Social Choice Theory

Antony Eagle

Choices, Models and Morals » Lecture 12

Contents

Social Choice and Aggregation

From Individual to Social Choice

- So far, we've mostly examined theories of **rational individual choice**: which **outcomes** should I try to promote, given what I believe, and what I value.
- If my values are informed by **moral considerations**, then I might be minded to promote outcomes that are best not merely for me, but in general.
- If a moral theory is **consequentialist**, **agent-neutral** and **maximising**, it will recommend the **same outcome** to every moral agent: if I ought to promote A, so should everyone else.
- Thus such theories are both accounts of individual choice, but also collective choice – or **social choice**.
- For example, as a theory of collective decision, utilitarianism could not be simpler: because whether an action is to be performed by an individual or by a group, it ought to be done iff it maximises total well-being among available actions.

Making Social Choice Too Easy?

- This makes social choice look deceptively easy, because the difficult part – **reconciling differences within the group** – all happens off-stage, or not at all.
- Utilitarianism doesn't assign a role to people's **beliefs** in determining outcome value, so factual disagreements within the group don't change the ranking of outcomes – so the utilitarian can set aside problems about how to aggregate beliefs.
  - Such disageements will need to be addressed when you have to persuade each member of the group to do their part in what the group ought to do!
- Utilitarianism presupposes a global scale of well-being that is **interpersonally comparable**, and which can be measured numerically so that may simply be **summed** (or averaged) to get a value for any possible outcome.
- A convenient assumption, but is it true that there is just one scale which already takes into consideration everyone's aims and interests so that trade-offs between them can be handled purely numerically?

Other Questions of Social Choice

- These features of utilitarian social choice mean that there are certain questions it does not address, for example:
    1. How should we pool our opinions to come up with a **group belief**?
    2. How should we decide what to do in circumstances of moral disagreement when we want to ensure **buy-in** or cooperation from group members?
    3. How should we decide if well-being isn't numerically comparable – e.g., if **preferentism** is correct?
    4. Can we answer **descriptive** questions (not normative ones) about what groups will choose or prefer, given facts about what members will choose or prefer?
- These are different questions, though a good theory of social choice might answer several of them.

Pooling Belief and Desire

- One idea: construct group beliefs and desires, and calculate the maximum group expected value – just as in individual choice.
- A tricky issue for belief (Hausman, McPherson, and Satz 2017: 263). Suppose Alice, Bob, and Carla each have opinions in the truth or falsity of P, $P \to Q$ and Q as follows:

Table 1: The Discursive Dilemma

	Alice	Bob	Carla	**Group**
P	T	F	T	**T**
$P \to Q$	T	T	F	**T**
Q	T	F	F	**F**

- Suppose the group opinion is fixed by **majority vote** over individual opinions.
- That takes three individually coherent belief states into a **logically inconsistent** group belief state.
  - The problem recurs for many approaches to belief aggregation (List and Pettit 2002), though some sophisticated approaches to pooling levels of confidence exist (Pettigrew 2020a: ch. 9).

Preference Aggregation

- Suppose then we start with preferences, the individual rankings of outcomes by expected value.
  - That is congenial to those who think belief and desire are real mental states, as well as to constructivists who think preference is primary.
- We don't need to worry about aggregating belief now, since factual disagreements will be represented, insofar as they have any significance, in differences in preferences.
- How then can we **collectivise individual preferences**?
- In social choice theory, this is the problem of constructing a **social welfare function**: a coherent preference ranking of outcomes – i.e., complete, transitive, and asymmetric (Peterson 2017: 292) – as a **function** of individual preference rankings.
- To solve this problem, you might begin by adopting this apparently plausible principle on aggregation:

  Majority Rule: The group should prefer a to b iff most people in G prefer a to b (Peterson 2017: 287).

The Voting Paradox

- A problem: Majority Rule **doesn't yield a social welfare function**. Consider our three people again, and their preferences between a, b and c:

Table 2: Voting Paradox

Preference	Agree	Disagree
$a \succ b$	Alice, Bob	Clara
$b \succ c$	Alice, Clara	Bob
$c \succ a$	Bob, Clara	Alice

- Aggregated by Majority Rule, this generates a preference for a over b, for b over c, and for c over a; if this were a ranking, transitivity would then give us that a is strictly preferred to itself, which is **incoherent**.
- Are there other principles we could use?

Arrow's Impossibility Theorem

Pareto Again

- It seems plausible that if enough people in the group prefer a to b, so should the group. Majority Rule takes 'enough' to mean: more than half, which didn't work out.
- But what if we take 'enough' to mean everyone? More precisely:

  Weak Pareto: If everyone in G prefers a to b, then the group should prefer a to b (Hausman, McPherson, and Satz 2017: 248; Pettigrew 2020a: §6.1).
- Surely this is defensible, as the weakest possible version of the 'enough people' principle.
  - In the kinds of welfarist approaches which take preference as basic, Pareto is already imposed as a condition on acceptable allocations of welfare (lecture 9), so it is not surprising to see it adopted as a principle on social choice too.

No Dictator

- Another plausible idea: the group preferences should **recognise disagreement** within the group, if there is any.
  - So if Alice and Bob have the same preference ranking, and Carla disagrees with them on something, the final ranking **shouldn't just ignore** Carla – it should not be identical with Alice's and Bob's preferences regardless of what Carla thinks.
- The weakest version of this: if there is disagrement within the group, then the group preference shouldn't ignore everyone except one person. For short: **there should be no dictator**.
- More precisely:

No Dictator There is no individual δ in G such that the social welfare function of G is always the same as δ's, no matter what the other individuals in G prefer.

  - This doesn't preclude the social welfare **happening** to be identical to some individual's. The output SWF may be the same as Alice's. But Bob could have agreed with Carla; then the SWF would not been Alice's. Then Alice is no dictator.

Independence of Irrelevant Alternatives

- A final plausible idea: the group ranking of a and b should depend only on preferences between a and b, and not on preferences between a (or b) and any other irrelevant alternative option.
- Suppose we had some group of individual preference rankings $G = \succ_1, \succ_2, \ldots, \succ_n$. An a, b-**stable variant** of G is a group of preferences G' obtained from G by modifying each $\succ_i$ by switching the ordering of any outcomes except keeping the relative order of a and b the same.

  IIA If G' is an a, b-stable variant of G, then the order of a and b in the social welfare function determined by G is the same as in the social welfare function determined by G' (Peterson 2017: 294).

- Violating IIA allows preferences for options other than a and b determine the final ranking of a and b. So systems which violate IIA incentivise **tactical** preferences.
  - Suppose for you $c \succ b \succ a$. You might **act as if** you prefer a to c to try and ensure that the group prefers b to a!

Arrow's Theorem

- In his PhD thesis, Arrow (Arrow 1951) proved this ground-breaking result:

  Arrow No procedure for constructing social welfare functions from individual preference orderings could satisfy IIA, No Dictator, and Weak Pareto (if there are more than two outcomes and the group has more than one member).
  A proof can be found in Peterson (Peterson 2017: 295–97).

> Any method that satisfies Weak Pareto and No Dictatorship will make the aggregate ordering of two possible acts depend on the individual orderings of other acts; any method that satisfies Weak Pareto and the Independence of Irrelevant Alternatives will give rise to a dictator; and any method that satisfies No Dictator and Independence of Irrelevant Alternatives will sometimes fail to preserve a unanimous consensus that one act is better than another. (Pettigrew 2020b, p. §6.1)

- In the voting paradox, Majority Rule yields pairwise preferences compatible with Weak Pareto, No Dictator and IIA – so doesn't yield an ordering (Hausman, McPherson, and Satz 2017: 250).

Responding to Arrow

- The conditions we've discussed seem like attractive ones.
- On the other hand, they can't be jointly satisfied as constraints when we try to aggregate individual preference into group preference, and yet we can sometimes do this fairly satisfactorily. Won't that show the conditions aren't after all as plausible as they seemed?
- For example, consider **voting**. Many attractive methods of voting that seem democratically acceptable and to reflect more or less successfully 'the will of the electorate' exist; each must violate one of Arrow's conditions.
- The most common condition to be violated in plausible voting systems is IIA; let's see a couple.
    - Peterson discusses how the **Borda count** violates IIA (Peterson 2017: 294), and this sort of procedure might also be what Housmann, *et al*. mean by a 'utilitarian social welfare function' (2017: 253).

First Past the Post

- In First Past the Post voting, each voter nominates their most preferred candidate, and the candidate with the most nominations wins. For example, suppose we have four candidates, α, β, γ and δ, and five voters, A, B, C, D and E. Compare two elections:

Table 3: Two different outcomes in FPTP

Election	A	B	C	D	E	Winner
1	α	α	β	γ	δ	α
2	α	α	β	β	β	β

- In election 1, α wins. In election 2, no voters change their preferences between α and β; but D and E change their preferences between β and the two minor candidates; but now β wins, violating IIA.
  - An example of **tactical voting**. Suppose D prefers $\gamma \succ \beta \succ \alpha$; knowing what the polls say, D pretends to prefer β to γ to try and ensure a less bad outcome.

Preferential Voting

- FPTP has many bad features. Lots of Australians like the system we call preferential voting, or **instant-runoff voting**.
- Here, voters give a complete preference ranking over candidates, which is successively trimmed by the iterated elimination of least-preferred candidates.
- Suppose voters A, B, C, D, and E fill in ballots for α, β and γ as follows:

Table 4: A preferential vote

Stage	A	B	C	D	E	Note
1	$\alpha \succ \gamma \succ \beta$	$\beta \succ \alpha \succ \gamma$	$\gamma \succ \beta \succ \alpha$	$\alpha \succ \gamma \succ \beta$	$\beta \succ \alpha \succ \gamma$	γ eliminated
2	$\alpha \succ \beta$	$\beta \succ \alpha$	$\beta \succ \alpha$	$\alpha \succ \beta$	$\beta \succ \alpha$	β wins

  - A natural way to read this order of elimination and election is that the group preference is $\beta \succ \alpha \succ \gamma$.

Preferential Voting Violates IIA

- Note however what happens if voter E changes their preference about γ, moving it from last to first. IIA says: because no one's preferences between α and β have changed, the group ranking of those candidates should remain unchanged. But this is violated:

Table 5: A preferential vote in an α, β-stable variant group

Stage	A	B	C	D	E	Note
1*	$\alpha \succ \gamma \succ \beta$	$\beta \succ \alpha \succ \gamma$	$\gamma \succ \beta \succ \alpha$	$\alpha \succ \gamma \succ \beta$	$\boldsymbol{\gamma \succ \beta \succ \alpha}$	β eliminated
2*	$\alpha \succ \gamma$	$\alpha \succ \gamma$	$\gamma \succ \alpha$	$\alpha \succ \gamma$	$\gamma \succ \alpha$	α wins

  - A natural way to read this order of elimination and election is that the group preference is $\alpha \succ \gamma \succ \beta$: which reverses the preference between α and β even though no one changed their mind about the rank of those two candidates!

Motivating the Denial of IIA

- Preferential voting improves on FPTP by incorporating some information about **strength of preference**.
- The idea: if Alice's preferences are $a \succ b \succ c$, and Bob's $a \succ c \succ b$, then while both prefer a to c, Alice does so more, because there are more options intervening between them in her preferences.
  - We don't have full interpersonal comparison – maybe I still enjoy even those things at the very bottom of my preference ranking! – but given how people usually are, we can get a rough idea of cardinal utility by looking at **rank order**. This is even more explicit in the Borda count.
- This gives us a rationale for denying IIA:

> while the group ordering of a and b really does just depend on the individual attitudes to a and b, the individual orderings of c, d, and e relative to a and b give information about the strength of the individual attitudes to a and b – information that is left out if we look just at whether each individual prefers a or b. (Pettigrew 2020a: §6.1)

Sen’s Impossibility Theorem

Liberalism

- Many ways to construct social welfare functions are **authoritarian**: no **individual** can, through their own preferences, guarantee any outcome.
  - Consider FPTP, suppose A and B opt for an option where A and B get half the cake and C gets nothing, while C opts for an equal share. C's preference is swamped by A's and B's.
- A **liberal** aggregation rule is one where some individuals have the **right** to secure some outcomes: 'to have a right to do X is for one's preferences to be socially decisive between social states that differ only with respect to whether one does X' (Hausman, McPherson, and Satz 2017: 259).
  - In the cake example, for C to have a right to an equal share would be for his preference to receive no less than 1/3 of the cake to guarantee that the group prefers that he receive that share.

Minimal Liberalism 'There are at least two individuals in society such that for each of them there is at least one pair of alternatives with respect to which she is decisive, that is, there is a pair *a* and *b* such that if she prefers *a* to *b*, then society prefers *a* to *b* (and society prefers *b* to *a* if she prefers *b* to *a*).' (Peterson 2017: 299)

Lewd and Prude

- Let's illustrate this idea. Suppose Lewd and Prude are the two members of our society G, and they are trying to figure out who their society should prefer reads a banned book (Hausman, McPherson, and Satz 2017: 260).
  - Lewd wants everyone to read it, but if only one person was to read it, prefers it to be Prude, to free his narrow mind:

    > Both $\succ_L$ Only Prude $\succ_L$ Only Lewd $\succ_L$ Neither.

  - Prude wants no one to read it, but if only one person was to read it, prefers it to be Prude, to save Lewd from further corruption:

    > Neither $\succ_P$ Only Prude $\succ_P$ Only Lewd $\succ_P$ Both.

- **Weak Pareto** tells us: because everyone agrees that Only Prude is preferable to only Lewd, the group preference will reflect that: Only Prude $\succ_G$ Only Lewd.

Minimally Liberal Rights to Read

- Suppose Lewd has the right to read the book if he wants, so that if an outcome differs only in whether Lewd reads the book or not, he can make society conform to his preference.
  - So between Only Prude and Both, which differ just in whether Lewd reads, Lewd's preference wins: Both $\succ_G$ Only Prude.
- Likewise, suppose Prude has the right to refrain from reading if he wants, so that if an outcome differs only in whether Prude reads the book or not, he can make society conform to his preference.
  - So between Both and Only Lewd, which differ just in whether Prude reads, Prude's preference wins: Only Lewd $\succ_G$ Both.
- Putting this together, with transitivity (assuming $\succ_G$ is an ordering), Only Lewd $\succ_G$ Only Prude.
- But this **contradicts** what Weak Pareto told us about the group preference ordering.

Pareto and Liberalism

- This is not a coincidence, not just a byproduct of some idiosyncratic feature of these preferences.

  Sen No social welfare function can aggregate preferences into an ordering in a way that satisfies Minimal liberalism and Pareto (Sen 1970).

- The proof is quite straightforward (Peterson 2017: 300): for any SWF, we can come up with some preferences such that individual decisiveness about part of the preference ordering, together with everyone's decisiveness about other parts, can be stitched together by transitivity into an incoherent ranking.
  - Note the **quantifier order**: it just says that you can undermine any attempt to aggregate with some specific collection of individual preferences, not that no collection of preferences can be coherently aggregated – maybe we can have a SWF that does pretty well on most cases.

A Liberal Response

- Sen's theorem renders liberalism as a thesis about group preference – that sometime individual preferences should decide group preference.
- But one response is to conceive of liberalism as saying that some outcomes for individuals are not a matter for the group **at all**:

> each person may exercise his rights as he chooses. The exercise of these rights fixes some features of the world. Within the constraints of these fixed features, a choice can be made by a social choice mechanism based upon a social ordering, if there are any choices left to make! (Nozick 1974: 165)

- That is to deny, in effect, that the social welfare function needs to be **complete**: to assert that while individuals may have preferences between a and b, it is not part of the job of society to choose between them, and hence a preference between them need not included in the group preference.
  - A commitment to Liberalism in this sense may be compatible with Weak Pareto.

Alternatives to Preference Aggregation

IIA, Pareto, Liberalism

- Sen's and Arrow's results together show that Weak Pareto (given ordering) entails that any plausible aggregation of preferences will deny at least one of (IIA, No Dictator) and also deny Minimal Liberalism.
  - This too is no accident: in a sense, Minimal Liberalism says that either my dictatorship over my rights makes me a total dictator, or my exercise of my rights actually cannot treated as an irrelevant alternative when determining group preference.
- You could thus save minimal liberalism and IIA by **denying Weak Pareto**.
- A high cost, since Pareto relies only on the **least controversial** part of aggregation, pooling agreeing opinions.
- On the other hand, maybe it is not so uncontroversial. After all, if A prefers a to b and ranks them bottom of his ordering, and B also prefers a to b and ranks them both near the top, their pairwise agreement on **relative value** actually masks a deep disagreement about the **absolute value** of those options. And couldn't absolute value matter?

Looking Below Preference

- Suppose we have two pairs of three voters, with these subjective values assigned to outcomes:

Table 6: Votes and Values

Option	A	B	C	**Average**	A*	B*	C*	**Average**
α	10	4	7	**7**	4	0	5	**3**
β	1	5	3	**3**	3	10	2	**5**

- Between A, B and C, α is preferred by 2/3, and also has much higher average value. But between A*, B* and C*, even though the preferences are the same, the average values are not. And why should we **ignore** such information when aggregating people's wants?

> when we have the cardinal information from which the ordinal information in the preference ordering is extracted, and upon which it is based, we should use that cardinal information and aggregate that, rather than directly aggregating the preferences. (Pettigrew 2020a: §6.1)

Manipulation: From Social Choice to Game Theory

- Another direction is to look at alternative ways to model individual choice.
- Another famous theorem in social choice theory is Gibbard's theorem:

  Gibbard If an aggregation procedure satisfies No Dictator and there are more than two options, then it is **manipulable**: an individual may secure an individually desired preference for the group by choosing their preferences stategically. (Gibbard 1973)

- In this light, what Arrow's theorem shows is that voting – and other collective decision problems – isn't about expressing your sincere preferences and lumping them together with everyone else's sincere preferences.
- It is about playing a **strategy**, in light of your expectations about what other individuals will do, in order to best secure your desired outcomes given the aggregation system in place.
- And that actually turns our attention back to **game theory** – lecture 4.

References

References

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

Gibbard, Allan (1973) 'Manipulation of Voting Schemes: A General Result', *Econometrica* **41**: 587. doi:10.2307/1914083.

Hausman, Daniel, Michael McPherson, and Debra Satz (2017) *Economic Analysis, Moral Philosophy, and Public Policy*, 3rd edition. Cambridge University Press.

List, Christian and Philip Pettit (2002) 'Aggregating Sets of Judgments: An Impossibility Result', *Economics and Philosophy* **18**: 89–110. doi:10.1017/S0266267102001098.

Nozick, Robert (1974) *Anarchy, State and Utopia*. Blackwell.

References (cont.)

Peterson, Martin (2017) *An Introduction to Decision Theory*, 2nd edition. Cambridge University Press.

Pettigrew, Richard (2020b) 'Radical Epistemology'.

Pettigrew, Richard (2020a) *Choosing for Changing Selves*. Oxford University Press.

Sen, Amartya (1970) 'The Impossibility of a Paretian Liberal', *Journal of Political Economy* **78**: 152–57. doi:10.1086/259614.