

Arrow's Impossibility Theorem and Experimental Tests of Preference Aggregation

Todd Davies
Symbolic Systems Program
Stanford University

joint work with Raja Shah, Renee Trochet, and
Katarina Ling

Decision Behavior: Theory and Evidence
(Symsys 170/270)

May 26, 2010

Questions

1. Does intuitive preference aggregation robustly violate Arrow's criteria of IIA and collective rationality?
2. Can we characterize intuitive preference aggregation judgments in a descriptive framework?
3. What is the relationship between intuitive preference aggregation and normative approaches to social choice?

DEFINITION: **Preference relations** R , P , and I respectively exist between two outcomes whenever the outcomes can be matched with labels x and y such that:

- $x R y$, meaning x is preferred or indifferent to y (**weak preference**).
- $x P y$, meaning x is preferred to y , and holds iff not $y R x$. (**strict preference**).
- $x I y$, meaning x is indifferent to y , and holds iff $x R y$ and $y R x$ (**indifference**).

DEFINITION: A preference relation R is **weakly ordered** iff for all outcomes x , y , and z : in a set of outcomes X :

- $x R y$ or $y R x$ (**completeness**).
- $x R y$ and $y R z$ imply $x R z$ (**transitivity**).

DEFINITION: A **social welfare function** is a function f such that if

- X is the set of all conceivable social outcomes or states ($|X| \geq 3$),
- N is a set of individuals ($2 \leq |N| = n$), and
- D is a domain of preference profiles $R^N = \langle R_1, R_2, \dots, R_n \rangle$, such that
- for all i in N , R_i is a weakly ordered preference relation for individual i over X (**individual rationality**),

then f maps D into a range R of possible social preference relations R_N on X for group N .

DEFINITION: A social welfare function f is an **Arrow social welfare function** if

- the domain D of f contains all possible preference profiles R^N (**universal domain**),
- and
- the range R of f is the set of weakly ordered social preference relations R_N (**collective rationality**).

THEOREM: Arrow's Impossibility Theorem

(Arrow, 1951/1963). There can be no Arrow social welfare function f satisfying all of the following:

- for all social outcomes x and y in X , if a preference profile \mathbf{R}^N obeys $x P_i y$ for each individual i in N , then $f(\mathbf{R}^N)$ yields $x R_N y$ (**weak Pareto efficiency**);
- for all preference profiles \mathbf{R}^N and $\mathbf{R}^{N'}$ in \mathbf{D} , and all social outcomes x and y in X , if \mathbf{R}^N and $\mathbf{R}^{N'}$ obey $x R_i y$ iff $x R_i' y$ for all individuals i in N , then $f(\mathbf{R}^N)$ and $f(\mathbf{R}^{N'})$ yield $x R_N y$ iff $x R_{N'} y$ (**independence of irrelevant alternatives**);

and

- there is no individual d in N such that for all preference profiles \mathbf{R}^N in \mathbf{D} , and all social outcomes x and y in X , if $x P_d y$ then $f(\mathbf{R}^N)$ yields $x P_N y$ (**non-dictatorship**).

Proof. See Arrow (1963) and Mas-Colell, Whinston and Green (1995).

Arrow on IIA

“The essential argument in favor of this principle is its direct appeal to intuition.” (1952)

“Stricter than desirable” but has practical benefits, necessary to keep the lid on the need to gather limitless information on unavailable options (1967)

Dead candidate example:

If an election were held and one of the candidates then died: "Surely the social choice should be made by taking each of the individual's preference lists, blotting out the dead candidate's name, and considering only the orderings of the remaining names..." (1951/1963).

DEFINITION: A **social choice rule** C maps the domain \mathbf{D} of preference profiles, together with environments $S \subseteq X$, into subsets of S .

SIDE NOTE: The above definition does not assume that C obeys an ordering. Arrow did assume this, however:

DEFINITION: A social choice rule C is an **Arrow social choice function** if it is determined by an Arrow social welfare function f and for all environments $S \subseteq X$, $C(S, f(\mathbf{R}^N)) = \{x: x \text{ is in } S \text{ and, for all } y \text{ in } S, f(\mathbf{R}^N) \text{ yields } x R_N y\}$.

DEFINITION: Given a set X of social outcomes, a preference profile \mathbf{R}^N , and a social choice rule C , $x R_N^* y$ (meaning there is a **revealed social preference** for x in relation to y) iff there is some environment $S \subseteq X$, such that x and y are in S and x is in $C(S, \mathbf{R}^N)$.

DEFINITION: **Weak axiom of revealed social preference.** Given a set X of social outcomes, a preference profile \mathbf{R}^N , and a social choice rule C , a social preference relation R_N satisfies **inter-menu consistency**¹ (**IMC**) iff it is a revealed social preference relation under C for all pairs of social outcomes in X , and the following condition holds: For all environments $S \subseteq X$, if x and y are in S and x is in $C(S, \mathbf{R}^N)$, then for all environments $S' \subseteq X$ such that x and y are in S' , if y is in $C(S', \mathbf{R}^N)$, then x is in $C(S', \mathbf{R}^N)$.

THEOREM: If R_N^* is a revealed social preference relation associated with a set X of social outcomes, a preference profile \mathbf{R}^N , and a social choice rule C , and R_N^* violates inter-menu consistency, then it violates collective rationality.

PROOF: Suppose not. Then there is a social welfare function f such that $f(\mathbf{R}^N) = R_N^*$ and f is collectively rational. By hypothesis, R_N^* violates inter-menu consistency, so there

¹ This term is from Sen (1993).

exist $x, y, S,$ and S' such that x is in $C(S, \mathbf{R}^N),$ y is in $C(S', \mathbf{R}^N),$ x is not in $C(S', \mathbf{R}^N).$ Thus, by the definition of a revealed social preference relation: $x R_N^* y, y R_N^* x,$ and there is a z in S' such that not $x R_N^* z.$ So $z P_N^* x$ by the definition of strict preference. But since y is in $C(S', \mathbf{R}^N), y R_N^* z.$ By the definition of collective rationality, R_N^* is weakly ordered. Applying transitivity and the definition of strict preference: $y R_N^* z, z P_N^* x,$ and $x R_N^* y$ imply $y P_N^* y,$ implying not $y R_N^* y,$ which violates the completeness condition for $x = y.$

DEFINITION: A social choice rule $C,$ defined on a set X of social outcomes and a domain \mathbf{D} of preference profiles satisfies **independence of unavailable alternatives (IUA)** iff for all environments $S \subseteq X,$ and all preference profiles \mathbf{R}^N and $\mathbf{R}^{N'}$ in $\mathbf{D},$ the following condition holds: If for all x, y in S, \mathbf{R}^N and $\mathbf{R}^{N'}$ obey $x R_i y$ iff $x R_i' y$ for all individuals i in $N,$ then $C(S, \mathbf{R}^N) = C(S, \mathbf{R}^{N'}).$

Figure 1 – Profile triplets (Rank-form presentation)

In each triplet, picking X in top profile and Y in middle profile violates IMC. Picking Y in middle profile and X in bottom profile violates IIA. Triplets differ in extent of “Borda reversal” between profiles (number of X-Y rank shifts)

5-2		Voter 1	Voter 2	Voter 3	Voter 4	Voter 5
Ranked Alternatives		X	X	X	Y	Y
		Y	Y	Y	X	X
5-3		Voter 1	Voter 2	Voter 3	Voter 4	Voter 5
Ranked Alternatives		Z	X	X	Y	Y
		X	Y	Y	Z	Z
		Y	Z	Z	X	X
5-3'		Voter 1	Voter 2	Voter 3	Voter 4	Voter 5
Ranked Alternatives		X	X	X	Z	Y
		Z	Z	Y	Y	X
		Y	Y	Z	X	Z
4-3		Voter 1	Voter 2	Voter 3	Voter 4	
Ranked Alternatives		X	X	Y	Y	
		Y	Z	X	X	
		Z	Y	Z	Z	
4-6		Voter 1	Voter 2	Voter 3	Voter 4	
Ranked Alternatives		U	X	Y	Y	
		X	Z	W	V	
		Y	Y	V	U	
		Z	U	U	W	
		V	V	X	X	
		W	W	Z	Z	
4-6'		Voter 1	Voter 2	Voter 3	Voter 4	
Ranked Alternatives		U	X	Y	Y	
		X	Z	X	X	
		Z	U	W	V	
		V	V	V	U	
		W	W	U	W	
		Y	Y	Z	Z	

3-2

	Voter 1	Voter 2	Voter 3
Ranked Alternatives	X	X	Y
	Y	Y	X

3-4

	Voter 1	Voter 2	Voter 3
Ranked Alternatives	X	X	Y
	Y	Y	Z
	Z	W	W
	W	Z	X

3-4'

	Voter 1	Voter 2	Voter 3
Ranked Alternatives	X	X	Z
	Z	W	Y
	W	Y	X
	Y	Z	W

7-2

	Voter 1	Voter 2	Voter 3	Voter 4	Voter 5	Voter 6	Voter 7
Ranked Alternatives	X	X	X	X	Y	Y	Y
	Y	Y	Y	Y	X	X	X

7-3

	Voter 1	Voter 2	Voter 3	Voter 4	Voter 5	Voter 6	Voter 7
Ranked Alternatives	Z	Z	X	X	Y	Y	Y
	X	X	Y	Y	Z	Z	Z
	Y	Y	Z	Z	X	X	X

7-3'

	Voter 1	Voter 2	Voter 3	Voter 4	Voter 5	Voter 6	Voter 7
Ranked Alternatives	X	X	X	X	Y	Y	Y
	Z	Z	Z	Z	X	X	X
	Y	Y	Y	Y	Z	Z	Z

Experimental Variations

Between Subjects

Each survey taker aggregates only one profile in a given triplet, is asked “Which alternative should be chosen for the group?”

Within Subjects

Each survey taker’s aggregations are tested for adherence to IIA and/or collective rationality

Ordering Task

Each survey taker asked to rank order the alternatives for the group instead of choosing just one

Pairwise Format

Profile is presented to survey taker as a matrix of individuals’ pairwise preferences instead of in the rank format shown above

Restricted Availability

Independence of unavailable alternatives (IUA). Survey taker is shown profile but told that only options S and T are available.

Indifference Option

Survey taker is told they s/he can circle more than one option for a group of voters if both options are equally desirable given the profile.

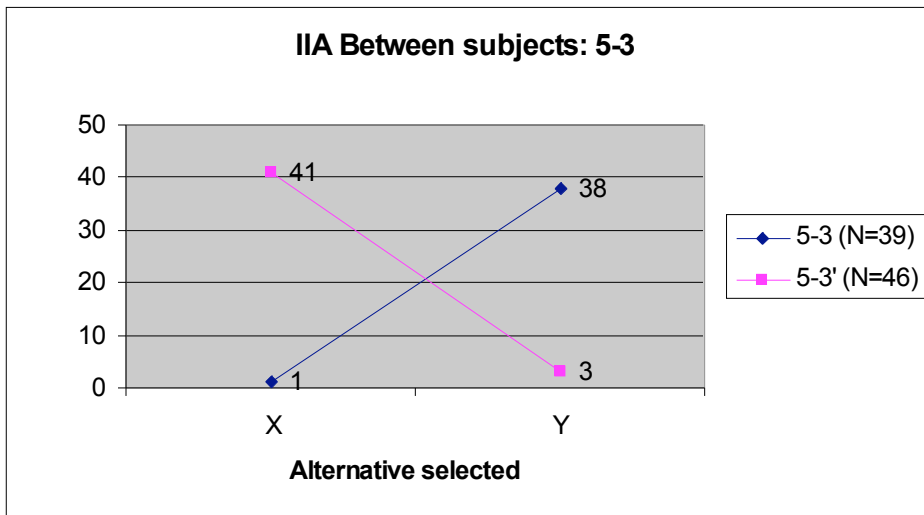
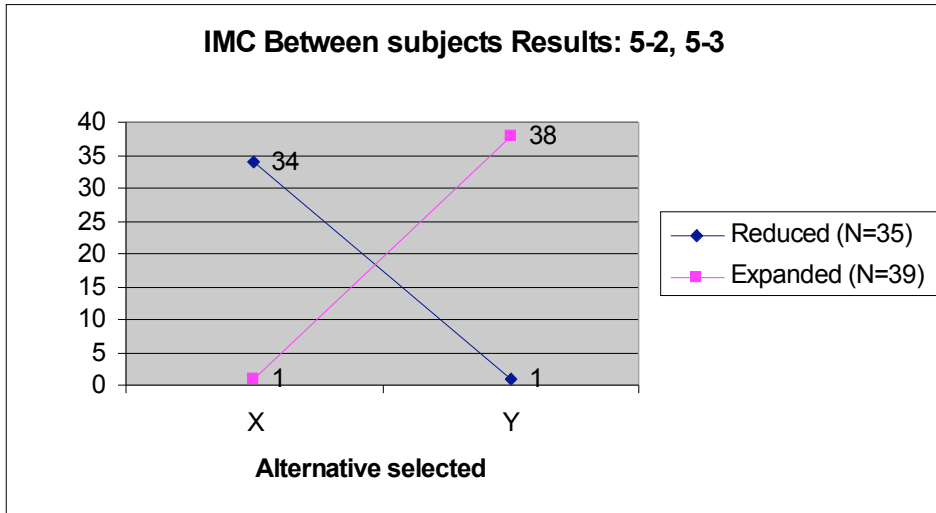
Presentation Order Variations

Example: Some survey takers see one profile in a triplet before another, with other survey takers seeing them in the reverse order.

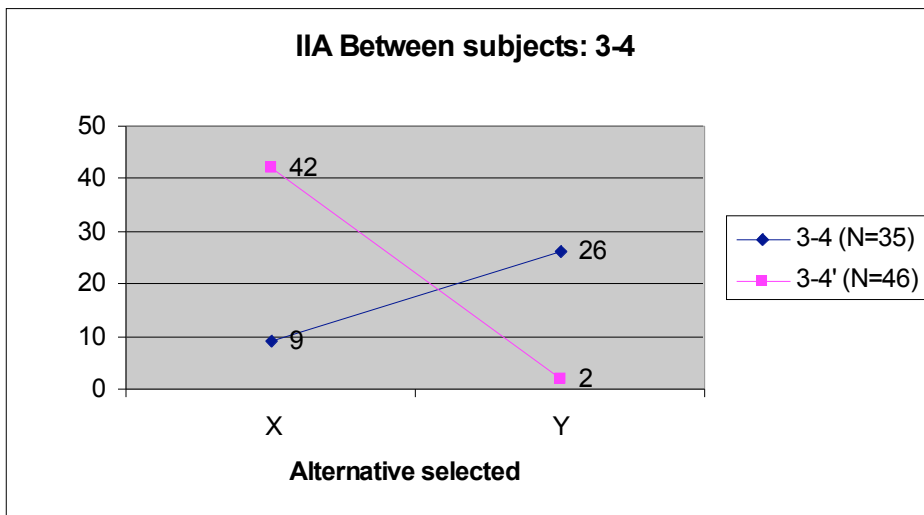
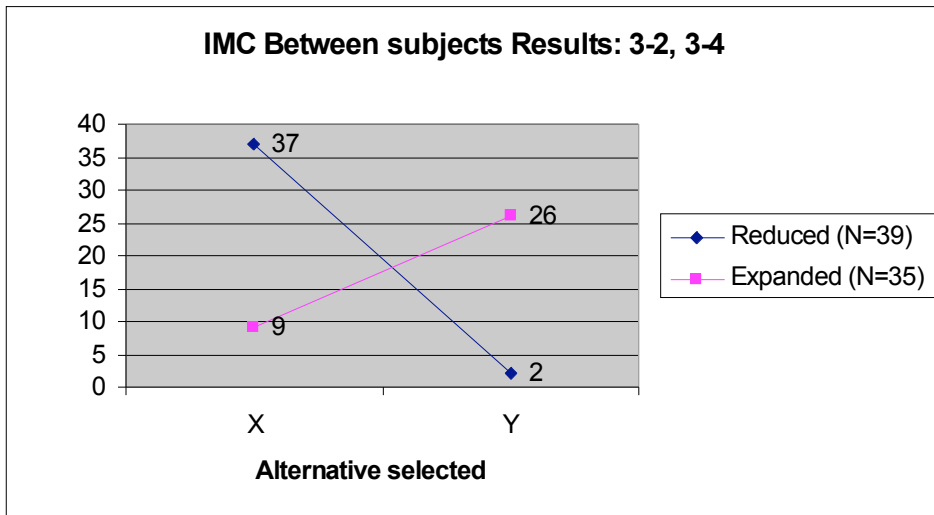
Profile Relationship Variations

Different profiles are presented to a survey taker as representing different voters and/or different alternatives, or changes in the menu of options.

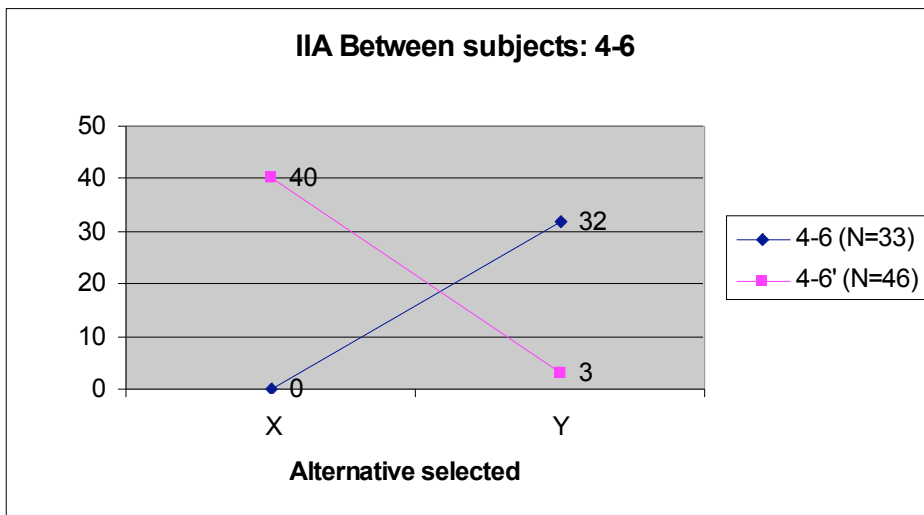
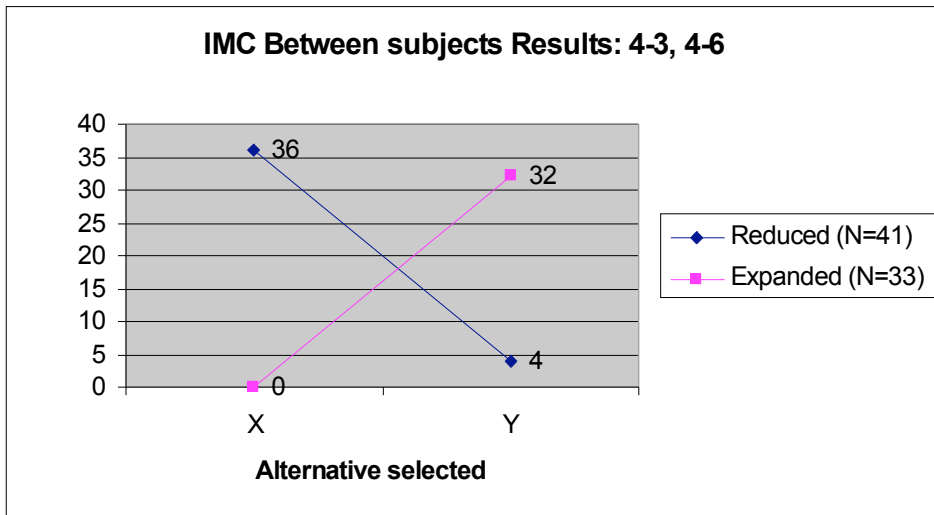
Graphs 1.1-1.3: IMC and IIA Between subjects (strict choice only) results for 5-voter profile pair



Graphs 2.1-2.3: IMC and IIA Between subjects (strict choice only) results for 3-voter profile pair



Graphs 3.1-3.3: IMC and IIA Between subjects (strict choice only) results for 4-voter profile pair



Ordering Task:

	51%	49%
Ranked Alternatives	L	P
	O	L
	M	O
	P	N
	N	Q
	Q	M

Out of 59 subjects, 55 ranked P ahead of M, while only 4 ranked M ahead of P. Assuming that they would have ranked M ahead of P if only shown the voting blocs' preferences over these two alternatives, we can say that the subject group, as a whole, is violating the IMC criterion.

Within-subjects: Tests for 5 voters x 3 alternatives pairings, rank format presentation

Test	Violators	Nonviolators
IMC	84%	16%
IIA	92%	8%

Question 1: Does intuitive preference aggregation robustly violate Arrow's criteria of IIA and collective rationality?

Answer: Yes.

Rank Format results summary:

For each of the four profile triplets, when presented in rank format, a majority of survey takers' aggregations violate both collective rationality (CR) and IIA, both between and within subjects.

Examples: In the 5-voter triplet...

- Between subjects, 34/35 pick S in 5-2, 38/39 pick T in 5-3, and 41/44 pick S in 5-3'
- Within-subjects test finds 84% violate CR and 92% violate IIA.

Pairwise presentation:

5-3 (Pairwise)

	X vs. Y	Y vs. Z	Z vs. X
Voter A	X	Z	Z
Voter B	X	Y	X
Voter C	X	Y	X
Voter D	Y	Y	Z
Voter E	Y	Y	Z

5-3' (Pairwise)

	X vs. Y	Y vs. Z	Z vs. X
Voter F	X	Z	X
Voter G	X	Z	X
Voter H	X	Y	X
Voter I	Y	Z	Z
Voter J	Y	Y	X

Presentation format results summary:

For the 4-voter and 7-voter triplets, survey takers robustly violate both CR and IIA when profiles are presented in pairwise format as well, but pairwise format can induce adherence to CR and IIA when the Borda reversal is weak (e.g. the 3- and 5-voter triplets). Examples: Between subjects...

- 90/90 pick T in 4-6, and 70/90 pick S in 4-6'
- 100/100 pick S in 5-3', and 55/100 pick S in 5-3

Other findings:

Availability of indifference option does not significantly increase adherence to IIA or CR.

Order of profile presentation does not affect results when Borda swings are large enough, but can matter when they are small.

Attendees of the Public Choice Society annual meeting show the same pattern of results as naïve survey-takers.

Question 2: Can we characterize intuitive preference aggregation judgments in a descriptive framework?

Answer: Survey-takers appear to apply a variety of principles to preference aggregation, none of which is fully determinative when they conflict:

- **PRINCIPLE 1:** *Independence of irrelevant voters (IIV)*. Social choices should be based only on the preferences of individuals affected by the choice.
- **PRINCIPLE 2:** *Interpersonal comparisons of utility (ICU)*. Social choices should take into account, as much as possible, information about the relative strengths of preference that individuals in the choice group have for one alternative over another.
- **PRINCIPLE 3:** *Inter-menu independence (IMI)*. Social choices should take into account only the alternatives that are available, and not utilize information about how individuals rank these available alternatives among other sets which include unavailable outcomes.
- **PRINCIPLE 4:** *Independence of unavailable alternatives (IUA)*. Social choices should be the same whenever the set of available outcomes does not change, even if individuals change their preferences involving unavailable outcomes.
(weakly supported)

Adherence to IUA is somewhat stronger than adherence to either IIA or CR, and for weak Borda reversals shows majority adherence when IIA and CR are violated

There probably exist other empirically supportable principles as well.

Question 3: What is the relationship between intuitive preference aggregation and normative approaches to social choice?

Two suggested answers:

A. No principle of social choice should be regarded as indispensable if there are situations in which the fully-informed opinions of a large majority of intelligent people think that the best social decision is one that entails violating the principle.

B. Surveys that systematically assess neutral subjects' aggregation across a large number of abstract preference profiles could be the basis for inducing a weighted, multi-criterion hybrid social choice procedure that would optimally characterize people's intuitions about fair social choice.