Notes on Social Choice and Welfare

Alejandro Saporiti

Social Choice and Welfare

Reference: Jehle, G., Reny, P.: *Advanced Microeconomic Theory*. Pearson, 3rd ed. (2011). Ch. 6.

Social choice and welfare (SCW) is to a great extent a normative field, in contrast with most of microeconomics, which is mainly positive, (i.e., it is concerned with characterizing and predicting individual and social behavior in economic environments without judging the nature of that behavior).

To judge whether a certain social state (e.g. the election of a political candidate, the division of a pie, a market allocation, etc.) is "good" or "bad", or "better" or "worse," it is necessary first to agree upon some (ethical) principles to rank the social states.

SCW makes explicit those principles and explores their logical implications.

Social Choice and Welfare

Consider, for instance, a 2×2 Edgeworth box economy.

- Each point, a way of dividing resources.
- Individual preferences in conflict over states.
- Social choice problem: which is best for society?
- Allocations outside *CC* rule out? (i.e., best state must be Pareto efficient?)
- Unequal allocations (e.g. $\bar{x}$) rule out?
- What else? How we trade off agents' well-being?

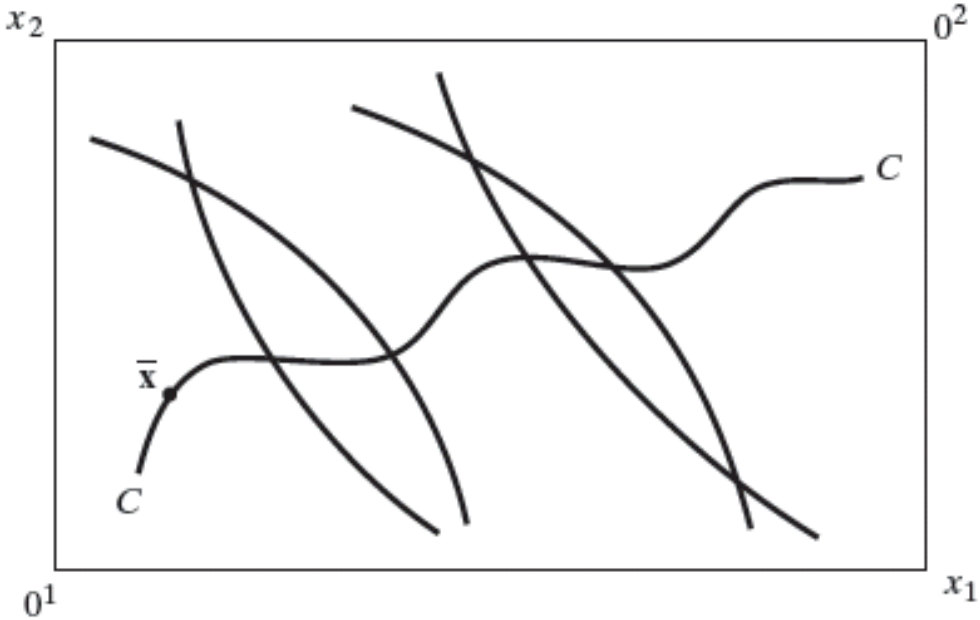

Figure 1 : 2×2 Edgeworth box economy

Social Choice Framework

Individuals are typically assumed to make choices based on some underlying *values* or preferences.

Social choices are then derived through some aggregation of individual preferences (for instance, via majority rule).

To which extent individual preferences can be aggregated into social preferences (or social choices) in a "satisfactory" way?

To study this problem, consider the following framework:

- $N = \{1, \ldots, n\}$ finite set of agents, $n \geq 2$;
- X finite set of (mutually exclusive) alternatives, $|X| > 2$;
- $\mathcal{B}$ set of all complete binary relations on X;
- $\mathcal{R}$ set of all weak (complete and transitive) orders on X (i.e., $\mathcal{R} \subset \mathcal{B}$);
- Each $i \in N$ endowed with weak order $R_i \in \mathcal{R}$;
- $\mathcal{R}^n$ set of society's preferences $\rho = (R_1, \ldots, R_n)$.

Social Welfare Function

Definition 1 (Social welfare function)

A social welfare function (swf) is a mapping $f : \mathcal{D} \subseteq \mathcal{R}^n \to \mathcal{B}$.

$R = f(R_1, \ldots, R_n)$ is interpreted as the social preference relation. Notice that it doesn't need to be rational (i.e., complete & transitive).

An example of a swf is majority rule, where for all $x, y \in X$,

$$x P_M y \Leftrightarrow |P(x, y; \rho)| > n/2.$$

As Fig 2 illustrates, majority voting may lead to cyclical social preferences!

Person 1	Person 2	Person 3
x	y	z
y	z	x
z	x	y

$\Longrightarrow$ $x P_M y P_M z P_M x$; i.e., majority preference relation violates transitivity in this case!

Figure 2 : Condorcet paradox

Arrow's Conditions

Consistency (transitivity) is not the only desirable property. Arrow suggested the following additional conditions:

Definition 2 (Swf's properties)

A social welfare function $f : \mathcal{D} \to \mathcal{B}$

- has unrestricted domain (UD) if $\mathcal{D} = \mathcal{R}^n$;
- is transitive (T) if for all $\rho \in \mathcal{D}$, $f(\rho)$ is transitive;
- is nondictatorial (ND) if there does not exist $i \in N$ such that, for every $\rho \in \mathcal{D}$, and every $x, y \in X$, $x P_i y \Rightarrow x P y$;
- is weakly Paretian (PA) if, for every $\rho \in \mathcal{D}$, and every $x, y \in X$, $x P y$ whenever $x P_i y$ for all $i \in N$;
- is independent of irrelevant alternatives (IIA) if, for every $\rho, \rho' \in \mathcal{D}$, and every $x, y \in X$, $\rho|_{\{x,y\}} = \rho'|_{\{x,y\}}$ implies $f(\rho)|_{\{x,y\}} = f(\rho')|_{\{x,y\}}$.

Arrow's Theorem

Are there any aggregation rules satisfying all these properties?

Theorem 1 (Arrow's impossibility theorem)

If a social welfare function satisfies UD, T, PA and IIA, then it fails to be ND.

In other terms, with three or more alternatives, every method of aggregating individual preferences must violate at least one of Arrow's conditions.

The choice of any aggregation rule necessarily involves trading off at least one condition against the others.

If one thinks that UD, PA and IIA are not negotiable, then collective rationality comes at the price of democracy!

At the collective level, there is a tension between rationality and power decentralization.

Arrow's Theorem

A note of caution: Arrow's theorem doesn't imply that democracy is not possible. What it shows is that we should not expect a group of agents to behave with the kind of coherence that we may hope from an individual.

It also points out that expressions like "Vox Populi, Vox Dei" or "people's will" do not have much content, as "people" (as opposite to an individual) do not necessarily possess a well defined voice or will.

Arrow's theorem shows that the irrational collective behavior exhibits by the Condorcet's paradox goes to the heart of the matter, affecting not only majority voting but also any other reasonable aggregation rule.

Single-Peakedness

Escaping Arrow's result requires sacrificing one or more of his conditions. In some situations a natural candidate is UD.

That's because certain economic problems impose some structure on individual preferences; and as a result not all complete and transitive binary orderings are reasonable.

A popular domain restriction that provides positive results within social choice is single-peakedness (SP).

Definition 3 (Single-peaked preferences)

A preference profile $\rho = (R_1, \ldots, R_n) \in \mathcal{R}^n$ is single-peaked if for every triple $x, y, z \in X$, there exists one alternative, say x, such that for all $i \in N$, either $x P_i y$ or $x P_i z$ (i.e., x is bottom for none of the agents!).

In unidimensional problems, SP is equivalent to strict quasi-concavity.

Single-Peakedness

Notice that Fig. 2 violates Def. 3. Here are other examples.

R_1	R_2	R_3
b	b	d
c	a	c
d	c	b
a	d	a

Figure 3 : Single-peaked

R_1	R_2	R_3
b	b	d
c	a	a
d	c	b
a	d	c

Figure 4 : Not SP (try a, b, d)

Theorem 2 (Black's theorem)

If the number of agents n is odd, and individual preferences are single-peaked, then the majority preference relation R_M is transitive.

Theorem 3 (Median voter theorem)

If preferences are single-peaked and n is odd, then the peak of R_M coincides with the median of individuals' most preferred alternatives.

Median Voter

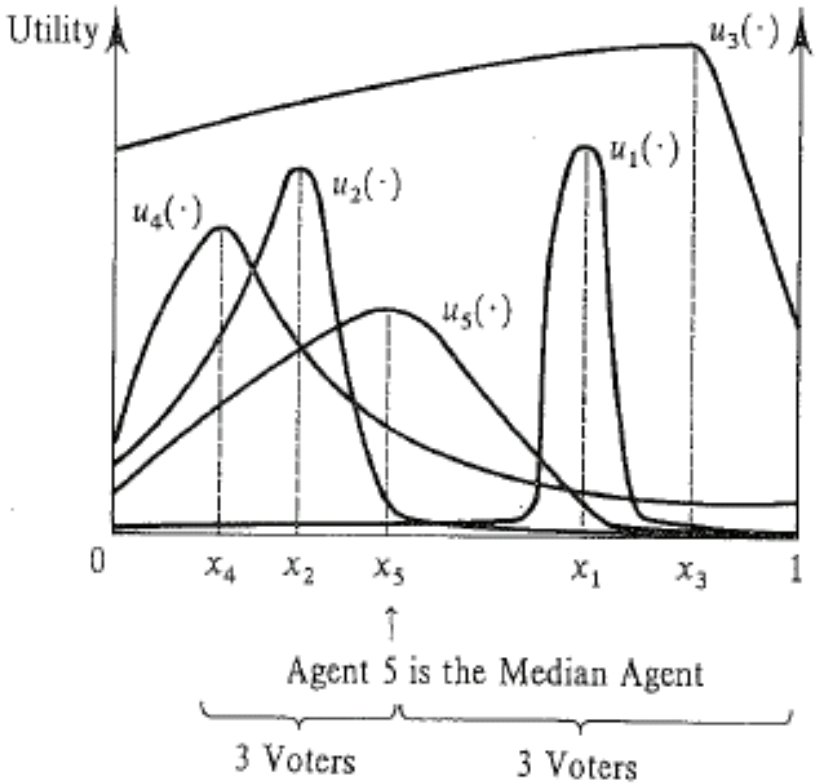

Figure 5 : Median voter theorem

Measurability & Comparability

Arrow's framework makes use only of ordinal information about individual preferences. In particular, for all $i, j \in N$ and any $x, y \in X$,

- No information is used about the strength of agent i's preference for x in comparison to i's preference for y;
- No information is used about how much i favors x over y in comparison to how much j favors x over y.

These kind of comparisons of intensities of preferences (especially the latter across agents) is controversial at best. (Recall utility numbers have no meaning in (neo)classical economic theory.)

Suppose however that this can be done in a meaningful way, and let's explore the implications of this assumption.

Measurability & Comparability

Assume the set of social states $X \subset \mathbb{R}^m$, $m \geq 1$, X convex.

Suppose the swf f maps profiles of continuous individual utility functions $u(\cdot) = (u_1(\cdot), \ldots, u_n(\cdot)) \in \mathcal{U}^n$ into a continuous social utility function $f_u(\cdot) \in \mathcal{U}$. (Bear in mind that for any x, $f_u(x) = f(u_1(x), \ldots, u_n(x))$.)

Let f satisfy UD, PA, and IIA on $\mathcal{U}^n$; and let's impose on f the Pareto indifference principle (PI) as well.

Definition 4 (PI)

A social welfare function $f : \mathcal{U}^n \to \mathcal{U}$ satisfies PI if for all $x, y \in X$, $f_u(x) = f_u(y)$ whenever $u_i(x) = u_i(y)$ for all $i \in N$.

Jehle & Reny call a swf that satisfies UD, PA, IIA and PI strict welfarism, in the sense that it relies only on ordinal information.

Anonymity & Equity

Suppose we add to strict welfarism the following two ethical principles.

The first one, anonymity (AN), simply says that people should be treated symmetrically, in the sense that social rankings shouldn't depend on people's names, but just on their welfare levels.

Definition 5 (AN)

A swf $f : \mathcal{U}^n \to \mathcal{U}$ is anonymous if for all $u, u' \in \mathcal{U}^n$, $f(u) = f(u')$ whenever u' is a utility profile obtained from u after some permutation of its elements.

The second one, Hammond equity (HE), expresses the idea that society has a preference for decreasing the dispersion of utilities across individuals.

Definition 6 (HE)

A swf $f : \mathcal{U}^n \to \mathcal{U}$ satisfies Hammond equity if for all $i, j \in N$ and any $u, u' \in \mathcal{U}^n$, such that $u_\ell = u'_\ell$ for all $\ell \in N \backslash \{i, j\}$ and $u'_i < u_i < u_j < u'_j$, it follows that $f(u) \geq f(u')$.

Rawls' Principle

Strict welfarism together with AN & HE provide some remarkable results.

The first one is the Ralwsian principle (RP) of social justice, according to which collective choices must be based on the welfare of society's worse-off members (a principle frequently invoked in the political debate). Formally,

$$f(u) = \min\{u_1, \ldots, u_n\}. \tag{1}$$

Theorem 4 (Rawls' Principle)

A swf $f : \mathcal{U}^n \to \mathcal{U}$ satisfies the Rawlsian principle if and only if it satifies strict welfarism and Hammond equity.

Theorem 4 says that adopting RP, that is, that society's decision-making must follow the welfare of its worse-off members, is equivalent to accepting strict welfarism and Hammond equity.

N.B. (i) AN isn't necessary, but it is implied by (1); (ii) HE requires measuring and comparing utility levels both for and across individuals.

Utilitarian Principle

A second very well-known principle of distributive justice, the utilitarian principle (UP), prays that it is the greatest happiness of the greatest number that is the measure of right and wrong in social decision-making.

The utilitarian swf is therefore based on the linear sum of individuals' utilities:

$$f(u) = \sum_{i=1}^{n} u_i. \tag{2}$$

Thus, when raking two states x and y, it is the linear sum of individual utility differences between the states that is the key factor (e.g., with $n = 2$, $f_u(x) \geq f_u(y)$ iff $u_1(x) - u_1(y) \geq u_2(y) - u_2(x)$ – i.e., the gain of 1 in moving from x to y is greater than or equal to the loss of 2!).

Once again, measuring and comparing utility levels both for and across individuals must be meaningful.

Utilitarian Principle

A swf is said to be utility-difference invariant if it is permitted to depend only on utility differences both for and across individuals.

Theorem 5 (Utilitarian Principle)

A swf $f : \mathcal{U}^n \to \mathcal{U}$ satisfies the utilitarian principle if and only if it is utility-difference invariant and it satisfies strict welfarism and anonymity.

If AN is dropped, we get generalized utilitarianism, where individuals' welfare can be weighted differently (i.e., $f(u) = \sum_{i=1}^{n} a_i u_i, a_i \geq 0$).

The UP can also be thought as the decision criterion adopted by a group of expected-utility-maximizer agents under the veil of ignorance (that is, uncertain ex-ante about who each of them will end up having to be).

Since a person might end up with any of n equally possible identities (Harsanyi's principle of insufficient reason), in the "original position" any state x would receive an expected utility $\sum_{i \in N} \frac{u_i(x)}{n}$; thus, the comparison between x and y reduces to $\sum u_i(x) \gtrless \sum u_i(y)$ – the utilitarian criterion.

Harsanyi vs. Rawls

Rawls views the social choice problem in the original position as one under complete ignorance (not even the probability of being person i is known).

Assuming people are risk averse, he argues social states x and y should be ordered according to how individuals would view them were they to end up as society's worst-off members; that is

$$\min\{u_1(x), \ldots, u_n(x)\} \gtrless \min\{u_1(y), \ldots, u_n(y)\}. \tag{3}$$

Rawls' maximim criterion, far from being incompatible with Harsanyi's utilitarianism, can be seen as a especial case of it, namely, the one that arises when individuals are infinitely risk averse.

Strategic Behavior

Up to now we studied the problem of aggregating the preferences of many individuals into a single preference for the group.

Implicit in the analysis is that the true individual preferences are known and ready to be aggregated.

But how exactly does society find out the preferences of its individual members?

One possibility is to ask them to report their rankings of social states.

However, how do we know that they are not going to report a false preference if by doing so they arrive to a preferred social state?

In sum, apart from the problem of aggregating individual preferences, there is a more basic problem of getting them in the first place.

Social Choice Function

Let's assume once again that X is finite, and let's allow each agent $i \in N$ to have any complete and transitive preference relation $R_i \in \mathcal{R}$ on X.

Instead of looking for a full social ranking of X, let's just consider the choice (presumably the top) that society would make for each preference profile $\rho = (R_1, \ldots, R_n)$. We call that a social choice function (scf).

Definition 7 (scf)

A scf $c(\cdot)$ is a single-valued mapping from $\mathcal{R}^n$ into X.

We will assume that the range of $c(\cdot)$ is X; that is, for every $x \in X$ there exists a preference profile $\rho = (R_1, \ldots, R_n) \in \mathcal{R}^n$ such that $c(R_1, \ldots, R_n) = x$.

If that were not the case, then we could simply eliminate x from X, as it would not be chosen under any circumstance.

Strategy-Proofness

What property would a scf need to have so that under no circumstance would any individual have an incentive to misreport his true preferences?

Definition 8 (Strategy-proofness (SP))

A scf $c(\cdot)$ is said to be strategy-proof if for all $i \in N$, all $R_i, R_i' \in \mathcal{R}$ and all $R_{-i} \in \mathcal{R}^{n-1}$, it follows that $c(R_i, R_{-i})\, R_i\, c(R_i', R_{-i})$.

If a scf is SP, no individual, no matter what his true preferences might be, can ever *strictly* gain by misreporting his preferences no matter what the other report (even if they lie)!

SP ensures that it is in the best interest of each agent to *always* report his true preferences sincerely.

Besides being extremely appealing, in the universal domain we are working on, this property is only achieved at a huge cost!

Gibbard-Satterthwaite Theorem

Definition 9 (Dictatorship)

A scf $c(\cdot)$ is said to be dictatorial if there exists $i \in N$ such that for all $x \in X$, if $x = c(R_i, R_{-i})$ for some $(R_i, R_{-i}) \in \mathcal{R}^n$, then $x\, R_i\, y$ for all $y \in X$.

Theorem 6 (Gibbard-Satterthwaite Theorem)

If $|X| > 2$, then every strategy-proof scf is dictatorial.

The message you should take away from GST is that, in a rich enough setting (i.e., with at least three social states and enough social diversity or heterogeneity), it is impossible to design a nondictatorial mechanism (institution) in which social choices are made based upon self-reported preferences and individuals have no incentives to lie.

In collective decision-making, the type of incentive compatibility demanded by strategy-proofness is not compatible with democracy.

Democratic institutions are subject to individual manipulation!