Overcoming Arrow's impossibility: Honeybee sidestep

Ben Smyth

University of Birmingham, UK

April 2, 2022

"Equality—impossible," cry social choice theorists. First-past-the-post plagued by vote splitting. Borda dismissing, championing ranked voting. "Bound to lead to error," muses Condorcet. Arrow discovering far more sinister defects. This note reviews their seminal results. Walks through Arrow's impossibility theorem—teaching ranked voting is fallible. And concludes with Smith sidestepping Arrow, guided by dancing honeybees, proclaiming range/score voting to be "a larger improvement in 'democracy' than the entire invention of democracy."

Defective, unsafe at any speed: Plurality voting

Borda (1781) denounced plurality voting systems "seriously defective" (Grazia, 1953) upon discovering *vote splitting*, whereby like-minded voters cast votes for similar candidates, rather than rallying behind a single candidate, creating an advantage for any dissimilar candidate. For example, when two left-leaning candidates run against a right-leaning candidate, the right-leaning candidate has an advantage, because left-leaning voters must pick between two candidates. This is known as the *spoiler effect* when it causes an opponent of similar candidates to win and as a *spoiler* when caused by a minor candidate. Poundstone proclaims "[f]ive presidential elections were probably decided by spoilers...At least two others...are questionable cases. In still another race...four-way vote splitting and the electoral college created such ambiguity that it was a factor in precipitating civil war" (2008, p91). He concludes, "[w]ere the plurality vote a car or an airliner, it would be recognized for what it is—a defective consumer product, unsafe at any speed."

Borda offered a solution: A voting system, now known as *Borda count*, which instructs voters to rank candidates, e.g., $A>B>C$. Tallying involves assigning a descending number of points to each position of each ranking, e.g., a voter's first preference might receive 3 points, the second might receive 2, and the last might receive just 1. The candidate with the most points wins. For instance, given preferences $A>B>C$, $B>A>C$, and $C>B>A$, candidate A receives 6 points, B receives 7, and C receives 5, candidate B wins. Equivalently, voters can rank candidates by putting numbers next to candidate names and the candidate with the least points wins. Henceforth, we'll consider this version of Borda count.

Some voters are more equal: Ranked voting

Condorcet (1785) claims Borda count is "bound to lead to error...because it takes into account elements which should be ignored" (McLean, 2003, p126). Let's consider this claim with an example: Three voters prefer $A>B>C$ and two voters prefer $B>C>A$. Candidate A is awarded 9 points (one point is awarded by each of the three voters that favour candidate A and three points are awarded by each of the remaining two voters) and B is awarded 8 (one point from each of the two voters favouring candidate B and two points from each of the remaining three voters). Candidate B wins. Borda would reason that B appears one place behind A on three ballots and two places ahead of A on two, hence, B is preferred (McLean, 2003, p11). By comparison, Condorcet reasons that candidate A is favoured over candidate B by 3 out of 5 voters, hence, A is preferred. Or, more generally, any winning candidate should beat every other candidate in a two-way race. Such a winner is known as a *Condorcet winner* and a voting system that ensures this criterion is known as a *Condorcet method*.

Condorcet went on to identify a general issue with ranked voting systems, which we consider with another example. Suppose two voters prefer candidate A to candidate B, and only one voter prefers B to A. A fallacy of composition allows one to infer that the collective preference favours candidate A. However, a truth about partial preferences is insufficient to derive a collective preference. Indeed, amongst the two voters that prefer candidate A to B, one may prefer C to both A and B, and the other may prefer C the least. Moreover, the voter that prefers B to A may prefer B the most and A the least. That is, the voters' preferences might be $C>A>B$, $A>B>C$, and $B>C>A$. Hence, the collective preference is intransitive. The majority prefer A to B, B to C, and C to A. Despite transitivity of individual preferences. This is known as the *Condorcet paradox*, which demonstrates ranked voting systems can fail to select a winner.

Black (1948) postulated positioning voter preferences along a linear spectrum: Starting with the voter's favourite, who is closest to the voter's ideology, and continuing through the voter's preferences, becoming increasingly distant from that ideology. Hence, a voter that prefers a left-wing candidate A will always favour a centre-left candidate B over a right-wing candidate C. Moreover, a voter that prefers B, will favour A over C. Similarly, a voter that prefers candidate C, will always favour B over A. It follows that only three of the six possible preferences will arise. This can be verified by positioning candidates A, B and C along a line, and attempting to draw a line between candidates. (Candidate A should be closer to B than B is to C.) It is rational to draw lines through A, B, C; B, A, C; and C, B, A; but illogical to draw lines through A, C, B; B, C, A; and C, A, B. (This is presumably somewhat akin to Black's observation: "I wrote down a single diagram and saw...that the diagram...could be interpreted as referring to a committee using a simple-majority procedure," 1996, pp43.) This suffices to preclude individual preferences that lead to intransitive collective preferences, i.e., the Condorcet paradox cannot occur, and leads to *the median voter theorem*, which asserts majority rule voting systems select the candidate favoured by the median voter. (Black's results are theoretical, rather than practical; "political landscapes...are inherently multidimensional and cannot be reduced to a single left-right dimension, or even to a two-dimensional space," Alós-Ferrer and Granić 2015.)

Having rediscovered the Condorcet paradox and that linear ideologies avoid it (Poundstone, 2008, pp38–39), Arrow discovered far more sinister defects of voting systems (1950).

Arrow's impossibility theorem

Arrow states two fairness conditions (Pareto efficient and independence of irrelevant alternatives) and proves that any ranked voting system (including plurality voting, which simply ignores all but the first-ranked candidate) satisfying those two conditions fails to achieve a third (non-dictatorial), in the context of any mathematical function that inputs a set of voter preferences and outputs the corresponding election result. The consequence: Ranked voting is unfair. (Or we tolerate dictators.) For any ranked voting system that satisfies our seemingly innocuous fairness conditions, a single voter controls the election result!

Theorem. *Given a social welfare function that is Pareto efficient and independent of irrelevant alternatives, the social welfare function is dictorial, assuming at least three candidates.*

The theorem shows there exists scenarios resulting in a dictatorial voter. Let's work through the details. (What follows is rather technical and you may prefer to skip this section, perhaps watching—`https://youtu.be/Q6OZXoXP6Hg`—an explanation of the key ideas instead, maybe watching regardless, as a primer.)

We start from a situation in which all voters rank a candidate last, e.g.,

$$[\succ]_0 = \begin{Bmatrix} a \succ_1 c \succ_1 \mathbf{b} \\ a \succ_2 c \succ_2 \mathbf{b} \\ c \succ_3 a \succ_3 \mathbf{b} \end{Bmatrix}$$

(In matrices we highlight in bold any candidate about whom we make an assumption.) Next, we start flipping voter preferences for that candidate from last to first, whilst preserving all other relative rankings, e.g.,

$$[\succ]_1 = \begin{Bmatrix} \mathbf{b} \succ_1 a \succ_1 c \\ a \succ_2 c \succ_2 \mathbf{b} \\ c \succ_3 a \succ_3 \mathbf{b} \end{Bmatrix} \quad [\succ]_2 = \begin{Bmatrix} \mathbf{b} \succ_1 a \succ_1 c \\ \mathbf{b} \succ_2 a \succ_2 c \\ c \succ_3 a \succ_3 \mathbf{b} \end{Bmatrix} \quad [\succ]_3 = \begin{Bmatrix} \mathbf{b} \succ_1 a \succ_1 c \\ \mathbf{b} \succ_2 a \succ_2 c \\ \mathbf{b} \succ_3 c \succ_3 a \end{Bmatrix}$$

We flip voter preferences until we discover a situation with a pivotal preference such that flipping the preference causes the social welfare function's rank of the candidate to increase from last to first, i.e., the function ranks the candidate last for set of preferences $[\succ]_{i-1}$ and first for $[\succ]_i$. That search cannot be in vain, because (PE) requires the function to rank the candidate last for $[\succ]_0$ and first when all voters rank the candidate first.

A social welfare function W is *Pareto efficient*, if the function ranks a candidate above another candidate when all voters rank the former candidate above the latter:

$$\forall a, b \in O \;.\; \bigwedge_{\succ \in [\succ]} a \succ b \Longrightarrow a \succ_{W([\succ])} b \tag{PE}$$

That is, given a pair of candidates a and b from the set of candidates O and given a set of voter preferences $[\succ]$, such that $a \succ b$ for every voter preference $\succ$ in $[\succ]$, it follows that $a \succ_{W([\succ])} b$.

Pareto efficiency teaches us that a social welfare function must rank a candidate last when voters unanimously do so (e.g., as per $[\succ]_0$), but is insufficient to teach us the social welfare function rank of the candidate when ranked either first or last by voters (e.g., as per $[\succ]_1$ and $[\succ]_2$). Coupling Pareto efficiency with independence of irrelevant alternatives allows us to proceed.

A social welfare function W is *independent of irrelevant alternatives*, if the function computes relative rankings between candidates based only on the voters' relative rankings of those candidates:

$$\forall a, b \in O \;.\; \left(\bigwedge_{\succ \in [\succ]} a \succ b \text{ iff } \bigwedge_{\succ' \in [\succ']} a \succ' b \right) \Longrightarrow \left(a \succ_{W([\succ])} b \text{ iff } a \succ_{W([\succ'])} b \right) \tag{IIA}$$

That is, given a pair of candidates a and b from the set of candidates O and given sets of voter preferences $[\succ]$ and $[\succ']$, such that $a \succ b$ for every voter preference $\succ$ in $[\succ]$ iff $a \succ' b$ for every voter preference $\succ'$ in $[\succ']$, we have $a \succ_{W([\succ])} b$ iff $a \succ_{W([\succ'])} b$.

Independence of irrelevant alternatives teaches us that a social welfare function must rank a candidate above another when voters unanimously prefer that ranking. A voter changing their rank of a further candidate doesn't alter that fact, e.g., if you favour candidate a over b, then your preference for candidate c is irrelevant to the fact that you prefer a to b. Taken together, Pareto efficiency and independence of irrelevant alternatives teach us that a social welfare function must rank a candidate first or last, when voters rank the candidate either first or last, giving way to the following lemma.

Lemma. *Social welfare function W ranks a candidate first or last when every voter preference $\succ \in [\succ]$ ranks the candidate first or last.*

Proof. Suppose every voter ranks a candidate first or last, e.g.,

$$[\succ] = \begin{Bmatrix} \mathbf{b} \succ_1 a \succ_1 c \\ a \succ_2 c \succ_2 \mathbf{b} \\ c \succ_3 a \succ_3 \mathbf{b} \end{Bmatrix}$$

Now suppose to the contrary that the social welfare function ranks the candidate neither first nor last, hence, there exists higher- and lower-ranked candidates, e.g., $a \succ_{W([\succ])} b \succ_{W([\succ])} c$. Let set of voter preferences $[\succ']$ be derived from $[\succ]$ by moving one such lower-ranked candidate immediately above one such higher-ranked candidate in every voter's preference, e.g., by moving candidate c immediately above a in every preference, i.e.,

$$[\succ'] = \begin{Bmatrix} \mathbf{b} \succ_1 \mathbf{c} \succ_1 a \\ \mathbf{c} \succ_2 a \succ_2 \mathbf{b} \\ \mathbf{c} \succ_3 a \succ_3 \mathbf{b} \end{Bmatrix}$$

Only the relative ranking between those two candidates are modified, so by (IIA), the social welfare function ranking must not change, e.g., $a \succ_{W([\succ'])} b \succ_{W([\succ'])} c$, hence, $a \succ_{W([\succ'])} c$ by transitivity. Yet, that contradicts (PE), since all voters rank one of those two candidates higher than the other, e.g., each voter ranks c above a. □

It follows naturally that a pivotal preference can be identified.

Corollary. *Given a social welfare function and a candidate, there exists a set of voter preferences for which the function's rank of the candidate increases from last to first when a single voter preference is modified.*

Proof. By (PE), the social welfare function ranks a candidate last when every voter preference does so. Moreover, from our lemma, it follows that there exists a set of voter preferences such that flipping a single voter's preference for that candidate from last to first increase the social welfare function's ranking of that candidate to first. □

Since a pivotal preference can be identified, we learn that there exist scenarios in which a single voter preference determines a candidate's fate. To demonstrate that the voter controlling that preference is a dictator, we will show they can pick the relative ranking of all other pairs of candidates, that is, the voter's preference coincides with the social choice function's rank.

Having identified a candidate that the social welfare function ranks last for set of voter preferences $[\succ]_{i-1}$ and first for set $[\succ]_i$, we know the function must rank every other candidate higher in the former instance. The function must also rank one of those other candidates higher when we modify a pivotal preference to favour one such candidate, because the relative rankings between the two candidates remain the same. Indeed, one of the candidates is ranked last for voters $1, \dots, i-1$ and first for voters $i+1, \dots, |[\succ]|$, moreover, that candidate is below the other candidate in the modified pivotal preference, hence, by (IIA), the social welfare function must rank the other candidate higher. E.g., suppose we modify $[\succ]_2$ to derive $[\succ]_2^a$ by replacing pivotal preference $b \succ_2 a \succ_2 c$ in $[\succ]_2$ with $a \succ_2 b \succ_2 c$, i.e.,

$$[\succ]_2^a = \begin{Bmatrix} b \succ_1 a \succ_1 c \\ \mathbf{a} \succ_2 \mathbf{b} \succ_2 c \\ c \succ_3 a \succ_3 b \end{Bmatrix}$$

Since the social welfare function ranks a higher than b in $[\succ]_1$, and since the relative rankings between a and b remain the same in both $[\succ]_1$ and $[\succ]_2^a$, the social welfare function must rank a higher than b in $[\succ]_2^a$. Moreover, the function must continue to rank that other candidate higher when we modify all other voter preferences (excluding the aforementioned pivotal preference) by switching that other candidate's rank with the rank of some other candidate, because the relative ranking between our initial two candidates are unchanged. E.g.,

$$[\succ]_2^{\hat{a}} = \begin{Bmatrix} b \succ_1 \mathbf{c} \succ_1 \mathbf{a} \\ \mathbf{a} \succ_2 \mathbf{b} \succ_2 c \\ \mathbf{a} \succ_3 \mathbf{c} \succ_3 b \end{Bmatrix}$$

Similarly, the candidate that ranks first for $[\succ]_i$ must rank higher than a further additional candidate, moreover, the first-ranked candidate must still be favoured after applying both of the aforementioned modifications to $[\succ]_i$, e.g., in $[\succ]_2^{\hat{a}}$. It follows by transitivity that the first additional candidate must rank higher than the second additional candidate.

Our reasoning is rather contrived—we've orchestrated a scenario in which the voter controlling the pivotal preference controls the fate of a pair of candidates. Let's eliminate the scaffolding, generalise to show the controlling voter is a dictator: Suppose we arbitrarily modify ranking of b in $[\succ]_2^{\hat{a}}$ for all preferences, and by arbitrarily modifying the rank of a in the pivotal preference under the constraint that a remains higher than c. E.g.,

$$[\succ]_2^{\bar{a}} = \begin{Bmatrix} c \succ_1 b \succ_1 a \\ b \succ_2 \mathbf{a} \succ_2 \mathbf{c} \\ a \succ_3 c \succ_3 b \end{Bmatrix}$$

It follows that voter preferences are all arbitrary, except the pivotal preference which ranks a higher than c, hence, the voter controlling the pivotal preference controls the fate of that pair of candidates. By (IIA), we can further generalise to all pairs, i.e., the voter controlling the pivotal preference is a dictator.

> A social welfare function W is *dictatorial*, if the function computes the result from a single voter's preference:
>
> $$\exists \succ \in [\succ] \,.\, \forall a, b \in O \,.\, a \succ b \Longrightarrow a \succ_{W([\succ])} b \qquad \text{(D)}$$
>
> That is, there exists a set of voter prefences $[\succ]$ containing voter preference $\succ$, such that for each pair of candidates a and b from the set of candidates O, if $a \succ b$, then $a \succ_{W([\succ])} b$.

Lemma. *The voter that modifies their preference in our corollary—to increase the social welfare function's rank of a candidate from last to first—is a dictator over all other pairs of candidates.*

Proof. Let $[\succ]_{i-1}$ be the set of voter preferences that exists by our corollary, and let b be the candidate whose rank increases from last to first. Suppose

$x, y \in O \setminus \{b\}$, e.g., when $x = a$ and $y = c$. Let $[\succ]_i^x$ be derived from $[\succ]_i$ by modifying a pivotal preference to rank x first, whilst preserving all other relative rankings, e.g., $[\succ]_2^a$. Moreover, let $[\succ]_i^{\hat{x}}$ be derived from $[\succ]_i^x$ by modifying the relative rankings of x and y for all preferences except the pivotal preference, whilst preserving b in its extremal position, e.g., $[\succ]_2^{\hat{a}}$. From our corollary, we know the social welfare function ranks b last, hence, $x \succ_{W([\succ]_{i-1})} b$ and, by (IIA), we have $x \succ_{W([\succ]_i^{\hat{x}})} b$, because the relative rankings between x and b are the same for all voters. Indeed, b is ranked first for voters $1, \ldots, i-1$ and last for voters $i+1, \ldots, |[\succ]|$, moreover, b is ranked last in the unmodified pivotal preference and below x after modification. Similarly, we also know $b \succ_{W([\succ]_i)} y$ and, by (IIA), we have $b \succ_{W([\succ]_i^{\hat{x}})} y$, because relative rankings between b and y are the same in $[\succ]_i$ and $[\succ]_i^{\hat{x}}$. It follows by transitivity that $x \succ_{W([\succ]_i^{\hat{x}})} y$.

Let $[\succ]_i^{\bar{x}}$ be derived from $[\succ]_i^{\hat{x}}$ by arbitrarily modifying the rank of b for all preferences, and by arbitrarily modifying the rank of x in the pivotal preference under the constraint that x remains higher than y, e.g., $[\succ]_2^{\bar{a}}$. We started from the set of voter preferences $[\succ]_i$, which only made assumptions on the position of candidate b; we derived $[\succ]_i^x$ by ranking x first in the pivotal preference and $[\succ]_i^{\hat{x}}$ by modifying the relative rankings of x and y for all preferences except the pivotal preference; finally, we derived $[\succ]_2^{\bar{x}}$ by arbitrarily modifying the rank of b for all preferences, and by arbitrarily modifying the rank of x in the pivotal preference under the constraint that x remains higher than y. It follows that voter preferences are all arbitrary, except the pivotal preference which ranks x higher than y, e.g., a beats c.

We have $x \succ_{W([\succ]_i^{\hat{x}})} y$ and, since the relative rankings between x and y are unchanged between $[\succ]_i^{\hat{x}}$ and $[\succ]_i^{\bar{x}}$, it follows by (IIA) that $x \succ_{W([\succ]_i^{\bar{x}})} y$. Yet, we have only assumed that x is ranked higher than y in the pivotal preference. Thus, the voter that modifies their preference in the above corollary is a dictator over x and y, and for that matter all pairs of candidates that exclude b. □

Proof of Arrow's impossibility Theorem. By our corollary, given a candidate z, there exists a set of voter preferences for which the social welfare function's rank of candidate z increases from last to first when a single voter's preference is modified. It follows by the above lemma that the single voter is a dictator over all other pairs of candidates $x, y \in O \setminus \{z\}$. □

Our proof is inspired by Kevin Leyton-Brown's excellent lecture: `https://youtu.be/QLi_5LCwJ20`.

Arrow's impossibility theorem teaches us that ranked voting is fallible, an ideal ranked-voting system simply cannot exist; there is a possibility that a single voter may yield dictorial control over the result. That's not to say such scenarios are probably, only that such scenarios can exist.

Range voting: Hot or Not?

I was (unknowingly) inducted to *Hot or Not* by my bestie, at uni, c. 2003—she wanted to establish her superior attractiveness. Receiving an almost two-decade later reminder, from Poundstone, was unexpected! *Hot or Not* (aka range voting), wherein voters score a candidate (or photo) on a scale (e.g., ten for hot, one for not), provides a means around Arrow's impossibility result (which only applies to ranked voting systems, rather than rating-based systems such as Hot or Not). Vote splitting and spoilers are avoided; similar candidates can be scored the same: "Arrow's Nobel-winning 1951 'impossibility theorem' misdirected the entire field of voting systems for 50 years," bashed Smith (2007), "Arrow's theorem is not nearly as important as it at first seems" (2005). Having previously argued the superiority of range voting over ranked voting, on the basis that range voting creates less "human unhappiness" than ranked voting (2000; c. 2008), Smith concludes "switching to range voting would be a larger improvement in 'democracy' than the entire invention of democracy" (2004)—Churchill, we have a solution, range voting, which perhaps pre-dates the human race, possibly originating from the nesting behaviour of honeybees (Smith, 2007).

References

Carlos Alós-Ferrer and Đura-Georg Granić. Political space representations with approval data. *Electoral Studies*, 39:56–71, 2015.

Kenneth Joseph Arrow. A Difficulty in the Concept of Social Welfare. *Journal of Political Economy*, 58(4):328–346, 1950.

Duncan Black. On the Rationale of Group Decision-Making. *Journal of Political Economy*, 56(1):23–32, 1948.

Duncan Black. Some notes on the development and structure of the theory of committees. In Gordon L Brady and Gordon Tullock, editors, *Formal Contributions to the Theory of Public Choice: The Unpublished Works of Duncan Black*, chapter 5. Springer, 1996.

Jean-Charles de Borda. Mémoire sur les Élections au Scrutin. *Mémoires de l'Académie Royale des Sciences*, pages 657–665, 1781.

Nicolas de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* Cambridge University Press, 2014. First published in 1785.

Alfred de Grazia. Mathematical Derivation of an Election System. *Isis*, 44: 42–51, 1953. Translation of Borda (1781).

Iain McLean. The Reasonableness of Independence: A Conversation from Condorcet and Borda to the present day. Technical Report 2003-W6, Nuffield College, University of Oxford, March 2003.

William Poundstone. *Gaming the vote: Why elections aren't fair (and what we can do about it)*. Macmillan, 2008.

Warren D. Smith. Range Voting, 2000.

Warren D. Smith. Smith's replies to attempts to criticize Range Voting, 2004.

Warren D. Smith. The voting impossibilities of Arrow, Gibbard & Satterthwaite, and Young, 2005.

Warren D. Smith. Ants, Bees, and Computers agree Range Voting is best single-winner system, 2007.

Warren D. Smith. Bayesian regret for dummies, c. 2008.