The Normativity Objection to Sen's 'Escape Route' to Arrow's Impossibility Theorem

Edward Windle

October 2026

Abstract

Sen escapes Arrow's impossibility theorem by enriching the information that Independence of Irrelevant Alternatives operates over. I argue that this doesn't meet Arrow's normative concern that an aggregation rule should translate individual preferences into a sensible collective ranking. Sen's rules cannot tell spurious unanimity from genuine unanimity. I show that this cannot be overcome with welfarism unless we drop IIA as a condition. I take this to amount to a failure of Sen in escaping Arrow's theorem normatively, despite his success in escaping it formally.

How should we aggregate the preferences of the individual citizens of a democracy to understand the preferences they have as a collective? With a social welfare function: a method of transforming a set of individuals' preferences into a single rational group preference ordering.

This hits the nail on the head, but we want this method to have some features. Kenneth Arrow (1951) proposed five innocuous-looking features of reasonable preference aggregation methods, namely Universal Domain, Collective Rationality, Weak Pareto, Independence of Irrelevant Alternatives (IIA), and No Dictator. Here, only Weak Pareto, IIA, and No Dictator shall be considered. Regarding these three, Arrow's theorem shows that, given the first two, any method also satisfying Weak Pareto and IIA is a dictatorship: the preference of a single individual is copied by the social welfare function, irrespective of the preferences of the other individuals. And if we satisfy No Dictator and IIA, then Weak Pareto isn't satisfied. And the same for IIA if Weak Pareto and No Dictator are satisfied. That these three axioms cannot be jointly satisfied is the so-called *Arrow's Impossibility Theorem* (§1)[1].

One family of escape routes drops Arrow's assumption that welfare information is only ordinally measurable and interpersonally non-comparable (see Roberts 1980, d'Aspremont 1985). And in this paper we'll look at Sen's (1977, 2017) version. Sen's strategy to take the edge off Arrow's theorem is to have a finer understanding of the information each person's welfare carries about the options being compared, admitting cardinal and interpersonal comparisons, and keeping IIA's restriction to those options (§2). The result is a class of social welfare functionals that satisfy both Weak Pareto and IIA without violating No Dictator.

I think enriching the information base is the correct move. But in this paper, I'd like to persuade you that Sen's account cannot meet the normative concern of Arrow's conditions that an aggregation rule can translate individual preferences into a sensible collective ranking. For, once we use Sen's welfarism to avoid Arrow's result, Weak Pareto preserves *spurious unanimity* (Mongin 2016): agreement at the welfare layer that masks disagreement at the layer of reasons. If we have spurious unanimity at the aggregate level, then the aggregation rule doesn't meet Arrow's normative concern. And even though welfarism provides the information to acknowledge spurious unanimity, IIA doesn't

[1]This line of argument extends also to combinations including Universal Domain and Collective Rationality also. However, Sen's argument focusses on IIA, and my counterargument focusses on Weak Pareto. I'll leave my justification for only exploring three at that.

allow aggregation rules to consider this (§3). To deal with spurious unanimity and satisfy Arrow's normative concern, we have to violate IIA. Therefore, whilst escaping Arrow formally, Sen fails to escape Arrow normatively. Let's see how.

1 Arrovian Impossibility

We'll start in the standard Arrovian setting. Let N and X be nonempty finite sets of, respectively, *voters* and *options*. We assume $|X| \geq 3$. Each individual $i \in N$ has a complete and transitive preference relation $\preceq_i$ on X. A *preference profile* is a sequence $\langle \preceq_i \rangle = \langle \preceq_1, \preceq_2, \ldots, \preceq_n \rangle$ of these relations. A *social welfare function* F takes a profile $\langle \preceq_i \rangle$ and returns the *aggregate preference* $\preceq_G$ on X, itself complete and transitive.

Let's now meet Arrow's three aforementioned conditions for this function (Gaertner 2009):

- **Weak Pareto**. F satisfies Weak Pareto if for all $x, y \in X$ and profiles $\langle \preceq_i \rangle$, if $y \prec_i x$ for all $i \in N$, then $y \prec_G x$.
- **Independence of Irrelevant Alternatives** (IIA). F satisfies (IIA) if for all $x, y \in X$ and any two profiles $\langle \preceq_i \rangle$, $\langle \preceq'_i \rangle$, if $y \preceq_i x$ iff $y \preceq'_i x$, then $y \preceq_G x$ iff $y \preceq'_G x$.
- **No Dictator** There is no individual k such that, for any $x, y \in X$ and any profile $\langle \preceq_i \rangle$, $y \prec_k x$ implies $y \prec_G x$.

And from these, we get Arrow's Impossibility Theorem:

> *Theorem 2.1* (Arrow 1951, pp.96-7) Assume $|X| \geq 3$. If F is a social welfare function satisfying Weak Pareto and IIA, then there's a dictator for F.[2]

Thus, suppose we are considering the preferences of a group over three political candidates (options): Farage, Polanski, and Starmer. Weak Pareto says that if all voters prefer Polanski to Farage, then so does the group. IIA says that the group's ranking of Farage against Polanski should depend exclusively on these two, not on where anyone places Starmer. No Dictator says that no individual's strict preferences are guaranteed to determine the group's. And to satisfy all three is impossible on this current framework.

Notice a presupposition? The informational constraint for F is exclusively ordinal. There are two sources. First, F only takes preference relations as inputs and encodes only order. Secondly, IIA is a *positive dependence principle* (Pettigrew 2019, p.52) which tells us some feature of the aggregate preference–for us, the ranking of x, y–should only depend on certain features of individual rankings–the individual rankings of x and y . That is, there are two distinct restrictions. First, there's an informational constraint where the inputs render all cardinal and interpersonal information inadmissible such that each person's ordering can be redescribed by any increasing transformation of her own, so no comparison between persons survives. Second, IIA excludes information about options outside the pair being ranked, including the positional information the Borda count uses (Borda 1781). Importantly, Sen's route relies upon lifting the first restriction and keeping the second.

The problem: we want Weak Pareto and IIA without forcing a dictator. So, is IIA culpable because rules such as the Borda count violate it[3]? Or is it the informational poverty of ordinal, non-comparable inputs that's the issue? For welfare economics, non-comparable ordinalism already has a bad rap. And Sen maintains this, blaming the latter. So, Sen (1977)'s proposal aims to enrich the informational base that IIA operates over, rather than just abandon it wholesale. Let's turn to this proposal now.

[2]For a proof, see Sen (2017).

[3]See Holliday & Kelley (2025) for an ordinal account that preserves the what's good about IIA but revises what's wrong with it.

2 Sen's Solution

So, what if we abandon the ordinal preference setting, and instead (i) cardinalise preferences, and (ii) allow for interpersonal comparability–might this dodge Arrow? Let's take these in turn.

Cardinalisation's aim is to aggregate *welfare functions*, which are numerical representations of the welfare of each individual. Sen's tool for doing so is the *social welfare functional* (SWFL): a map Φ that takes a profile $\langle W_i \rangle$ of real-valued personal welfare functions, $W_i : X \to \mathbb{R}$ to a complete and transitive aggregate $\preceq_G$ on X, such that $\preceq_G = \Phi(\langle W_i \rangle)$.

Next, we need to fix the relevant numerical difference of W_i. So Sen introduces an *invariance requirement* for *comparability sets* $\overline{L} \subseteq L$, where L is the family of personal welfare functions, and $\overline{L}$ is the comparability set of welfare functions 'which give the *same* welfare signals interpreted in terms of the chosen assumptions of measurability and interpersonal comparability' (Sen 1977, p.1541). The invariance requirement is: if $\langle W_i \rangle$ and $\langle W_i' \rangle$ are in the same *comparability set* $\overline{L}$, then $\Phi(\langle W_i \rangle) = \Phi(\langle W_i' \rangle)$. A stronger informational assumption yields narrower $\overline{L}$ and weaker invariance. So, the three conditions restated in this context give us:

- **Weak Pareto'**: If $W_i(y) < W_i(x)$ for all i, then $y \prec_G x$.
- **IIA'**: If $W_i(x) = W_i'(x)$ and $W_i(y) = W_i'(y)$ for all i, then $\Phi(\langle W_i \rangle)$ and $\Phi(\langle W_i' \rangle)$ rank x and y the same way.
- **No Dictator'**: There is no k such that, for all x, y, and all profiles $W_k(y) < W_k(x)$ implies $y \prec_G x$

While IIA still only looks at x and y, what varies between Arrow's conditions and Sen's conditions is how much information about x and y the invariance requirement permits for Φ.

What this leaves us with is a taxonomy of several *alternative measurability-comparability frameworks*. I'll list three of interest for us here. For any welfare profile $\langle W_i^* \rangle \in \overline{L}$, $\overline{L}$ consists of exactly all welfare profiles $\langle W_i \rangle$ such that:

1. *Ordinal non-comparability* (ONC): There exists an n-tuple $\langle \Psi_i \rangle$ of positive monotonic transformations with $W_i = \Psi_i(W_i^*)$ for all i.
2. *Ordinal Level Comparability* (OLC). There exists a positive monotonic transformation Ψ with $W_i = \Psi(W_i^*)$ for all i.
3. *Cardinal Unit Comparability* (CUC). There exist a positive real b and an n-vector $\mathbf{a}$ with $W_i = a_i + bW_i^*$ for all i.

ONC is Arrow's framework. As you'll notice, it includes an n-tuple $\langle \Psi_i \rangle$–the comparability set $\overline{L}$ admits different transformations per individual i. And these transformations mean Φ doesn't include interpersonal information. Moreover, recall that IIA restricts the Φ to information of the binary ranking of x, y. This returns us to Arrow's setting and shows that any aggregation rule over exclusively binary ranking generates the impossibility result. All's in order thus far.

What's more, cardinalisation alone doesn't help. Take CNC, whereby there exists an n-vector a and positive reals $b_1, ..., b_n$, with $W_i = a_i + b_i W_i^*$ for all i. Here, the impossibility survives (see Sen 1977). Knowing the size of each person's gains in her own units says nothing about how to weigh them against another individual's cardinal ranking—interpersonal comparability is required to do this.

However, under both OLC and CUC, interpersonal comparison information is admitted. For OLC, $\overline{L}$ admits ordinal level comparability (who's worse off than whom), whilst on CUC it admits unit comparability (whose gain exceeds whose loss). Both admit different kinds of information (OLC admits leximin; CUC admits utilitarianism), but both mean impossibility doesn't result. That is, all three conditions can be jointly satisfied.

It's worth noting what this buys for Sen. Under OLC, statements such as 'person 1 at x is worse off than person 2 at y' are meaningful, so we can use leximin: rank x above y if the worst-off person at x is better off than the worst-off person at y, moving to the next worst off in the case of a tie. This satisfies Weak Pareto. For if everyone is better off at x, then the lowest welfare at x exceeds the lowest welfare at y. It also satisfies IIA by consulting only welfare at x and y. And finally, it satisfies No Dictator, since who is worst off varies from profile to profile, and thus no singular voter decides.

Under CUC, both gains and losses are now comparable across persons, so we can also use utilitarian rules, such as: rank x above y if $\sum_i W_i(y) < \sum_i W_i(x)$. We achieve this because the transformation imposed by a_i changes both sums by the same amount, and the common factor b rescales them. That is, ranking is invariant. Weak Pareto and IIA follow immediately, and No Dictator follows because a voter's mild preference is easily outweighed by another's larger gains.

In short, it is interpersonal comparability that lets us escape Arrow's theorem. Furthermore, we also see that the two leading traditions in distributive ethics are coherent aggregation rules for groups. Arrow's impossibility is thus located in the informational poverty of ONC rather than in the aggregation rules themselves. Nevertheless, Φ's ranking of x and y depends on only the welfare at x and y. Sen enriches the information about the pair; he doesn't let information about other options in. Let's see why Arrow's normative concern survives.

3 The Limits of Enrichment

For Sen's account, everything would be fine if this is where we stopped. But it's not. First, it's worth noting that Sen does successfully escape the theorem by leaving the Arrovian setting of (§1). Yet a deeper normative worry remains (Mongin 2016): Weak Pareto preserves spurious unanimity. And on Sen's account detailed here, he cannot respond to it because of the restriction to x and y his account inherits from IIA. Let's see how.

Weak Pareto is a *unanimity preservation principle*: for some feature a welfare function might have–for instance, the feature of ranking x over y–that, if each individual has a welfare function with such feature, then so does the social ordering $\Phi(\langle W_i \rangle)$ (Pettigrew 2019, p.52). And its goal is to translate individual welfare functions, say, strict pairwise agreement, into the social observer's ranking. That is, if $W_i(y) < W_i(x)$ for every i, then, on aggregate, $y \prec_G x$. Simple and innocuous enough. That is until we consider where Weak Pareto gets its normative force, namely that unanimous preference is based in unanimous reasons (Mongin 2016, p.5), where reasons cover both beliefs and values.

Spurious unanimity surfaces from this normative force. Consider Brexit Referendum as in 2016: a binary decision: Leave the EU or Remain in the EU. Consider also two Brexiteers, Isabelle and Angelo, who both vote Leave over Remain. Assume Isabelle wants P, namely a Brexit for strong workers' protections; Angelo wants to leave because he wants T, namely more trade liberalisation. Both vote Leave. However, their preferred implementations of Leave are clearly incompatible. Moreover, Leave is not a fully specified outcome. It might deliver Isabelle's desired outcome, or Angelo's. Suppose Isabelle and Angelo dread one another's outcome. Let Isabelle's welfare be 10 at P, 6 at Remain, and 0 at T, while Angelo's is 10 at T, 6 at Remain, and 0 at P. Say both are 80% confident that Brexit will deliver their desired outcomes. This gives us 8 for leave, and 6 for remain. So, $W_i(\text{Remain}) < W_i(\text{Leave})$. However, the agreement derives from their disagreement. If they share a credence p that Leave delivers P, then Isabelle would prefer Leave only if $p > 0.6$, and Angelo would prefer it only if $p < 0.4$. And if $p = 0.5$, then they both prefer Remain. That is, Weak Pareto preserves only the higher-level information of the welfare function such that $(\text{Remain}) \prec_G (\text{Leave})$, thus leading to spurious unanimity.

Now this is nothing more than Mongin's claim with Sen's framework. Why can't this framework resolve this, and why can't any further enrichment resolve it? However, the natural response from

the Sen camp is simply to broaden the informational base further. But this doesn't work. Following Baccelli (forthcoming), Pareto combines two fundamentally different ideas. The first is true Unanimity, namely 'requirement that if all agents evaluate all options in the same way, then society, too, should evaluate them in that way'. The second is a special case of IIA, where to socially evaluate x and y, it suffices to know how all individuals evaluate *these two options*, and no other option' (p.13). Pareto and IIA are logically independent, but given Universal Domain, Baccelli's two components together entail Weak Pareto. Take any profile in which everyone prefers x to y. Build a second that agrees with it on x and y, but in which everyone ranks every option in the same order. The independence component says society ranks x and y the same in both profiles. Unanimity says that in the second, society follows the common order puts x above y. So, independence carries Pareto's verdict from genuine to spurious unanimity.

To illustrate, let's return to the Brexit example. Table 1 sets out two profiles, $\mathbf{W} = \langle W_i \rangle$ and $\mathbf{W}' = \langle W_i' \rangle$. The first is the case from above, and in the second, Angelo shares Isabelle's tastes and beliefs (i.e., has her welfare function):

	$\mathbf{W}$ (spurious)		$\mathbf{W}'$ (genuine)	
	Isabelle	Angelo	Isabelle	Angelo
Leave (L)	8	8	8	8
Remain (R)	6	6	6	6
P (workers' protections)	10	0	10	10
T (trade liberalisation)	0	10	0	0

Table 1: Welfare in a spurious and genuine profile

You'll notice that the two profiles are perfectly identical at both Leave and Remain, and differ at P and T.

Now recall that IIA requires that if two profiles give every person the same welfare at x and y, then Φ must rank x and y the same way in both. Everyone has 8 at Leave and 6 at Remain in both $\mathbf{W}$ and $\mathbf{W}'$, so Φ must rank Leave and Remain the same in both. In $\mathbf{W}'$, there's genuine agreement, so Unanimity puts Leave above Remain. In $\mathbf{W}$, their agreement results from their disagreement, whereby each prefers Leave because they expect it to deliver the respective desired outcome. However, no matter the outcome Leave delivers, one of them ends up with 0 welfare, which is lower than the welfare delivered by Remain. So, the mere fact that both prefer Leave is no reason for society to prefer it. Neither of them is being irrational. People choose an option that they believe is likely to deliver a desired outcome, even if it might not deliver that desired outcome. Their beliefs conflict, and that's fine. The issue I am highlighting is social: because IIA forces Φ to follow spurious unanimity in $\mathbf{W}$, Φ gets $\mathbf{W}$ wrong and cannot respond to the difference between genuine and spurious unanimity.

Is the trouble that Φ lacks the information it needs to tell two profiles apart? No. Consider the subjective expected utility (SEU) reading, where their reasons and credences captured by P and T. So, a person's welfare at an uncertain option is her expected welfare. Consider p_i for person i's credence that Leave delivers P. Then we can calculate SEU with the following:

$$W_i(L) = p_i \cdot W_i(P) + (1 - p_i) \cdot W_i(T) \tag{1}$$

And subtracting $W_i(T)$ from both sides gives $W_i(L) - W_i(T) = p_i(W_i(P) - W_i(T))$. So, to calculate their credence:

$$p_i = \frac{W_i(L) - W_i(T)}{W_i(P) - W_i(T)} \tag{2}$$

And we can thus get each person's credence from her welfare at L, P, and T. In $\mathbf{W}$, Isabelle's is $(8-0)/(10-0) = 0.8$, and Angelo's is $\frac{8-10}{0-10} = 0.2$. In $\mathbf{W}'$, both are 0.8. So, the two profiles differ

in virtue of what makes one spurious and the other genuine, namely the input. That is, the relevant information for the difference between spurious and genuine results is actually contained within the welfare measurements.

A natural objection is that this depends on how we scale welfare. But it doesn't. Rescale a person's welfare to $a_i + b_i \cdot W_i$, with $b_i > 0$. Then, the numerator of the credence calculation becomes $(a_i + b_i \cdot W_i(L)) - (a_i + b_i \cdot W_i(T))$, or $b_i(W_i(L) - W_i(T))$. Likewise, the denominator picks up a factor of b_i and the two factors cancel, leaving p_i the same. Therefore, the credences I have shown here are meaningful under Sen's cardinal assumptions[4].

So, we know that Sen's welfare profiles contain everything that makes the unanimity in $\mathbf{W}$ spurious, namely opposing preferences and credences at P and T. Why, then, can't Φ use this information? Simply: IIA. It treats P and T as irrelevant alternatives, even though on the SEU reading, they are the very outcomes Leave might produce. A stronger informational assumption lets Φ make more use of the welfare numbers at Leave and Remain, yet not P and T. And since those numbers are identical in $\mathbf{W}$ and $\mathbf{W}'$, enriching the information base further doesn't provide anything.

Why is this an issue? Recall the normative concern: an aggregation rule rule should be able to translate individual preferences into a sensible collective ranking. I don't think it's sensible to follow an agreement that is guaranteed to give one party the outcome they dread, so a rule that meets the concern should be able to tell spurious unanimity from genuine unanimity. And unless we give up IIA, Sen's cannot. It may well be too demanding to ask every aggregation rule to draw this distinction, but I think it's fair for Sen's. By introducing cardinality, the profile itself contains the information that reveals spurious unanimity.

4 Conclusion

As we saw in §2, Sen clearly mitigates the initial formal worries posed by Arrow's theorem. Once we admit interpersonal comparisons, we can jointly satisfy Weak Pareto, IIA, and No Dictator, and leximin under OLC and utilitarianism under CUC allow us to express distributive ideals. Yet by admitting interpersonal comparisons through welfare judgements, one also admits comparisons between expectations, which can lead to spurious unanimity. Yet the normative concern is unmet. Admitting cardinal welfare information makes each person's credences recoverable from the profile, so everything that reveals spurious unanimity is accessible. Sen, however, cannot use this information unless he violate IIA by admitting information about options other than the two being ranked. That is, the issue is in IIA's restriction to the pair, and not the information.

And I think we can also draw a fairly intuitive conclusion from what I've shown here: the binariness encoded by IIA requires us to abandon welfarism in the face of Arrow's theorem. This leads to the more general conclusion that welfarism, and thus any 'escape route' motivated in the name of welfarism, cannot resolve Arrow's impossibility result.

References

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

Baccelli, J. (forthcoming), 'Spurious unanimity revisited', *Journal of Philosophy* .

Borda, J.-C. d. (1781), 'Mémoire sur les Élections au scrutin', *Histoire de l'Académie Royale des Sciences* pp. 657–664.

[4]This isn't the case with ONC or OLC, since any non-linear transformation can change the fraction

d'Aspremont, C. (1985), Axioms for social welfare orderings, *in* L. Hurwicz, D. Schmeidler & H. Sonnenschein, eds, 'Social Goals and Social Organizations: Essays in Memory of Elisha Pazner', Cambridge University Press, Cambridge, pp. 19–76.

Gaertner, W. (2009), *A Primer in Social Choice Theory: Revised Edition*, Oxford University Press, Oxford, GB.

Holliday, W. & Kelley, M. (2025), 'Escaping arrow?s theorem: The advantage-standard model', *Theory and Decision* **98**(2), 165–204.

Mongin, P. (2016), 'Spurious unanimity and the pareto principle', *Economics and Philosophy* **32**(3), 511–532.

Pettigrew, R. (2019), *Choosing for Changing Selves*, Oxford University Press, Oxford, UK.

Roberts, K. W. S. (1980), 'Interpersonal comparability and social choice theory', *The Review of Economic Studies* **47**(2), 421–439.
URL: *https://ideas.repec.org/a/oup/restud/v47y1980i2p421-439..html*

Sen, A. (1977), 'On weights and measures: Informational constraints in social welfare analysis', *Econometrica* **45**(7), 1539–1572.
URL: *https://onlinelibrary.wiley.com/doi/abs/0012-9682(197710)45:7lt;1539:OWAMICgt;2.0.CO;2-T*

Sen, A. (2017), *Collective Choice and Social Welfare: An Expanded Edition*, Harvard University Press.
URL: *http://www.jstor.org/stable/j.ctv2sp3dqx*