PE3101P: Decision and Social Choice

Lecture Notes Weeks 10–12 — The Liberal Paradox and Arrow's Theorem

Instructor: Tomi Francis

1 From wellbeing to preferences

In Lectures 7–9 we did welfarist social choice. We often compared how well off different people are, and asked how much a gain for one person counts against a loss for another.

Now drop that information. Instead, we look at the social aggregation of people's *preferences*: given what each person prefers, what should society choose?

There are two main topics. The first is about Sen's liberal paradox: a clash between respecting unanimous preferences and respecting people's liberty. The second is about Arrow's theorem: no way of turning voters' rankings into a social ranking meets a few (apparently) very plausible conditions, unless one voter is a dictator.

In order to talk about Sen's liberal paradox, we'll need the concept of a choice function. And along the way, we'll cover what are typically known as *internal consistency conditions*. Think of these as patterns on the way you choose. They don't tell you substantively what to pick, but they tell you how what you pick in one situation relates to what you pick in another—similarly to principles like transitivity or the Independence or Continuity axioms of expected utility theory, which you may recall from the first half of the course.

These internal consistency conditions are quite important in rational choice theory in general, for similar reasons to the axioms of expected utility theory. They're apparently quite minimal, but they impose a lot of structure on a choice function. They encode certain important ways in which a person's (or society's) way of choosing may be "well-behaved".

2 Choice functions

2.1 Social states, preferences and menus

A *social state* is a complete description of how things are for everyone. Each person i has a preference relation $\succsim_i$ over social states. This is an ordinary preference relation, as in the decision theory half of the course: $x \succsim_i y$ means that person i finds x at least as good as y, $x \succ_i y$ means that i strictly prefers x to y, and $x \sim_i y$ means that i is indifferent between them.

Unlike where we were dealing with wellbeing, where effectively what was good for one person was tacitly restricted to the things that only affect them, when it comes to individual *preferences*, they can be about other people's business. You can prefer one social state to another because of what somebody else does in it. This will matter a lot in Section 3; indeed, Sen's liberal paradox can't get going without some people being "busybodies" in this sort of way.

Back to terminology. A *menu* S (I sometimes call this an *option set*) is a set of available social states (or, more generally, a set of available options). We will only consider menus with finitely many elements, because infinite decision theory is a hassle. What we're mostly interested in is which options on a menu may be chosen: the options which aren't ruled out by some social choice procedure.

Definition 1 (Choice function)**.** A *choice function* C gives, for every non-empty menu S, a non-empty subset $C(S)$ of S: the options that may be chosen from S.

I will often call this a *social choice function*. But only when we're doing social choice.

Notice an important condition there: $C(S)$ is never empty. This matters in several proofs below. For now, we leave open precisely what "may be chosen" means. Section 3.5 looks at different ways of reading it. But it will generally not do too much harm to read it as stating which options may permissibly be chosen from the perspective of some hypothetical social planner. Mathematically speaking, though, *any* function that meets the conditions can be a (social) choice function. It doesn't need to be a *plausible* one to count, even though as philosophers we're ultimately interested in finding the social choice function which encodes correct social decision-making (if there is such a thing).

2.2 From a relation to a choice function

A relation $\succsim$ over the options gives us a choice function in a natural way: from any menu, choose the *maximal* options, the ones that nothing on the menu is strictly better than. In symbols:

$$C(S) = \{x \in S : \text{there is no } y \in S \text{ with } y \succ x\}.$$

An *optimal* option, by contrast, is one that is at least as good as everything on the menu. If $\succsim$ is complete, the maximal options and the optimal options are the same: if nothing on the menu is strictly better than x, then by completeness x is at least as good as everything on it. Without completeness they can come apart, and then maximality is the better test. Suppose there are just two options, and neither is at least as good as the other: a trip to New York and a trip to Sydney, say, each good in its own way. Neither is optimal, so choosing the optimal options would allow us to choose neither. That seems crazy: surely at least one of them may be chosen. Choosing the maximal options allows either.

Will every menu have a maximal option? Not if the relation has cycles: if $x \succ y$, $y \succ z$ and $z \succ x$, then nothing on the menu $\{x, y, z\}$ is maximal. But if there are no cycles, every finite menu has a maximal option. We don't need transitivity to show this: acyclicity is enough.

Acyclicity: There is no sequence of options $x_1, x_2, \ldots, x_n$ with $x_1 \succ x_2 \succ \cdots \succ x_n \succ x_1$.

Claim. If $\succ$ satisfies Acyclicity, every finite non-empty menu has a maximal option.

The idea of the proof is simple. If every option on the menu had something strictly better than it, we could keep stepping up from one option to a better one. Since the menu is finite, we would eventually come back to an option we had already visited, and that would be a cycle.

Proof. Take a finite non-empty menu S, and suppose for contradiction that no option on S is maximal. Then every option on S has something on S strictly better than it. Start with any x_1 on S. Something on S is strictly better than x_1: call it x_2, so $x_2 \succ x_1$. Something on S is strictly better than x_2: call it x_3, so $x_3 \succ x_2$. We can keep going for as long as we like. But S has only finitely many options, so eventually we reach an option that we have already visited: some x_k that is the same option as an earlier x_j. Then $x_k \succ x_{k-1} \succ \cdots \succ x_{j+1} \succ x_j = x_k$, which is a cycle. This contradicts Acyclicity, so S has a maximal option. □

Not every choice function comes from a relation in this way. The point of working with choice functions is that they don't assume this. In particular, they don't assume that what else is on the menu is irrelevant to the choice between two options. The next three subsections give conditions that say how a choice function's choices from different menus fit together.

2.3 Contraction consistency

Contraction consistency (Property α): If x is chosen from a menu S, and x is on a smaller menu T inside S, then x is chosen from T. In symbols: if $x \in C(S)$ and $x \in T \subset S$, then $x \in C(T)$.

In words: taking other options off the menu can't stop x being chosen. Figure 1 illustrates it.

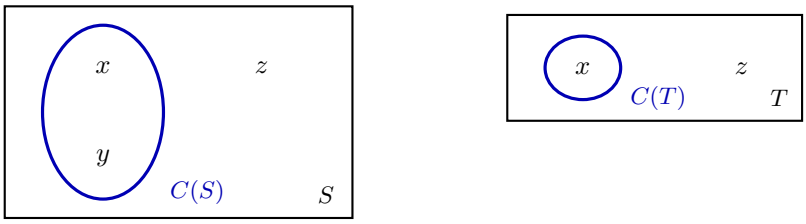


Figure 1: Left: the rectangle is the menu $S = \{x, y, z\}$, and the circle is what may be chosen, $C(S) = \{x, y\}$. Right: shrink the menu to $T = \{x, z\}$. Since x was chosen from S and is still on T, contraction consistency says that x is chosen from T. It says nothing about z.

A choice function that comes from a relation, by choosing maximal options, always satisfies contraction consistency. If nothing on S is strictly better than x, then nothing on a smaller menu inside S is strictly better than x either.

Here is a case in which contraction consistency fails.[1] A diner is offered apple pie (a) or blueberry pie (b), and orders apple. The waiter comes back: there is also cherry pie (c). "In that case, I'll have the blueberry."

Menu	He chooses
$\{a, b\}$	a
$\{a, b, c\}$	b

Blueberry is chosen from $\{a, b, c\}$, and it is on the smaller menu $\{a, b\}$, but it isn't chosen there. So contraction consistency fails. It is hard to see what reason the cherry pie could give him to switch.

But contraction consistency can fail without anything irrational going on. A guest at dinner sees one apple left in the fruit basket.[2] She could take it (y) or take nothing (x). To be polite, she takes nothing. If there had been two apples (y and z), she could have taken one without being rude.

Menu	She chooses
$\{x, y\}$	x, nothing
$\{x, y, z\}$	y or z, an apple

The apple y is chosen from $\{x, y, z\}$, and y is on the smaller menu $\{x, y\}$, but it isn't chosen there. So contraction consistency fails. But nothing here seems irrational: she is just being polite.

2.4 Expansion consistency

Expansion consistency (Property γ): If x is chosen from a menu S_1, and x is chosen from a menu S_2, then x is chosen from the menu that combines them. In symbols: if $x \in C(S_1)$ and $x \in C(S_2)$, then $x \in C(S_1 \cup S_2)$.

In words: if x survives against everything on S_1, and against everything on S_2, then it survives against them all together.

Sen states this condition for any number of menus at once.[3] For finitely many menus, which is all we need, the version for two menus is enough: to combine three menus, combine two of them, and then combine the result with the third, and so on. Figure 2 pictures it.

[1] Attributed to Sidney Morgenbesser.

[2] Sen, Amartya (1993). "Internal Consistency of Choice". In: *Econometrica* 61.3, pp. 495–521: 501.

[3] Sen, Amartya (1971). "Choice Functions and Revealed Preference". In: *The Review of Economic Studies* 38.3, pp. 307–317: 314.

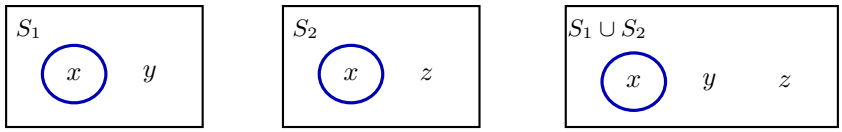


Figure 2: x may be chosen from $S_1 = \{x, y\}$ and from $S_2 = \{x, z\}$. Expansion consistency says that x may be chosen from $S_1 \cup S_2 = \{x, y, z\}$ too.

The guest with the fruit basket breaks expansion consistency as well. With one apple, y, she takes nothing; with the other apple alone, z, she takes nothing too; with both apples, she takes one.

Menu	She chooses
$\{x, y\}$	x, nothing
$\{x, z\}$	x, nothing
$\{x, y, z\}$	y or z, an apple

So x is chosen from $\{x, y\}$ and from $\{x, z\}$, but not from $\{x, y\} \cup \{x, z\} = \{x, y, z\}$.

2.5 Property β

Property β: If x and y are both chosen from a menu T, and T is inside a menu S, then y is chosen from S if and only if x is. In symbols: if $x, y \in C(T)$ and $T \subset S$, then $y \in C(S)$ if and only if $x \in C(S)$.

In words: if x and y are both chosen from a smaller menu, then on a bigger menu they stand or fall together. From the bigger menu, either both are chosen or neither is. What Property β rules out is one of them being chosen without the other. Figure 3 pictures the three possibilities.

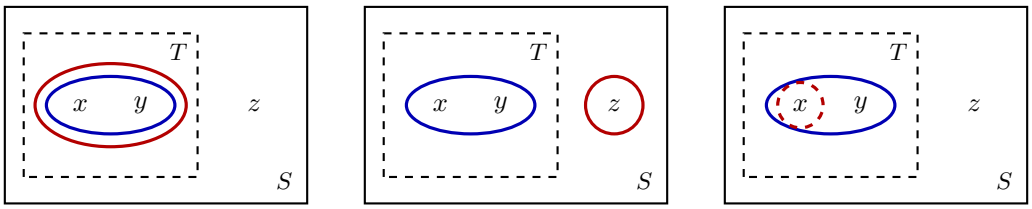


Figure 3: In each picture, x and y are both chosen from the smaller menu T (dashed): the blue circle is $C(T)$. T is inside the bigger menu S, and the red circle is $C(S)$. Left: x and y are both chosen from S; allowed. Middle: neither is chosen from S; also allowed. Right: x is chosen from S without y; ruled out by Property β.

Here is a case in which Property β fails. Go back to the trips to New York (x) and Sydney (y) from Section 2.2. Each is good in its own way, and either would be fine. Then a third option turns up: Sydney, plus a free dollar (y^+). Now you shouldn't choose Sydney without the dollar, since Sydney with the dollar is better. But New York is still fine. Suppose also that from $\{y, y^+\}$ only y^+ may be chosen, and from $\{x, y^+\}$ either may be chosen.

Menu	May be chosen
$\{x, y\}$	x or y
$\{y, y^+\}$	y^+
$\{x, y^+\}$	x or y^+
$\{x, y, y^+\}$	x or y^+

x and y are both chosen from $\{x, y\}$, but from $\{x, y, y^+\}$ only x of the two is. So Property β fails. Notice what is going on in this case. New York and Sydney are not equally good: they are good in different ways,

and neither is at least as good as the other. That is why a small improvement to Sydney doesn't make it better than New York. So the failure of Property β here comes from incompleteness.

2.6 From a choice function to a social preference relation

We have seen how a relation gives a choice function. We can also go the other way: from any choice function, we can read off a relation, by looking at what it chooses from pairs. When the choices are society's, we call the relation a *social preference relation*. The name builds in no assumptions about it: it need not be transitive, for example.

Definition 2 (The relation $\succsim_C$)**.** $x \succsim_C y$ if and only if $x \in C(\{x, y\})$.

So $x \succ_C y$ when x alone is chosen from $\{x, y\}$, and $x \sim_C y$ when both are. $\succsim_C$ is complete, by construction: something is always chosen from $\{x, y\}$, so $x \succsim_C y$ or $y \succsim_C x$.

Every choice function gives a social preference relation in this way. But not every choice function is given by a social preference relation, in the way described in Section 2.2.

Definition 3 (Binary representability)**.** A choice function C is *binary-representable* when there is a relation $\succsim$ (a social preference relation, when the choices are society's) such that, from every menu S, $C(S)$ is the set of options on S that nothing on S is strictly better than. Then $\succsim$ *represents* C.[4]

If a complete social preference relation represents C, it has to be $\succsim_C$. For it has to agree with C about every pair: from the menu $\{x, y\}$, a complete relation chooses x exactly when x is at least as good as y, and C chooses x exactly when $x \succsim_C y$.

The diner's choices are not binary-representable. Since only apple is chosen from $\{a, b\}$, a relation that represents his choices would have to rank apple strictly above blueberry. But then it would never choose blueberry from $\{a, b, c\}$, since apple is on that menu too.

The trip choices are binary-representable: they are represented by $\succsim_C$. Check it menu by menu: $\succsim_C$ has $x \sim_C y$, $x \sim_C y^+$ and $y^+ \succ_C y$, and choosing the maximal options by this relation gives back every row of the table above. Notice that $\succsim_C$ counts New York and Sydney as equally good, since both are chosen from $\{x, y\}$. Built from choices between pairs, it can't tell the difference between two options being equally good and two options neither of which is at least as good as the other.

2.7 What the conditions give us

The three conditions fit together in useful ways. We prove five results. The first says that contraction consistency and Property β together give expansion consistency. The second and third say that contraction consistency and expansion consistency together hold exactly when the choice function is binary-representable. The fourth and fifth say that contraction consistency and Property β give binary representability by a transitive relation. The main results here are Sen's.[5]

From α and β to γ. If a choice function satisfies contraction consistency and Property β, it satisfies expansion consistency.

Here is the idea. We know that x is chosen from S_1 and from S_2, and we want to show that x is chosen from $S_1 \cup S_2$. Something is chosen from $S_1 \cup S_2$: call it y. That y came from one of the two smaller menus, so by contraction consistency it is chosen there too, alongside x. Property β then says that x and y stand or fall together on the bigger menu $S_1 \cup S_2$. Since y stands, so does x.

Proof. Suppose $x \in C(S_1)$ and $x \in C(S_2)$. We show that $x \in C(S_1 \cup S_2)$.

[4] Sen calls such choice functions "normal" (Sen 1971: 309) or "binary" (Sen 1993: 500).

[5] Sen, Amartya (1971). "Choice Functions and Revealed Preference". In: *The Review of Economic Studies* 38.3, pp. 307–317: 310, 314.

$C(S_1 \cup S_2)$ is not empty, so it contains some option y. If y is x, we are done. Otherwise, y is in S_1 or in S_2: say S_1 (the other case is the same). Now y is chosen from $S_1 \cup S_2$, and y is in S_1, which is inside $S_1 \cup S_2$. So by contraction consistency, $y \in C(S_1)$.

So x and y are both chosen from S_1, and S_1 is inside $S_1 \cup S_2$. By Property β, y is chosen from $S_1 \cup S_2$ if and only if x is. Since $y \in C(S_1 \cup S_2)$, we have $x \in C(S_1 \cup S_2)$, which is what we wanted to show. □

Contraction consistency is needed: Property β alone does not give expansion consistency. Suppose every choice set is a single option, with $C(\{x, y\}) = \{x\}$, $C(\{x, z\}) = \{x\}$ and $C(\{x, y, z\}) = \{y\}$. Property β is satisfied automatically, since two different options are never both chosen from the same menu. But expansion consistency fails: x is chosen from $\{x, y\}$ and from $\{x, z\}$, but not from $\{x, y, z\}$. (Contraction consistency fails too: y is chosen from $\{x, y, z\}$ but not from $\{x, y\}$.)

From α and γ to binary representability. If a choice function C satisfies contraction consistency and expansion consistency, then $\succsim_C$ represents C.

We have to show two things about any option x on any menu S. First, if x is chosen from S, then nothing on S is strictly better than x by $\succsim_C$: this uses contraction consistency, to get from the whole menu down to pairs. Second, if nothing on S is strictly better than x, then x is chosen from S: this uses expansion consistency, to get from pairs back up to the whole menu, which is just all the pairs $\{x, y\}$ put together.

Proof. Take any menu S and any x on S. We show that $x \in C(S)$ if and only if nothing on S is strictly better than x by $\succsim_C$.

Suppose first that $x \in C(S)$, and take any y on S. $\{x, y\}$ is inside S, so by contraction consistency, $x \in C(\{x, y\})$. That is, $x \succsim_C y$, so y is not strictly better than x. Since y was any option on S, nothing on S is strictly better than x.

Now suppose that nothing on S is strictly better than x by $\succsim_C$, and take any y on S. If x were not in $C(\{x, y\})$, then $C(\{x, y\})$ would be $\{y\}$, since it is not empty; and then $y \succ_C x$, which is false. So $x \in C(\{x, y\})$, for every y on S. Putting together the menus $\{x, y\}$, for every y on S, gives the menu S. So by expansion consistency (applied again and again, since S is finite), $x \in C(S)$.

So $C(S)$ is exactly the set of options on S that nothing on S is strictly better than, which is what we wanted to show. □

Binary representability gives α and γ. If any social preference relation represents a choice function C, then C satisfies contraction consistency and expansion consistency.

This direction is easier. Choosing the maximal options means that x is chosen exactly when nothing on the menu beats it. Removing options can't add something that beats x, and putting two menus together can't either, if nothing on either of them beat x.

Proof. Suppose $\succsim$ represents C.

Contraction consistency: suppose $x \in C(S)$, and x is on a menu T inside S. Since $\succsim$ represents C, nothing on S is strictly better than x. Everything on T is on S, so nothing on T is strictly better than x either. So $x \in C(T)$.

Expansion consistency: suppose $x \in C(S_1)$ and $x \in C(S_2)$. Then nothing on S_1 is strictly better than x, and nothing on S_2 is strictly better than x. Everything on $S_1 \cup S_2$ is on S_1 or on S_2, so nothing on $S_1 \cup S_2$ is strictly better than x. So $x \in C(S_1 \cup S_2)$.

So C satisfies both conditions, which is what we wanted to show. □

Putting the first two results together gives another.

From α and β to binary representability. If a choice function C satisfies contraction consistency and Property β, then $\succsim_C$ represents C.

Proof. Contraction consistency and Property β give expansion consistency. Contraction consistency and expansion consistency give binary representability by $\succsim_C$. □

From α and β to transitivity. If a choice function C satisfies contraction consistency and Property β, then $\succsim_C$ is transitive.

We need to show that if $x \succsim_C y$ and $y \succsim_C z$, then $x \succsim_C z$: that is, x is chosen from $\{x, z\}$. By contraction consistency, it is enough to show that x is chosen from the bigger menu $\{x, y, z\}$. Something is chosen from that menu. The idea is that, whatever it is, Property β drags x in with it. If y is chosen, x comes with it, since x and y are both chosen from $\{x, y\}$. If z is chosen, y comes with it, since y and z are both chosen from $\{y, z\}$; and then x comes with y.

Proof. Suppose $x \succsim_C y$ and $y \succsim_C z$. If two of x, y and z are the same option, then $x \succsim_C z$ is immediate, so suppose they are all different. We show that $x \in C(\{x, y, z\})$. Then, since $\{x, z\}$ is inside $\{x, y, z\}$, contraction consistency gives $x \in C(\{x, z\})$: that is, $x \succsim_C z$.

First, if y is chosen from $\{x, y, z\}$, so is x. For suppose $y \in C(\{x, y, z\})$. By contraction consistency, $y \in C(\{x, y\})$. And $x \in C(\{x, y\})$, since $x \succsim_C y$. So x and y are both chosen from $\{x, y\}$, which is inside $\{x, y, z\}$. By Property β, since y is chosen from $\{x, y, z\}$, so is x.

Second, if z is chosen from $\{x, y, z\}$, so is y. For suppose $z \in C(\{x, y, z\})$. By contraction consistency, $z \in C(\{y, z\})$. And $y \in C(\{y, z\})$, since $y \succsim_C z$. So by Property β, since z is chosen from $\{x, y, z\}$, so is y.

Now, $C(\{x, y, z\})$ is not empty, so at least one of x, y and z is chosen from it. If it is x, we are done. If it is y, then x is chosen too, by the first step. If it is z, then y is chosen too, by the second step, and so x is chosen too, by the first step. Whichever it is, $x \in C(\{x, y, z\})$, which is what we wanted to show. □

3 The liberal paradox

3.1 Profiles and social choice rules

Now let the choices be society's. Society is made up of some people (at least two), each with their own preferences over social states.

Definition 4 (Profile)**.** A *profile* is a list of preference relations, one for each person: $(\succsim_1, \ldots, \succsim_n)$.

For a fixed profile, society's choice function says what may be chosen from each menu. But we also want to ask how society's choices should depend on people's preferences. For that we need something that gives a social choice function for each profile.

Definition 5 (Social choice rule)**.** A *social choice rule* gives, for each profile it applies to, a social choice function: society's choice function, saying what may be chosen from each menu.

Sen's result is about social choice rules. It uses one condition on which profiles the rule has to deal with, and two conditions on what it chooses.

Unrestricted Domain: The social choice rule applies to every profile: any preferences at all are allowed.

The first condition on what is chosen is a version of the Pareto principle from Lecture 8. This version is about preferences, not wellbeing, and it is written with society's choice function C.

Pareto: If everyone strictly prefers x to y, then y is not chosen from any menu containing x. In symbols: if $x \succ_i y$ for every person i, and x and y are both on S, then $y \notin C(S)$.

3.2 Liberty in personal matters

Some choices seem to be nobody's business but your own.[6] If you prefer your walls pink rather than white, society should let you have them pink, even if a majority would like to see them white. Whether you sleep on your back or on your belly should be up to you, even if a majority would rather you slept on your back.

Sen states this kind of liberty in a very weak form. It uses the idea of a person being decisive over a pair of social states.

Definition 6 (Decisiveness)**.** A person is *decisive* over a pair of social states $\{x, y\}$ when, at every profile, whichever of the two they strictly prefer, the other is not chosen from any menu containing both. That is: if $x \succ_i y$, then $y \notin C(S)$ for every menu S containing x and y; and if $y \succ_i x$, then $x \notin C(S)$ for every such menu.

Minimal Liberalism: There are at least two people who are each decisive over at least one pair of social states.

Informally, the two states in a person's pair should differ only in a personal matter: the colour of that person's walls, say. Formally, the pair can be any pair. The person wins in either direction: whichever of the two they prefer. And the condition is "minimal" because it asks for so little: only two people, and only one pair each.

We will often say that one state *excludes* another: x excludes y when y is not chosen from any menu containing x. So Pareto says that if everyone prefers x to y, x excludes y. And if a person is decisive over $\{x, y\}$ and prefers x to y, then x excludes y.

3.3 The book case

Here is a case in which Pareto and Minimal Liberalism clash.[7] There is one copy of *Lady Chatterley's Lover*, and two people, Prude and Lewd.[8] There are three social states.

State	What happens
x	Prude reads it
y	Lewd reads it
z	Nobody reads it

Prude would most like nobody to read it. But rather than Lewd reading it, he would read it himself. Lewd would most like Prude to read it; next, to read it himself; and last, for nobody to read it.

	Prude	Lewd
Best	Nobody reads (z)	Prude reads (x)
Middle	Prude reads (x)	Lewd reads (y)
Worst	Lewd reads (y)	Nobody reads (z)

Each has a right over whether he reads the book. Prude is decisive over $\{x, z\}$: whether he reads it. Lewd is decisive over $\{y, z\}$: whether he reads it.

Now go through the menus one at a time.

– Menu $\{x, z\}$. Prude prefers nobody reading it to reading it himself. By his right, z excludes x. So only Nobody reads (z) may be chosen.

– Menu $\{y, z\}$. Lewd prefers reading it himself to nobody reading it. By his right, y excludes z. So only Lewd reads (y) may be chosen.

[6] Sen, Amartya (1970). "The Impossibility of a Paretian Liberal". In: *Journal of Political Economy* 78.1, pp. 152–157: 152.
[7] Sen, Amartya (1970). "The Impossibility of a Paretian Liberal". In: *Journal of Political Economy* 78.1, pp. 152–157: 155.
[8] The names are from Sen, Amartya (2017). *Collective Choice and Social Welfare: An Expanded Edition.* Cambridge, MA: Harvard University Press: 434.

– Menu $\{x, y\}$. Prude and Lewd both prefer Prude reading it to Lewd reading it. By Pareto, x excludes y. So only Prude reads (x) may be chosen.

– Menu $\{x, y, z\}$. Now every state is excluded: x by Prude's right, y by Pareto, and z by Lewd's right. So nothing can be chosen.

But a choice function has to choose something from every menu. So, with these preferences, nothing can be chosen from $\{x, y, z\}$ without going against Pareto or one of the two rights.

> **An aside.** *Lady Chatterley's Lover*, D. H. Lawrence's novel of 1928, was banned in Britain for decades as obscene. In 1960, Penguin Books published the full text, and was prosecuted under the Obscene Publications Act: *R v Penguin Books Ltd.* The trial is best remembered for the prosecuting counsel, Mervyn Griffith-Jones, asking the jury whether it was a book "you would even wish your wife or your servants to read". After three hours of deliberation, the jury returned a unanimous verdict of *not guilty.*

3.4 Sen's result

The book case is one instance of a general result.[9]

Sen's liberal paradox. No social choice rule satisfies Unrestricted Domain, Pareto and Minimal Liberalism.

The idea of the proof is to use Unrestricted Domain to pick people's preferences, so that every option on some menu is excluded, either by someone's right or by Pareto. Then nothing can be chosen from that menu. Minimal Liberalism gives us two people, each decisive over a pair. How we pick the preferences depends on how the two pairs overlap: they are the same pair, or they share one option, or they share none.

Proof. Minimal Liberalism gives two people: person 1, decisive over one pair, and person 2, decisive over a pair (perhaps the same one). There are three cases.

Case 1: the same pair. Both people are decisive over $\{x, y\}$. Give them opposite preferences: person 1 has $x \succ y$, and person 2 has $y \succ x$. Then on the menu $\{x, y\}$:

Option	Excluded by
~~y~~	person 1's right: $\boldsymbol{x} \succ_1 \boldsymbol{y}$
~~x~~	person 2's right: $\boldsymbol{y} \succ_2 \boldsymbol{x}$

Nothing can be chosen from $\{x, y\}$. This case doesn't need Pareto.

Case 2: one shared option. Person 1 is decisive over $\{x, z\}$, and person 2 is decisive over $\{y, z\}$. Since any preferences are allowed, take these:

- person 1: $z \succ x \succ y$;
- person 2: $x \succ y \succ z$;
- everyone else: anything with x above y.

Then on the menu $\{x, y, z\}$:

Option	Excluded by
~~x~~	person 1's right: $\boldsymbol{z} \succ_1 \boldsymbol{x}$
~~z~~	person 2's right: $\boldsymbol{y} \succ_2 \boldsymbol{z}$
~~y~~	Pareto: everyone has $\boldsymbol{x}$ above $\boldsymbol{y}$

[9] Sen 1970: 154.

Nothing can be chosen from $\{x, y, z\}$. These are Prude's and Lewd's preferences: the book case.

Case 3: no shared option. Person 1 is decisive over $\{x, y\}$, and person 2 is decisive over $\{z, w\}$. Take these preferences:

- person 1: $w \succ x \succ y \succ z$;
- person 2: $y \succ z \succ w \succ x$;
- everyone else: anything with w above x and y above z.

Then on the menu $\{x, y, z, w\}$:

Option	Excluded by
~~y~~	person 1's right: $\boldsymbol{x} \succ_1 \boldsymbol{y}$
~~w~~	person 2's right: $\boldsymbol{z} \succ_2 \boldsymbol{w}$
~~x~~	Pareto: everyone has $\boldsymbol{w}$ above $\boldsymbol{x}$
~~z~~	Pareto: everyone has $\boldsymbol{y}$ above $\boldsymbol{z}$

Nothing can be chosen from $\{x, y, z, w\}$.

In every case, some menu has every option on it excluded. But a social choice function must choose something from every menu. So no social choice rule satisfies all three conditions. □

Notice how little the proof uses. It doesn't assume that society's choices come from a relation, or contraction consistency, or transitivity. It uses only that something can be chosen from every menu.

3.5 When is it a paradox?

In Section 2, we left open what "may be chosen" means. We can read "y is not chosen when x is available" in (at least) four ways.

(1) *Outcome evaluation*: x is a better state of affairs than y.[10]

(2) *Deontic*: y ought not to be chosen when x is available.

(3) *Normative choice*: decision-making in society ought to be organised so that y is not chosen when x is available. This is a two-level claim, about how decision-making is set up, rather than about what ought to be chosen.

(4) *Descriptive choice*: decision-making in society is organised so that y will not be chosen when x is available.[11]

In Lecture 8's terms, (1) is the telic reading, and (2) is the deontic one. Only the first reading requires society's choices to come from a relation: on it, they come from the betterness relation. On the deontic reading, what ought to be chosen may or may not come from a relation. And Sen points out that the two choice readings, (3) and (4), don't require it either.[12]

Sen's result holds on each reading. But to see whether it is really a paradox on a given reading, we have to ask whether its conditions are all plausible on that reading. We have to assess the whole theorem under each interpretation: not just Minimal Liberalism, but Unrestricted Domain and Pareto too.

[10] Sen, Amartya (1983). "Liberty and Social Choice". In: *The Journal of Philosophy* 80.1, pp. 5–28: 6.
[11] Sen 1983: 7.
[12] Sen 1983: 7.

The betterness reading. On the first reading, the theorem is not much of a paradox, because Minimal Liberalism isn't very plausible as a claim about betterness. Principles of liberty are best thought of as claims about people's rights.

Take a busybody, who prefers something that is none of their business, or even gets wellbeing from it. There may be principled reasons not to take their preferences into account when deciding what happens. But it is hard to deny that satisfying those preferences, or increasing that wellbeing, makes things better. Suppose we could costlessly give Lewd the thought that Prude is reading the book. Lewd doesn't care whether it is true; he just likes the idea. That does seem to make the outcome better. To deny it, you need a messier kind of consequentialism, on which some kinds of preference satisfaction or wellbeing don't count; and that doesn't make for a tidy theory.

In any case, hardly anyone puts Minimal Liberalism forward as a claim about betterness. Rights are a standard example of something deontic rather than axiological: they are about what may be done, not about what makes things better. Only a consequentialist about rights would hold Minimal Liberalism as a claim about betterness, and that is not a popular view. (If you are one, though, the paradox should trouble you.)

The deontic reading. On the second reading, each condition of the theorem is hard to give up. Unrestricted Domain is hard to deny: you might simply be in a situation like the book case. Pareto is very plausible: if everyone agrees in preferring x to y, it seems that y ought not to be chosen when x is available. It is supposed to be an expression of liberalism, or of welfarism. And it seems right that we ought to respect people's liberty in some way. So on this reading, it looks like a genuine paradox. Whether Minimal Liberalism is the right way to state liberty is the question of Sections 3.6 and 3.7.

The other two readings. On the normative choice reading, the result says that no good way of organising decision-making, one that always chooses something whatever people's preferences, satisfies both Pareto and Minimal Liberalism.[13] On this reading, Minimal Liberalism is weaker than giving people control over the choice.[14] It says only that decision-making should be set up so that the person's preference over their pair wins, however that is achieved.

On the descriptive choice reading, the result says that no actual way of making decisions that always chooses something satisfies both. This is interesting as a paradox about descriptive choice, because it tells us that there is no way of setting things up so that all of these principles are satisfied. For example, leave Prude and Lewd free to decide whether to read the book, and Lewd reads it.[15] But that is a state that both of them rank below Prude reading it.

Ways out. There are three main ways out.

(1) Restrict which preferences count: for example, count only "tolerant" or "non-meddlesome" preferences.[16]

(2) Weaken Pareto. This is Sen's own response. A unanimous preference should decide only if everyone wants their preference to count; and a person who respects other people's rights won't want their meddlesome preferences to count.[17] If even one person respects the rights of others, this weaker Pareto principle is consistent with liberty, though some ordinary Pareto preferences are then overridden.[18]

(3) Reformulate liberty. The rest of this section looks at this way out.

[13] Sen 1983: 10.
[14] Sen 1983: 9.
[15] Sen 2017: 435.
[16] Sen 1983: 12.
[17] Sen, Amartya (1976). "Liberty, Unanimity and Rights". In: *Economica* 43.171, pp. 217–245: 236.
[18] Sen 1976: 236–237.

3.6 Gibbard's paradox

Sen understands rights as individual decisiveness: for you to have a right over a pair of social states is for your preference between them to settle what society may choose. Gibbard showed that this way of understanding rights is inconsistent by itself.[19] Take a case in which people clearly do have rights, and read their rights as decisiveness. Then, with Unrestricted Domain, the rights clash with each other, without any help from Pareto.

This is a paradox of a different kind from Sen's. In Sen's paradox, several plausible conditions clash, and it isn't clear which of them has to go. In Gibbard's paradox, it is clear what has to go. If you thought that individual decisiveness was the correct way to understand liberalism, you would have to conclude that liberalism is inconsistent, and that is a really strange view. It is much more natural to think that the idea of individual rights is consistent, and that this is just the wrong way to understand it. That is the point of the paradox: rights shouldn't be understood as individual decisiveness.

Here is the case. Nick and Carl each have the right to paint their own walls white (W) or yellow (Y). Write the social states with Nick's walls first and Carl's second: WW is the state in which both are white, WY the state in which Nick's are white and Carl's yellow, and so on. Read their rights as decisiveness: of two states that differ only in Nick's walls, the one Nick prefers excludes the other; and of two states that differ only in Carl's walls, the one Carl prefers excludes the other.

Nick is a nonconformist: he wants his walls to be a different colour from Carl's. Carl is a conformist: he wants his walls to be the same colour as Nick's. So:

– Nick prefers YW to WW, and WY to YY;

– Carl prefers WW to WY, and YY to YW.

Now look at the four states.

State	Excluded by
~~WW~~	Nick's right: **YW** and **WW** differ only in Nick's walls, and Nick prefers YW
~~WY~~	Carl's right: **WW** and **WY** differ only in Carl's walls, and Carl prefers WW
~~YY~~	Nick's right: **WY** and **YY** differ only in Nick's walls, and Nick prefers WY
~~YW~~	Carl's right: **YY** and **YW** differ only in Carl's walls, and Carl prefers YY

Every state is excluded, by rights alone, so nothing can be chosen from the menu of all four. But in real life, there is nothing paradoxical about Nick and Carl. Each of them just paints his walls. They may be unhappy with the result, but nobody's right has been violated.

3.7 Rights as choices

Minimal Liberalism treats rights as automatically exercised: a person's preference over their pair is simply imposed on society. But a right is something you may exercise or not, and perhaps exercise strategically. Players in a game don't always do what is immediately better for them.

Sugden takes this further, and objects to the way Sen states liberty.[20] Sen's condition ties social choice to people's preferences over whole social states: to what each person would choose, if they could choose between two social states. But people usually don't choose between social states. They choose features of them, such as the colour of their own walls. So the choices Sen's condition is about are only hypothetical. On the *control* view of liberty, liberty is about what people are actually left free to do. In Sugden's words, "it is his choice that must be respected, not his preference".[21]

To state liberty this way, Sugden uses game forms.[22]

[19] Gibbard, Allan (1974). "A Pareto-Consistent Libertarian Claim". In: *Journal of Economic Theory* 7.4, pp. 388–410: 389.
[20] Sugden, Robert (1985). "Liberty, Preference, and Choice". In: *Economics and Philosophy* 1.2, pp. 213–229: 214–218.
[21] Sugden 1985: 217.
[22] Sugden 1985: 219.

Definition 7 (Game form)**.** A *game form* says who the players are, what strategies each player has, and which outcome results from each combination of strategies.

A game form says nothing about anyone's preferences. A right, on this view, is a strategy that guarantees some feature of the outcome, whatever the others do.[23] So rights are a property of the game form, and people have them whatever their preferences are.

The walls as a game form. Nick picks a row, and Carl picks a column. Whatever the other does, each can make his own walls white or yellow. In Table 1, each cell gives the state, and then the payoffs, Nick's first: 1 if a player gets what he wants (Nick differing from Carl, Carl matching Nick), and 0 if not.

Table 1: The walls as a game form. Payoffs: (Nick, Carl).

	Carl: white	Carl: yellow
Nick: white	WW $(0,1)$	WY $(1,0)$
Nick: yellow	YW $(1,0)$	YY $(0,1)$

Suppose we impose each player's preferred move, as Sen's condition does. In WW, Nick gets 0, and would switch to yellow, so YW excludes WW. In YW, Carl gets 0, and would switch to yellow, so YY excludes YW. In YY, Nick gets 0, and would switch to white, so WY excludes YY. In WY, Carl gets 0, and would switch to white, so WW excludes WY. These are the four exclusions from Section 3.6. Every cell is ruled out, so nothing can be chosen. But as a game, it still has an outcome: each player just picks a colour. Both have their rights, whatever they prefer.

Annie, Bill and Charlie. The next case shows that Sen's way of stating rights can misdescribe a perfectly ordinary exercise of a right.[24] There are three states: x, nobody marries; y, Annie marries Bill, and Charlie stays single; z, Annie marries Charlie, and Bill stays single.

	Annie	Bill	Charlie
Best	y	x	z
Middle	z	y	y
Worst	x	z	x

Everyone has the right to stay single. (People also have the right to marry, but it is not a unilateral right: nobody can exercise it alone, since a marriage takes two people who both choose it. So there seem to be group rights too: rights that no one individual can exercise, but that two or more people can exercise together.) On Sen's way of stating rights, Bill's preferring x to y is enough for x to exclude y. So y should not be chosen when x is available.

Now think of it as a game: the men propose, or not, and Annie accepts or refuses. Charlie's best choice is to propose. Annie's is to accept Bill or, failing him, Charlie. So if Bill proposes, Annie marries him (y); if not, she marries Charlie (z). Bill prefers y to z, so his best choice is to propose, and the outcome is y.

On Sen's way of stating rights, this fails to respect Bill's liberty: x was available, Bill prefers it to y, and y is chosen. But nobody invaded Bill's personal sphere. Bill chose to marry Annie: he exercised his right.

Exercising a right, not waiving it. Cases like Bill's are cases of exercising a right, not waiving it.[25] You have the right to choose the colour of your T-shirt. If you choose the blue T-shirt, you have exercised your right to choose your T-shirt's colour. You haven't waived a right to the green one. Likewise, Bill hasn't waived his right to stay single: his right was to choose whether to marry, and he chose.

[23] Sugden 1985: 225–226.
[24] Sugden 1985: 220–222.
[25] I owe this point to Ralf Bader.

Prude and Lewd as a game. Now give Prude and Lewd rights in this sense: each can choose whether he himself reads the book.

	Lewd reads	Lewd doesn't
Prude reads	Both read	Prude only
Prude doesn't	Lewd only	Nobody reads

Compare two situations.

(i) All four combinations are available. The cooperative solution is that Prude reads, and Lewd doesn't.

(ii) The only options are "Prude reads" and "nobody reads". This is a degenerate game: Prude is the only player whose choice makes any difference. And Prude doesn't read.

Prude reads is chosen from the bigger menu in (i), and it is on the smaller menu in (ii), but it is not chosen there. So this solution breaks contraction consistency. It breaks Sen's rights condition too: in (i), Prude reads is chosen although Nobody reads was available, and Prude prefers Nobody reads.

Liberty can still be inefficient. Game forms remove the paradox: there is always an outcome. But procedures that respect rights can still produce Prisoner's Dilemma outcomes: outcomes that everyone ranks below another available outcome.[26] For example, take part in politics: joining a party, going to meetings, perhaps voting itself.[27] We all want democratic politics to continue, but we would each rather other people did the work. Perhaps we would all be better off if we all took more part. But a law compelling everyone to take part would invade their liberty. As on the descriptive reading in Section 3.5, leaving people free can lead to a state that everyone ranks below another.

But notice that this happens when each person chooses for themselves, straightforwardly following individual rationality. We already know that choosing that way often leads to Pareto violations: that is the whole point of the Prisoner's Dilemma, the classic case in which the Nash equilibrium is Pareto inefficient. So in that respect it doesn't seem to be a big problem for liberty. What we are interested in is whether liberty brings a special problem of its own. If we strip out what would cause Pareto to be violated anyway, do we still get Pareto violations? If people can decide collectively, and cooperate, it seems that we need not. But that question goes beyond what we are doing in this course.

4 Arrow's theorem

4.1 Majority voting and Condorcet's paradox

From now on, we call the people *voters*, and the social states *options*. Each voter has a preference *ordering* of the options: a complete and transitive preference relation.

With just two options, x and y, there is an obvious way to decide: society ranks x above y when more voters prefer x to y. With two options, majority voting looks ideal. With three or more options, we can try the same thing: compare each pair of options by majority. But then the social ranking can go round in a circle. This was pointed out by Condorcet.[28]

Voter	Ranking
Voter 1	$x \succ y \succ z$
Voter 2	$y \succ z \succ x$
Voter 3	$z \succ x \succ y$

How does majority voting rank each pair?

[26] Sugden 1985: 227.

[27] Sugden 1985: 228.

[28] Condorcet, Marquis de (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie Royale.

– x against y: voters 1 and 3 prefer x. So $x \succ y$.

– y against z: voters 1 and 2 prefer y. So $y \succ z$.

– z against x: voters 2 and 3 prefer z. So $z \succ x$.

So society has $x \succ y$, $y \succ z$ and $z \succ x$: a cycle. There is no transitive social ranking, and (in the terms of Section 2.2) nothing on the menu $\{x, y, z\}$ is maximal.

4.2 Social ordering functions

Arrow's theorem is about a different kind of rule from Sen's result. A social choice rule gives a social choice function for each profile. What Arrow asks for is a social *ordering* for each profile.

Definition 8 (Social ordering function)**.** There are finitely many voters (at least two). A *social ordering function* gives, for each profile of voters' orderings that it applies to, a social ordering $\succsim$ of the options: a complete and transitive relation.

Majority voting is not a social ordering function: for Condorcet's profile, it gives no ordering.

4.3 Arrow's conditions

Arrow's theorem uses four conditions on a social ordering function. The first three are below; the fourth, Non-dictatorship, is in Section 4.4.

Unrestricted Domain: The social ordering function applies to every profile: voters may hold any orderings at all.

Pareto: If every voter strictly prefers x to y, then $x \succ y$.

Independence of Irrelevant Alternatives: Take any two profiles, and any two options x and y. If every voter ranks x against y the same way in both profiles, then society ranks x against y the same way in both.

"The same way" means: $x \succ y$ in both, or $y \succ x$ in both, or $x \sim y$ in both. So only how each voter ranks x against y matters for how society ranks x against y. Where anyone puts any other option is irrelevant.

Here is Independence of Irrelevant Alternatives in action. Take these two profiles, P and Q:

	Voter 1	Voter 2	Voter 3
P	$x \succ z \succ y$	$y \succ x \succ z$	$x \succ y \succ z$
Q	$z \succ x \succ y$	$y \succ z \succ x$	$x \succ y \succ z$

In Q, voters 1 and 2 have moved z. But every voter ranks x against y the same way in P and in Q: voters 1 and 3 prefer x, and voter 2 prefers y. So by Independence of Irrelevant Alternatives, society ranks x against y the same way in P and in Q. Where z is doesn't matter.

This is what Independence of Irrelevant Alternatives lets us do. We can change a profile however we like. As long as nobody changes how they rank x against y, society's ranking of x against y stays the same. So we can take what society says about x and y in one profile, and carry it across to another. The proof of Arrow's theorem does this again and again, and each time we will name the two profiles we are comparing.

4.4 Decisiveness and dictators

In Section 3.2, a person was decisive over a pair when, whichever of the two they preferred, the other was not chosen from any menu containing both. Here is the same idea for a social ordering function.

Definition 9 (Decisiveness)**.** A voter i is *decisive* over a pair $\{x, y\}$ when, at every profile, whichever of the two they strictly prefer, society strictly prefers too: if $x \succ_i y$, then $x \succ y$; and if $y \succ_i x$, then $y \succ x$.

Definition 10 (Dictator)**.** A *dictator* is a voter who is decisive over every pair.

Non-dictatorship: No voter is a dictator.

For example, take the social ordering function "society's ordering is always voter i's ordering", for some fixed voter i. Voter i is decisive over every pair, so voter i is a dictator. This social ordering function satisfies the other three conditions. Every profile is allowed. If every voter prefers x to y, so does voter i, so society does too. And society's ranking of x against y depends only on voter i's ranking of x against y, so Independence of Irrelevant Alternatives holds.

4.5 The theorem

Arrow's theorem. With finitely many voters and at least three options, no social ordering function satisfies Unrestricted Domain, Pareto, Independence of Irrelevant Alternatives and Non-dictatorship.[29]

Put another way: every social ordering function satisfying Unrestricted Domain, Pareto and Independence of Irrelevant Alternatives has a dictator. That is how we will prove it.

The three options matter. With only two options, majority voting satisfies all four conditions.

4.6 How the proof works

The proof has four steps. Steps 1–3 follow Geanakoplos's first proof.[30] Throughout, we assume that the social ordering function satisfies Unrestricted Domain, Pareto and Independence of Irrelevant Alternatives.

(1) The Extreme Option Lemma: if every voter puts b at the very top or the very bottom, so does society.

(2) Moving b from bottom to top one voter at a time, we find a pivotal voter for b.

(3) A pivotal voter for b is decisive over every pair not involving b.

(4) Every option has a pivotal voter, and they are all the same voter. (That is just what it is for this voter to be a dictator: a voter who is pivotal for every option is decisive over every pair.)

Along the way we keep moving options around in voters' orderings. Whenever we take one option out of a voter's ordering and put it back somewhere else, the result is still an ordering, and every pair not involving the moved option is ranked as before. Unrestricted Domain says that the new profile is allowed.

4.7 Step 1: the Extreme Option Lemma

The Extreme Option Lemma. Suppose a social ordering function satisfies Unrestricted Domain, Pareto and Independence of Irrelevant Alternatives, and every voter ranks b strictly above all other options, or strictly below all other options. Then society ranks b strictly above all other options, or strictly below all other options.

This holds even if some voters put b at the top and the rest put b at the bottom.

Here is the idea. Suppose society did put b somewhere in the middle, with some option a at least as high as b and some option c no higher than b. Now let every voter move c up to just above a. Because b is at an extreme for every voter, c never passes b, so nobody's ranking of a against b, or of b against c, changes. By Independence of Irrelevant Alternatives, society still has a at least as high as b, and b at least as high as c; so by transitivity, a is at least as high as c. But now every voter prefers c to a, so by Pareto, society prefers c to a. That is a contradiction.

First, an example with three voters. In the profile P, every voter puts b at the very top or the very bottom. Suppose, for contradiction, that society has $a \succsim b$ and $b \succsim c$. Build the profile P' from P: every voter moves c to just above a. (Voter 3 already has c just above a, so nothing changes for voter 3.)

[29] Arrow, Kenneth J. (1951). *Social Choice and Individual Values*. New York: Wiley.

[30] Geanakoplos, John (2005). "Three Brief Proofs of Arrow's Impossibility Theorem". In: *Economic Theory* 26.1, pp. 211–215: 212–213.

	Voter 1	Voter 2	Voter 3
P	$b \succ a \succ c$	$a \succ c \succ b$	$c \succ a \succ b$
P'	$b \succ \boldsymbol{c} \succ a$	$\boldsymbol{c} \succ a \succ b$	$c \succ a \succ b$

Compare P and P'. Every voter ranks a against b the same way in both, and b against c the same way in both. So by Independence of Irrelevant Alternatives, society has $a \succsim b$ and $b \succsim c$ in P' too. So, since the social ordering is transitive, $a \succsim c$ in P'. But every voter prefers c to a in P', so by Pareto, $c \succ a$ in P'. That is a contradiction.

And now for the general case:

Proof. Suppose every voter ranks b strictly at the top or strictly at the bottom in a profile P, and suppose for contradiction that society does not put b strictly at the top or strictly at the bottom. Then there are options a and c, other than b, with $a \succsim b$ and $b \succsim c$. We can take a and c to be different options. (If a and c are the same option, take a third option d, other than a and b. Since the social ordering is complete, $d \succsim b$ or $b \succsim d$. So d can play one of the two roles, and a the other.)

Build the profile P' from P by having every voter move c to just above a. Since every voter has b at the very top or the very bottom, c never passes b. So in P', every voter ranks a against b, and b against c, the same way as in P. And in P', every voter prefers c to a.

Compare P and P'. By Independence of Irrelevant Alternatives, $a \succsim b$ and $b \succsim c$ in P', as in P. By transitivity, $a \succsim c$ in P'. But by Pareto, $c \succ a$ in P'. This is a contradiction. So society puts b strictly at the top or strictly at the bottom. □

4.8 Step 2: a pivotal voter

Number the voters $1, 2, \ldots, n$. Start from a profile P_0 in which every voter puts b strictly at the bottom. Then let the voters move b to the very top, one at a time, in order: voter 1 first, giving P_1; then voter 2, giving P_2; and so on, until in P_n every voter puts b strictly at the top. Nothing else moves.

Here is an example with three voters.

	Voter 1	Voter 2	Voter 3	Society
P_0	$c \succ a \succ b$	$a \succ c \succ b$	$c \succ a \succ b$	b at the bottom
P_1	$b \succ c \succ a$	$a \succ c \succ b$	$c \succ a \succ b$	b at the bottom
P_2	$b \succ c \succ a$	$b \succ a \succ c$	$c \succ a \succ b$	b at the top
P_3	$b \succ c \succ a$	$b \succ a \succ c$	$b \succ c \succ a$	b at the top

In P_0, every voter has b at the bottom, so by Pareto society does too. In P_3, every voter has b at the top, so by Pareto society does too. In between, in every profile, every voter has b at the very top or the very bottom, so by the Extreme Option Lemma, society has b strictly at the top or strictly at the bottom. So at some move, society's b jumps all the way from the bottom to the top. The Society column in the table is an example: all we know is that there is a jump somewhere. Say it jumps when voter 2 moves. Then voter 2 is a *pivotal voter* for b. We call the profile just before the jump b-BOTTOM (here, P_1), and the profile just after it b-TOP (here, P_2).

In general, the first voter whose move takes society's b from the bottom to the top is a pivotal voter for b. There must be one, since society's b starts at the bottom and ends at the top. In b-BOTTOM, the voters numbered below the pivotal voter put b at the top, and the pivotal voter and the voters numbered above put b at the bottom; society puts b strictly at the bottom. In b-TOP, the pivotal voter has moved b to the top too; society puts b strictly at the top.

Notice that we say *a* pivotal voter, not *the* pivotal voter. For all we know so far, if the voters moved in a different order, or if we started from a different P_0, a different voter might turn out to be pivotal. It is only at the end of the proof that we find out that there was only ever one.

4.9 Step 3: a pivotal voter is decisive over every pair not involving b

Take any two options a and c, other than b. We show that a pivotal voter for b is decisive over $\{a, c\}$. Stay with the example, in which voter 2 is pivotal.

Let ORIGINAL be any profile in which voter 2 ranks a above c. The other voters can rank a and c however they like. In this example, voters 1 and 3 rank c above a. We show that society goes with voter 2, and ranks a above c, whatever everyone else thinks.

	Voter 1	Voter 2	Voter 3
ORIGINAL	$c \succ b \succ a$	$a \succ c \succ b$	$b \succ c \succ a$

The plan is this. We build a new profile, b-MIDDLE, from ORIGINAL by moving b only. In b-MIDDLE, voter 2 puts b between a and c, and everyone else puts b where they put it in b-BOTTOM and b-TOP.[31] Then b-MIDDLE looks like b-BOTTOM as far as a against b goes, so society must rank a above b in b-MIDDLE. And it looks like b-TOP as far as b against c goes, so society must rank b above c in b-MIDDLE. Then transitivity gives a above c in b-MIDDLE; and since only b moved, the same holds in ORIGINAL.

Building b-MIDDLE. Start from ORIGINAL and move b only. Voter 1 (numbered below voter 2) moves b to the top. Voter 2 moves b to between a and c. Voter 3 (numbered above voter 2) moves b to the bottom.

	Voter 1	Voter 2	Voter 3
ORIGINAL	$c \succ b \succ a$	$a \succ c \succ b$	$b \succ c \succ a$
b-MIDDLE	$\boldsymbol{b} \succ c \succ a$	$a \succ \boldsymbol{b} \succ c$	$c \succ a \succ \boldsymbol{b}$

Only b moved, so every voter ranks a against c the same way in ORIGINAL and b-MIDDLE. So by Independence of Irrelevant Alternatives, society ranks a against c the same way in ORIGINAL and b-MIDDLE. It is enough, then, to show that society ranks a above c in b-MIDDLE.

Society ranks a above b in b-MIDDLE. Start from b-MIDDLE, and let voter 2 move b to the bottom. In this example, the result is exactly b-BOTTOM, in which society puts b at the bottom: so society has $a \succ b$ in b-BOTTOM. Now compare b-MIDDLE and b-BOTTOM, looking only at a against b:

	Voter 1	Voter 2	Voter 3	Society
b-MIDDLE	$\boldsymbol{b} \succ c \succ \boldsymbol{a}$	$\boldsymbol{a} \succ \boldsymbol{b} \succ c$	$c \succ \boldsymbol{a} \succ \boldsymbol{b}$	$a \succ b$
b-BOTTOM	$\boldsymbol{b} \succ c \succ \boldsymbol{a}$	$\boldsymbol{a} \succ c \succ \boldsymbol{b}$	$c \succ \boldsymbol{a} \succ \boldsymbol{b}$	$a \succ b$

Every voter ranks a against b the same way in both: voter 1 prefers b, and voters 2 and 3 prefer a. So by Independence of Irrelevant Alternatives, since $a \succ b$ in b-BOTTOM, $a \succ b$ in b-MIDDLE.

Society ranks b above c in b-MIDDLE. Start from b-MIDDLE again, and this time let voter 2 move b to the top. In this example, the result is exactly b-TOP, in which society puts b at the top: so society has $b \succ c$ in b-TOP. Now compare b-MIDDLE and b-TOP, looking only at b against c:

	Voter 1	Voter 2	Voter 3	Society
b-MIDDLE	$\boldsymbol{b} \succ \boldsymbol{c} \succ a$	$a \succ \boldsymbol{b} \succ \boldsymbol{c}$	$\boldsymbol{c} \succ a \succ \boldsymbol{b}$	$b \succ c$
b-TOP	$\boldsymbol{b} \succ \boldsymbol{c} \succ a$	$\boldsymbol{b} \succ a \succ \boldsymbol{c}$	$\boldsymbol{c} \succ a \succ \boldsymbol{b}$	$b \succ c$

[31] b-BOTTOM and b-TOP are the profiles found in Section 4.8, where only b moved, starting from P_0. b-MIDDLE, here in Section 4.9, is built from ORIGINAL instead, so the options other than b may be arranged differently in it. So if, starting from b-MIDDLE, voter 2 moves b to the bottom, the result need not be the very same profile as the b-BOTTOM of Section 4.8; and likewise, if voter 2 moves b to the top, the result need not be the b-TOP of Section 4.8. But it does not differ from it in how any voter ranks b against any other option. That is because how a voter ranks b against anything depends only on whether b is at the very top or the very bottom, not on how the other options are arranged. So, by Independence of Irrelevant Alternatives, society ranks every pair involving b the same way in both.

Every voter ranks b against c the same way in both: voters 1 and 2 prefer b, and voter 3 prefers c. So by Independence of Irrelevant Alternatives, since $b \succ c$ in b-TOP, $b \succ c$ in b-MIDDLE.

Society ranks a above c in ORIGINAL. In b-MIDDLE, $a \succ b$ and $b \succ c$, so by transitivity, $a \succ c$. And society ranks a against c the same way in ORIGINAL as in b-MIDDLE. So $a \succ c$ in ORIGINAL: society goes with voter 2, although voters 1 and 3 both prefer c.

ORIGINAL was any profile in which voter 2 ranks a above c, and the same argument works with a and c swapped. So voter 2 is decisive over $\{a, c\}$.

Here is the general argument. It is the same as in the example, except that moving the pivotal voter's b need not give exactly b-BOTTOM or b-TOP, since the other options may be arranged differently. So we compare b-MIDDLE with b-BOTTOM and b-TOP directly.

Proof. Let V be a pivotal voter for b, found as in Step 2, and let a and c be any two options other than b. Let ORIGINAL be any profile in which V ranks a strictly above c. We show that $a \succ c$ in ORIGINAL.

Build b-MIDDLE from ORIGINAL by moving b only: the voters numbered below V put b strictly at the top; V puts b strictly between a and c, so that $a \succ_V b \succ_V c$; and the voters numbered above V put b strictly at the bottom. Only b moved, so every voter ranks a against c the same way in ORIGINAL and b-MIDDLE. By Independence of Irrelevant Alternatives, society ranks a against c the same way in ORIGINAL and b-MIDDLE.

Society has $a \succ b$ in b-MIDDLE. Compare b-MIDDLE with b-BOTTOM. The voters numbered below V put b at the very top in both, so they prefer b to a in both. V prefers a to b in both: in b-MIDDLE, V puts b below a, and in b-BOTTOM, V puts b at the very bottom. The voters numbered above V put b at the very bottom in both, so they prefer a to b in both. So every voter ranks a against b the same way in b-MIDDLE and b-BOTTOM. In b-BOTTOM, society puts b strictly at the bottom, so $a \succ b$. So by Independence of Irrelevant Alternatives, $a \succ b$ in b-MIDDLE.

Society has $b \succ c$ in b-MIDDLE. Compare b-MIDDLE with b-TOP. The voters numbered below V put b at the very top in both, so they prefer b to c in both. V prefers b to c in both: in b-MIDDLE, V puts b above c, and in b-TOP, V puts b at the very top. The voters numbered above V put b at the very bottom in both, so they prefer c to b in both. So every voter ranks b against c the same way in b-MIDDLE and b-TOP. In b-TOP, society puts b strictly at the top, so $b \succ c$. So by Independence of Irrelevant Alternatives, $b \succ c$ in b-MIDDLE.

Society has $a \succ c$ in ORIGINAL. In b-MIDDLE, $a \succ b$ and $b \succ c$, so $a \succ c$ by transitivity. So $a \succ c$ in ORIGINAL.

ORIGINAL was any profile in which V ranks a above c, and the same argument works with a and c swapped. So V is decisive over $\{a, c\}$. Since a and c were any two options other than b, V is decisive over every pair not involving b. □

4.10 Step 4: a dictator

So far, we have a pivotal voter for b who is decisive over every pair not involving b. But we don't yet know about pairs involving b, such as $\{a, b\}$.

There was nothing special about b: the same argument works with any option in place of b. So for each option x, we can find a pivotal voter for x, by moving x from bottom to top one voter at a time, and that voter is decisive over every pair not involving x. For each option x, fix one such voter, and call them $V(x)$.

If these voters are all the same voter, that voter is a dictator. For take any pair $\{x, y\}$. There are at least three options, so there is a third option z, and $V(z)$ is decisive over $\{x, y\}$, since that pair doesn't involve z. So it is enough to show that the voters $V(x)$ are all the same voter.

Here is the idea. Take three options a, b and c. $V(c)$ is decisive over $\{a, b\}$, $V(a)$ over $\{b, c\}$, and $V(b)$ over $\{a, c\}$. Give them these preferences:

Voter	Decisive over	Give them
$V(c)$	$\{a, b\}$	$a \succ b$
$V(a)$	$\{b, c\}$	$b \succ c$
$V(b)$	$\{a, c\}$	$c \succ a$

Then society would have $a \succ b$, $b \succ c$ and $c \succ a$: a cycle, which is not an ordering. If there are at least two different voters among $V(a)$, $V(b)$ and $V(c)$, nothing stops them having these preferences. Each voter then needs at most two of the three, and any two of them fit into an ordering. So there would be a profile at which society cycles, which is impossible. The only way out is for $V(a)$, $V(b)$ and $V(c)$ to be just one voter. Then the three preferences would all belong to one voter, and they form a cycle, which no voter can hold, since every voter's preferences are an ordering.

Proof. Take any three options a, b and c, and suppose for contradiction that $V(a)$, $V(b)$ and $V(c)$ are not all the same voter. Build a profile as follows. $V(c)$ ranks a above b; $V(a)$ ranks b above c; and $V(b)$ ranks c above a. Since the three are not all one voter, each voter is given at most two of these three preferences, and any two of them fit into an ordering: $a \succ b$ and $b \succ c$ fit into $a \succ b \succ c$; $b \succ c$ and $c \succ a$ fit into $b \succ c \succ a$; and $c \succ a$ and $a \succ b$ fit into $c \succ a \succ b$. Every other voter can rank the options however they like. By Unrestricted Domain, this profile is allowed.

At this profile, $V(c)$ is decisive over $\{a, b\}$, so $a \succ b$. $V(a)$ is decisive over $\{b, c\}$, so $b \succ c$. $V(b)$ is decisive over $\{a, c\}$, so $c \succ a$. By transitivity, $a \succ b$ and $b \succ c$ give $a \succ c$, which contradicts $c \succ a$. So $V(a)$, $V(b)$ and $V(c)$ are all the same voter.

This holds for any three options. Now take any two options x and y, and a third option z: by what we have just shown, $V(x) = V(y) = V(z)$. So all the voters $V(x)$ are one voter, V. Take any pair $\{x, y\}$, and a third option z. $V = V(z)$ is decisive over $\{x, y\}$, since that pair doesn't involve z. So V is decisive over every pair: V is a dictator. □

Since $V(x)$ could have been any pivotal voter for x, every pivotal voter, for any option, is this one voter. So there was only ever one pivotal voter, whatever order the voters moved in.

This completes the proof of Arrow's theorem: every social ordering function satisfying Unrestricted Domain, Pareto and Independence of Irrelevant Alternatives has a dictator.

5 Voting systems

Arrow's theorem says that every social ordering function breaks at least one of his conditions. For familiar voting systems, we can ask which condition breaks.

5.1 Plurality

Plurality: rank the options by how many voters rank them first.

Plurality violates Independence of Irrelevant Alternatives. This is the *spoiler effect*. Take these two profiles (the numbers on the left say how many voters have each ranking).

Profile 1

Voters	Ranking
45	$x \succ y \succ z$
40	$y \succ x \succ z$
15	$z \succ y \succ x$

Profile 2

Voters	Ranking
45	$x \succ y \succ z$
40	$y \succ x \succ z$
15	$y \succ z \succ x$

In Profile 1, the first places are x 45, y 40, z 15, so $x \succ y \succ z$. In Profile 2, the 15 voters have moved z down, below y. The first places are now y 55, x 45, z 0, so $y \succ x \succ z$. Nobody changed how they rank x against y: 55 prefer y, and 45 prefer x, both times. But society's ranking of x against y flipped. So plurality

violates Independence of Irrelevant Alternatives. Option z is the spoiler: the 15 voters who put it first take their first places away from y, and that costs y the win.

5.2 Instant-runoff

Instant-runoff: remove the option with the fewest first places, and repeat, each time counting each voter's first place among the options that are left. Rank the options in reverse order of removal.

Profile A

Voters	Ranking
6	$x \succ z \succ y$
5	$y \succ x \succ z$
4	$z \succ y \succ x$

Profile B

Voters	Ranking
6	$x \succ z \succ y$
3	$y \succ x \succ z$
6	$z \succ y \succ x$

From A to B, two voters change their ranking from $y \succ x \succ z$ to $z \succ y \succ x$. In A, the first places are x 6, y 5, z 4. Remove z; then y beats x, 9 to 6. So $y \succ x \succ z$. In B, the first places are x 6, y 3, z 6. Remove y; then x beats z, 9 to 6. So $x \succ z \succ y$. In both profiles, 6 voters prefer x to y and 9 prefer y to x. But $y \succ x$ in A and $x \succ y$ in B. So instant-runoff violates Independence of Irrelevant Alternatives.

5.3 The Borda count

The Borda count: each voter gives each option 1 point for every option it is ranked above (so with three options, 2 points for first place, 1 for second and 0 for third). Rank the options by their total points.

Profile A

Voters	Ranking
3	$x \succ y \succ z$
2	$y \succ x \succ z$

Profile B

Voters	Ranking
3	$x \succ y \succ z$
2	$y \succ z \succ x$

From A to B, two voters move z up. In A, x gets $6 + 2 = 8$ and y gets $3 + 4 = 7$, so $x \succ y$. In B, x gets 6, y gets $3 + 4 = 7$, and z gets 2, so $y \succ x$. Nobody changed how they rank x against y: 3 prefer x and 2 prefer y, both times. But the verdict flipped. So the Borda count violates Independence of Irrelevant Alternatives.

5.4 Majority voting, approval voting and dictatorship

Majority voting: rank x above y when more voters prefer x to y. It violates none of Arrow's four conditions: it satisfies Unrestricted Domain, Pareto, Independence of Irrelevant Alternatives and Non-dictatorship. But, as Condorcet's profile shows, it can give a cycle rather than an ordering. So it is not a social ordering function at all.

Approval voting: each voter approves as many options as they like, and the option with the most approvals wins. The ballot is not a ranking, so Arrow's theorem does not apply to it directly.

Dictatorship: society's ordering is always voter i's ordering, for some fixed voter i. It violates only Non-dictatorship. As we saw in Section 4.4, it satisfies Unrestricted Domain, Pareto and Independence of Irrelevant Alternatives.

5.5 Summary

System	What fails
Plurality	Independence of Irrelevant Alternatives
Instant-runoff	Independence of Irrelevant Alternatives
Borda count	Independence of Irrelevant Alternatives
Majority voting	No social ordering (cycles)
Approval voting	Outside Arrow's framework
Dictatorship	Non-dictatorship

Looking back, every one of Arrow's conditions matters. Drop any one of them, or the demand that the output be an ordering, and something satisfies the rest.

Drop	And this satisfies the rest
Unrestricted Domain	Majority voting, if only profiles at which it gives an ordering are allowed
Pareto	The same fixed ordering, whatever anyone prefers
Independence of Irrelevant Alternatives	The Borda count
Non-dictatorship	"Society's ordering is voter 1's"
A social ordering as the output	Majority voting

References

Arrow, Kenneth J. (1951). *Social Choice and Individual Values*. New York: Wiley.

Condorcet, Marquis de (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie Royale.

Geanakoplos, John (2005). "Three Brief Proofs of Arrow's Impossibility Theorem". In: *Economic Theory* 26.1, pp. 211–215.

Gibbard, Allan (1974). "A Pareto-Consistent Libertarian Claim". In: *Journal of Economic Theory* 7.4, pp. 388–410.

Sen, Amartya (1970). "The Impossibility of a Paretian Liberal". In: *Journal of Political Economy* 78.1, pp. 152–157.

– (1971). "Choice Functions and Revealed Preference". In: *The Review of Economic Studies* 38.3, pp. 307–317.

– (1976). "Liberty, Unanimity and Rights". In: *Economica* 43.171, pp. 217–245.

– (1983). "Liberty and Social Choice". In: *The Journal of Philosophy* 80.1, pp. 5–28.

– (1993). "Internal Consistency of Choice". In: *Econometrica* 61.3, pp. 495–521.

– (2017). *Collective Choice and Social Welfare: An Expanded Edition*. Cambridge, MA: Harvard University Press.

Sugden, Robert (1985). "Liberty, Preference, and Choice". In: *Economics and Philosophy* 1.2, pp. 213–229.