_2^1) = x_1 x_2$$

and

$$U^1(x_1^1, x_2^1) = \sqrt{x_1^1 x_2^1}$$

represented the same preferences for person 1. The reason they represented the same preferences was that U^1 is the square root of u^1, which means that U^1 is a monotonic transformation of u^1. These two utility functions give rise to the exact same set of indifference curves for person 1. And they will give rise to the exact same demand functions for person 1, if she maximizes her utility over all the consumption bundles in her budget set.

But in doing welfare economics, we have to address the very sticky issue of how well–off a person is from some allocation. We have to somehow consider how much happier some change in consumption might make one person. In moving up and to the right along the contract curve, in a 2–person exchange economy, we are making person #1 better off and person #2 worse off. In order to rank allocations, we have to somehow pretend that we can think of how much better off person #1 is, and how much worse off person #2 is.

That's what a cardinal utility function is. Now $u^1(x_1^1, x_2^1, \ldots, x_N^1)$ is not just a score for the bundle $\mathbf{x} \equiv (x_1^1, x_2^1, \ldots, x_N^1)$, to be compared with the scores from some other bundles which might be available for her. Now $u^1(x_1^1, x_2^1, \ldots, x_N^1)$ is a measure of how happy person #1 would be if she consumed this bundle — to be compared with how happy other people are with their own bundles.

So using a Bergson–Samuelson social welfare function means we are somehow comparing the increase in happiness to person #1, when we move up the contract curve, with the decrease in happiness of person #2.

So I'm doing 2 things. First, I am pretending that I can measure the happiness of each person in the economy. And second, I am using these measures of individual happiness to construct a measure of overall happiness for the whole many–person economy.

When I do that, I have a social welfare function, which gives an overall score for each feasible allocation. This overall score is based on the utility level of all the different people in the society.

So, formally, a social welfare function is a function

$$W(u^1, u^2, \ldots, u^H)$$

which is defined on the H–vectors of individuals' utility, when we have H people in the country. [It's a Bergson–Samuelson social welfare function if the individual utility of each person is based — only — on her own consumption bundle.]

So (repeating) there are two key properties which make $W(\cdot)$ a Bergson–Samuelson social welfare function : (1) it depends (only) on individual people's utilities ; (2) those individual utilities come from the same utility functions people use when they make consumption decisions in a market economy.

One more things : social welfare should increase with individual happiness. Or at least it wouldn't decrease. So for $W(u^1, u^2, \ldots, u^H)$ to be a social welfare function, I certainly am also going to require that

$$\frac{\partial W}{\partial u^h} \geq 0$$

for each person h.

There are a lot of different functions which satisfy those assumptions. A couple of examples :

$$W(u^1, u^2, \ldots, u^H) \equiv u^1 + u^2 + \cdots + u^H \tag{Ben}$$

$$W(u^1, u^2, \ldots, u^H) \equiv u^1 u^2 \cdots u^H$$

$$W(u^1, u^2, \ldots, u^H) \equiv u^1 + 2u^2 + 3\log(u^3 + u^4 + \ldots + u^H)$$

$$W(u^1, u^2, \ldots, u^H) \equiv \min(u^1, u^2, \ldots, u^H) \tag{mm}$$

The first example, the arithmetic sum of people's utilities, is often referred to as a "utilitarian", or "Benthamite" social welfare function. The fourth example is often referred to by economists as a "Rawlsian", or "max–min" social welfare function.

What is the "right" welfare function? That's not the right question. The social welfare function is supposed to be a way of representing the decision–making of someone who is ranking the various possible allocations. This someone could be a politician, administrator, outside evaluator, voter, or anyone who has an opinion about the best policy for the economy in question.

It's a good way of representing that decision–making if the person in question respects the behaviour of the people in the economy : if person 1 thinks that the bundle $\mathbf{x} \equiv (x_1^1, x_2^1, \ldots, x_N^1)$ is better for her than the bundle $\mathbf{z} \equiv (z_1^1, z_2^1, \ldots, z_N^1)$, then the decision maker agrees with this ranking. (So if $\mathbf{x}$ contains a lot of alcohol and drugs, and bundle $\mathbf{z}$ contains a lot of kale and antioxidants, the decision–maker is not going to overrule the person's own ranking.)

So what welfare function represents what decision–maker? Clearly, if some decision–maker likes one person, or group of people, more than another, the decision–maker would give that (first) person's utility more weight in the social welfare function. My third example might be consistent who gave person #2's happiness twice as much importance as person #1's.

The other big factor is how "redistributive" the decision maker is. Think of a policy change which makes the best–off person even better off than before, and makes the worse–off person even worse off. Would this be a good idea? Of course it depends on how much the increase is for the best–off person, and how big the decrease for the worst off. What if the change increased the best–off person's utility by 1000 and decreased the worst–off's utility by 1? Some possible decision–makers might approve of that change, even though it increases inequality, because the "overall" gain was so big. Others might disapprove, feeling that the value to society of further gains by the best–off are very very small.

Consider three of the welfare function examples I had above, for the simpler case in which there are only 2 people. With the Benthamite social welfare function,

$$\frac{\partial W}{\partial u^1} = 1 = \frac{\partial W}{\partial u^2}$$

The value — to society — of one person's increase in utility is the same as another's no matter how well–off or badly–off the people are. With the Benthamite social welfare function, all that matters is the sum of utilities, not how the utilities are distributed. A Benthamite decision–maker would certainly approve a change which increased u^1 by 1000 and decreased u^2 by 1, no matter how high is u^1 initially, or how low is u^2.

The second example I had of a social welfare function was $W(u^1, u^2) = u^1 u^2$ so that

$$\frac{\partial W}{\partial u^1} = u^2 \quad ; \quad \frac{\partial W}{\partial u_2} = u^1$$

If person #1 is better off than person #2 (that is if $u^1 > u^2$) then, for this decision maker, increases to the worse–off person #2's utility are valued more than increases in the better–off person #1's. For the example I had above, of increasing u^1 by 1000 and decreasing u^2 by 1, then this decision–maker's measure of overall welfare would change by

$$1000u^2 - u^1$$

This decision–maker would approve the change (which increased the well–being of the best–off by a lot, and decreased the well–being of the worst off by a little) if and only if $u^1 < 1000u^2$. In other words, this second decision maker cares about income distribution, and will disapprove of policies which make that distribution less equal — if they do not increase overall income enough, and if the initial income distribution was already very skewed.

In place of my third example above, I'll give another social welfare function,

$$W(u^1, u^2) = 10 - \frac{1}{(u_1)^2} - \frac{1}{(u^2)^2}$$

This is a social welfare function : despite the minus signs, W does increase with both u_1 and u_2 since

$$\frac{\partial W}{\partial u^1} = \frac{1}{(u^1)^3} > 0$$

$$\frac{\partial W}{\partial u^2} = \frac{1}{(u^2)^3} > 0$$

In this case, the proposed change (raising u^1 by 1000, lowering u^2 by 1) changes overall welfare by

$$\Delta W = \frac{1000}{(u^1)^3} - \frac{1}{(u^2)^3} = \frac{1}{(u^1)^3}[1000 - (\frac{u^1}{u^2})^3]$$

so that the decision–maker would approve the change only if the best–off person #1 was no more than 10 times better off than the worst–off person #2. Decision–maker 3 is more averse to inequality than decision–maker 2 : if u^1/u^2 were between 10 and 1000, decision–maker 2 would like the policy which increases inequality (and overall income) but decision maker 3 would not.

In my final example, overall welfare is

$$W(u^1, u^2) = u^2$$

if $u^1 > u^2$: a "max–min" decision–maker cares only about the well–being of the worst–off person. This does fit my definition of a social welfare function, since I required only that $\frac{\partial W}{\partial u^1} \geq 0$, not that it be strictly positive. As long as $u^1 > u^2$, this decision maker would reject the policy change (which increases both overall income and inequality). Any policies which lowers the well–being of the worst–off is a bad policy for this decision maker.

So the functional forms for my examples of social welfare functions would be constructed from finding out how averse to inequality the decision–maker was.

[In my definition of a Bergson–Samuelson social welfare function, I did limit how far this aversion to inequality could go. I did not allow my social welfare function to be a decreasing function of any one person's utility. That rules out people who have such a strong aversion to inequality that they would be willing to approve policies which made everyone worse off, provided inequality were reduced enough. Such preferences could be represented by a welfare function which depended on the difference $u^1 - u^2$ between the utility of the best–off and the worst–off. There is nothing wrong with that sort of preference over inequality ; it just happens not to fit the definition of a Bergson–Samuelson social welfare function.]

The nice thing about Bergson–Samuelson social welfare functions is that they are consistent with Pareto efficiency. If my decision–maker gets to pick the allocation, from all the feasible allocations, which maximizes social welfare, then she will always choose a Pareto efficient allocation.

The proof is straightforward : suppose a decision maker chose, as the best allocation, some allocation which gave the two people utilities of u^1 and u^2 ; suppose now that some other feasible allocation yielded utilities v^1 and v^2, with $v^1 > u^1$ and $v^2 > u^2$; then this other feasible allocation must yield a higher W to the decision–maker ; so if an allocation is Pareto–dominated, it cannot be the one which maximizes W ; therefore the allocation which maximizes this decision–maker's W must be Pareto–efficient.