Non-Imposition Non-Imposed: Social Choice without Weak Pareto or Non-Imposition

Edward J. R. Elliott

School of Philosophy, Religion and History of Science
University of Leeds

Abstract

Arrow's General Possibility Theorem for social choice theory makes use of three key constraints: Weak Pareto, Non-Dictatorship, and Binary Independence. Wilson's later version of the theorem drops the Weak Pareto constraint, replacing it with Non-Imposition. In this paper, we show how to have a version of Wilson's theorem without Non-Imposition. Doing so requires a slight—but independently very reasonable—strengthening of the remaining conditions; most notably, we motivate a very natural strengthening of the Non-Dictatorship constraint.

1 Introduction

Arrow's (1963) General Possibility Theorem tells us that, where our goal is to aggregate the preferences of some population of individuals into a social preference ranking, there can exist no general method for doing so that will satisfy a small number of (purportedly) desirable constraints.

Following Arrow's work, a variety of similar results have since been discovered and studied. (See Arrow et al. 2002, 2011, for a thorough overview.) Of particular relevance for this paper is Wilson's (1972) theorem. Wilson drops Arrow's Weak Pareto condition, and replaces it with Non-Imposition.[1]

[1] Many actual voting systems violate Weak Pareto, doing so under the presupposition that they probably won't end up selecting a Pareto-dominated alternative often. (See Mbih et al. 2008 for discussion.) For some philosophical reasons to doubt Weak Pareto, see (Sen 1970; 1976; 1979), (Gilboa et al. 2004), (Mongin 2016), and (Sher 2020). In general, though, I take the perspective that an impossibility result which foregoes Weak Pareto and Non-Imposition will be interesting even if those conditions are reasonable given the standard contexts in which the results are applied. This is in part because it teaches us of the significance of the remaining constraints (see esp. Malawski and Zhou 1994), but also because the Arrovian framework has a huge variety of potential applications and it is not remotely obvious that Weak Pareto and/or Non-Imposition will be plausible for all such applications.

Roughly, Non-Imposition tells us that for any two alternatives x and y, it could be the case that x is socially ranked at least as high as y.

I'll show how we can have a version of Wilson's result which drops Non-Imposition, given some light adjustments to the remaining conditions—most notably, the non-(inverse-)dictatorship conditions. I'll also argue that the strengthening is independently very reasonable, in the sense that essentially the same reasons that would usually support the weaker non-dictatorship conditions can in general be expected to motivate the stronger versions that I require. In a sense, then, I'll be arguing that the original conditions are too weak: it is entirely possible to satisfy the weaker non-dictatorship constraints while still having what basically amounts to a dictator in all but the very strictest possible sense.

2 Arrow's and Wilson's Theorems

We let $\mathbf{X} = \{x, y, \dots\}$ be a set of *alternatives*, and $\mathbf{N} = \{1, 2, \dots, n\}$ a finite set of *individuals*. We say that a binary relation R on $\mathbf{X}$ is a *preference relation* just when it's transitive and connected (xRy or yRx for all x, y in $\mathbf{X}$), and we use I and P to designate the symmetric ('indifference') and asymmetric ('strict preference') parts of R respectively.

We assume that every individual i in $\mathbf{N}$ determines a preference relation over $\mathbf{X}$. We can interpret xR_iy to mean that i takes x to be at least as good as y; xP_iy and xI_iy are interpreted similarly, *mutatis mutandis*. A *profile* is an ordered n-tuple $\langle R_1, R_2, \dots, R_n\rangle$ of preference relations on $\mathbf{X}$. We use π, π' as variables for profiles, and I'll assume throughout that the *profile domain* $\mathbf{\Pi}$ satisfies:

Unrestricted Domain (UD) $\mathbf{\Pi}$ includes all logically possible profiles.

An *aggregation method* is a function $\mathcal{F}$ from profiles in $\mathbf{\Pi}$ to preference relations on $\mathbf{X}$. Given a method $\mathcal{F}$, we'll use R_π to designate $\mathcal{F}(\pi)$. We say that individual i is a *dictator* just in case, for any profile π and alternatives x and y, if xP_iy then $xP_\pi y$; and i is an *inverse dictator* just in case, for any profile π and alternatives x and y, if xP_iy then $yP_\pi x$. Finally, we say that an aggregation method is *null* iff, for any π and all x and y, $xI_\pi y$.

Given all that, where:

Weak Pareto (WP) For any $\pi \in \mathbf{\Pi}$ and all $x, y \in \mathbf{X}$, xP_iy for all $i \in \mathbf{N}$ only if $xP_\pi y$.

Independence of Irrelevant Alternatives (IIA) For any $\pi, \pi' \in \mathbf{P}$ and all $x, y \in \mathbf{X}$: if π and π' restricted to $\{x, y\}$ are identical, then $xR_\pi y$ iff $xR_{\pi'}y$.

Non-Dictatorship (ND) There is no dictator in $\mathbf{N}$.

Non-Inverse-Dictatorship (NID) There is no inverse dictator in $\mathbf{N}$.

Non-Imposition (NI) For all $x, y \in \mathbf{X}$, there exists a $\pi \in \mathbf{\Pi}$ such that $xR_\pi y$.

Non-Nullity (NN) For some profile π and alternatives $x, y \in \mathbf{X}$, $xP_\pi y$.

Then Arrow's result can be expressed as:

Theorem 1. (Arrow 1963) *Given UD, no aggregation method satisfies WP, IIA, and ND.*

While Wilson's is:

Theorem 2. (Wilson 1972) *Given UD, no aggregation method satisfies IIA, ND, NID, NI, and NN.*

What I want to do now is show that Non-Imposition is not really needed, so long as we slightly strengthen ND, NID, and NN.

3 Wilson without Non-Imposition

Given an method $\mathcal{F}$, say that an alternative x is a *winner* just when it sits at the top of the social preference ranking for some profile. (There may be more than one winner.) Say furthermore that x is *hopeful* just when there is some profile under which x gets to be a winner. If x isn't hopeful, then x is *hopeless*.

We say that an aggregation method is *trivial* if being hopeful implies being a winner. In other words, the method is trivial just when the winner (or winners) are always the same regardless of the preferences of the individuals in the population, so in that sense the method is utterly insensitive to variations in the individuals' preferences across profiles. It's clear that a trivial aggregation method is undesirable; thus:

Non-Triviality (NT) There exist profiles π and π' such that, for some $x \in \mathbf{X}$, $xR_\pi y$ for all $y \in \mathbf{X}$ and $yP_{\pi'}x$ for some $y \in \mathbf{X}$.

Obviously, NT is stronger than NN. But not by much, and not in a way that's very interesting. There are two main problems with a null aggregation method: first, it won't allow us to distinguish some alternatives as better or worse than others in the social preference ordering; and second, a null method will be insensitive to the preferences being aggregated. A method which is trivial but non-null won't face the first problem, but it will still face the second problem—and that's bad enough. If there are a large number of hopefuls, then a trivial method still leaves us with far too many equal-best alternatives; if there are only a very few hopefuls, then the method essentially tells us to select the same few winners at every possibility regardless of what the individuals think about those alternatives.

We'll also want to strengthen the Non-Dictatorship condition. But that's ok, because ND is *way* too easy to satisfy. To recall: a dictator is an individual such that, at *all* profiles, and for *any* two alternatives x and y, if they take x to be better than y, then the social preference ranking cannot place y above x. By that definition, an individual—let's call him Kim—wouldn't count as a dictator if, for exactly one profile, and for exactly two alternatives x and y at the very bottom of Kim's personal preference ordering, Kim takes x to be better than y but the social preference ordering places y above x (but still below everything else). For all other alternatives, and for all other profile, Kim will still be in all but the strictest sense a dictator—just not for *those* two alternatives and *that* specific profile. It's not going to be of much comfort to the other individuals in the population to know that Kim isn't a 'dictator' by that definition, because there is exactly one way things might be where they get to disagree with Kim about the overall ranking of the two worst alternatives. So we're going to strengthen ND very slightly.

Say that an individual i is a *dictator over the hopefuls* just in case, whenever x and y are hopefuls, if i takes x to be better than y then the social preference ordering cannot place y above x. Note, of course, that a dictator over the hopefuls gets to decide who can and cannot be a winner: if one exists, then to the extent that the rest of the population gets a say, it's can only be (a) to break ties between hopefuls regarding which the dictator over the hopefuls is indifferent, or (b) to decide the social preference ordering over the hopeless alternatives. The existence of a dictator over the hopefuls is undesirable for essentially the same reasons that a dictator is undesirable, so we'll want to rule them out:

> Strong Non-Dictatorship (SND). There is no dictator over the hopefuls in $\mathbf{N}$.

Being a dictator entails being a dictator over the hopefuls, but not vice versa. Consequently, SND is stronger than ND—but again, not by much, and not in a way that's very interesting.

Finally, we want to strengthen Non-Inverse-Dictatorship in the same way. The ideas here parallel those above: first we define an inverse dictator over the hopefuls, then we rule out the existence of an inverse dictator with the following:

> Strong Non-Inverse-Dictatorship (SNID). There is no inverse dictator over the hopefuls in $\mathbf{N}$.

And now we have everything needed for:

Theorem 3. *Given UD, no aggregation method satisfies IIA, SND, SNID, and NT.*

The proof strategy is straightforward. Suppose first of all that UD holds, and we have a method $\mathcal{F}$ which satisfies IIA, SND and SNID. We want to show that it cannot also satisfy NT. Now, either $\mathcal{F}$ satisfies NI or it doesn't. If $\mathcal{F}$ does satisfy NI, then Wilson's theorem entails that $\mathcal{F}$ cannot satisfy NN, and therefore $\mathcal{F}$ cannot satisfy NT. If $\mathcal{F}$ doesn't satisfy NI, then we want to consider the restriction of $\mathcal{F}$ to the set of hopefuls.

Specifically, given the set of hopefuls $\mathbf{H} \subseteq \mathbf{X}$, we let '$R|_{\mathbf{H}}$' be the restriction of the relation R to the alternatives in $\mathbf{H}$, and we define '$\pi|_{\mathbf{H}}$', '$\mathbf{\Pi}|_{\mathbf{H}}$' similarly. Since $\mathcal{F}$ satisfies IIA, if $\pi|_{\mathbf{H}} = \pi'|_{\mathbf{H}}$ then $R_\pi|_{\mathbf{H}} = R_{\pi'}|_{\mathbf{H}}$. This means that we can without fear of inconsistency let '$\mathcal{F}|_{\mathbf{H}}$' be the aggregation method on $\mathbf{\Pi}|_{\mathbf{H}}$ which agrees exactly with $\mathcal{F}$ as far as the hopefuls are concerned; i.e.,

$$\mathcal{F}|_{\mathbf{H}}(\pi|_{\mathbf{H}}) = R_\pi|_{\mathbf{H}}$$

Trivially, $\mathcal{F}|_{\mathbf{H}}$ satisfies NI and IIA, and if the profile domain is unrestricted for $\mathbf{\Pi}$, then it'll be unrestricted for $\mathbf{\Pi}|_{\mathbf{H}}$. If i is a dictator over the hopefuls relative to $\mathcal{F}$, then i will be a dictator *tout court* relative to $\mathcal{F}|_{\mathbf{H}}$; likewise for inverse dictators. Therefore, by Wilson's theorem, $\mathcal{F}|_{\mathbf{H}}$ must be null. But $\mathcal{F}|_{\mathbf{H}}$ is null just in case for all $\pi|_{\mathbf{H}}$ and any $x, y \in \mathbf{H}$, $xI_\pi|_{\mathbf{H}}y$, and so x and y are both winners at every profile. Or in other words: $\mathcal{F}|_{\mathbf{H}}$ is null just when being a hopeful entails being a winner; so if $\mathcal{F}|_{\mathbf{H}}$ is null then $\mathcal{F}$ is trivial.

4 Conclusion

In sum: given Unrestricted Domain, any aggregation method satisfying Independence of Irrelevant Alternatives, Strong Non-Dictatorship, and Strong Non-Inverse-Dictatorship is therefore trivial. Two points are noteworthy about this result.

First: there has been little discussion on what happens when we strengthen the anti-(inverse-)dictatorship constraints, especially in the way done here. Most of the work done on social choice theory without Weak Pareto, for example, involves either strengthening Independence of Irrelevant Alternatives, or strengthening Non-Imposition. (See Cato 2012 for a comprehensive overview of social choice theory without Weak Pareto.) I've shown that there is value to strengthening the non-dictatorship conditions as well, especially given that such strengthenings are entirely natural given the reasons typically used to motivated those conditions.

Second: while it's well-known that it will often be possible to drop one condition if appropriate additions are made elsewhere, such additions will typically require further motivation—they are not simply 'more of the same'. For example, Deitrich & List's (2007) result requires neither Weak Pareto and Non-Imposition, but it does require stronger constraints on preferences and strengthening Independence of Irrelevant Alternatives with a neutrality condition. But the reasons that support IIA cannot *in general* be expected to also support neutrality; for many contexts the latter will require further argument going substantially beyond the arguments used to support IIA. By contrast, I've argued that the very lightly strengthened anti-dictatorship conditions I require are naturally motivated by *the very same* reasons that would usually motivate the non-strengthened conditions; and similarly for Non-Triviality versus Non-Nullity. Nothing new is required.

So, in the same way that Wilson's theorem establishes that the essential significance of Arrow's theorem is not diminished if one abandons Weak Pareto, so too is the essential significance of Wilson's theorem not diminished if one abandons Non-Imposition.

References

Arrow, K. J. (1963). *Social Choice and Individual Values* (2nd ed.). New York: Wiley.

Arrow, K. J., A. Sen, and K. Suzumura (Eds.) (2002). *Handbook of Social Choice and Welfare, Volume I.* New York: Elsevier.

Arrow, K. J., A. Sen, and K. Suzumura (Eds.) (2011). *Handbook of Social Choice and Welfare, Volume II.* New York: Elsevier.

Cato, S. (2012). Social choice without the Pareto principle: a comprehensive analysis. *Social Choice and Welfare 39*, 869–889.

Dietrich, F. and C. List (2007). Arrow's theorem in judgment aggregation. *Social Choice and Welfare 29*(1), 19–33.

Gilboa, I., D. Samet, and D. Schmeidler (2004). Utilitarian Aggregation of Beliefs and Tastes. *Journal of Political Economy 112*(4), 932–938.

Malawski, M. and L. Zhou (1994). A note on social choice theory without the Pareto principle. *Social Choice and Welfare 11*(2), 103–107.

Mbih, B., I. Moyouwou, and J. Picot (2008). Pareto violations of parliamentary voting systems. *Economic Theory 34*, 331–358.

Mongin, P. (2016). Spurious Unanimity and the Pareto Principle. *Economics and Philosophy 32*(3), 511–532.

Sen, A. (1970). *Collective Choice and Social Welfare.* San Francisco: Holden-Day.

Sen, A. (1976). Liberty, Unanimity and Rights. *Economica 43*, 217–245.

Sen, A. (1979). Utilitarianism and Welfarism. *The Journal of Philosophy 76*, 463–489.

Sher, I. (2020). How perspective-based aggregation undermines the Pareto principle. *Politics, Philosophy & Economics 19*(2), 182–205.

Wilson, R. (1972). Social choice theory without the Pareto principle. *Journal of Economic Theory 5*, 478–486.