New Perspectives on Social Choice

Eric Pacuit, University of Maryland

Lecture 7: Evaluative Voting II

PHIL 808K

Grading

In many group decision situations, people use measures or grades from a **common language of evaluation** to evaluate candidates or alternatives:

- in figure skating, diving and gymnastics competitions;
- in piano, flute and orchestra competitions;
- in classifying wines at wine competitions;
- in ranking university students;
- in classifying hotels and restaurants, e.g., the Michelin $*$

M. Balinski and R. Laraki. *Majority Judgement: Measuring, Ranking and Electing*. The MIT Press, 2010.

- voters have more on their minds than merely comparing candidates, or approving of some and disapproving of others, and wish to express it,

- voters have more on their minds than merely comparing candidates, or approving of some and disapproving of others, and wish to express it,

- voters care about all the results of an election (the distributions of the votes, who is in second, third, down to last place, the spreads between candidates, and so on) and not merely who is the winner, and

- voters have more on their minds than merely comparing candidates, or approving of some and disapproving of others, and wish to express it,
- voters care about all the results of an election (the distributions of the votes, who is in second, third, down to last place, the spreads between candidates, and so on) and not merely who is the winner, and
- voters are dissatisfied with election results that do not reflect their true opinions.

Voting by Grading: Questions

- What grading language should be used? (e.g., $A - F$, $0 - 10$, $* - ****$)

Voting by Grading: Questions

- What grading language should be used? (e.g., $A - F$, $0 - 10$, $* - ****$)
- How should we *aggregate* the grades? (e.g., Average or Median)

Voting by Grading: Questions

- What grading language should be used? (e.g., $A - F$, $0 - 10$, $* - ****$)
- How should we *aggregate* the grades? (e.g., Average or Median)
- Should there be a "no opinion" option?

Measurement Theory

How to construct a scale is a science—measurement theory—that raises two key problems.

1. the faithful representation problem: What scale? "When measuring some attribute of a class of objects or events, we associate numbers with the objects in such a way that the properties of the attribute are faithfully represented as numerical properties" (Krantz et al. 1971, p. 1). For example, if the scale is a finite set of numbers from 0 to 20, should they be spaced evenly or otherwise?

Measurement Theory

How to construct a scale is a science—measurement theory—that raises two key problems.

1. the faithful representation problem: What scale? "When measuring some attribute of a class of objects or events, we associate numbers with the objects in such a way that the properties of the attribute are faithfully represented as numerical properties" (Krantz et al. 1971, p. 1). For example, if the scale is a finite set of numbers from 0 to 20, should they be spaced evenly or otherwise?
2. the meaningfulness problem: Given a faithful representation, what analyses of sets of measurements are valid? For example, if the scale consists of the integers 0 - 20 when is it justified to sum and take averages of measurements?

D. Krantz, D. Luce, P. Suppes, A. Tversky. *Foundations of Measurement: Vol. 1: Additive and Polynomial Representations*. Academic Press, New York, 1971.

- Pain, for example, is measured on an 11-point ordinal scale going from 0 to 10, each number endowed with a careful verbal description: it is not meaningful to sum or average such measures since an increase from (say) 2 to 3 cannot be equated with an increase from 8 to 9.

- Temperature, Celsius or Fahrenheit, is an interval scale because equal intervals have the same significance: sums and averages are meaningful but multiplication is not, for there is no absolute 0.

- Ounces, inches, and the Kelvin temperature scale are ratio scales: they are interval scales where 0 has an absolute sense and multiplication is meaningful as well.

"...voting or judging is measuring. The scale used by approval voting and first-past-the-post is not a faithful representation of voters' opinions; moreover, the semantics are confusing, one tick lumping all kinds of different meanings into one. Taking their sum is tantamount to declaring one mile+one meter+one inch = three and is at best a very imprecise measure.

"...voting or judging is measuring. The scale used by approval voting and first-past-the-post is not a faithful representation of voters' opinions; moreover, the semantics are confusing, one tick lumping all kinds of different meanings into one. Taking their sum is tantamount to declaring one mile+one meter+one inch = three and is at best a very imprecise measure. The semantics of rank-order inputs are perfectly clear, but they are far too limited to permit a faithful expression of opinion, deny the existence of any common scale, and lead to unacceptable methods.

"...voting or judging is measuring. The scale used by approval voting and first-past-the-post is not a faithful representation of voters' opinions; moreover, the semantics are confusing, one tick lumping all kinds of different meanings into one. Taking their sum is tantamount to declaring one mile+one meter+one inch = three and is at best a very imprecise measure. The semantics of rank-order inputs are perfectly clear, but they are far too limited to permit a faithful expression of opinion, deny the existence of any common scale, and lead to unacceptable methods. Point-summing methods exaggerate in the other direction, assuming the existence of a perfect scale of measurement—an interval scale—which is almost impossible to achieve and, in any case, leads to highly manipulable methods.

"...voting or judging is measuring. The scale used by approval voting and first-past-the-post is not a faithful representation of voters' opinions; moreover, the semantics are confusing, one tick lumping all kinds of different meanings into one. Taking their sum is tantamount to declaring one mile+one meter+one inch = three and is at best a very imprecise measure. The semantics of rank-order inputs are perfectly clear, but they are far too limited to permit a faithful expression of opinion, deny the existence of any common scale, and lead to unacceptable methods. Point-summing methods exaggerate in the other direction, assuming the existence of a perfect scale of measurement—an interval scale—which is almost impossible to achieve and, in any case, leads to highly manipulable methods. There is, however, a middle ground that asks for more than rank orders but less than an interval scale: an ordinal scale of merit."

M. Balinski and R. Laraki. *Judge:* Don't *Vote!*. Operations Research, Vol. 62, No. 3, May–June 2014, pp. 483–511.

Evaluative Voting: Examples

Approval Voting: voters can assign a single grade "approve" to the candidates. The candidates with the most approvals are the winner.

Dis&Approval Voting: voters can approve or disapprove of the candidates. The candidates with the greatest scores are the winners.

Score Voting: voters can assign any grade from a fixed set of grades to the candidates. The candidate with the greatest sum of the scores is the winner.

Majority Judgement: voters can assign any grade from a fixed set of grades to the candidates. The candidate with the greatest median score is the winner.

Score Voting/Range Voting

Fixe a common grading language consisting of, for example, the integers $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest *average* grade is declared the winner.

Score Voting/Range Voting

Fixe a common grading language consisting of, for example, the integers $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest *average* grade is declared the winner.

Suppose A's grades are $\{7, 7, 8, 8, 9, 9, 9, 10\}$. The average grade is 8.375

Suppose B's grades are $\{9, 9, 9, 9, 9, 10, 10, 10\}$. The average grade is 9.375

So, Score Vote (Range Vote) ranks B above candidate A.

Score Voting/Range Voting

Fixe a common grading language consisting of, for example, the integers $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest *average* grade is declared the winner.

Suppose A's grades are $\{7, 7, 8, 8, 9, 9, 9, 10\}$. The average grade is 8.375

Suppose B's grades are $\{9, 9, 9, 9, 9, 10, 10, 10\}$. The average grade is 9.375

So, Score Vote (Range Vote) ranks B above candidate A.

www.electology.org/score-voting and rangevoting.org

Majority Judgement

Fix a common grading language. For example, $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest median grade is declared the winner.

The *median* grade is the grade that is in the middle of the list when the grades are ordered (If there is an even number of judges, then the median grade is the lowest grade in the middle interval.)

Majority Judgement

Fix a common grading language. For example, $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest median grade is declared the winner.

The *median* grade is the grade that is in the middle of the list when the grades are ordered (If there is an even number of judges, then the median grade is the lowest grade in the middle interval.)

Majority Judgement

Fix a common grading language. For example, $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest median grade is declared the winner.

The *median* grade is the grade that is in the middle of the list when the grades are ordered (If there is an even number of judges, then the median grade is the lowest grade in the middle interval.)

Suppose that A's grades are $\{6, 6, 7, 7, 7, 8, 9, 10, 10\}$: The median grade is 7.

Majority Judgement

Fix a common grading language. For example, $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest median grade is declared the winner.

The *median* grade is the grade that is in the middle of the list when the grades are ordered (If there is an even number of judges, then the median grade is the lowest grade in the middle interval.)

Suppose that A's grades are $\{6, 6, 7, 7, 7, 8, 9, 10, 10\}$: The median grade is 7.

Suppose B's grades are $\{6, 6, 6, 6, 9, 9, 9, 10\}$: The median grade is 6.

Majority Judgement

Fix a common grading language. For example, $\{0, 1, 2, \ldots, 10\}$

The candidate with the largest median grade is declared the winner.

The *median* grade is the grade that is in the middle of the list when the grades are ordered (If there is an even number of judges, then the median grade is the lowest grade in the middle interval.)

Suppose that A's grades are $\{6, 6, 7, 7, 7, 8, 9, 10, 10\}$: The median grade is 7.

Suppose B's grades are $\{6, 6, 6, 6, 9, 9, 9, 10\}$: The median grade is 6.

Majority Judgement ranks B above A.

Majority Judgement: Tie-breaking rules

What happens when the median grades are the same?

Majority Judgement: Tie-breaking rules

What happens when the median grades are the same?

A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

Majority Judgement: Tie-breaking rules

What happens when the median grades are the same?

A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

The second median grade is found:
A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

Majority Judgement: Tie-breaking rules

What happens when the median grades are the same?

A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

The second median grade is found:
A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

The third median grade is found:
A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

Majority Judgement: Tie-breaking rules

What happens when the median grades are the same?

A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

The second median grade is found:
A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

The third median grade is found:
A's grades: $\{7, 9, 9, 11, 11\}$
B's grades: $\{8, 9, 9, 10, 11\}$

So, A is ranked above B.

Example

Suppose that there are five voters, $1, \ldots, 5$ and three candidates I, II, and III.

The grades are A, B, C, D, or F (from best to worst).

Example

Suppose that there are five voters, $1, \ldots, 5$ and three candidates I, II, and III. The grades are A, B, C, D, or F (from best to worst).

	1	2	3	4	5
I	A	A	C	D	D
II	B	B	F	B	F
III	D	C	B	A	D

Example

Suppose that there are five voters, $1, \ldots, 5$ and three candidates I, II, and III. The grades are A, B, C, D, or F (from best to worst).

	1	2	3	4	5	Median
I	A	A	C	D	D	C
II	B	B	F	B	F	B
III	D	C	B	A	D	C

Candidate II is the majority judgement winner.

Example

Suppose that there are five voters, $1, \ldots, 5$ and three candidates I, II, and III. The grades are A, B, C, D, or F (from best to worst).

	1	2	3	4	5	Median
I	A	A	C	D	D	C
II	B	B	F	B	F	B
III	D	C	B	A	D	C

Candidate II is the majority judgement winner. *If asked about their preference, 4 voters would rank candidate I above candidate II*

Example

Suppose that there are five voters, $1, \ldots, 5$ and three candidates I, II, and III. The grades are $A = 4$, $B = 3$, $C = 2$, $D = 1$, or $F = 0$ (from best to worst).

	1	2	3	4	5	Average
I	4	4	2	1	1	2.4
II	3	3	0	3	0	1.8
III	1	2	3	4	1	2.2

Candidate II is the Majority Judgement winner. Candidate I is the Score Voting winner

# voters	1	1	1
A	20/20	9/20	9/20
B	11/20	0/20	10/20

Majority Judgement Winner: B

# voters	1	1	1
A	20/20	9/20	9/20
B	11/20	0/20	10/20

Majority Judgement Winner: B

2 out of 3 voters prefer A to B

# voters	50	50	1
A	$20/20$	$9/20$	$9/20$
B	$11/20$	$0/20$	$10/20$

Majority Judgement Winner: B

# voters	50	50	1
A	20/20	9/20	9/20
B	11/20	0/20	10/20

Majority Judgement Winner: B

100 out of 101 voters prefer A to B

S. Brams and R. Potthoff. *The paradox of grading systems*. Public Choice, 165, pp. 193 - 210, 2015.

Grades: $\{0, 1, 2, 3\}$

Candidates: $\{A, B, C\}$

3 Voters

# voters	1	1	1	Avg
A	3	2	0	
B	0	3	1	
C	0	3	1	

Grades: $\{0, 1, 2, 3\}$

Candidates: $\{A, B, C\}$

3 Voters

# voters	1	1	1	Avg
A	3	2	0	$5/3$
B	0	3	1	$4/3$
C	0	3	1	$4/3$

Average Grade Winner: A

Grades: $\{0, 1, 2, 3\}$

Candidates: $\{A, B, C\}$

3 Voters

# voters	1	1	1	Avg
A	3	2	0	
B	0	3	1	
C	0	3	1	

Average Grade Winner: A

$$B \succ A$$

Grades: $\{0, 1, 2, 3\}$

Candidates: $\{A, B, C\}$

3 Voters

# voters	1	1	1	Avg
A	3	2	0	
B	0	3	1	
C	0	3	1	

Average Grade Winner: A

$$C \sim B \succ A$$

Grades: $\{0, 1, 2, 3\}$

Candidates: $\{A, B, C\}$

3 Voters

# voters	1	1	1	Avg
A	3	2	0	
B	0	3	1	
C	0	3	1	

Average Grade Winner: A

$$C \sim B \succ A$$

Grades: $\{0, 1, 2, 3\}$

Candidates: $\{A, B, C\}$

3 Voters

# voters	1	1	1	Avg
A	3	2	0	
B	0	3	1	
C	0	3	1	

Average Grade Winner: A

Superior Grade Winners: C, B

Grades: $\{0, 1, 2, 3, 4, 5\}$

Candidates: $\{A, B, C\}$

5 Voters

# voters	1	4	Avg
A	5	0	$5/5$
B	0	1	$4/5$
C	0	1	$4/5$

Average Grade Winner: A

Superior Grade Winner: B, C

“...we have identified a paradox of grading systems, which is not just a mirror of the well-known differences that crop up in aggregating votes under ranking systems. Unlike these systems, for which there is no accepted way of reconciling which candidate to choose when, for example, the Hare, Borda and Condorcet winners differ, AV provides a solution when the AG and SG winners differ. ”

Theorem (Brams and Potthoff). When there are two grades, the AG and SG winners are identical.

AG, SG, Median Winners

Grades: $\{0, 1, 2\}$

Candidates: $\{A, B, C\}$

9 Voters

# voters	2	3	4	Avg	Median
A	2	1	1		
B	1	2	0		
C	1	2	1		

AG, SG, Median Winners

Grades: $\{0, 1, 2\}$

Candidates: $\{A, B, C\}$

9 Voters

# voters	2	3	4	Avg	Median
A	2	1	1	$11/9$	1
B	1	2	0	$10/9$	2
C	1	2	1	$10/9$	1

Average Grade Winner: A

Median Grade Winner: B

Superior Grade Winner: C

weak paradox of grading systems: Profiles in which the AG and SG winners overlap but do not coincide.

strong paradox of grading systems: Profiles in which the AG and SG winners do not overlap.

Probability of Weak Paradox of Grading Systems

Table 2 Probability of weak paradox of grading systems for sincere (strategic) models with three grades ($g = 3$), as a function of the numbers of candidates (c) and voters (v)

#Candidates → #Voters ↓	$c = 3$	$c = 4$	$c = 5$	$c = 6$	$c = 9$	$c = 10$
$v = 3$	0.2216 (0.2037)	0.2797 (0.2995)	0.3339 (0.3501)	0.3747 (0.4017)	0.4340 (0.4490)	0.4523 (0.4508)
$v = 4$	0.2443 (0.2874)	0.3168 (0.3447)	0.3487 (0.3740)	0.3933 (0.4081)	0.4687 (0.4698)	0.4732 (0.4889)
$v = 5$	0.2526 (0.2315)	0.3157 (0.3393)	0.3663 (0.3757)	0.4179 (0.4194)	0.4847 (0.4807)	0.4970 (0.4909)
$v = 6$	0.2627 (0.2671)	0.3321 (0.3391)	0.3839 (0.3758)	0.4057 (0.4194)	0.4897 (0.4909)	0.5084 (0.5045)
$v = 7$	0.2633 (0.2394)	0.3225 (0.3330)	0.3746 (0.3724)	0.4129 (0.4103)	0.4863 (0.4918)	0.4977 (0.4944)
$v = 8$	0.2658 (0.2555)	0.3284 (0.3372)	0.3854 (0.3708)	0.4097 (0.4055)	0.4900 (0.4891)	0.5016 (0.5131)
$v = 9$	0.2709 (0.2426)	0.3249 (0.3314)	0.3780 (0.3681)	0.4116 (0.4120)	0.4857 (0.4880)	0.4869 (0.5069)
$v = 25$	0.2518	0.3071	0.3448	0.3873	0.4441	0.4557
$v = 50$	0.2310	0.2919	0.3253	0.3455	0.4087	0.4273

Probability of Weak Paradox of Grading Systems

Table 3 Probability of weak paradox of grading systems for sincere model, as a function of the numbers of candidates (c), voters (v), and grades (g)

#Grades → #Candidates/#Voters ↓	$g = 3$	$g = 5$	$g = 7$	$g = 9$
$c = 3$:				
$v = 3$	0.2216	0.3276	0.3372	0.3344
$v = 4$	0.2443	0.3763	0.4183	0.4534
$v = 9$	0.2709	0.3369	0.3614	0.3572
$v = 25$	0.2518	0.3059	0.3310	0.3378
$c = 4$:				
$v = 4$	0.3168	0.4661	0.5103	0.5476
$v = 5$	0.3157	0.4307	0.4385	0.4385
$v = 9$	0.3249	0.4196	0.4394	0.4475
$v = 25$	0.3071	0.3733	0.3962	0.4452
$c = 5$:				
$v = 5$	0.3663	0.4947	0.5008	0.5016
$v = 6$	0.3839	0.5000	0.5275	0.5562
$v = 9$	0.3780	0.4657	0.4997	0.5021
$v = 25$	0.3448	0.4399	0.4526	0.4631

Concluding Remarks - AV vs. Grading Systems

- The strategic model, in which all voters give at least one candidate a bottom grade and at least one candidate a top grade, seems a more realistic model than the sincere model without such a constraint...Unless voters dichotomize their grades by grading up some middling candidates to the maximum grade, and grading down other middling candidates to the minimum grade—which is in effect to reduce the grades to two (as under approval voting, AV)—then a paradox can occur.

Concluding Remarks - AV vs. Grading Systems

- The strategic model, in which all voters give at least one candidate a bottom grade and at least one candidate a top grade, seems a more realistic model than the sincere model without such a constraint...Unless voters dichotomize their grades by grading up some middling candidates to the maximum grade, and grading down other middling candidates to the minimum grade—which is in effect to reduce the grades to two (as under approval voting, AV)—then a paradox can occur.

- Among grading systems, we favor AV ($g = 2$), not only because it ensures that the AG and SG winners are the same, so no paradox can arise, but also because it is sensitive to how individual voters dichotomize their middling candidates.

Concluding Remarks - AV vs. Grading Systems

- If voters can assign more than two grades, they will have a strong incentive to give a maximum grade to only one candidate—at least if he or she is viewed as a contender—and a minimum grade to all the other candidates. If a voter's favorite is not viewed as a contender, then he or she will be motivated to give a maximum grade to his or her (noncompetitive) favorite and a favorite among the contenders, and a minimum grade to everybody else. In either case, these strategies are tantamount to those that would be used under AV.

Concluding Remarks - AV vs. Grading Systems

- If voters can assign more than two grades, they will have a strong incentive to give a maximum grade to only one candidate—at least if he or she is viewed as a contender—and a minimum grade to all the other candidates. If a voter's favorite is not viewed as a contender, then he or she will be motivated to give a maximum grade to his or her (noncompetitive) favorite and a favorite among the contenders, and a minimum grade to everybody else. In either case, these strategies are tantamount to those that would be used under AV.
- Although AV may not be the ideal system to use in figure-skating contests or wine-tasting competitions—wherein voters have little or no incentive to dichotomize their grades—we think it has much to commend in political contests in which more than two grades can result in the paradox, even when voters are strategic.

J.-F. Laslier. *The Strange "Majority Judgement"*. https://hal.archives-ouvertes.fr/hal-01965227.

D. Felsenthal and M. Machover. *The Majority Judgement voting procedure: a critical evaluation*. Homo oeconomicus, 25, pp. 319 - 334.

S. Brams. *Grading Candidates*. Book Review, American Scientist, 2012.

M. Balinski. *A letter regarding Steven J. Brams's review of Majority Judgment* . American Scientist.

Majority Judgement: Spoilers

49%	3%	48%
Gore: 20/20	Nader: 20/20	Bush: 20/20
Nader: 6/20	Gore: 4/20	Gore: 2/20
Bush: 0/20	Bush: 0/20	Nader: 0/20

In the absence of Nader, Gore would win according to any reasonable voting method. What is more, note that Gore would also win in the absence of the Nader voters, and that all Nader voters prefer Gore to Bush.

Majority Judgement: Spoilers

49%	3%	48%
Gore: 20/20	Nader: 20/20	Bush: 20/20
Nader: 6/20	Gore: 4/20	Gore: 2/20
Bush: 0/20	Bush: 0/20	Nader: 0/20

In the absence of Nader, Gore would win according to any reasonable voting method. What is more, note that Gore would also win in the absence of the Nader voters, and that all Nader voters prefer Gore to Bush.

The median evaluations are 0 for Bush, 4 for Gore, and 6 for Nader. Consequently, Nader would be elected with the Majority Judgment method.

Majority Judgement: Spoilers

This example has the benefit, among others, of showing that it is not true that all systems of evaluative voting avoid the spoiler effect. It all depends on how the ballot papers are counted. In this society, with Majority Judgment, Gore electors would be wrong to believe that they can rate Nader 6/20 without causing damage.

Majority Judgement: No Show Paradox

	Excellent	Very Good	Good	Quite good	Acceptable	Mediocre	Rejected
A	1	2	0	3*	3	0	0
B	0	2	2	0	2*	1	2

Majority Judgement winner: *A*.

Majority Judgement: No Show Paradox

	Excellent	Very Good	Good	Quite good	Acceptable	Mediocre	Rejected
A	1	2	0	3^*	3	0	0
B	0	2	2	0	2^*	1	2

Majority Judgement winner: A.

Suppose two voters decide to submit their vote: A is *Excellent* and B is *Very Good*.

	Excellent	Very Good	Good	Quite good	Acceptable	Mediocre	Rejected
A	3	2	0	3^*	3	0	0
B	0	4	2^*	0	2	1	2

Majority Judgement winner: B.

Voters who are indifferent between all candidates may decide to abstain from voting, or to express their indifference actively.

Voters who are indifferent between all candidates may decide to abstain from voting, or to express their indifference actively.

x	3	4	4
y	2	5	5

Winner: y

Voters who are indifferent between all candidates may decide to abstain from voting, or to express their indifference actively.

x	3	4	4
y	2	5	5

Winner: y

x	1	1	3	4	4
y	1	1	2	5	5

Winner: x

Strategic Grading

The *median grade* encourages sincere grading.

Strategic Grading

The *median grade* encourages sincere grading.

Suppose a student submits a paper to a panel of examiners.

Strategic Grading

The *median grade* encourages sincere grading.

Suppose a student submits a paper to a panel of examiners.

If an examiner were informed of the grades awarded by the other examiners, she would have no incentive to change her grade.

Strategic Grading

The *median grade* encourages sincere grading.

Suppose a student submits a paper to a panel of examiners.

If an examiner were informed of the grades awarded by the other examiners, she would have no incentive to change her grade.

If the examiner's grade coincides with the median, there is no incentive to change

Strategic Grading

The *median grade* encourages sincere grading.

Suppose a student submits a paper to a panel of examiners.

If an examiner were informed of the grades awarded by the other examiners, she would have no incentive to change her grade.

If the examiner's grade coincides with the median, there is no incentive to change

If the examiner's grade turns out to be lower than the median, there is no incentive to change: lowering the grade would not change the result, raising the grade may result in a *higher* median grade.

Strategic Grading

The *median grade* encourages sincere grading.

Suppose a student submits a paper to a panel of examiners.

If an examiner were informed of the grades awarded by the other examiners, she would have no incentive to change her grade.

If the examiner's grade coincides with the median, there is no incentive to change

If the examiner's grade turns out to be lower than the median, there is no incentive to change: lowering the grade would not change the result, raising the grade may result in a *higher* median grade.

Similarly, if the examiner's grade turns out to be higher than the median grade.

	1	2	3
x	A	C	C
y	D	A	D
z	B	D	A

	1	2	3
x	A	C	C
y	D	A	D
z	B	D	A

- If the judges act sincerely, then z is the winner

	1	2	3
x	A	C	C
y	D	A	D
z	B	D	A

- If the judges act sincerely, then z is the winner
- Both 1 and 2 would prefer that x *wins*

	1	2	3
x	A	C	C
y	D	A	D
z	B	D	A

- If the judges act sincerely, then z is the winner
- Both 1 and 2 would prefer that x *wins*
- 1 can down grade z to D, then x will win

	1	2	3
x	A	C	C
y	D	A	D
z	B	D	A

- If the judges act sincerely, then z is the winner
- Both 1 and 2 would prefer that x *wins*
- 1 can down grade z to D, then x will win
- 2 can upgrade x from C to A, then x will win

	1	2	3
x	A	C	C
y	D	A	D
z	B	D	A

- If the judges act sincerely, then z is the winner
- Both 1 and 2 would prefer that x *wins*
- 1 can down grade z to D, then x will win
- 2 can upgrade x from C to A, then x will win
- 3 cannot do anything about this.

	1	2	3
x	A	C	C
y	D	A	D
z	B	D	A

- If the judges act sincerely, then z is the winner
- Both 1 and 2 would prefer that x *wins*
- 1 can down grade z to D, then x will win
- 2 can upgrade x from C to A, then x will win
- 3 cannot do anything about this.

M. Nunez and J.-F. Laslier. *Preference intensity representation: strategic overstating in large elections*. Social Choice and Welfare, 42, pp. 313 - 340, 2014.

Baujard et al. (2013) report on field work on EV with various scales, and observe that voters often say that they appreciate the possibility of voicing their opinions more finely than what a plurality vote allows.

M. Nunez and J.-F. Laslier. *Preference intensity representation: strategic overstating in large elections*. Social Choice and Welfare, 42, pp. 313 - 340, 2014.

Baujard et al. (2013) report on field work on EV with various scales, and observe that voters often say that they appreciate the possibility of voicing their opinions more finely than what a plurality vote allows.

But does extending the set of ballots available under a given voting rule modify the set of *voting equilibria*?

Nunez and Laslier

- Focus on additive voting rules, in which a ballot is a list of points that the voter is affording to the candidates, and where points for each candidate are simply added. We analyze the issue in the framework of strategic voting, that is assuming that voters strategically cast their votes in order to maximize their (expected) utility. We study equilibria and consider that two voting rules are strategically equivalent if they have the same equilibrium outcomes.
- Two main conclusions:
  1. Approval Voting and Evaluative Voting are strategically equivalent.
  2. Cumulative Voting and Plurality Voting are strategically equivalent.

Nunez and Laslier

- Strategic incentives in large and small electorates may differ and have been modeled in different ways. To study small electorates, we use a standard refinement of Nash equilibrium (perfectness) and provide an example that shows that voters need not overstate at equilibrium.
- The more flexibility a voting rule allows to voters, the more it becomes similar, from a strategic point of view, to Approval Voting. Small elections (that is elections with few voters) raise new questions. For instance, the information available to voters might be much more detailed in a small election than in a mass election.

Characterizing *Scoring Rules*

Voting Methods

$$F : \mathcal{D} \to \mathcal{R}$$

Voting Methods

$F : \mathcal{D} \to \mathcal{R}$

- What is the range? $\mathcal{R} = O(X)$ or $\mathcal{R} = \wp(X) - \emptyset$

Voting Methods

$F : \mathcal{D} \to \mathcal{R}$

- What is the range? $\mathcal{R} = O(X)$ or $\mathcal{R} = \wp(X) - \emptyset$
- What is the domain? $\mathcal{D} = O(X)^n$, $\mathcal{D} = L(X)^n$ or $\mathcal{D} = \mathcal{B}^n$

Voting Methods

$F : \mathcal{D} \to \mathcal{R}$

- What is the range? $\mathcal{R} = O(X)$ or $\mathcal{R} = \wp(X) - \emptyset$

- What is the domain? $\mathcal{D} = O(X)^n$, $\mathcal{D} = L(X)^n$ or $\mathcal{D} = \mathcal{B}^n$

- Do we focus on a fixed set of voters?

Variable Population Model

Let $\mathbb{N}$ be the set of "potential" voters.

Let $\mathcal{V} = \{V \mid V \subseteq \mathbb{N}$ and V is finite$\}$ be the set of all voting blocks.

For $V \in \mathcal{V}$, a **profile** for V is a function $\pi : V \to \mathcal{P}$ where $\mathcal{P}$ is $O(X)$, $L(X)$ or some set $\mathcal{B}$ of "ballots".

Let $\Pi^{\mathcal{P}}$ be the set of all profiles based on $\mathcal{P}$.

Variable Population Model

Two profiles $\pi : V \to \mathcal{P}$ and $\pi' : V' \to \mathcal{P}$ are *disjoint* if $V \cap V' = \emptyset$

If $\pi : V \to \mathcal{P}$ and $\pi' : V' \to \mathcal{P}$ are *disjoint*, then $(\pi{+}\pi') : (V{\cup}V') \to \mathcal{P}$ is the profile where for all $i \in V{\cup}V'$,

$$(\pi + \pi')(i) = \begin{cases} \pi(i) & \text{if } i \in V \\ \pi'(i) & \text{if } i \in V' \end{cases}$$

Variable population model: $F : \Pi^{\mathcal{B}} \rightarrow (\wp(X) - \emptyset)$ where $\mathcal{B}$ is some fixed set of "ballots".

Five characterization results:

1. Positional Scoring Functions ($\mathcal{B} = L(X)$)
2. Borda Count ($\mathcal{B} = L(X)$)
3. Approval Voting ($\mathcal{B} = \wp(X) - \emptyset$)
4. Plurality Rule ($\mathcal{B} = L(X)$)
5. Range Voting ($\mathcal{B} = \mathcal{S}$)

Positional Scoring Rules

$F : \Pi^{L(X)} \to (\wp(X) - \emptyset)$

Suppose there are m candidates. Let $\{s_1, \ldots, s_m\}$ be a set of scores,

$\boldsymbol{N}_\pi(j, A) = \{i \mid i \text{ ranks } A \text{ in the } j\text{th position}\}$ and
$Score_\pi(A) = \sum_{j=1}^{m} s_j \times |\boldsymbol{N}_\pi(j, A)|$

F is a scoring function if $F(\pi) = \{A \mid Score_\pi(A) \geq Score_\pi(B) \text{ for all } B \in X\}$

Positional Scoring Rules

$F : \Pi^{L(X)} \to (\wp(X) - \emptyset)$

Suppose there are m candidates. Let $\{s_1, \ldots, s_m\}$ be a set of scores,

$\boldsymbol{N}_\pi(j, A) = \{i \mid i \text{ ranks } A \text{ in the } j\text{th position}\}$ and
$Score_\pi(A) = \sum_{j=1}^{m} s_j \times |\boldsymbol{N}_\pi(j, A)|$
F is a scoring function if $F(\pi) = \{A \mid Score_\pi(A) \geq Score_\pi(B) \text{ for all } B \in X\}$

Consistency: Suppose that π and π' are disjoint profiles. If $F(\pi) \cap F(\pi') \neq \emptyset$, then $F(\pi + \pi') = F(\pi) \cap F(\pi')$.

Positional Scoring Rules

$F : \Pi^{L(X)} \to (\wp(X) - \emptyset)$

Suppose there are m candidates. Let $\{s_1, \ldots, s_m\}$ be a set of scores,

$\boldsymbol{N}_\pi(j, A) = \{i \mid i$ ranks A in the jth position$\}$ and
$Score_\pi(A) = \sum_{j=1}^{m} s_j \times |\boldsymbol{N}_\pi(j, A)|$
F is a scoring function if $F(\pi) = \{A \mid Score_\pi(A) \geq Score_\pi(B)$ for all $B \in X\}$

Overwhelming Majority: Suppose that a group of voters V elects a candidate A and a disjoint group of voters Y elects a different candidate B. Then there must be some number m such that the population consisting of the subgroup V together with m copies of Y will elect B.

Positional Scoring Rules

Theorem. A social choice correspondence $F : \Pi^{L(X)} \to (\wp(X) - \emptyset)$ satisfies anonymity, neutrality, consistency and overwhelming majority if and only if F is a scoring rule.

H. P. Young. *Social Choice Scoring Functions*. SIAM Journal of Applied Mathematics, 28, pgs. 824 - 838, 1975.

J. H. Smith. *Aggregation of Preferences with Variable Electorate*. Econometrica, 41, pgs. 1027 - 1041, 1973.

Borda Count

Borda Count is a scoring rule with scores $s_j = m - j$, where m is the number of candidates.

Cancellation: For all profiles π, if $|\boldsymbol{N}_\pi(A\ P\ B)| = |\boldsymbol{N}_\pi(B\ P\ A)|$ for all pairs of candidates $A, B \in X$, then $F(\pi) = X$.

Faithfulness: If there is only one voter ($V = \{i\}$) and $\pi : \{i\} \to L(X)$, then $F(\pi) = \{A\}$ where A is voter i's top-ranked candidate according to the linear order $\pi(i) \in L(X)$.

Borda Count

Theorem. A social choice correspondence $F : \Pi^{L(X)} \to (\wp(X) - \emptyset)$ satisfies anonymity, neutrality, consistency, cancellation and faithfulness if and only if F is Borda Count.

H. P. Young. *An axiomatization of Borda's rule*. Journal of Economic Theory, 9, pgs. 43 - 52, 1974.

S. Nitzan and A. Rubinstein. *A further characterization of Borda ranking method*. Public Choice, 36, pgs. 153 - 158, 1981.

Approval Voting

$F : \Pi^{\mathcal{B}} \to (\wp(X) - \emptyset)$, where $\mathcal{B} = \{B \subseteq X \mid B \neq \emptyset\}$.

Theorem. A social choice correspondence $F : \Pi^{\mathcal{B}} \to (\wp(X) - \emptyset)$ satisfies anonymity, neutrality, consistency*, cancellation* and faithfulness* iff F is approval voting.

C. Alós-Ferrer. *A simple characterization of approval voting*. Social Choice and Welfare, 27:3, pgs. 621 - 625, 2006.

P. Fishburn . *Axioms for Approval Voting: Direct Proof*. Journal of Economic Theory, 19, pgs. 180 - 185, 1978.

Plurality Rule

$F : \Pi^{L(X)} \to (\wp(X) - \emptyset)$

For a linear ordering $R \in L(X)$, $T(R)$ is the top ranked candidate.

For a profile π, let $T(\pi) = (T(\pi_i))_{i \in V}$, where V is the domain of π.

For $A \in X$, let $Top_\pi(A) = \{i \mid T(\pi(i)) = A\}$.

F is plurality rule if $F(\pi) = \{A \mid |Top(A)| \geq |Top(B)| \text{ for all } B \in X\}$

Tops-Only: For all profiles π and π' on the same set V of voters, if $T(\pi) = T(\pi')$ then $F(\pi) = F(\pi')$

Plurality Rule

Theorem. A social choice correspondence $F : \Pi^{L(X)} \rightarrow (\wp(X) - \emptyset)$ satisfies anonymity, neutrality, consistency, faithfulness and tops-only if and only if F is plurality rule.

Y. Sekiguchi. *A Characterization of Plurality Rule*. Economic Letters, 116:3, pgs. 330-332, 2012.

S. Ching. *A Simple Characterization of Plurality Rule*. Journal of Economic Theory, 71, pgs. 298 - 302, 1996.

Range Voting

$F : \Pi^{\mathcal{S}} \to (\wp(X) - \emptyset)$, where $\mathcal{S}$ is a (possibly infinite) set of 'signals'.

F is **at least as expressive as** G if the voters can express any profiles of opinions via F which they could have expressed via F.

F **admits minority overrides** if regardless of the size of the set of voters in a profile and the weight of existing public opinion, a single voter can always cast a vote that changes the outcome.

Range Voting

Theorem. Range voting is the most expressive voting rule which satisfies consistency, neutrality, over-whelming majority and does not admit minority overrides.

M. Pivato. *Formal utilitarianism and range voting*. Mathematical Social Sciences, 2013.

Additional Readings

V. Merlin. *The Axiomatic Characterizations of Majority Voting and Scoring Rules*. Mahtematical Social Sciences, 161, pp. 81 - 109, 2003.

M. Pivato. *Variable-population voting rules*. Journal of Mathematical Economics, 49, pp. 210 - 221, 2013.

Grades vs. Utilities

In the real world the deep preferences or utilities of a judge or a voter are a very complicated function that depens on a host of factors....We contend that the deep preference of judges or voters *cannot* be the inputs of a practical model of voting. *A judge's input is imply a message, no more no less*. But her input chosen strategically, depends, on her deep preferences or utilities.

(Balinksi and Laraki, p. 184)

Grades vs. Utilities

The language of grades has nothing to do with utilities (viewed as measures of individual satisfaction). Grades are absolute measures of merit. In the context of voting and judging, utilities are relative measures of satisfaction. Formally, a distinction must be made between two different types of scales of measurement. An *absolute* scale measures each entity individually (height, area, merit). A *relative* scale measure each entity with respect to a collective of like entities (velocity, satisfaction). Were voters to be asked their satisfaction as inputs, adjoining or eliminating candidates would alter their answers, provoking the possibility of Arrow's paradox. A common language must be an absolute scale of measurement. (Balinksi and Laraki, p. 185)

A. Dhillon and J.-F. Mertens. *Relative Utilitarianism*. Econometrica, Vol. 67, No. 3, pp. 471 - 498, 1999.

Axiomatize **relative utilitarianism**: normalize the nonconstant individual von Neumann-Morgenstern utility functions to have infimum zero and supremum one, and taking the sum as social utility.

Sen: Informational Basis of Social Choice

A. Sen (1999). *The possibility of social choice.* American Economic Review 89: 349 - 78.

...social choice is impossible in the absence of interpersonally comparable indices of well-being.

Utility Functions

A **utility function** on a set X is a function $u : X \to \mathbb{R}$

Utility Functions

A **utility function** on a set X is a function $u : X \to \mathbb{R}$

A preference ordering is **represented** by a utility function iff x is (weakly) preferred to y provided $u(x) \geq u(y)$

L. Narens and B. Skyrms . *The Pursuit of Happiness Philosophical and Psychological Foundations of Utility*. Oxford University Press, 2020.

Let X and V be nonempty sets with $|X| \geq 3$ and V finite.

Let $\mathcal{U}(X)$ be the set of all functions $u : X \to \mathbb{R}$

A **profile** is a function $\boldsymbol{U} : V \to \mathcal{U}(X)$, write $\boldsymbol{U}_i$ for voter i's utility function on X in profile $\boldsymbol{U}$.

A **Social Welfare Functional (SWFL)** is a function f mapping profiles of utilities to asymmetric relations on X. So for each profile $\boldsymbol{U}$, $f(\boldsymbol{U})$ is the social preference order on X.

U	x	y	z
a	3	1	8
b	3	2	1
c	1	4	1

P	a	b	c
	z	x	y
	x	y	x z
	y	z	

U	x	y	z
a	3	1	8
b	300	2	1
c	1	4	1

P	a	b	c
	z	x	y
	x	y	x z
	y	z	

U	x	y	z
a	3	1	8
b	300	2	1
c	1	400	1

P	a	b	c
	z	x	y
	x	y	x z
	y	z	

Von Neumann-Morgenstern Axioms

Let $\mathcal{L}$ be the set of lotteries on a set X, P a strict preference relation on $\mathcal{L}$, and I an indifference relation on $\mathcal{L}$.

Preference	preference over $\mathcal{L}$ are complete and transitive
Compound Lotteries	The decision maker is indifferent between every compound lottery and its *simplification*
Independence	For all $L_1, L_2, L_3 \in$ Ł and $a \in (0, 1]$, $L_1 \ P \ L_2$ iff $[L_1 : a, \ L_3 : (1-a)] \ P \ [L_2 : a, \ L_3 : (1-a)]$ $L_1 \ I \ L_2$ iff $[L_1 : a, \ L_3 : (1-a)] \ I \ [L_2 : a, \ L_3 : (1-a)]$
Continuity	For all $L_1, L_2, L_3 \in$ Ł, if $L_1 \ P \ L_2 \ P \ L_3$, then there exists $a \in (0, 1)$ such that $[L_1 : a, \ L_3 : (1-a)] \ I \ L_2$

Von Neumann-Morgenstern Theorem

$u : Ł \to \mathbb{R}$ is **linear** provided for all $L = [L_1 : p_1, \ldots, L_n : p_n] \in Ł$,

$$u(L) = \sum_{i=1}^{n} p_i \times u(L_i)$$

Von Neumann-Morgenstern Representation Theorem Suppose that (P, I) is a rational preference on the set $\mathcal{L}$ of lotteries. Then, (P, I) satisfies Compound Lotteries, Independence and Continuity if, and only if, (P, I) is represented by a **linear utility function**.

Von Neumann-Morgenstern Theorem

$u : Ł \to \mathbb{R}$ is **linear** provided for all $L = [L_1 : p_1, \ldots, L_n : p_n] \in Ł$,

$$u(L) = \sum_{i=1}^{n} p_i \times u(L_i)$$

Von Neumann-Morgenstern Representation Theorem Suppose that (P, I) is a rational preference on the set $\mathcal{L}$ of lotteries. Then, (P, I) satisfies Compound Lotteries, Independence and Continuity if, and only if, (P, I) is represented by a **linear utility function**.

Moreover, $u' : Ł \to \mathbb{R}$ also represents (P, I) if, and only if, there are real numbers $c > 0$ and d such that for all lotteries $L \in \mathcal{L}$, $u'(\cdot) = cu(\cdot) + d$. ("$u$ is unique up to linear transformations.")

IIA says two things. First, it says that the comparison of two options should depend only on people's preferences, as opposed to the numerical values of utility. Second, it adds that the comparison must rely only on pairwise preferences about the two contemplated options.

IIA says two things. First, it says that the comparison of two options should depend only on people's preferences, as opposed to the numerical values of utility. Second, it adds that the comparison must rely only on pairwise preferences about the two contemplated options.

Ordinal Non-Comparability: Social preferences should remain the same when the profile of individual utility functions changes without altering individual preferences.

Independence of Irrelevant Utilities: Social preferences on two options should only depend on individual utilities at these two options.

Ordinal Invariance: Given two profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, let $\boldsymbol{U} \sim_{OM} \boldsymbol{U}'$ if for all $i \in V$ and $x, y \in X$, $\boldsymbol{U}_i(x) \geq \boldsymbol{U}_i(y)$ if and only if $\boldsymbol{U}'_i(x) \geq \boldsymbol{U}'_i(y)$.

For all profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, if $\boldsymbol{U} \sim_{OM} \boldsymbol{U}'$, then $f(\boldsymbol{U}) = f(\boldsymbol{U}')$.

Ordinal Invariance: Given two profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, let $\boldsymbol{U} \sim_{OM} \boldsymbol{U}'$ if for all $i \in V$ and $x, y \in X$, $\boldsymbol{U}_i(x) \geq \boldsymbol{U}_i(y)$ if and only if $\boldsymbol{U}'_i(x) \geq \boldsymbol{U}'_i(y)$.

For all profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, if $\boldsymbol{U} \sim_{OM} \boldsymbol{U}'$, then $f(\boldsymbol{U}) = f(\boldsymbol{U}')$.

Cardinal Measurability: Given two profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, let $\boldsymbol{U} \sim_{CM} \boldsymbol{U}'$ if for all $i \in V$, there are $\alpha_i, \beta_i \in \mathbb{R}$ with $\beta_i > 0$ such that for all $x \in X$, $\boldsymbol{U}_i(x) = \alpha_i + \beta_i \boldsymbol{U}'_i(x)$.

For all profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, if $\boldsymbol{U} \sim_{CM} \boldsymbol{U}'$, then $f(\boldsymbol{U}) = f(\boldsymbol{U}')$.

Ordinal Invariance: Given two profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, let $\boldsymbol{U} \sim_{OM} \boldsymbol{U}'$ if for all $i \in V$ and $x, y \in X$, $\boldsymbol{U}_i(x) \geq \boldsymbol{U}_i(y)$ if and only if $\boldsymbol{U}'_i(x) \geq \boldsymbol{U}'_i(y)$.

For all profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, if $\boldsymbol{U} \sim_{OM} \boldsymbol{U}'$, then $f(\boldsymbol{U}) = f(\boldsymbol{U}')$.

Cardinal Measurability: Given two profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, let $\boldsymbol{U} \sim_{CM} \boldsymbol{U}'$ if for all $i \in V$, there are $\alpha_i, \beta_i \in \mathbb{R}$ with $\beta_i > 0$ such that for all $x \in X$, $\boldsymbol{U}_i(x) = \alpha_i + \beta_i \boldsymbol{U}'_i(x)$.

For all profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, if $\boldsymbol{U} \sim_{CM} \boldsymbol{U}'$, then $f(\boldsymbol{U}) = f(\boldsymbol{U}')$.

Cardinal Unit Comparability: Given two profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, let $\boldsymbol{U} \sim_{CUC} \boldsymbol{U}'$ if there is a $\beta \in \mathbb{R}$ with $\beta > 0$ such that for all $i \in V$, there are $\alpha_i \in \mathbb{R}$ such that for all $x \in X$, $\boldsymbol{U}_i(x) = \alpha_i + \beta \boldsymbol{U}'_i(x)$.

For all profiles $\boldsymbol{U}$ and $\boldsymbol{U}'$, if $\boldsymbol{U} \sim_{CUC} \boldsymbol{U}'$, then $f(\boldsymbol{U}) = f(\boldsymbol{U}')$.

To prove his impossibility theorem, Arrow assumed OM invariance (and Sen generalized it to CM invariance)

Arrow's Epistemological Objection

> The viewpoint will be taken here that interpersonal comparison of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility...
>
> (Social Choice and Individual Values, p. 9)

Arrow's Epistemological Objection

> Even if...we should admit the measurability of utility for an individual, there is still the question of aggregating the individual utilities.

Arrow's Epistemological Objection

> Even if...we should admit the measurability of utility for an individual, there is still the question of aggregating the individual utilities. At best, it is contended that, for an individual, his utility function is uniquely determined up to a linear transformation; we must still choose one out of the infinite family of indicators to represent the individual, and the values of the aggregate (say a sum) are dependent on how the choice is made for each individual.

Arrow's Epistemological Objection

> Even if...we should admit the measurability of utility for an individual, there is still the question of aggregating the individual utilities. At best, it is contended that, for an individual, his utility function is uniquely determined up to a linear transformation; we must still choose one out of the infinite family of indicators to represent the individual, and the values of the aggregate (say a sum) are dependent on how the choice is made for each individual. In general, there seems to be no method intrinsic to utility measurement which will make the choices compatible.
>
> (Social Choice and Individual Values, pp. 10-11)

Arrow's Epistemological Objection

It requires a definite value judgment not derivable from individual sensations to make the utilities of different individuals dimensionally compatible and still a further value judgment to aggregate them according to any particular mathematical formula....

Arrow's Epistemological Objection

> It requires a definite value judgment not derivable from individual sensations to make the utilities of different individuals dimensionally compatible and still a further value judgment to aggregate them according to any particular mathematical formula....If we look away from the mathematical aspects of the matter, it seems to make no sense to add the utility of one individual, a psychic magnitude in his mind, with the utility of another individual. Even Bentham had his doubts on this point.
>
> (Social Choice and Individual Values, pp. 10-11)

Why not take interpersonally comparable utility functions as primitives?

"...the mythical character of any numbers behind preferences: they are just a construct in the "observing mathematician"'s mind, and without any uniqueness property in addition. But if until Arrow this ordinalist position was almost the consensus, apparently his theorem itself, together with the very influential work of Harsanyi, turned the tide partially, and led to the conclusion that interpersonal comparability was a must to obtain SWF's. The present theorem proves this conclusion false." (p. 473)

To summarize, we consider only preferences as empirically meaningful primitives, and hence axiomatize their aggregation coming thus back to the original formulation of social choice theory (Arrow 1963) as a guide to policy recommendations, not as the ultimate foundations of ethics. So a priori the question of interpersonal comparisons is not even meaningful in our framework. Following this approach requires that the set A consist of all feasible and just alternatives: the ethical question lies in the choice of A and it is this that will determine whatever implicit "interpersonal comparisons" occur. The questions of justice being. accounted for by the choice of A, the choice within A is to be handled by an appropriate generalization of majority voting: the SWF. (p. 477)

M. Pivato. *Condorcet meets Bentham*. Journal of Mathematical Economics, 59, pp. 58 - 65, 2015.

...given a mild assumption (called "concordance") about the statistical distribution of voter's preferences, we will show that the Condorcet winner actually maximizes utilitarian social welfare. We will then show that, if the voters' utility functions arise from certain plausible random processes, then a sufficiently large population of voters will have a concordant distribution of utility functions, with very high probability. In other words, in a large population satisfying certain statistical regularities, not only is the Condorcet winner almost guaranteed to exist, but it is almost guaranteed to also be the utilitarian social choice. So for such populations, Condorcet and Bentham agree. (p. 85-6)