A Modified Version of Arrow's IIA Condition

E. Maskin

Harvard University

This paper was prepared for a special issue of *Social Choice and Welfare* in memory of Kenneth Arrow. I draw on material published in my "Foreword" to the third (2012) edition of Arrow's classic monograph *Social Choice and Individual Value*s and on my *New York Times* essay with Amartya Sen "How Majority Rule Might Have Stopped Donald Trump." I thank Amartya Sen and Kotaro Suzumura for very helpful comments on a previous version of the paper.

0. Introduction

In this short paper I argue that Arrow's independence of irrelevant alternatives condition (IIA) has considerable appeal because it rules out the phenomenon of vote-splitting in elections. However, it is open to criticism for failing to take into account useful information about the intensities of voters' preferences. I therefore propose a modified version of IIA that prevents vote-splitting while allowing election outcomes to reflect intensities.

I begin, in section I, with some personal memories of Ken Arrow. Section II lays out the axioms behind Arrow's Impossibility Theorem, in particular IIA. Section III explains the weakness in IIA that I mention above, while section IV points out IIA's strength in stopping vote-splitting. Section V concludes by presenting the modified definition of IIA.

I. Personal Reminiscences

I first encountered Kenneth Arrow at Harvard when I was an undergraduate there in the early 1970s. Although I was a math major, my advisor suggested that I take Ken's graduate course on information economics. The course was a mixture of some cutting-edge subjects he was thinking about then – e.g., mechanism design, adverse selection, communication in organizations, and Shannon's information theory – and turned out to be life-changing for me (it made me switch direction and do a PhD with Ken). Still, I can't say that Ken – despite his interest in the theory of organizations – was an organized teacher. His lectures had a somewhat jumbled quality - - perhaps because he decided what to talk about on his way over to the classroom (and sometimes in the classroom itself). On one rare occasion, he actually planned a lecture in advance – a technical presentation of the Gibbard-Satterthwaite Theorem (not yet published at the time) – but then forgot to bring his notes in. It was fascinating watching him work out a detailed proof of the theorem in real time.

One feature that made the course so engaging was Ken's ability to pack so much material into one lecture. This was partly because he talked exceptionally fast. But even his quick tongue was no match for his mind. So, his sentences were full of ellipses -- he would leave out the last few words so that he could begin the next thought. We had to stay on our toes to fill in the gaps.

We students were amazed by the extent of Ken's knowledge, which went way beyond economic matters. He would be talking about some aspect of asymmetric information when an appropriate quotation from Shakespeare or an analogy from physics would occur to him. Ken seemed to know more on nearly any subject than any of us. There was a story that a group of junior faculty once concocted a plan by which they could finally appear to outshine their erudite senior colleague. They read up on the most arcane topic they could think of: the breeding habits of gray whales. On the appointed day they gathered in the coffee room and waited for Ken to come in. Then they started talking about the elaborate theory of a marine biologist named Turner on how gray whales find their way back to the same breeding spot year after year. Ken was silent … they had him at last! With a delicious sense of triumph they continued to discuss Turner, while Ken looked increasingly perplexed. Finally, he couldn't hold back: "But I though Turner's Theory was discredited by Spenser, who showed that the purported homing mechanisms couldn't possibly work."

II. The Impossibility Theorem

The publication of the Impossibility Theorem – first, in a journal article (Arrow 1950) and then, in expanded form, in the celebrated monograph *Social Choice And Individual Values* (Arrow 1951, 1963, 2012) – was a landmark in twentieth-century social thought. Arrow laid out a general framework for translating individuals' preferences into social choices and then showed, disconcertingly, that such a translation cannot work as well as we would want.

Arrow begins with a set of *social alternatives* X from which a society of n individuals must make a choice (Depending on the context, X might consist of the possible candidates for some political office; the different public projects that the society could undertake; or the alternative ways that stocks of natural resources might be allocated to different individuals. There is an unlimited variety of other possible applications). Each individual i $(i = 1, \ldots, n)$ has a *preference ordering* $\succsim_i$ over X (the expression $x \succsim_i y$ means that i finds x to be at least as good as y), where $\succsim_i$ lies in a set of possible orderings $\Re_i$. A *social welfare function* (SWF) F is mapping

$$F : \Re_1 \times \cdots \times \Re_n \to \Re,$$

where $\Re$ is also a set of orderings. Thus, a SWF F assigns a social ranking of alternatives $\succsim_s (= F(\succsim_1, \ldots, \succsim_n))$ to each possible profile of individuals' orderings $(\succsim_1, \ldots, \succsim_n)$; it determines how society should choose among X on the basis of its members' preferences.[1]

Arrow imposed four conditions on SWFs:

Unrestricted Domain (UD): The SWF must determine a social ranking for all possible preference orderings that individuals might have. Formally, for all $i, \Re_i$ consists of *all* logically possible orderings of X.

[1] Arrow insisted, for a pragmatic reason, that a SWF should determine a social *ranking* of alternatives rather than merely the social *choice* of an optimal alternative. He imagined that which alternatives in X would turn out to be feasible might not be known in advance, and so a social ranking serves as a contingency plan: if the top-ranked alternative is not available, choose the next alternative, and so on.

Pareto Property (P): If all individuals strictly prefer *x* to *y*, then *x* must be strictly socially preferred. Formally, for all $(\succsim_1,\ldots,\succsim_n)\in\Re_1\times\cdots\times\Re_n$ and for all $x,y\in X$, if, for all *i*, $x\succ_i y$ [2] then $x\succ_s y$, where $\succsim_s = F(\succsim_1,\ldots,\succsim_n)$.

Independence of Irrelevant Alternatives (IIA): Social preferences between *x* and *y* should depend only on individuals' preferences between *x* and *y*, and not on their preferences concerning some third alternative. Formally, for all $(\succsim_1,\ldots,\succsim_n),(\succsim_1',\ldots,\succsim_n')\in\Re_1\times\cdots\times\Re_n$ and all $x,y\in X$, if for all i, $\succsim_i$ ranks *x* and *y* in the same way that $\succsim_i'$ does, then $\succsim_s$ ranks *x* and *y* in the same way that $\succsim_s'$ does, where $\succsim_s = F(\succsim_1,\ldots,\succsim_n)$ and $\succsim_s' = F(\succsim_1',\ldots,\succsim_n')$

and

Nondictatorship (ND): There exists no individual who always gets his way in the sense that, if he prefers *x* to *y*, then *x* must be socially preferred to *y*, regardless of others' preferences. Formally, there does *not* exist i^* such that, for all $(\succsim_1,\ldots,\succsim_n)\in\Re_1\times\cdots\times\Re_n$ and all $x,y\in X$, if $x\succ_{i^*} y$ then $x\succ_s y$, where $\succsim_s = F(\succsim_1,\ldots,\succsim_n)$.

We can now state:

Impossibility Theorem: If *X* contains at least three alternatives, there exists no SWF satisfying UD, P, IIA, and ND.

III. A Shortcoming of IIA

Of the four axioms assumed in the Impossibility Theorem, IIA has been by far the most controversial. Quarreling with the other three seems difficult: UD requires simply that social preferences always be defined; P rules out the perverse possibility that everyone prefers *x* to *y*

[2] $x\succ_i y$ means that individual *i* strictly prefers *x* to *y*, i.e., $x\succsim_i y$ but $y\not\succsim_i x$.

and yet y is chosen over x socially; ND is the weak requirement that social preferences should not be determined by a single all-powerful individual.

However, IIA is not as obviously compelling as the others, and indeed is violated by an especially well-regarded SWF, rank-order voting (the Borda count). For this SWF, an alternative gets m points (assuming there are m alternatives in all) for every individual who ranks it first, $m-1$ points for a second-place ranking, etc. Alternatives are then socially ranked according to their total point score. To see that rank-order voting fails to satisfy IIA, see Example 1, in which there are three candidates x, y, z and two groups of voters, one (45% of the electorate) with preferences $x \succ z \succ y$ and the other (55%) with preferences $y \succ x \succ z$. Calculating the point totals, we see that the social ranking is $x \succ_s y \succ_s z.$ However, if the first group's preferences are replaced by $x \succ y \succ z,$ the social ranking now becomes $y \succ_s x \succ_s z$. This violates IIA because, in making the replacement, we have not changed any individual's ranking of x and y, and yet the social ranking changes from $x \succ_s y$ to $y \succ_s x.$

45%	**55%**	
x	*y*	
z	*x*	
y	*z*	

Under rank-order voting,

x gets 3×45 + 2×55 = 245 points

y gets 3×55 + 1×45 = 210 points

z gets 2×45 + 1×55 = 145 points

So, the social ranking is $\begin{array}{c} x \\ y \\ z \end{array}$

Now, consider

45%	**55%**
x	*y*
y	*x*
z	*z*

Under rank-order voting, the social ranking is now $\begin{array}{c} y \\ x \\ z \end{array}$, a violation of IIA as applied to *x* and *y*.

Example 1

The standard rationale for IIA (see Arrow 2012, p.26) is that in making a choice between candidates *x* and *y*, society need take account only of how individuals rank *x* and *y*; how they feel about candidate *z* is irrelevant. But a Borda-count proponent might retort: The position of candidate *z* in Example 1 may provide useful information about the intensity of group 1 voters' preferences between *x* and *y*. In the first case, *z* lies between *x* and *y* – suggesting that the gap between *x* and *y* may be fairly large. In the second case, *z* lies below both *x* and *y*, and so perhaps the difference between *x* and *y* is not so large. Thus, even if *z* is not a serious candidate himself, how individuals rank him vis à vis *x* and *y* might be pertinent.

Still, despite this shortcoming, IIA embodies a feature with considerable appeal, as I discuss in the next section.

IV. IIA and Vote Splitting

For Donald Trump to have been elected president of the United States in 2016, he first had to secure the Republican nomination. He did this by winning a succession of Republican primaries in which the non-Trump vote was substantially larger that the pro-Trump vote, but split among various other candidates. Example 2 (modified from Maskin and Sen 2016) provides a stylized illustration of this. There are three candidates – Donald Trump, Ted Cruz, and John Kasich – and three groups of voters. In the top scenario, one group (40%) has ranking Trump $\succ$ Kasich $\succ$ Cruz; the second (35%) has ranking Cruz $\succ$ Kasich $\succ$ Trump; and the third (25%) Kasich $\succ$ Trump $\succ$ Cruz. Most Republican primaries used *plurality rule* (the candidate who is top-ranked by the most voters wins), and under this voting rule Trump is the winner, with 40% of the vote. But, in fact, a strong majority of voters (groups 2 and 3) prefer Kasich to Trump, and so electing Trump seems quite anti-democratic (Note that there is nothing to suggest that group 1's preference for Trump over Kasich is especially strong; the Borda proponent's argument from Example 1 doesn't apply).

Observe that this outcome is possible only because, under plurality rule, Cruz and Kasich *split* the anti-Trump vote. To make the concept of vulnerability to vote splitting clear, let me suppose that there are two situations. In one (see, for instance, the bottom scenario of Example 2), there is a group of individuals (the middle group in Example 2) who have the ranking Kasich $\succ$ Trump $\succ$ Cruz, and Kasich is ranked above Trump socially. In the other (see the top scenario of Example 2), all individuals' preferences are the same as before except that now the group in question has ranking Cruz $\succ$ Kasich $\succ$ Trump and the social ranking is Trump $\succ$ Kasich. That is, moving a candidate (Cruz, in this case) up to the top of the group's preferences causes him to siphon off support for another candidate (Kasich here), allowing a third candidate (Trump) to win.

40%	**35%**	**25%**	
Trump	Cruz	Kasich	Trump is the plurality (and runoff) winner
Kasich	Kasich	Trump	
Cruz	Trump	Cruz	

40%	**35%**	**25%**	
Trump	Kasich	Kasich	Kasich is the plurality (and runoff) winner
Kasich	Trump	Trump	
Cruz	Cruz	Cruz	

Example 2

Example 2 also illustrates that *runoff voting*[3] (as practiced in French, Brazilian, and Russian presidential elections) is vulnerable to vote-splitting as well. Observe that Kasich wins outright in the bottom scenario. But, in the top scenario, Trump and Cruz advance to the runoff, where Trump wins (and this discrepancy holds even though all individuals rank Trump and Kasich in the same way in both scenarios).

As I already observed, the "Borda defense" that can be made in favor of the social ranking in Example 1 does not apply to Example 2. In Example 1, z lies *between* x and y in group 1's preferences in the top scenario (implicitly giving us information about the intensity of the

[3] In runoff voting, there are two rounds. Each individual votes for one alternative in the first round, and if some alternative gets a majority, it ranked first socially. If no alternative gets a majority, the top two vote-getters face each other in a runoff round, and the majority winner there is ranked first socially.

preference for *x* over *y*). But in Example 2, Cruz is always above or below both Kasich and Trump - - never in between.

Example 2 makes apparent that vulnerability to vote-splitting leads to the "wrong" candidate winning. IIA rules out such vulnerability but does so in too a heavy-handed a way - - it also rules out the Borda count.

V. Modified IIA

Thus, I suggest that the following form of IIA preserves its strength while jettisoning its weakness:

Modified IIA (MIIA): For all $(\succsim_1,\ldots,\succsim_n)$, $(\succsim_1',\ldots,\succsim_n') \in \Re_1 \times \cdots \times \Re_n$ and all $x,\ y \in X$, if for all i, $\succsim_i$ ranks *x* and *y* in the same way that $\succsim_i'$ does and, for all $z \in X - \{x, y\}$, *z* lies between *x* and *y* in $\succsim_i$ if and only if *z* lies between *x* and *y* in $\succsim_i'$, then $\succsim_s$ ranks *x* and *y* in the same way that $\succsim_s'$ does, where $\succsim_s = F(\succsim_1,\ldots,\succsim_n)$ and $\succsim_s' = F(\succsim_1',\ldots,\succsim_n')$.

It is evident that plurality rule and runoff voting violate MIIA, but that the Borda count satisfies it. Indeed, in Maskin (Forthcoming), I show that if we impose (i) MIIA, (ii) UD (unrestricted domain), (iii) the axioms that May (1952) invoked to characterize majority rule in the case of two alternatives [i.e., anonymity (voters are treated symmetrically), neutrality (candidates are treated symmetrically), and positive association (if *x* improves relative to *y* in some individual's preference ordering, then *x* improves relative to *y* in the social ordering)], and (iv) the requirement that votes be counted linearly, then the Borda count is the *unique* social welfare function to satisfy them all.[4]

[4] I also show that the Borda count continues to be uniquely characterized if we replace MIIA and positive association with a modified version of strategy-proofness (strategy-proofness is analyzed in Gibbard 1973 and Satterthwaite 1975).

References

Arrow K.J. 1950. A Difficulty in the Concept of Social Welfare. *J. Political Econ.* 58(4):328-46.

Arrow K.J. 1951, 1963, 2012. *Social Choice and Individual Values*. 1951, New York: Wiley. 1st edition; 1963, New Haven: Yale Univ. Press. 2nd edition; 2012, New Haven: Yale Univ. Press. 3rd edition.

Gibbard A. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica.* 41(4):587-601.

Maskin, E. Forthcoming. The Borda Count Reconsidered.

Maskin E and Sen A. 2016. How Majority Rule Might Have Stopped Donald Trump. New York Times, 28 April; available at https://www.nytimes.com/2016/05/01/opinion/sunday/how-majority-rule-might-have-stopped-donald-trump.html

May, K. O. 1952. A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decisions. *Econometrica.* 20:680-684.

Satterthwaite M. 1975. Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory*. 10(2): 187-217.