Arrow's Impossibility Theorem

Alexander Tabarrok

Department of Economics

George Mason University

Tabarrok@gmu.edu

March 4, 2015

1. Arrow's Impossibility Theorem

In the previous chapter we gave many examples which showed that common voting systems have surprising or paradoxical properties. Examples, however, can only take us so far. We have examined only a handful out of an infinite number of possible voting systems. The systems we looked at may be unusual or perhaps they are typical but there exists nevertheless systems which are paradox-free. To arrive at general conclusions we need a more general method. In 1951 Kenneth

Arrow applied the axiomatic method to the problems of voting theory. A voting system can be thought of as a black box into which individual preference orderings are inputted and a social preference ordering is outputed. It's sometimes useful to substitute the term *social choice mechanism* for voting system because Arrow's theorem concerns any 'mechanism' into which individual preferences are fed and out of which comes a 'social preference.' The market, for example, is a social choice mechanism. Individual preferences are the input and we can interpret the market equilibrium, the output, as a sort of social preference. Figure 1.1 illustrates the main ideas.

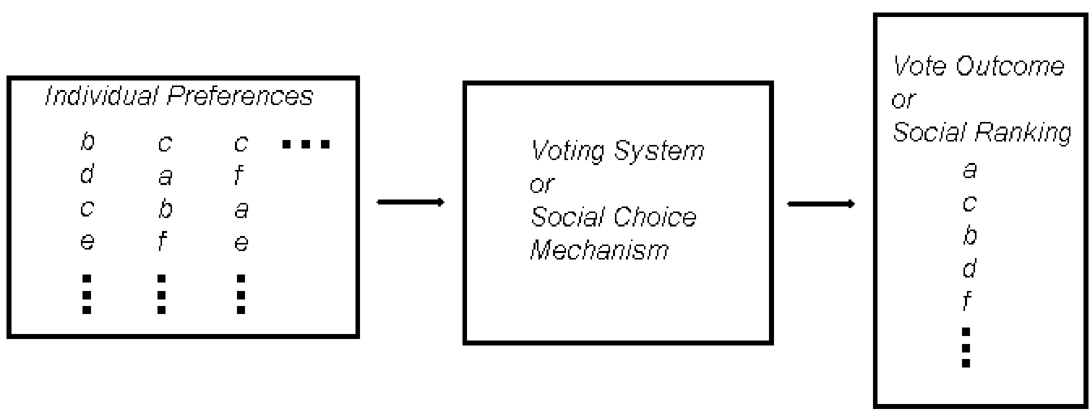

Figure 1.1: A voting system or social choice mechanism aggregates individual preference orderings into a social preference ordering.

We assume in the discussion that follows that the individual preferences in figure 1.1 are complete and transitive, or in other words that each of the individuals in our society is rational. We can't expect groups to have rational preferences when individuals have irrational preferences! We also assume that there are three or more choices to be voted upon. If there are only two choices to be made the scope for voting paradoxes is greatly narrowed and in fact in this simple situation groups using majority rule will act as if they had rational preferences.[1]

Arrow argued that any good voting system should possess certain desirable properties. Arrow's desirable properties or axioms come in three types. There are axioms which restrict the inputs to a voting system, axioms which restrict the outputs and axioms which put restrictions on how the voting system transforms inputs into ouputs. Arrow's theorem says that no voting system can ever possesses all of the properties he deemed desirable.

1.1. Arrow's Axioms

The first of Arrow's axioms is a restriction on the inputs.

1) *Universal Domain*: All individually rational preference orderings are allowed as inputs into the voting system.

A voting system should be able to transform any set of individual preferences

[1] By only two choices we mean that there only two choices in total. Pairwise voting A v B, winner v C is subject to paradoxes as we saw in the previous chapter.

into a social preference ordering. A voting system which works only when individuals are unanimous, for example, is not much of a voting system. The universal domain assumption says we can't beg the question by assuming that all individuals have preferences of a certain type.

The second axiom is a restriction on the output of the voting system.

2) *Completeness and Transitivity*: The derived social preference ordering should be complete and transitive.

The completeness axiom requires that whatever the input, the voting system returns a definite output. In other words given any question of the form 'Is X socially preferred to Y or is Y socially preferred to X or are X and Y socially indifferent?' the voting system must return a definite answer. The transitivity axiom says that the answers the voting system returns must be consistent. A voting system which returns a 'does not compute' message is not very useful. But neither is a voting system which returns $X \succ Y$, $Y \succ Z$, and $Z \succ X$. We want a voting system to aggregate preferences in a way which well help us make social choices. When completeness fails the voting system doesn't answer our questions. When transitivity fails the voting system answers our questions ambiguously.

Arrow's axioms are normative which means that we will accept them only if we beleive that a voting system *should* have certain properties. The completeness axiom, for example, is valuable only if we beleive that all questions of the form 'Is X socially preferable to Y....' *should* have answers. But suppose that X is the outcome, "tax Peter to pay Paul," and Y the outcome "tax Paul to pay Peter." A libertarian would argue that the question 'Is X socially preferable to Y' has no answer (Rothbard 1956). In an ideal libertarian society the only legitimate exchanges are between individuals who agree to those exchanges. A 'voting system' for such a society is nothing more than the market.[2] We can interpret Figure 1.1 as a group of individuals taking their 'preferences to market', trading, and arriving at outcome B (with no other outcomes listed). B is thus the socially preferred choice. The libertarian believes that the only meaning that 'X is socially preferred to Y 'can have is 'X was arrived at by voluntary exchange from Y'. In the libertarian view, the fact that non-voluntary exchanges cannot be ranked is not a fault of the market as a social choice mechanism it is rather an expression of the fact that *there is no social preference ordering between non-voluntary exchanges*. Whether we accept the completeness axiom depends on our values.

If we abandon the completeness axiom there are perfectly sound social choice

[2]Recall that by voting system we mean any method of aggregating individual preferences to create a social preference ordering.

mechanisms. An example of a social choice mechanism which non-controversially satisfies all the axioms except completeness is the Pareto rule. The Pareto rule says that if someone prefers X to Y and no one prefers Y to X then socially $X \succ Y$. The Pareto rule is incomplete because it can't rank order X and Y when some people prefer X to Y and others prefer Y to X.[3]

The four remaining axioms all restrict the ways in which individual preferences are transformed into social preferences.

3) Positive Association: Suppose that at some point the voting rule outputs the social preference $X \succ Y$, then it should continue to output $X \succ Y$ when some individuals raise X in their preference orderings.

Positive association requires that individual preference orderings and social preference orderings be positively connected. If someone raises their ranking of X and no one reduces their ranking of X then it seems entirely reasonable that this should never cause the social ranking to change from $X \succ Y$ to $X \prec Y$.[4]

4) Independence of Irrelevant Alternatives: The social ranking of X and Y

[3] We might say that when neither $X \succ Y$ nor $Y \succ X$ that X is socially indifferent to Y but doing so will lead to intransitivities involving the indifference relation. It is quite possible, for example, that $X \succ Y$ and $Y \sim Z$ but X is not preferred to Z. The Pareto rule, however, does satisfy quasi-transitivity which we discuss further below.

[4] Positive association does not require that X increase in social ranking when it increases in some individual's ranking. Suppose, for example, that the voting system is majority rule and X beats Y by 7 to 3 votes. If one individual raises X in his ranking it is appropriate that X still beats Y.

should depend only on how individuals rank X and Y (and not on how individuals rank some 'irrelevant alternative' W relative to X and Y).

Independent of irrelevant alternatives (IIA) is the most subtle and controversial of Arrow's axioms because it has two implications depending on whether the alternative W is part of the choice set or not. Suppose first that voters must choose between X, Y and W and that when they do so the social ranking indicates $X \succ Y$. Now let some individuals raise W in their preference rankings *without changing the ranking of* X *relative to* Y. An individual, for example, might change his ranking from $\begin{pmatrix} Y \\ X \\ W \end{pmatrix}$ to $\begin{pmatrix} W \\ Y \\ X \end{pmatrix}$ or $\begin{pmatrix} Y \\ W \\ X \end{pmatrix}$. IIA says that this change in individual preferences cannot change the social ranking of $X \succ Y$ (it might of course change the social ranking of W and X or W and Y).[5] The second implication of IIA occurs when the choice set changes. Assume for example that voter's must choose between X ,Y and W and that when they do so the social ranking indicates $X \succ Y$. Now assume that W is dropped from the choice set. IIA says that the social ranking of X, Y must continue to have $X \succ Y$.[6]

[5] In a sense IIA is quite similar to positive association (PA). PA says loosely that certain changes in individual preference orderings must be positively associated with changes in social orderings. IIA says that certain changes in individual preference orderings must never be associated with changes in social orderings.

[6] Arrow (1951) caused a great deal of confusion by mathematically defining IIA so that it

The IIA requirement has two substantive effects. First, IIA implies that a voting system can only respond to ordinal information about preferences. Depending on one's point of view this may be a reason for accepting or rejecting IIA. Suppose an individual changes his ranking from $\begin{pmatrix} Y \\ X \\ W \end{pmatrix}$ to $\begin{pmatrix} Y \\ W \\ X \end{pmatrix}$. We might interpret this change in ranking as indicating that W increased in value or that X decreased in value. Under the latter interpretation, it seems natural to say that the individual's preference for Y over X is more intense when $Y \succ W \succ X$ than when $Y \succ X \succ W$. (Taking this one step further we might say that an individual with the ranking $Y \succ W \succ Z \succ Q \succ X$ prefers Y to X *very* much more than someone with the ranking $Y \succ X \succ W \succ Z \succ Q$.) The ranking of W relative to X and Y, thus provides information about the *intensity* of the X, Y ranking. It would be quite reasonable, given this interpretation, if the social preference changes from $X \succ Y$ to $Y \succ X$ when preferences change from $\begin{pmatrix} Y \\ X \\ W \end{pmatrix}$ to $\begin{pmatrix} Y \\ W \\ X \end{pmatrix}$, because the latter ranking indicates a more intense preference for Y relative to X.

covered only the first implication but illustrating the meaning of IIA with an example of the latter implication. Different authors focus on different implications of IIA without indicating that both implications are covered. Feldman (1980) is one author who focuses on implication 1 while Mueller (1989) focuses on implication 2.

If the relative ranking of W provides information about the intensity of X, Y preferences, then IIA should be dropped because under IIA $Y \succ X \succ W$ means exactly the same thing as $Y \succ W \succ X$ (when determining the social ranking of X v. Y.) Defenders of IIA argue that the relative ranking of W does *not* provide any information about the intensity of preference. Earlier we noted that the change in ranking could be interpreted as a fall in the value of X or an increase in the value of W. Which of these interpretations we make seems arbitrary but under the latter interpretation there is no increase in the intensity of preference! If W increases in value it would be absurd on this account to raise Y (relative to X) in the social ranking.

If the only inputs to the voting system are *ordinal rankings* it is impossible to distinguish between X falling and W rising. Instead of providing rankings we could ask voters to assign utility numbers to their choices in which case we could tell when X fell in value and when W increased in value. The difficulty with this procedure is that voter's would have little incentive to tell the truth about their rankings. As Arrow once put it, "A man sufficiently intense about being greedy would get everything." (Cite). Moreover, if it is difficult to measure an individual's intensity of preference it is near impossible to compare the intensities of preference of two different individuals. If Jones has ranking $Y \succ X \succ W$ and

Smith has ranking $Y \succ W \succ X$ we can't logically claim that Smith prefers Y to X more than Jones does. Perhaps Smith is nearly indifferent between Y, W and X while Jones greatly prefers Y to either of X or W. This problem only gets worse if we ask voters to assign utility numbers to choices. If I assign the number 1563 to X and you assign the number 287 does this mean that I prefer X more than you do? If we beleive that these problems are insurmountable then perhaps we should impose IIA (but see below).

The second defense of IIA focuses on its interpretation when the choice set changes. It seems paradoxical and somehow wrong that when choosing among X, Y, and W a voting system crowns X as the winner yet when choosing among the pair (X, Y), Y wins.[7] How can X be superior to Y when W is available yet inferior when W drops out? We would like to have a voting system where the social ranking of X and Y is decided by the relative merits of X and Y and not by whether some other irrelevant choice is available or not. IIA ensures that the rankings of pairs is always consistent with the rankings of triples.[8]

[7] These absurd outcomes can occur in practice. It's quite possible, for example, that George Bush would have won the 1992 US Presidential election had Ross Perot not been on the ballot. If B stands for Bush, C for Clinton and P for Perot then the B, C, P ranking had $C \succ B \succ P$ but the C, B ranking might have had $B \succ C$.

[8] One reason choosing X from (X, Y, W) and Y from (X, Y) seems paradoxical is that such choices are inconsistent with strong utilitarianism. Strong utilitarianism says that there is a set of true utility numbers which exist in the minds of voters. One argument for a point-score voting system is that the points represent these true utilities, if only approximately. Now when

The last two of Arrow's axioms are straightforward and are used mainly to avoid voting systems which are in some sense 'trivial.'

5) Non-Imposition: An outcome is not to be imposed which is independent of voter preferences.

6) Non-Dictatorship: The voting rule cannot be based solely on one person's preferences.

An example of an imposed outcome is $X \succ Y$ regardless of voter preferences. If there were enough impositions we could always find a voting system which would satisfy all the other axioms but it would be trivial and not worth discussing. Similarly, the voting system where I always get my way regardless of other people's preferences would satisfy all the other axioms but most people would consider it trivial although I am willing to discuss such a system.

W drops out of the choice set the true utility numbers assigned to X and Y *do not change* (X, Y,and W and independent). If the social ranking of X, Y changes with a change in choice set this must mean that our *method of measuring* preference intensity has changed. But why should the measuring stick change when the choice set changes? If we are correctly measuring utilities when the choice set is (X, Y, W) and X is chosen then we cannot be measuring utilities correctly when Y is chosen among (X, Y) . Justifying point-score voting systems using utilitarian arguments is therefore very difficult if not impossible.

2. Arrow's Impossibility Theorem

Arrow's impossibility theorem says that the six axioms, *1) Universal Domain, 2) Completeness and Transitivity, 3) Positive Association, 4) Independence of Irrelevant Alternatives, 5) Non-Imposition, and 6) Non-Dictatorship* are inconsistent.[9] Inconsistency of the axioms means that all six axioms can never be true at the same time. If any five axioms are true then the sixth axiom must be false. If a voting system satisfies, for example, universal domain, completeness and transitivity, positive associaton, IIA and non-imposition then it must be a dictatorship.

It is worthwhile to review the voting systems we examined in the previous chapter. None of these voting systems was dictatorial or imposed so they each must violate at least one and perhaps several of Arrow's other axioms. Pairwise voting with majority rule violates the Transitivity axiom (ie. majority rule can create intransitive group preferences). Positive Association is violated by runoff procedures. Positional vote systems like plurality rule, the Borda count violate the Independence of Irrelevant Alternatives axiom.

[9] Arrow orginally called his theorem the Possibility theorem but the literature has for the most part adopted the more descriptive impossibility term.

2.1. Escape from Arrow's Theorem?

Arrow's Theorem tells us we can't have everything we desire in a voting system - something must be given up. We certainly don't want to give up the Non-Imposition and Non-Dictatorship axioms. Of the remaining four axioms, Positive Association seems the most desirable one to maintain. Position Association and Non-Imposition together imply the weak Pareto principle which says that if every individual prefers X to Y then the social ranking must have $X \succ Y$.[10] The weak Pareto principle seems very desirable so Positive Association should remain. We are left with rejecting at least one of Universal Domian, Completeness, Transitivity, or the Independence of Irrelevant Alternatives axiom.

Suppose that we give up Universal Domain (UD). Giving up UD is the same as looking for a voting system which will work well for some but not all distributions of individual preference rankings. If everyone has identical preferences, for example, then majority rule is a perfectly acceptable voting system (ie. it will satisfy the remaining axioms). But a voting system which works well only when everyone has identical preferences is not very useful. We are thus interested in knowing how much homogenity we need to impose on preference orderings if we

[10]Many presentations of Arrow's theorem replace Positive Association and Non-Imposition with the weak Pareto principle. I stick to the older formulation to easier connect the paradoxes in chapter X with the theorem axioms.

want a voting system which satisfies the remaining 5 axioms. Realistically the answer is that quite a lot of homogeneity is required but perhaps not so much to be uninteresting. If everyone's preferences are single peaked on the same single dimension then majority rule satisfies the remaining 5 axioms. We explain and take up this restriction further in the next chapter on the median voter theorem.

Voting systems like the Pareto rule satisfy all the axioms but completeness. At the current time, however, most people are not willing to restrict democracy to the subset of issues which could be decided by these principles. Most authors therefore take completeness to be a necessary requirement of any voting system.

We could drop transitivity in which case majority rule is an adequate voting system. Majority rule is indeed what democracies use for a wide variety of decisions. It seems a shame, however, that we can't do better. Majority rule can lead to vote cycles and in an actual decision process it can easily violate Pareto optimality, as we showed in the previous chapter. Majority rule satisfies the weak Pareto principle in the sense that if everyone prefers X to Y *and we have a vote* between X and Y then X will win. Actual voting processes, however, do not guarantee that every alternative is matched up against every other alternative. Majority rule as actually used, therefore, can violate the weak Pareto principle and this is a strong mark against majority rule as a voting system.

Instead of dropping Transitivity altogether we could weaken it to Quasi-Transitivity (QT). Recall the definition of transitivity is that if $X \succeq Y$ and $Y \succeq Z$ then $X \succeq Z$. Quasi-Transitivity says that if $X \succ Y$ and $Y \succ Z$ then $X \succ Z$. Unlike transitivity, quasi-transitivity is compatible with $X \sim Y$ and $Y \sim Z$ but $X \succ Z$. If we replace Transitivity with Quasi-Transitivity then the Pareto rule discussed earlier satisfies all the other axioms. In particular, with QT the Pareto rule is a complete social choice mechanism - but this is not a substantive improvement. A voting rule which says society is indifferent between all Pareto optimal positions is hardly better and perhaps worse (because less honest) than a voting rule which can't decide between Pareto optimal positions. There are other rules which satisfy QT and the remaining axioms but these all have a particularly bad property, they result in oligarchies. Allan Gibbard (1969) showed that any social choice mechanism which satisfies QT and the remaining axioms produces an oligarchy - where an oligarachy is defined as a group of individuals each of whom can veto any outcome and who when united can determine the social outcome. (The Pareto rule satisfies QT and is an extreme example of Gibbard's theorem. Under the Pareto rule *any* individual can veto an outcome and when *all* individuals act together they determine the social outcome - thus in the case of the Pareto rule the oligarchy is all of society.)

Why quasi-transitivity should lead to oligarchy is not at all obvious. It is easier to see, however, why oligarchy necessitates quasi-transitivity. In figure 2.1 we plot A's utility on the Y axis and B's utility on the X axis. Both A and B increase their utility levels when society moves from Z to X so they vote accordingly and $X \succ Z$ socially. In a choice of Y vs Z individual A will veto Z so that $Y \succeq Z$ (Veto power lets A force Z to be socially not preferred to Y but does not give A the power to make Y preferred to Z, thus $Y \succeq Z$ means Y is at least as good as Z.) Individual B will veto Y in the choice of Z vs Y so that $Z \succeq Y$. But if Y is at least as good as Z and Z is at least as good as Y then it must be the case that 'society' is indifferent between Y and Z written $Y \sim Z$.[11] Similarly, individual A will veto a move from Y to X and B will veto the opposite move so $X \sim Y$. We thus have $X \succ Z$, and $Z \sim Y$ but nevertheless $X \sim Y$ rather than $X \succ Y$ which would be required by transitivity.

Notice also that above analysis explains why the group of individuals with veto power has dictatorial powers when they act together. A single member of the oligarchy can force $Y \succeq Z$ and *only* another member can force $Z \succeq Y$, together creating $Z \sim Y$. Thus, if the members of the oligarchy all act together so that $Y \succeq Z$ and *not* $Z \succeq Y$, it must be the case that $Y \succ Z$. In other words, once a

[11] If $Z \succeq Y$ and $Y \succeq Z$ it follows that $Y \sim Z$ where $\sim$ is read "is indifferent to."

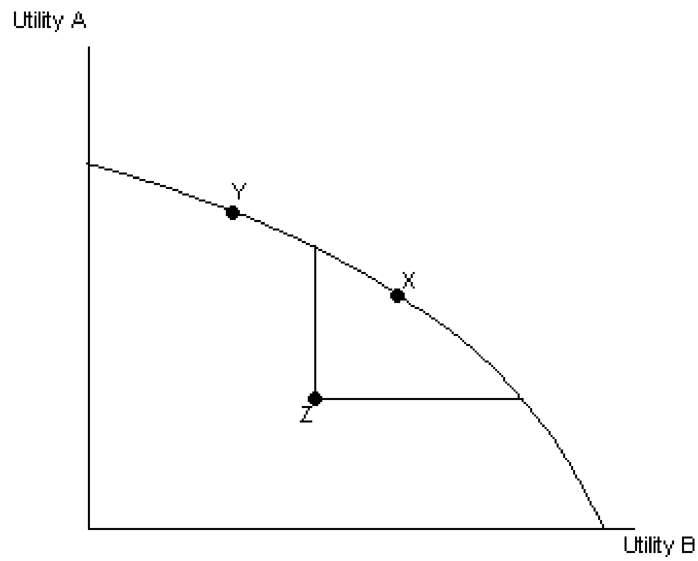

Figure 2.1: The Pareto Rule Implies Quasi-Transitive Preferences: Although $X \succ Z$ and $Z \sim Y$, $X \sim Y$ which violates transitivity but not quasi-transitivity.

veto is in place there is a presumption that $Y \succ Z$ and the only thing which can neutralize that presumption is another veto in the opposite direction

Further weakenings of transitivity are possible and these weaken dictatorship even more than oligarchy does but the spectre of group rule always remains. Weakening transitivity does not appear to be a plausible method of escape from Arrow's Theorem.

If we eliminate IIA we must face squarely the fact that our voting system will be making judgements about relative intensities of preference, both for a given individual and between individuals. The Borda Count (BC) assigns $m-1$ points to a top ranked choice, $m-2$ points to a second ranked choice and so forth down to 0 points for a least favoured choice (where m is the number of candidates). The BC implicitly assumes that the difference in utility between an nth ranked candidate and an $(n-1)th$ ranked candidate is the same as the difference between an $(n+t)th$ ranked candidate and $(n+t-1)th$ ranked candidate (where t is any number). Moreover, each voter is implicitly assumed to have the same utility differences! These assumptions seem extreme and also arbitrary. Why not argue that the difference between a voter's first choice and a voter's second choice is the *truly* critical difference and therefore lend support to a point system like $10, 4, 3$ or $100, 12, 2$? We might take refuge in the principle of insufficient reason which suggests that in a situation of ignorance an assumption of equal measures is the best. The principle could be used to defend both the constancy of utility differences and the assumed equality of intensity of differences across individuals. Unfortunately, the principle of insufficient reason is subject to serious reservations and even if we were accept the principle it seems a week reed on which to defend the BC.

There is an alternative defense of the Borda Count. Let us accept as a lost cause the attempt to measure intensities of preference and return our attention to the simplest vote, that between two choices, X and Y. In this situation, majority rule has a strong claim to the title of best voting system. Majority rule satisifes all of Arrow's axioms and without any information about intensities of preference it's difficult to justify a higher voting standard such as a two-thirds rule. The difficulty with majority rule is that with three or more choices it fails transitivity and so returns ambiguous answers to questions of social preference. Ideally, we would like a voting sytem to be consistent with pairwise voting and at the same time result in transitive rankings over 3 or more choices. (Recall that the second justification of the independence of irrelevant alternatives condition was to impose consistency of the vote system with the pairwise votes). The Arrow theorem tells us that this is impossible - we cannot have transitivity and consistency with the pairwise votes if we maintain Arrow's other axioms. We can, however, try to find that voting system which is *most* consistent with majority rule over pairwise choices. The voting system which is most consistent with majority rule over pairwise votes is the Borda Count (Saari 1994). Proving this result would take us too far afield but we will give some intuition for the result by illustrating the intimate connection between pairwise voting and the Borda Count.

Consider another voting procedure which Saari (1994) calls the aggregated version of pairwise voting. With three choices there are three possible pairwise votes XvY, XvZ, and YvZ. The aggregated pairwise vote adds up the votes on each pairing and then uses the total to define a social preference. Suppose X beats Y by 10 to 5, X beats Z by 8 to 7 and Z beats Y by 14 to 1. The aggregate point allocations are then $X = 18$ $(10 + 8)$, $Y = 6$ $(5 + 1)$ and $Z = 21$ $(7 + 14)$. The social preference is $Z \succ X \succ Y$. The aggregated pairwise vote appears to be a natural way of extending pairwise voting. Moreover, since the aggregate is derived directly from the pairwise votes it is evident that the aggregate will preserve many of the pairwise relationships. Amazingly, the aggregate pairwise procedure is identical to the Borda Count! A way of seeing the identity is to consider how a voter with preference $X \succ Y \succ Z$ contributes to the aggregated vote tally - this is illustrated in Table 2.1.

Table : Vote Contributions from Voter with Preferences $X \succ Y \succ Z$

	Vote Contributions		
Pairwise Vote	X	Y	Z
XvY	1	0	0
XvZ	1	0	0
YvZ	0	1	0
Sum	2	1	0

Notice that using the aggregate pairwise vote a voter with preferences $X \succ$

$Y \succ Z$ contributes 2 votes to his top ranked candidate X, 1 vote to his second ranked candidate Y and 0 votes to the last ranked candidate Z. But this is exactly the vote scoring system used by the Borda Count. Going through the same calculations for the other possible rankings we conclude that the aggregate pairwise vote system and the Borda Count are identical. Since the Borda Count can be understood as a natural extension of pairwise voting it's not surprising that the BC and pairwise voting should be relatively consistent with one another. If we value IIA because we want votes over triples to be consistent with votes over pairs then the Borda Count best supports that value.

3. Conclusions

4.