Applied Economics 8004
Applied Microeconomic Analysis
III. Social Choice

> Social choice theory is a subject in which formal and mathematical techniques have been very extensively used. Those who are suspicious of formal (and in particular, of mathematical) modes of reasoning are often skeptical of the usefulness of discussing real-world problems in this way. Their suspicion is understandable, but it is ultimately misplaced. The exercise of trying to get an integrated picture from diverse preferences or interests of different people does involve many complex problems in which one could be seriously misled in the absence of formal scrutiny. – Amartya Sen, Nobel Lecture, *American Economic Review*, 89 (1999), 349–378.

1 Introduction

This note contains a brief introduction to social choice. The main focus is two famous theorems, by Arrow and by Gibbard-Satterthwaite. Problems in social choice can thought of as much more general than this, though. The question of which allocation along the contract curve in an Edgeworth box is a social-choice problem. How should society pick one? This is not an easy question. The problem of social choice can be motivated through the use of a much simpler setting though. Consider a well-known early result on voting known as Condorcet's paradox. The Marquis de Condorcet wrote about this paradox in the late 1700's. It points up the difficulty of making collective decisions, even in simple situations, by use of majority voting.

Suppose a committee of three people, $j = 1, 2, 3$, must decide how to order three alternatives, x, y, and z. Each has strict preferences P_j over the alternatives, where for any two alternatives a and b we say that aP_jb if voter j strictly prefers a to b. It so happens that the preferences are these:

$$\begin{aligned} \text{Voter 1:} &\quad xP_1yP_1z \\ \text{Voter 2:} &\quad yP_2zP_2x \\ \text{Voter 3:} &\quad zP_3xP_3y \end{aligned}$$

The committee's by-laws state that decisions shall be made according to majority rule. By some means, the order of voting has been set as follows. First a vote is taken between alternatives x and y. The alternative receiving a majority of votes is then set against alternative z. How will this vote proceed?

Let P denote the committee's collective strict preference over the alternatives. First we know that

$$xPy,$$

because 1 and 3 will vote for x over y. Next x is set against z, with the result that

$$zPx.$$

Voters 2 and 3 vote for z over x. If the voting procedure ends here, z is selected. But voter 2 complains, saying that y was never given a chance to go against z. The committee agrees to hold this vote, which results in

$$yPz,$$

because voters 1 and 2 will vote for y over z. The problem is this: collective preferences are not *transitive*:

$$xPyPzPx,$$

which in this literature is considered a fatal flaw. Transitivity is a requirement so basic that it is not given the exalted status of an axiom in Arrow's theorem. Among the many results that are fundamental to the theory of voting is the power held by any person who is allowed to set the *agenda*, the order in which proposals are considered by the committee. It would be difficult to exaggerate the degree to which the literature on social choice is negative in nature. Not completely, but substantially negative.

2 Preliminaries and notation

We start with a finite set of alternatives X and a finite set of individuals $J = \{1, \ldots n\}$, with generic element j. (But note that many of our results have analogs for cases with an infinite number of alternatives.) Each individual j has an *ordering* P_j over X, with $P_j \subset X \times X$. A vector of preference orderings, one for each individual, is a *profile* and is denoted $\{P_j\}_{j=1}^n$. For this proof, suppose that preferences are strict. Arrow's social welfare function (SWF), denoted $f(\{P_j\}_{j=1}^n)$, maps a profile into an ordering P said to embody the preferences of the society.

We first define Arrow's four axioms, applied to a social welfare function.

Definition 1. *A social welfare function satisfies* ***unrestricted domain*** *if the domain of f includes all logically possible profiles. This condition is denoted (U).*

Definition 2. *A social welfare function is* ***non-dictatorial*** *if there is no individual j such that for all profiles, for every pair of alternatives x and y in X, xP_jy implies xPy. If f does not satisfy this condition, we say that f is* **dictatorial***, denoted (D).*

Definition 3. *A social welfare function satisfies the* ***Pareto principle*** *if for every pair $x, y \in X$, $[xP_jy \forall j] \Rightarrow [xPy]$. This condition is denoted (P).*

Definition 4. *A social welfare function is* ***independent of irrelevant alternatives*** *if for every pair $x, y \in X$, the social ordering of x and y depends only on the individuals' pair-wise ordering of x and y. This condition is denoted (I).*

Now we define the concepts of *semi-decisive* and *decisive*, which are used in the proof.

Definition 5. *The set of individuals $D \subset J$ is* ***semi-decisive*** *for alternatives x and y if $[\forall j \in D, xP_jy]$ and $[\forall j \notin D, yP_jx]$ together imply $[xPy]$.*

Definition 6. *The set of individuals $D \subset J$ is* ***decisive*** *for alternatives x and y if $[\forall j \in D, xP_jy]$ implies $[xPy]$.*

Decisive is the stronger condition because a decisive group gets its way in the social ranking of x and y regardless of what non-members think. A semi-decisive group gets its way only in the event that everyone else disagrees with them in the ranking.

3 Arrow's Theorem, first proof

Given this set-up, the theorem may be stated quite simply.

Arrow's Impossibility Theorem. *Suppose that preferences P_j are transitive, and that $\# \mid X \mid \geq 3$. Then any social welfare function $f(\{P_j\}_{j=1}^n)$ satisfying (I), (U), and (P) is dictatorial.*

Proof. We proceed in two steps. In the first, we show that there is some pair $a, b \in X$ such that some person $j \in J$ is semidecisive (SD) over a and b. In the second, we show that this person j is decisive over all $x, y \in X$ (that is, j is a dictator).

Step 1. By (P), J is SD $\forall\, x, y \in X$. Because $\# \mid J \mid < \infty$, $\exists$ some smallest D, say D^*, such that D^* is SD over a, b for some pair a, b in X. (This argument includes the case with $D^* = J$.) Choose some $j \in D^*$. We claim that $j = D^*$. By way of contradiction, suppose not, and consider any profile $\{P_j\}_{j=1}^n$ with $x \in X \backslash \{a, b\}$, with

$$\begin{aligned} &xP_jaP_jb \\ &aP_kbP_kx \qquad \forall\, k \in D^* \backslash \{j\} \\ &bP_ixP_ia \qquad \forall\, i \in J \backslash D^*. \end{aligned}$$

Then because D^* is SD, aPb. Also, we must have:

i. xPa, or otherwise $D^* \backslash \{j\}$ is SD over x, a, contradicting that D^* is the smallest SD set, and

ii. bPx, or otherwise j is SD over b, x, again contradicting that D^* is the smallest SD group. But then $aPbPxPa$, contradicting that P is transitive. Having reached a contradiction as a result of supposing that $j \neq D^*$, we conclude that the claim must be true: There is a person $j \in J$ who is semidecisive over some a, b.

Step 2. Given the person j who, from step 1, is SD over a, b, take $x \in X \backslash \{a, b\}$. There are three possibilities for x: either *i.*) xP_jaP_jb (x is better than both a and b for person j), or *ii.*) aP_jbP_jx (x is worse than either a or b for person j), or *iii.*) aP_jxP_jb (x is between a and b). Now, x was chosen arbitrarily, so one of these three things must be true, and if we can show that the social ordering P depends only upon j's preferences in each case, then we're through.

i.) (xP_jaP_jb) By (U), any profile is permissible. Thus, we can consider the following profile.

$$\begin{aligned} & xP_jaP_jb \\ & xP_ka \text{ and } bP_ka \qquad \forall\, k \in J \backslash \{j\}. \end{aligned}$$

Then aPb (because j is SD over a, b), and xPa (by (P)), so xPb (by transitivity). Thus, because only j's preferences over x and b are known, but still j gets his or her way in ordering this pair, j is decisive over x and b. No SWF f can possibly avoid giving person j his or her way in choosing between x and b.

ii.) (aP_jbP_jx) By (U), any profile is permissible. Thus, we can consider

$$\begin{aligned} & aP_jbP_jx \\ & bP_kx \text{ and } bP_ka \qquad \forall\, k \in J \backslash \{j\}. \end{aligned}$$

Then aPb (because j is SD over a, b), and bPx (by (P)), so aPx (by transitivity). Thus, because only j's preferences over a and x are known, but still j gets his or her way in ordering this pair, j is decisive over x and a. These first two cases together demonstrate that from knowing only that j is SD over a and b, we can show that j has complete control over how *any other alternative* x is ordered if that alternative is not between a and b for person j.

iii.) (aP_jxP_jb) This is the trickiest case. We don't directly find a pair over which j is decisive. Instead, we have to take an intermediate step, showing that there is another pair in the trio $\{a, x, b\}$ over which j is SD, and using this fact we can still show that j is decisive over either x and a or x and b, and from there the argument in our first two steps is sufficient to show that j is decisive everywhere. Again using (U), consider the profile

$$\begin{aligned} & aP_jxP_jb \\ & bP_kx \text{ and } bP_ka \qquad \forall\, k \in J \backslash \{j\}. \end{aligned}$$

Then aPb (because j is SD over a, b). Now, a and x are the two alternatives whose ordering by $k \in J \backslash \{j\}$ is unknown. Either xPa or aPx. If xPa, then by transitivity xPb, which means that j is SD over x and b. In this case, start over and show cases *i.*) and *ii.*) from above. If, on the other hand, aPx, then j is decisive over a and x, because we don't know anyone else's preferences over these two alternatives. In either case, then, we've shown that from a starting point of "j SD over a and b", only j's preferences matter in ordering *any* pair in X.

This completes the proof of Arrow's theorem. □

I hope that the above development gives you a good handle on what's going on. The point of all of this is that our SWF f is required to satisfy a very hard condition. It must *always* work, no matter what kind of profile we end up with. In this proof we reason as follows: Okay, there are a huge number of possible mixes of P_i's, and f must work for every one of them. Now, a great many of them will present no problem for even a simple-minded SWF. If all P_i are identical, for example, then we're in: let P agree with P_i. But that's not what we're interested in. Rather, because f must work everywhere, a very demanding requirement, we need to find only one collection of P_i's that implies f must behave badly. This is the set of P_i's we use in step 2, and because it leads to decisiveness we know that whatever f is, if it satisfies (U), (P), and (I), then it is dictatorial.

4 Arrow's Theorem, second proof

In a 2005 paper, John Geanakoplos provided three alternative proofs of Arrow's theorem.[1] The idea of the proof, as expressed in the abstract, is to replace Arrow's decisive voter with "an extremely pivotal voter (a voter who by unilaterally changing his vote can move some alternative from the bottom of the social ranking to the top), thereby simplifying both steps in Arrow's proof." Here is a sketch of Geanakoplos's first proof, adapted from a textbook by Jehle and Reny, which proceeds in four steps rather than Arrow's two. We allow ties in individual rankings and the social rankings, so this proof will use R_j and R to denote individual and social orderings.

Proof. **Step 1.** Consider the alternative c and suppose each person places c at the very bottom of her ranking. Then by (P) the social ranking must also place c at the very bottom.

Step 2. Now move c to the top of person 1's ranking. This will lead to a new profile, but by (U) the new profile is permissible. Next move c to the top of person 2's ranking. Continue doing this for person 3, person 4, and so on up to person n. By the end, c will be at the top of everyone's ranking, so by (P) it must be at the top of the social ranking too. This means there must be a person, call her j^*, such that raising c to the top of j^*'s ranking caused c to move from last to something higher than last in the social ranking.

Geanakopolos claims that when c moves to the top of j^*'s ranking, it cannot move up just a notch or two, but must instead move to the top of the social ranking. To see this, suppose to the contrary that there are x and y with xRc and cRy. By transitivity we must have xRy. Now move y above x for every person, leaving c in its place either at the top or the bottom of each ranking. By (I) we know that the social ranking of c versus x and c versus y cannot have changed because no individual's ranking of these pairs changed. By (P), at the new profile with y moved above x we must have yRx, a contradiction. We conclude that at the moment c moved to the top of j^*'s ranking, c must jump all the way from the bottom to the top of the social ranking.

Step 3. Consider the two profiles depicted in Table 1, before and after c is moved to the top of j^*'s ranking. In the *Before* profile, c is ranked last by the social ranking R. In the *After* profile, c is ranked first. Now start with the *After* profile, in which j^* and all the people $j < j^*$ have c

[1] Geanakoplos, John, "Three brief proofs of Arrow's Impossibility Theorem," *Economic Theory*, 26 (2005), 211–215.

***Before* profile**

R_1	R_2	$\cdots$	R_{j^*}	$\cdots$	R_n	R
c	c	$\cdots$	a	$\cdots$	a	
	b	$\cdots$		$\cdots$	b	
a		$\cdots$	b	$\cdots$		
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
b	a	$\cdots$	c	$\cdots$	c	c

***After* profile**

R_1	R_2	$\cdots$	R_{j^*}	$\cdots$	R_n	R
c	c	$\cdots$	c	$\cdots$	a	c
	b	$\cdots$	a	$\cdots$	b	
a		$\cdots$	b	$\cdots$		
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
b	a	$\cdots$		$\cdots$	c	

Table 1: Profiles before and after c is moved to the top of j^*'s ranking. Individual rankings of a and b are indeterminate

ranked first and all the people $j > j^*$ have c ranked last. Consider arbitrary alternatives a and b different from c and change rankings to obtain a new profile, also permissible by (U):

i.) Move a above c for j^* so that $aP_{j^*}cP_{j^*}b$; and

ii.) For everyone else, rank a and b in any way, but always so that c remains at the top for $j < j^*$ and at the bottom for $j > j^*$. The only thing we specify about the rankings of people $j \neq j^*$ in the new profile is the position of alternative c.

In this *New* profile, shown in Table 2, the ranking of a and c is the same for everyone as it was in the *Before* profile, just before c was moved from the bottom to the top for person j^*. At that moment we had c at the bottom of the social ranking. By (I), with the *New* profile we must still have c ranked last in the social ranking. But then we must have aPc.

Similarly, in the *New* profile the ranking of c and b is the same for everyone as in the *After* profile, just after c was moved from the bottom to the top for person j^*. At that moment we had c at the top of the social ranking. By (I), with the *New* profile we must still have c ranked first in the social ranking. But then we must have cPb.

In the new profile we have aPc and cPb, so by transitivity we must also have aPb. But we haven't said anything about how anyone other than j^* ranks a and b, so we know that the social ranking of these two alternatives agrees with j^*'s ranking regardless of anyone else's ranking. By

***New* profile**

R_1	R_2	$\cdots$	R_{j^*}	$\cdots$	R_n	R
c	c	$\cdots$	a	$\cdots$	a	
	b	$\cdots$	c	$\cdots$	b	?
a		$\cdots$	b	$\cdots$		
$\vdots$	$\vdots$		$\vdots$		$\vdots$	?
b	a	$\cdots$		$\cdots$	c	

Table 2: New profile with a moved above c in j^*'s ranking

(I), and because a and b were arbitrary, we conclude that person j^* is decisive on all pairs of alternatives not involving c:

$$[aP_{j^*}b] \implies [aPb].$$

Step 4. The final step is to show that j^* is a dictator. That is, she is decisive also on any pair involving c. Take alternative a different from c. Repeat the previous steps with a now playing the role of c there. There must be a dictator on all pairs not involving a. But "all pairs not involving a" includes the pair b and c, and we know j^* is dictator on that pair. We conclude that no one other than j^* can possibly be a dictator on pairs not involving a and thus that j^* is a dictator. □

5 The liberal paradox

In a 6-page note in the 1970 *JPE*, Amartya Sen proved one of the most perplexing results in all of social choice.[2] The result has become known as the "liberal paradox." He first defines a *social decision function*, which means essentially that the outcome of the social rule is an alternative (perhaps more than one), and not a ranking as in Arrow's SWF. He suggests that a social decision function should allow at least a moderate amount of freedom, giving people some ability to decide for themselves among at least a couple pairs of outcomes. You can read Sen's story of the two people whose preferences have to do with whether they or the other are allowed or forced to read *Lady Chatterly's Lover*.

Another story, from Gibbard, is concerned with the color that two neighbors, Alice and Bob, paint their houses. Alice loves blue, hates red. Bob loves green, hates yellow. If they were allowed to choose the color of their own house, they would choose blue and green. But Alice, in addition to hating red, also hates Bob. He returns the favor by hating Alice. So Alice would rather suffer with a red home herself than allow Bob to get his preferred green. Bob, likewise, would rather suffer with a yellow home himself than allow Alice to get her preferred blue. There are four alternatives in this story, with the following preference ranking for each person. The first color in a pair is the color of Alice's house, the second Bob's.

Alice	Bob
(Blue,Yellow)	(Red,Green)
(Red,Yellow)	(Red,Yellow)
(Blue,Green)	(Blue,Green)
(Red,Green)	(Blue,Yellow)

A liberal society might find it desirable to allow both to choose the color of their own house, in which case the result would be (Blue,Green). But this is not Pareto optimal, because both prefer (Red,Yellow). Both would be happier if they did not both choose independently, but rather had another outcome chosen where each decides the color of the other's house.

[2]Sen, Amartya, "The Impossibility of a Paretian Liberal," *Journal of Political Economy*, 78 (1970), 152–157.

Sen's paradox is that if we insist on letting at least some people have their own way over some pairs of social outcomes, then we cannot also guarantee that the choice rule will obey the Pareto axiom. His axiom is the following, written here in his own words.

Definition 7. *A social decision function satisfies* ***minimal liberalism*** *if there are at least two individuals such that for each of them there is at least one pair of alternatives over which he is decisive, that is, there is a pair of $x, y \in X$, such that if he prefers x (respectively y) to y (respectively x), then society should prefer x (respectively y) to y (respectively x). This condition is denoted (L^*).*

In the notation we've used so far, the axiom may be expressed as follows.

Definition 8. *A social decision function satisfies* ***minimal liberalism*** *if there are at least two individuals i, j such that for each there exists a pair of outcomes $x_i, y_i \in X$ ($x_j, y_j \in X$) such that for any profile, $y_i \in f(\{P_j\}_{j=1}^n)$ only if $y_i \succsim_i x_i$ ($y_j \in f(\{P_j\}_{j=1}^n)$ only if $y_j \succsim_j x_j$).*

The function given here is like a G-S social choice function, but it does not need to choose a single outcome. Thus, it is a social choice correspondence. It's definitely not Arrow's SWF. Sen's theorem II says the following. (Theorem I is for a stronger liberalism restriction, so this version is the stronger theorem.)

Theorem 1. *There is no social decision function that can simultaneously satisfy conditions (U), (P), and (L^*).*

The strategy of the proof is to show that there are profiles, resembling the example of house colors, in which there is no best element according to the "social preference." The following is a sketch and explanation of the proof.

Sen considers three possibilities in turn. In the first, the two pairs of alternatives (x, y) over which 1 is decisive, and (z, w) over which 2 is decisive, are the same pair. This produces an immediate contradiction, because two people cannot both be decisive over the same pair.

In the second, the two pairs share an outcome, say $z = x$. Suppose 1 is decisive over (x, y) with ranking xP_1y, and 2 is decisive over (x, w) with ranking wP_2x. Everyone in the society prefers y to w. We know the following about preferences over the three alternatives (the last line is to emphasize that yP_iw is all we know for $i \neq 1, 2$):

$$\begin{aligned} &xP_1yP_1w \quad \text{for } 1 \\ &yP_2wP_2x \quad \text{for } 2 \\ &yP_iw \qquad \forall\, i \neq 1, 2. \end{aligned}$$

Unrestricted domain guarantees that these preferences are allowed. Minimal liberalism implies that we must have wPx and xPy, and transitivity of the social preference then requires wPy. Because everyone prefers y to w, the Pareto principle implies that we must have yPw. These two are incompatible. Sen concludes that "A choice function for the society does not therefore exist."

Wait, what? He concocted one specific case with one set of preferences. How does it allow him to say that no choice function can exist? Once again, it seems like a bit of cheating is going on. Here's why the logic is sound. The choice rule he's after must work for any set of preferences, including this set. But he found one situation with a particular arrangement of preferences for which minimal liberalism and Pareto are incompatible. That's enough. Any candidate f would have to work everywhere, including for this case. But *no f can work* for this case, so it's all hopeless.

The third case is the most general. The four alternatives are all distinct, with xP_1y and zP_2w. Everyone in society prefers w to x and y to z. We know the following about preferences:

$$\begin{aligned} &wP_1xP_1yP_1z && \text{for 1} \\ &yP_2zP_2wP_2x && \text{for 2} \\ &wP_ix \text{ and } yP_iz && \forall\, i \neq 1, 2. \end{aligned}$$

Once again, by (U) this set of preferences is allowed and f must be able to select an outcome. But L* says we must have

$$xPy \quad \text{and} \quad zPw,$$

and (P) say we must have

$$wPx \quad \text{and} \quad yPz.$$

Put these together to obtain another violation of transitivity:

$$wPxPyPzPw.$$

There is no best alternative and so once again a choice function cannot exist that always selects society's preferred outcome.

6 The Median-Voter Theorem

In some social-choice problems it makes sense to express the alternatives along a single dimension from least to greatest. They might be different versions of a public project, ordered from least expensive to most expensive. They might be political candidates, ordered from left to right on the ideological spectrum. Now, by letting go of Arrow's axiom of an unrestricted domain, simple majority voting works just fine.

The domain restriction we need is that of "single-peaked" preferences.

Definition 9. *Consider a set of alternatives $X = \{x^1, \ldots, x^m\}$, ordered on a single dimension so that $x^1 < x^2 < \cdots < x^m$. The preference ordering P_j is **single peaked** if there is x_j^*, the peak, such that (i.) $a < b \leq x_j^* \implies bP_ja$ and (ii.) $x_j^* \leq a < b \implies aP_jb$*

Say that a profile $\{P_j\}_{j=1}^n$ is single peaked if each P_j is single peaked.

Theorem 2. *Given X as in the previous definition, suppose the number of voters is odd and that each voter's preferences are single peaked. There is a unique Condorcet (majority-rule) winner for any single-peaked profile. It is the median of the x_j^*.*

Proof. (Moulin) Rank the peaks according to $x_1^* \leq x_2^* \leq \cdots \leq x_n^*$. Because n is odd, we know that $k = (n+1)/2$ is an integer and x_k^* is the median peak. Consider the set of voters with $x_j^* \leq x_k^*$, who number $k + 1 > 1/2$. They form a majority (the "rightists") who will defeat any alternative $x < x_k^*$. Similarly, consider the set of voters with $x_j^* \geq x_k^*$, who also number $k + 1 > 1/2$. They form a majority (the "leftists") who will defeat any alternative $x > x_k^*$. The median voter is both a rightist and a leftist. Alternative x_k^* will defeat any outcome $x > x_k^*$ because the leftists will defeat it. It will also defeat any outcome $x > x_k^*$ because the rightists will defeat it. □

7 The Gibbard-Satterthwaite Theorem

Arrow's theorem guarantees that any social welfare function satisfying (U), (I), and (P) will be dictatorial. The setup here is the same as it was there. We have a group of n people, $J = \{1, \ldots, n\}$, a finite set of alternatives X, and for each person a preference ordering P_j over X. These orderings come from the set of all permissible orderings, which we denote $\mathcal{P} \subset X \times X$. A profile is a collection $\{P_j\}_{j=1}^n \in \mathcal{P}^n$. Arrow's social welfare function f maps profiles into $\mathcal{P}$. That is, it takes everyone's P_j, and produces a single *ordering*.

Two things about Arrow's theorem are worth exploring further. The first is that for many problems it asks too much of f. Oftentimes one is interested only in determining society's favorite outcome, not the ordering of all possible outcomes. Maybe if this less demanding result—a selection of one element of X—is all that f were asked to yield, a well-behaved f could be found. The second is that Arrow's theorem ignores how things look to people in the model. They are never given a chance to decide whether to report their true preferences to the machine that calculates the answer to the problem. But we know that there is often an incentive to lie when the collective outcome depends upon the reports of all members of society.

The Gibbard-Satterthwaite theorem is similar to Arrow's theorem, but it changes the perspective slightly to account for these two alterations.[3] As in Arrow, the elements of the problem are people J, outcomes X, and people's preference orderings P_j over outcomes. A *social choice function* (distinct from a social *welfare* function) is a rule that maps profiles into outcomes:

$$f : \mathcal{P}^n \longrightarrow X.$$

It is important that f is required to yield only a single element of X. Indeed, this is what makes f a function as opposed to a correspondence. We also insist that f be *onto*. That is, it must select every $x \in X$ for *some* profile. For example, suppose that everyone in the society ranks $\hat{x}$ first.

[3]The treatment of G-S given here is taken from H. Moulin, *Axioms of Cooperative Decision Making*, (Cambridge University Press, 1988).

Then it is natural to suppose that $\hat{x}$ must be chosen by f. This isn't quite what's required for f to be onto, but it gives an idea.

Now we must lay out formally what it means to say that people in the model have no incentive to lie when they participate in the program that uses f to select a social outcome.

Definition 10. *The social choice function f is **strategyproof** (also called nonmanipulable) if for every profile $\{P_j\}_{j=1}^n$, and for every $j \in J$, we have that*

$$\forall \quad P'_j \in \mathcal{P}, \quad f(P_j, P_{-j})\, P_j\, f(P'_j, P_{-j}).$$

That is, each member of society prefers the outcome obtained by telling the truth ($f(P_j, P_{-j})$) to any outcome that could be obtained by lying unilaterally and reporting some P'_j, which would produce $f(P'_j, P_{-j})$.

The definition of a dictator is similar to that used in Arrow's setup. Say that some member j is a *dictator* if that person's preferred choice is selected as the social choice regardless of the preferences of others. That is, if j is a dictator then for every profile, $f(\{P_j\}_{j=1}^n)\, P_j x$ for every $x \in X$ with $x \neq f(\{P_j\}_{j=1}^n)$.

The main theorem needs two more definitions, which compare two profiles and the outcome selected by f under them. First is the notion of an outcome being "pushed up."

Definition 11. *Take an f, and consider two profiles, $\{P_j\}$ and $\{P'_j\}$, and some outcome $y \in X$. Outcome y is **pushed up** in going from P_j to P'_j if*

i. $\{P_j\}$ and $\{P'_j\}$ coincide on all $x \in X \backslash y$;
ii. y keeps or improves its relative position from $\{P_j\}$ to $\{P'_j\}$; and
iii. $\{P_j\} \neq \{P'_j\}$, so the position of y improves for at least one agent.

Second is the notion of strong monotonicity.

Definition 12. *A voting rule f is **strongly monotonic** if for all profiles $\{P_j\}$ and $\{P'_j\}$ and an outcome y, if $\{P'_j\}$ is deduced from $\{P_j\}$ by pushing up y, then either $f(\{P'_j\}) = f(\{P_j\})$ or $f(\{P'_j\}) = y$.*

The proof of the Gibbard-Satterthwaite result is straightforward if we make use of the following (more difficult) result, due to Muller and Satterthwaite (1977). Their proof is given on pp. 260–61 in Moulin.

Muller-Satterthwaite Theorem. *If $\# \mid X \mid \geq 3$, a social choice function f is strongly monotonic if and only if it is dictatorial.*

This theorem makes the proof of G-S easier because now we need to show only that strategyproofness and strong monotonicity amount to the same thing. We now present the proof of Gibbard-Satterthwaite.

Gibbard-Satterthwaite Theorem. *If* $\# \mid X \mid \geq 3$*, a social choice function* f *is strategyproof if and only if it is dictatorial.*

Proof. There are two parts to the proof. The first involves showing that if f is dictatorial then it is strategyproof. But this is obvious. To see it, consider first j, the person who is dictator. This person should never lie. Telling the truth will ensure that j gets his or her most preferred outcome, while lying will yield something different. Other people don't care one way or the other—they simply don't matter, so they have no incentive to lie. (They have no incentive to tell the truth, either, but the definition of strategyproof is still met.)

It is harder to show the second part, that strategyproofness guarantees that someone is dictatorial. By the M-S theorem, it is enough to show that strategyproofness implies that f is strongly monotonic (which by M-S we know implies that there is a dictator).

Assume that f is strategyproof, and take a profile $\{P_j\}_{j=1}^n$. Fix a person j, and suppose that P'_j is deduced from P_j by pushing x up, for some $x \in X$. (For everyone else, $k \neq j$, $P_k = P'_k$.) To show that f must be strongly monotonic, it is enough to show that $f(P'_j, P_{-j})$ must be equal to $f(\{P_j\})$ or else it must be x. There are two cases: either $f(\{P_j\}) = x$ or not.

Case I: $f(\{P_j\}) = y \neq x$. This means that y is chosen when j tells the truth and reports P_j. Now suppose, **by way of contradiction**, that when j has preferences P'_j, with x pushed up, still another outcome is chosen if j tells the truth: $f(P'_j, P_{-i}) = z \neq y, x$. Because the relative position of y and z have not changed for j, one of two things must be true:

$$yP_jz \quad \text{and} \quad yP'_jz \quad \text{or} \tag{1}$$

$$zP_jy \quad \text{and} \quad zP'_jy. \tag{2}$$

If (1) is true, then we must have

$$f(P_j, \{P_{-j}\})P'_j f(P'_j, \{P_{-j}\}) = z,$$

where the equality follows from the assumption that z is chosen under the profile $(P'_j, \{P_{-j}\})$. But if (1) is true, j prefers y to z under both P_j and P'_j, so she will manipulate and report P_j to achieve the preferred outcome y. If (2) is true, then we must have

$$f(P'_j, \{P_{-j}\})P_j f(P_j, \{P_{-j}\}) = y,$$

where the equality follows from the assumption that y is chosen under the profile $(P_j, \{P_{-j}\})$. But if (2) is true, j prefers z to y under both P_j and P'_j, so she will manipulate and report P'_j to achieve the preferred outcome z. In each of these arguments, relative to (1) as well as (2), we have shown that

$$\text{[not strongly monotonic]} \Rightarrow \text{[manipulable]}. \tag{3}$$

We conclude that for Case I, the conditions for strong monotonicity must be satisfied. Put another way, under the contradictory assumption, f cannot be strategyproof.

Case II: $f(\{P_j\}) = x$. This means that x is chosen when j tells the truth and reports P_j. Now suppose, **by way of contradiction**, that when j has preferences P'_j, with x pushed up, something other than x is chosen. Call that something $y \neq x$. One of the following must be true:

$$yP_jx \quad \text{or} \tag{4}$$

$$xP_jy \quad \text{and} \quad xP'_jy. \tag{5}$$

If (4) is true, it means that although x was pushed up j's ranking, it didn't rise high enough to displace y. In this case, j should manipulate and report P'_j, which assures that y is chosen. If (5) is true, j prefers x to y under both profiles and so should manipulate by reporting P_j (which assures that x is chosen) rather than her true ordering P'_j (which would assure that the less-preferred y is chosen).

In both cases, I and II, in both of the possible configurations, (1) and (2) for case I and (4) and (5) for case II, j can gain by lying. This is the definition of manipulable. We conclude that in all possible cases, the logical statement in (3) is true. The only conclusion remaining is that nonmanipulability implies strong monotonicity. This completes the proof. □

It is not terribly important that you master the proof of this theorem. What is important is that you understand that people in the model have a strong incentive to lie. This is a question that Arrow's theorem doesn't ask, but that pervades both the public goods literature and the mechanism design literature.

8 Voting methods: an example

A committee of seven voters, named A to G, has been charged with selecting a single soft drink to be made available in the student lounge. The choices are Coke, Sunkist, Root Beer, Milk, and OJ. The following table reports the members' preferences, where 1 represents the most-preferred and 5 the least-preferred choice.

Choice	Voter A	Voter B	Voter C	Voter D	Voter E	Voter F	Voter G
Coke	5	1	2	4	2	4	2
Sunkist	1	2	5	5	5	5	1
Root Beer	2	3	4	3	3	3	3
Milk	4	4	1	2	4	2	5
OJ	3	5	3	1	1	1	4

Our goal is to advise the committee on *how* to make its collective decision: either to rank the choices from top to bottom (this is the object in Arrow) or to choose the committee's single most-preferred soft drink.

Majority rule. Each person votes for her first choice only. The vote counts are OJ: 3; Sunkist: 2; and Coke and Milk: 1. OJ wins.

Condorcet method. The Condorcet winner is the candidate who would beat all other candidates in a pairwise vote. For each pair of candidates we determine which candidate is preferred. If there is a candidate who wins every pairwise vote, then this candidate is the winner. If there is no such candidate, then there is no Condorcet winner. (Note: we define a winning candidate as that candidate having a number of preferential votes which is *greater than or equal to* the number of preferential votes of all other candidates.)

The winners (column entries) are denoted by "X" in the entry for a pairwise loser (row entries) in the following matrix. There are no ties. OJ is the Condorcet winner.

	Winner				
Loser	Coke	Sunkist	Root Beer	Milk	OJ
Coke	-			X	X
Sunkist	X	-	X	X	X
Root Beer	X		-		X
Milk			X	-	X
OJ					-

Condorcet score. The Condorcet score is the number of times a candidate wins over an opponent minus the number of times a candidate loses to an opponent. For each candidate, its Condorcet score is given in the following matrix. Once again OJ is the Condorcet winner.

	Winner				
Loser	Coke	Sunkist	Root Beer	Milk	OJ
Coke	-	-1	-1	1	1
Sunkist	1	-	1	1	1
Root Beer	1	-1	-	-1	1
Milk	-1	-1	1	-	3
OJ	-1	-1	-1	-3	-

Another way to compute the Condorcet winner is to derive a single number for each candidate by counting the number of other candidates it defeats in pairwise voting. This approach also gives OJ as the winner because it wins all four pairwise votes.

Plurality with runoff. First, restrict the set to the two candidates having the most first choice votes. Call them candidates A and B. Then, for all voters, change the scores for candidate A to be 1 if the voter prefers A to B and 2 otherwise. The winner is the candidate who gets the most first place votes. If two (or more) candidates have the same number of first place votes, we look to their second-place (or if necessary, their third-place, etc.) votes to decide how to rank them. The top two vote getters are OJ and Sunkist, with 3 and 2 votes respectively. In a runoff between these

two candidates, OJ wins by a vote of 4–3 and so OJ is the final winner.

Instant runoff. The instant-runoff scheme eliminates candidates one by one. First, we eliminate the candidate with the fewest first-place votes. Then we make up a preference schedule in which only the remaining candidates are listed; renumber the ranking in every column so that the 4 remaining candidates have now ranks labeled by 1, 2, 3 and 4. Repeat the procedure: eliminate again the candidate with the fewest first-place rankings, and renumber for the remaining 3 contenders, and so on. Indicate who was eliminated first, second, etc. Indicate the winner.

We first eliminate Root Beer, which gets no first-place votes. The new set of rankings is:

Choice	Voter A	Voter B	Voter C	Voter D	Voter E	Voter F	Voter G
Coke	4	1	2	3	2	3	2
Sunkist	1	2	4	4	4	4	1
Milk	3	3	1	2	3	2	4
OJ	2	4	3	1	1	1	3

Milk and Coke each get only one first-place vote. One can eliminate both of them at this stage, leaving us with the following ranking:

Choice	Voter A	Voter B	Voter C	Voter D	Voter E	Voter F	Voter G
Sunkist	1	1	2	2	2	2	1
OJ	2	2	1	1	1	1	2

OJ is the winner with four first-place votes. Alternatively, one can eliminate only one of Milk and Coke (perhaps by comparing second-place votes) and continue from there. The ultimate outcome is still OJ as winner.

Borda count. Here each agent reveals his or her preference ranking and the social welfare function assigns 1 point to the top choice of the individual, 2 to the next choice, and so on, where N is the number of candidates. (One can also assign points in the opposite order, which is better if someone doesn't rate all candidates.) The points for each candidate are totaled, the candidates are sorted by point total, and the resulting ordering is society's preference ordering. One way to derive the Borda score is in a single step (tally all the scores, and pick the winner). The other is the iterated Borda (tally all the scores, eliminate the agent with the worst score, and repeat until a winner is chosen). Indicate the name of the candidate that is eliminated each time, and the Borda points of other candidate who are still in the running at the time the candidate is eliminated. Indicate the name of the winner.

The numbered columns in the following table give the sequence of Borda scores obtained in the iterated method. The column labeled "1st" gives the outcome for the one-short Borda method. OJ

wins with a low score of 20.

Choice	A	B	C	D	E	F	G	1st	2nd	3rd	4th
Coke	5	1	2	4	2	4	2	20	18	14	11
Sunkist	1	2	5	5	5	5	1	24			
Root Beer	2	3	4	3	3	3	3	21	18	15	
Milk	4	4	1	2	4	2	5	22	19		
OJ	3	5	3	1	1	1	4	18	15	13	10

Using the iterated method, we first eliminate Sunkist. (Note the Root Beer was eliminated first under instant runoff.) The "2nd" column gives the scores for the second round, and so on. After Sunkist eliminate Milk, then Root Beer. In the final comparison, OJ beats Coke.

Approval voting. Each voter designates a subset of the outcomes of which she approves. The winning option is the one that appears in the highest number of approval lists. For this input, we assume that the approval list is anything in the top half of the ordering. If there are five choices, anything in the top two slots is deemed "approved." The following table shows which two candidates are approved by each voter. Here at last we get a winner other than OJ. Coke takes four votes, of which three are second-place votes.

Choice	A	B	C	D	E	F	G	Score
Coke		X	X		X		X	4
Sunkist	X	X					X	3
Root Beer	X							1
Milk			X	X		X		3
OJ				X	X	X		3

It's interesting to note that approval voting is sensitive to the number of candidates that can be approved. If one, OJ wins. If three, Root Beer with its 5 third-place votes wins.

9 Manipulability of voting rules

Have Gibbard and Satterthwaite said something that is of practical relevance? I think so. People routinely vote strategically, casting a vote for a candidate other than their top choice, in order to improve the chance that another candidate, whom they prefer to the likely winner, will win. Consider the following table, which represents a society's preferences over alternatives A through C. The numbers at the top of each column tell us how many voters hold the ranking in that column, with the top entry representing the most-preferred alternative.

The true rankings are

3	2
A	B
B	C
C	A

If this society uses the Borda count, the scores are

$$A : 9 + 2 = 11$$
$$B : 6 + 6 = 12$$
$$C : 3 + 4 = 7$$

and B wins. Now what if the 3 voters in the first column, who prefer A, instead report a ranking of $A \succ C \succ B$. The scores change to

$$A : 9 + 2 = 11$$
$$B : 3 + 6 = 9$$
$$C : 6 + 4 = 10$$

and A, the preferred candidate of the 3, wins election. Borda is manipulable.

Now consider another example, which shows how Borda responds in a strange way to candidates dropping out of the race.

25	24	22	29
A	B	C	C
B	A	B	D
D	C	A	A
C	D	D	B

If this society uses the Borda count, we will get scores of

$$A : 100 + 72 + 44 + 58 = 274$$
$$B : 75 + 96 + 66 + 29 = 266$$
$$C : 25 + 48 + 88 + 116 = 277$$
$$D : 50 + 24 + 22 + 87 = 183$$

Candidate C wins. Now suppose D drops out of the race, with the candidates remaining in place in each ranking. The new rankings are

25	24	22	29
A	B	C	C
B	A	B	A
C	C	A	B

and the new Borda counts are

$$A : 75 + 48 + 22 + 58 = 203$$
$$B : 50 + 72 + 44 + 29 = 195$$
$$C : 25 + 24 + 66 + 87 = 202$$

Now A wins. But every candidate defeats D by pairwise comparison.

10 Fairness in exchange

An allocation is fair if it is both Pareto optimal and envy free. In an Edgeworth box, PO allocations are the familiar contract curve. The idea of an envy-free allocation is due to Foley from 1967. An allocation is envy free if no consumer wishes she had received another consumer's bundle. More formally, we have the following definition.

Definition 13. *Consider an exchange economy and let x be an allocation. x is **envy free** if for each i, $U_i(x_i) \geq U_i(x_j)$ for all $j \neq i$.*

If you like to think about such problems geometrically, an allocation x is envy free if its reflection about the midpoint of the Edgeworth box lies outside the lens formed by drawing an indifference curve through x_i for each person.

Fairness requires envy freeness and also Pareto optimality.

Definition 14. *Consider an exchange economy and let x be an allocation. x is **fair** if it is both Pareto optimal and envy free.*

There must exist an envy-free allocation in an exchange economy. Start at y^e, the midpoint of the Edgeworth box in Figure 1(b), representing equal division of each good. Draw two indifference curves through this point, one for each consumer. Any allocation along the portion of the contract curve that lies within the lens formed by the two indifference curves must be fair.

It is useful, perhaps, to look at how an allocation can fail to be fair. In Figure 1(a), consider x, which is Pareto optimal. Its reflection around the midpoint, x^e, is at x^T. We can see that consumer 2 prefers x_2^T to x_2. This means he prefers 1's bundle, x_1, to his own, x_2. Thus, x is not envy free. Allocation y in panel (b) is envy free, because the reflection allocation y^T lies outside the lens formed by indifference curves through y. And y is also PO, so it is fair. Neither envies the other and y is PO.

In case you might be interested in working through this example, it's one we've seen before. Preferences are given by $U_1(x_1^1, x_1^2) = x_1^1 x_1^2$ and $U_2(x_2^1, x_2^2) = 3 \ln x_2^1 + x_2^2$. The aggregate endowment is $\omega = (16, 12)$.

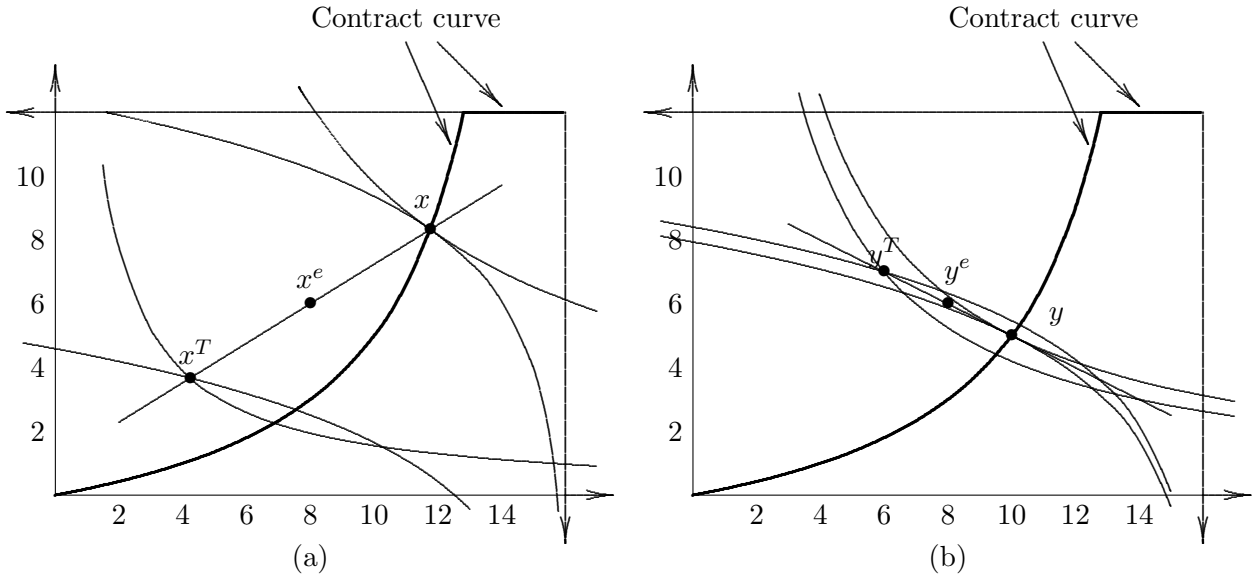

Figure 1: Allocation x in panel (a) is PO but not envy free (consumer 2 prefers x_2^T to x_2). Allocation y in panel (b) is envy free and PO. Thus, y is fair.

11 Egalitarian equivalence

Remember what fairness is meant to accomplish: to rule out some Pareto-optimal allocations, those that are least appealing from an eithical perspective, without bringing interpersonal comparisons into the analysis. In a 1978 paper, Pazner and Schmeidler returned to the question and proposed an alternative criterion, *egalitarian equivalence*.[4] This criterion, like fairness, rules out interpersonal comparisons of utility. Also like fairness, it can be used to discard a portion of the PO allocations. We will see, though, that the EE test is a low bar. It does not discard very many allocations.

The idea of an egalitarian-equivalent allocation (EEA) is this. Consider any allocation, and ask whether there is another aggregate resource base, possibly (in fact, likely) infeasible, that can be divided equally among all the agents and give them the same level of utility as at the initial candidate allocation. If so, that allocation is egalitarian equivalent. Any envy-free allocation is also an EEA.

Pazner and Schmeidler then join their notion of egalitarian equivalence to Pareto optimality. And allocation is a *Pareto-efficient egalitarian-equivalent allocation* (a PEEEA) if it both egalitarian equivalent and Pareto optimal.

I find the definition of an EEA, as stated, to be just a bit murky. Here is another explanation that might be a little more transparent. Start with an exchange economy and a feasible allocation x in that economy. Now unfold the Edgeworth box and place the two bundles, x_1 and x_2, against the same set of coordinate axes. Draw 1's indifference through x_1 and 2's indifference curve through x_2. If the two curves cross, then x is an EEA. That's all there is to it.

[4]Pazner, Elisha and David Schmeidler, "Egalitarian Equivalent Allocations: A New Concept of Economic Equity," *Quarterly Journal of Economics*, 92, (1978), 671–687.

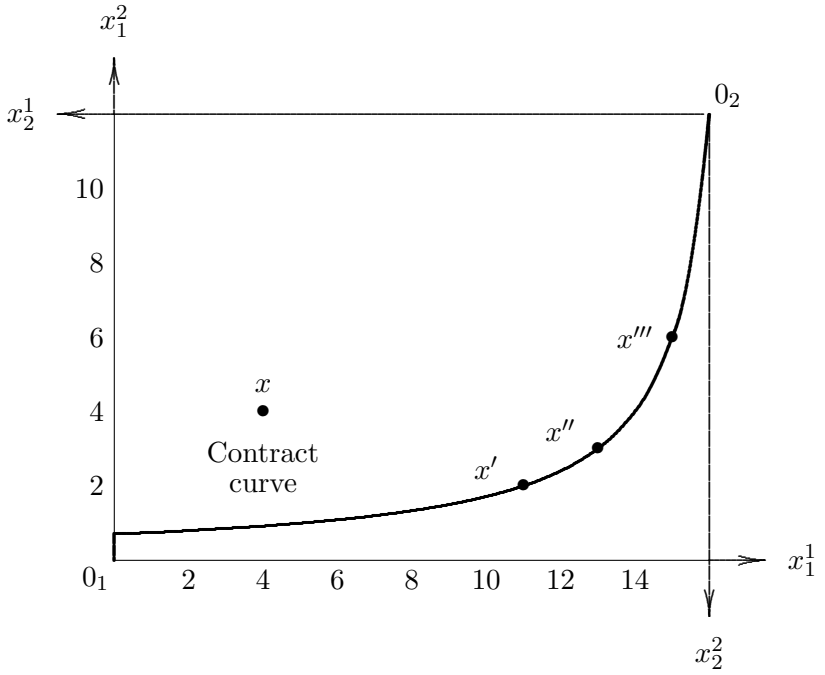

Figure 2: Contract curve in second example.

Let's consider another 2×2 example, similar to the first. The aggregate endowment is $\omega = (16, 12)$. Utility functions for consumers 1 and 2 are $U_1(x_1^1, x_1^2) = x_1^1 + \ln x_1^2$ and $U_2(x_2^1, x_2^2) = x_2^1 x_2^2$. (Subscripts j refer to people and superscripts i refer to goods.) To find the interior portion of the contract curve, set MRS's equal:

$$\text{MRS}_1 = -\frac{1}{1/x_1^2} = -\frac{x_2^2}{x_2^1} = \text{MRS}_2.$$

Use the resource constraints to obtain

$$x_1^2 = \frac{12}{17 - x_1^1}$$

for $x_1^1 \in [0, 16]$, with $x_2^1 = 16 - x_1^1$ and $x_2^2 = 12 - x_1^2$. This curve is depicted in Figure 2. A small portion of the contract curve runs up the left side of the box, from $(0, 0)$ to $(0, 0.71)$. We'll soon make heavy use of the marked points along the bolded contract curve. For now, though, consider the non-PO but feasible allocation denoted $x = (4, 4, 12, 8)$. Utility levels at this allocation are $U_1(4, 4) = 4 + \ln 4 = 5.3863$ and $U_2(12, 8) = 96$. In Figure 3 we see the two bundles, one for each consumer, and the indifference curve for each passing through the x_j. These curves never cross. (Could you show this mathematically?) Therefore, x is not an EEA. Of course, it isn't PO either. And neither is it envy free. Consumer 1 would rather have 2's bundle.

When using the Pazner-Schmeidler approach, we aren't interested in allocations that aren't PO. The whole point is to start with the set of PO allocations and determine which of them are also egalitarian equivalent. Return to Figure 2 and consider the PO allocation $x' = (11, 2, 5, 10)$. Figure 4 illustrates indifference curves for the two consumers at the bundles $x_1' = (11, 2)$ and

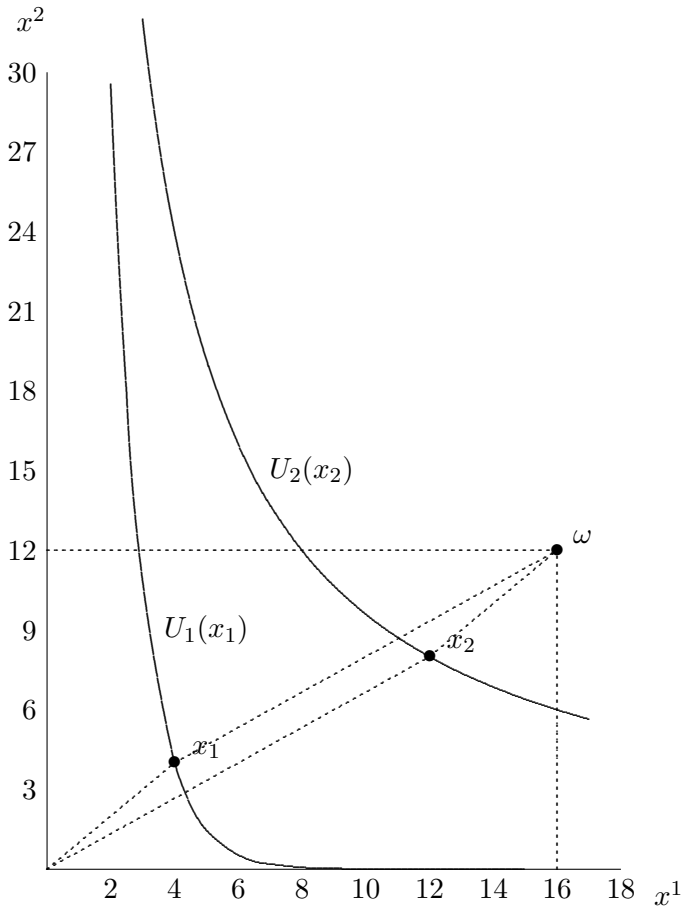

Figure 3: Allocation x is not egalitarian equivalent. Indifference curves never cross.

$x_2' = (5, 10)$, using $U_1(11, 2) = 11.69315$ and $U_2(5, 10) = 50$. The point of intersection of these two indifference curves is found by solving the intersection of the two equations. We know that $U_1(x_1') = 11.69315$, so that on 1's indifference curve we have $x^1 = 11.69315 - \ln x^2$. Plug this into the expression for x^1 in 2's indifference curve to get $50 = (11.69315 - \ln x^2)x^2$. I solved this numerically, using brute force, to get $x^2 = 4.95395$. This then gives us $x^1 = 10.09296$. The (hypothetical) endowment bundle $\omega^E = (20.18593, 9.90790)$ is the PEEEA we were looking for.

Notice a couple of things about Figure 4. One is that at x' neither consumer envies the other. Another is that the endowment ω^E is entirely fictional. That economy doesn't exist. But we call x' an EEA because, *if* ω^E were the endowment, *then* we could given both consumers the same utility as at x_j' *while also giving them idential bundles,* x^E.

Now select another allocation on the contract curve, say $x'' = (13, 3, 3, 9)$. Figure 5 illustrates indifference curves for the two consumers at $x_1 = (13, 3)$ and $x_2 = (3, 9)$, using $U_1(13, 3) = 14.09861$ and $U_2(3, 9) = 27$. It might be a useful exercise to solve this case for the intersection of the two indifference curves, where you will find the PEEEA corresponding to x''. Notice that here we have a PEEEA allocation even though person 2 envies 1's allocation (it lies above 2's indifference curve through x_2'). Thus, x' is not envy-free. The PEEEA criterion is less demanding than the fairness

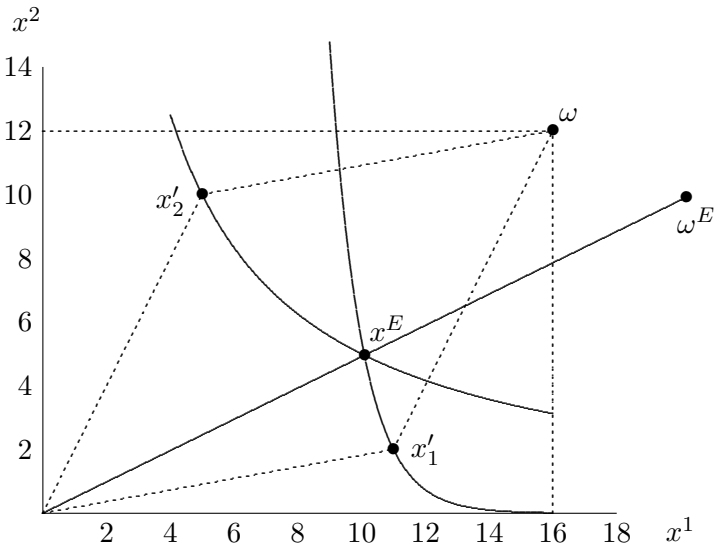

Figure 4: Allocation x' is egalitarian equivalent. Indifference curves cross at $x^E = (4.954, 10.093)$.

criterion. Indeed, the set of envy-free allocations is a subset of the EEA allocations: any envy-free allocation must be an EEA, but the opposite implication is not true.

This example illustrates that egalitarian equivalence does not do much to rule out allocations along the contract curve. In fact, everything along the contract cuve to the north and east of x'' is also a PEEEA. Figure 6 illustrates this for the allocation $x''' = (15, 6, 1, 6)$. You can see that these indifference curves do indeed cross, because 1's quasi-linear indifference curves approach the horizontal axis very quickly.

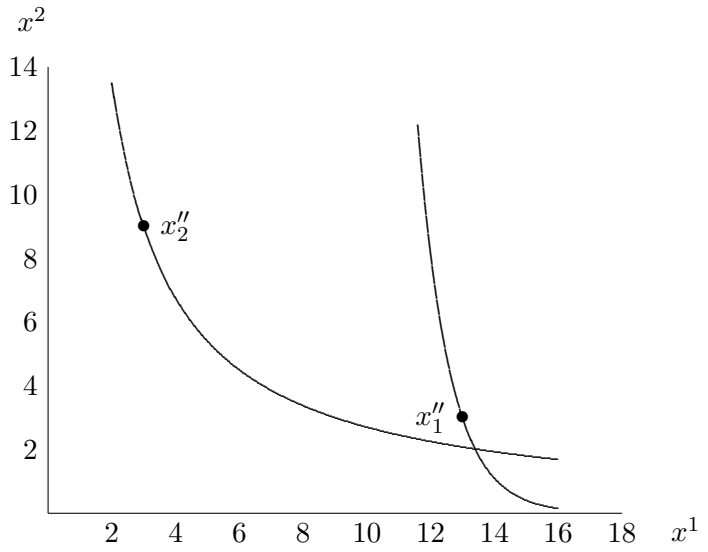

Figure 5: Allocation x'' is egalitarian equivalent, even though it is not envy free.

12 Exercises

1. Consider the following 2×2 competitive exchange economy, with consumers $j = 1, 2$ and goods x and y. The consumers' preferences are given by

$$U_1(x_1, y_1) = \ln x_1 + 2 \ln y_1 \qquad \text{and} \qquad U_2(x_2, y_2) = 2 \ln x_2 + \ln y_2.$$

Endowments are $\omega_1 = (4, 2)$ and $\omega_2 = (2, 4)$. Determine whether the initial endowment ω is egalitarian-equivalent. If so, find the endowment ω^E that can be divided equally to give each consumer the same utility level as at ω. Determine whether ω^E is feasible. If not, show why not.

Solution. An allocation is e-e if there is an aggregate endowment that gives each consumer the same bundle and also gives each consumer the same utility level as at the reference allocation, in this case ω. One finds it by computing the intersection point of indifference curves through ω_j and, in this case with two people, doubling it. See Figure 7. The solution is at x^E, found by solving

$$\frac{\bar{U} - \ln x_1}{2} = \bar{U} - 2 \ln x_2$$

for x, where $\bar{U} = 2.77$ is initial utility for both consumers. The required bundle is $\omega^E = 5.04$, which is feasible.

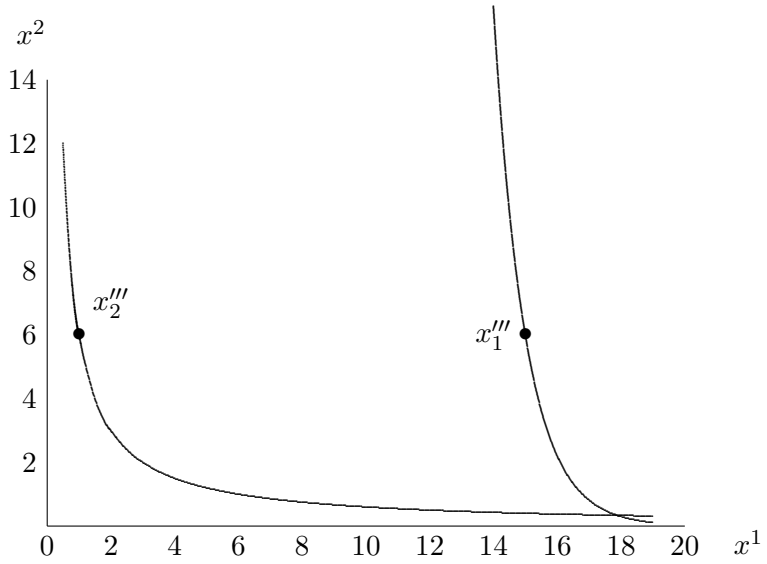

Figure 6: Allocation x''' is egalitarian equivalent, even though it is not envy free.

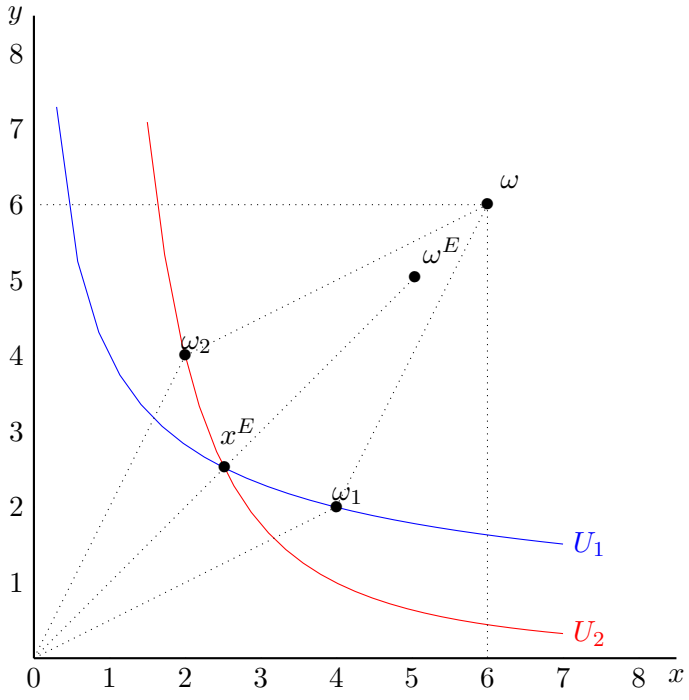

Figure 7: Egalatarian equivalence