Tutorial

Ranking Mechanisms in Games

Vanessa Volz and Boris Naujoks

CIG 2018, Maastricht

technische universität dortmund

Technology Arts Sciences TH Köln

UTOPIÆ Uncertainty Treatment and Optimisation in Aerospace Engineering

MARIE CURIE ALUMNI

SYNERGY Horizon 2020 GA No 692286

SYNERGY

Applications of Ranking Mechanisms

- **Hierarchy**
- Winner in a live event
- Matchmaking
- Handycapping
- Performance Thresholds
- In-game Decisions

Bundesliga standings

MATCHES NEWS STANDINGS STATS PLAYERS

	Club	MP	W	D	L	GF	GA	GD	Pts
1	Dortmund	34	23	6	5	67	22	45	75
2	Bayer	34	20	8	6	64	44	20	68
3	Bayern	34	19	8	7	81	40	41	65
4	Hannover 96	34	19	3	12	49	45	4	60
5	Mainz	34	18	4	12	52	39	13	58

Football (German Bundesliga 2011)

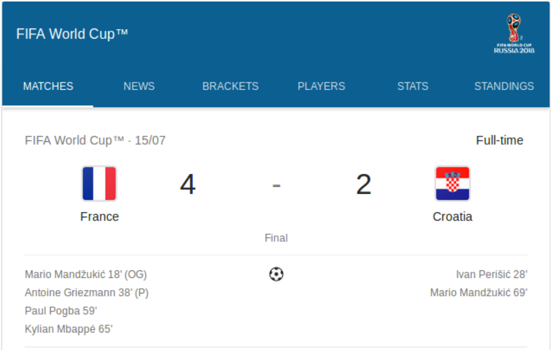
Applications of Ranking Mechanisms

- **Hierarchy**
- Winner in a live event
- Matchmaking
- Handycapping
- Performance Thresholds
- In-game Decisions

StarCraft II (Grandmaster League 18 Season 2)

Applications of Ranking Mechanisms

- Hierarchy
- **Winner in a live event**
- Matchmaking
- Handycapping
- Performance Thresholds
- In-game Decisions



Football (FIFA World Cup 2018)

Applications of Ranking Mechanisms

- Hierarchy
- **Winner in a live event**
- Matchmaking
- Handycapping
- Performance Thresholds
- In-game Decisions

AlphaGo vs. Lee Sedol (2016)

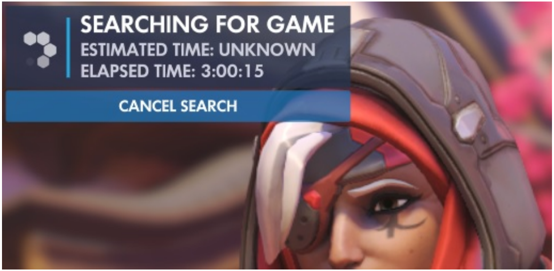
Applications of Ranking Mechanisms

- Hierarchy
- Winner in a live event
- **Matchmaking**
- Handycapping
- Performance Thresholds
- In-game Decisions

Hearthstone Queue

Applications of Ranking Mechanisms

- Hierarchy
- Winner in a live event
- **Matchmaking**
- Handycapping
- Performance Thresholds
- In-game Decisions



Overwatch Queue

Applications of Ranking Mechanisms

- Hierarchy
- Winner in a live event
- Matchmaking
- **Handycapping**
- Performance Thresholds
- In-game Decisions

Chess

Applications of Ranking Mechanisms

- Hierarchy
- Winner in a live event
- Matchmaking
- **Handycapping**
- Performance Thresholds
- In-game Decisions

Golf

Applications of Ranking Mechanisms

- Hierarchy
- Winner in a live event
- Matchmaking
- Handycapping
- **Performance Thresholds**
- In-game Decisions

National Collegiate Counter-Strike League

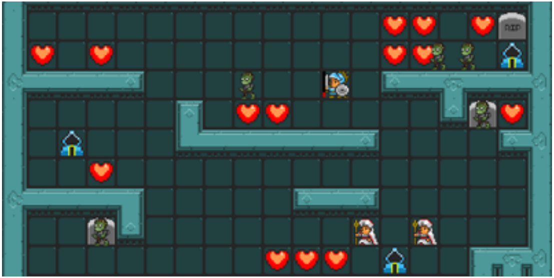
Applications of Ranking Mechanisms

- Hierarchy
- Winner in a live event
- Matchmaking
- Handycapping
- **Performance Thresholds**
- In-game Decisions

Smite Divisions

Applications of Ranking Mechanisms

- Hierarchy
- Winner in a live event
- Matchmaking
- Handycapping
- Performance Thresholds
- **In-game Decisions**



GVGAI

Ranking Mechanisms in CIG 18 Competitions

- Round Robin Tournament
  - Hearthstone AI
  - Fighting Game AI (Standard)
  - microRTS
  - StarCraft AI
- Average Score
  - Hanabi
  - Ms. Pac-Man vs. Ghost Team
  - Text-based adventure AI
  - Visual Doom AI (Deathmatch)
  - GVGAI
- Time to beat opponent
  - Fighting Game AI (Speedrun)
  - Visual Doom AI (Speedrun)
- Others
  - Short Video (Vote)
  - Hearthstone AI alt (Glicko2)
  - AI Birds: AI (Elim. tournament)
  - AI Birds: Level (Vote)

Why we're here!

- Various examples of ranking mechanisms in games
- But are they fair?

Social Choice Theory

- formalisation of characteristics
- recommendations for ranking mechanisms

Relations

- Relation R on a set X
  Subset of cartesian product $X \times X$:

$$R \subset X \times X$$

- Properties of relations
  - **reflexive**, if $\forall x \in X: \quad xRx$.
  - **symmetric**, if $\forall x, y \in X: \quad xRy \;\Rightarrow\; yRx$.
  - **anti-symmetric**, if $\forall x, y \in X: \quad xRy \;\wedge\; yRx \;\Rightarrow\; x = y$.
  - **transitive**, if $\forall x, y, z \in X: \quad xRy \;\wedge\; yRz \;\Rightarrow\; xRz$.

Examples for relations

Set of real number $\boldsymbol{R}$ and relation

"$<$" (less than)

- not reflexive ($x < x$ doesn't hold)
- not symmetric (from $x < y$ does not follow $y < x$)
- but anti-symmetric ($x < y$ and $y < x$ cannot hold both, hence implication is true)
- and transitive, (from $x < y$ and $y < z$ follows $x < z$)

"$\leqslant$" (less or equal)

- is reflexive ($x \leqslant x$ holds)
- not symmetric (in general $x \leqslant y$ does not imply $y \leqslant x$)
- but anti-symmetric ($x \leqslant y$ and $y \leqslant x$ implies $x = y$)
- and transitive (from $x < y$ and $y < z$ follows $x < z$)

Examples for relations

Set of real number $\boldsymbol{R}$ and relation

"$\neq$" (unequal)

- not reflexive ($x \neq x$ does not hold)
- but symmetric ($x \neq y \;\Rightarrow\; y \neq x$)
- not anti-symmetric ($x \neq y$ and $y \neq x$ do not imply $y = x$)
- and not transitive ($x \neq y$ and $y \neq z$ do not imply $x \neq z$; $x = z$ is still possible).

"$=$" (equal)

- is reflexive ($x = x$ holds)
- symmetrisch ($x = y \;\Rightarrow\; y = x$)
- anti-symmetric ($x = y$ and $y = x$ implies $x = y$)
- and transitive ($x = y$ and $y = z$ imply $x = z$)

Orders

- Relation R on set X is called **order** $:\Leftrightarrow$ R is
  - reflexive
  - anti-symmertric and
  - transitive
- Relation R on set X is called **linear** or **total order** $:\Leftrightarrow$ R is
  - an order
  - additionally:

$$\forall x, y \in X : \quad xRy \ \vee \ yRx$$

- Example
  - $(\boldsymbol{R}, <)$ is not an order, not reflexive
  - $(\boldsymbol{R}, \leqslant)$ is a total order

The social choice model

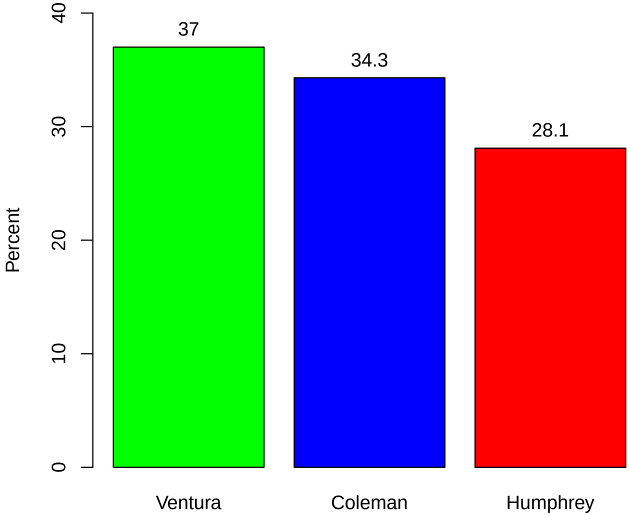
Social Choice Theory

- formalisation of characteristics
- recommendations for ranking mechanisms

How they correlate ...

- Finite set of n voters and finite set X of k choices or candidates
- In gaming competitions: n games and k players
- In racing competitions: n tracks and k drivers
- In algorithm comparision: n runs of k algorithms

1998 Minnesota governor election



Common Social Choice Example

Candidate	Votes
Jesse Ventura	37.0%
Norm Coleman	34.3%
Skip Humphrey	28.1%

Preference list			Perc. of voters
Coleman	Humphrey	Ventura	35%
Humphrey	Coleman	Ventura	28%
Ventura	Coleman	Humphrey	20%
Ventura	Humphrey	Coleman	17%

Ventura won, but

- 63% of voters liked him least!
- Coleman wins pairwise comparisons
  - 55% prefer Coleman to Humphrey
  - 63% prefer Coleman to Ventura

Easy example

- Imagine a racing competition featuring 7 tracks
- 3 drivers compete against each other: driver1, driver2, driver3

Preference list			Number of occurrences
driver1	driver2	driver3	3
driver2	driver1	driver3	2
driver3	driver2	driver1	2

- Who is the best driver?

And the winner is ...

Preference list			Number of occurrences
driver1	driver2	driver3	3
driver2	driver1	driver3	2
driver3	driver2	driver1	2

- **driver1 !**
  - Wins on most tracks
- **driver2 !**
  - Outperforms driver1 on 4 of 7 tracks

What?!? How ... ?!? 😱

The social choice model

- $L(X)$ set of all preference lists
  i.e. set of all possible strict linear orders of X (no ties allowed)
- $O(X)$ set of all preference lists
  i.e. set of all possible linear orders of X (ties allowed)
- Profile or election is element of cartesian product $L(X)^n$
  i.e. set of n preference lists, one from each voter (game, track)
- Ranking mechanism in games (social choice function or voting method) is function

  $$F : L(X)^n \rightarrow O(X).$$

  For given profile $R \in L(X)^n$, image $F(R)$ is called the ranking (social choice or societal ranking)

Examples of social choice functions, rankings

- Plurality (also called majority)
  - Candidates are ranked by number of first-place rankings
  - Winner(s) is/are candidate(s) with the most first-place rankings
  - Method is used in many elections
    including many local and state elections in US and partly German Bundestag
- Antiplurality
  - Candidate with least last-place rankings wins
  - Candidates ranked from last to first by the number of last-place rankings they receive

Club president election example

- Anne (A), Brigitte (B), Claus (C), and David (D) running for president of a club
- club has 27 members
- 24 possible preference lists, but for this example only 4 are used

Preference list	**Number of occurrences**
A B C D	12
B C D A	7
C D A B	5
D C B A	3
Other preferences	0

- Who is the elected for president?

And the winner is ...

Preference list	Number of occurrences
A B C D	12
B C D A	7
C D A B	5
D C B A	3
Other preferences	0

- **Anne !**
  - Plurality
  - 12 (most) first-place votes
- **Claus !**
  - Antiplurality
  - No (least) last-place votes

Examples of social choice function

- **Instant runoff**
  - Candidate(s) with the least first-place rankings is/are removed
  - New set of preference lists for a smaller set of candidates
  - Repeated until all candidates are eliminated
  - Social choice is formed by listing candidates in reverse order in which they were eliminated
  - Used for elections in Australia and for presidential elections in Ireland.

And the winner is ...

Preference list	Number of occurrences
A B C D	12
B C D A	7
C D A B	5
D C B A	3
Other preferences	0

- Instant runoff
  - David eliminated first

Preference list	No. of occu.
A B C	12
B C A	7
C A B	5
C B A	3
Other preferences	0

- continued ...
  - Brigitte is eliminated second

Preference list	No. of occu.
A C	12
C A	7
C A	5
C A	3
Other preferences	0

  - Anne is emilinated last

⇒ **Claus !**

Examples of social choice functions

- **Borda count**
  - With k candidates
    - ⋆ k - 1 points are given for a first place ranking
    - ⋆ k - 2 points for a second place ranking
    - ⋆ and so on ...
  - Candidates ranked by total sum of points they receive
  - Candidate(s) with the most points win(s)
  - Method (or derivatives) used frequently for sports-related polls

And the winner is ...

Preference list	Number of occurrences
A B C D	12
B C D A	7
C D A B	5
D C B A	3
Other preferences	0

- Borda count
  - Anne: $(12 \times 3) + (5 \times 1) = 41$
  - Brigitte: $(12 \times 2) + (7 \times 3) + (3 \times 1) = 48$
  - Claus: $(12 \times 1) + (7 \times 1) + (5 \times 3) + (3 \times 2) = 47$
  - David: $(7 \times 1) + (5 \times 2) + (3 \times 3) = 26$

$\Rightarrow$ **Brigitte !**

Wait ...

- Anne won wrt. Plurality
- Claus won according to Antiplurality
- Claus won again wrt. Instant runoff
- Brigitte won wrt. Borda count

What?!? How ... ?!? 😱

- Three different winners using four methods?
- So winner is depending on voting method?
- **Does this seem reasonable?**

Condorcet

- Marquis Nicolas de Condorcet (1743–1794)
- French liberal thinker in the era of the French Revolution
- philosopher, mathematician, and political scientist
- Pursued by the revolutionary authorities for criticizing them
- Died in prison
- Essay on the Application of Analysis to the Probability of Majority Decisions (1785):

Essay sur l’Application de l’Analyse á la Probabilité des Décisions Rendue á la Pluralité des Voix

Condorcet's 2 prominent insights

- **Condorcet's jury theorem**
  - Each member of jury has chance of making a correct judgment on whether a defendant is guilty
    - ⋆ equal and independent
    - ⋆ better than random
    - ⋆ worse than perfect

⇒ majority of jurors is more likely to be correct than each individual juror

⇒ Probability of correct majority judgment approaches 1 as jury size increases

⇒ Under certain conditions, majority rule is good at ‘tracking the truth’

Condorcet's 2 prominent insights

- **Condorcet's paradox**

  Majority preferences can be 'irrational' (intransitive)

  – even when individual preferences are 'rational' (transitive).

  – Example

Preference list	**No. of occu.**
A B C	1 / 3
B C A	1 / 3
C A B	1 / 3

⇒ there are majorities (of two thirds)

- ⋆ for A against B
- ⋆ for B against C
- ⋆ for C against A

⇒ **Cycle** violates transitivity

Condorcet

- **Condorcet winner**
  Candidate who beat all other candidates in head-to-head contests
  - Examples
    - ⋆ No Condorcet winner in Condorcet's paradox
    - ⋆ Coleman in 1998 Minnesota governor election example
- **Condorcet loser**
  Candidate who loses to all other candidates in head-to-head contests

Condorcet winner criterion

> Whenever there is a Condorcet winner,
> that candidate is the unique winner of the election.

- Plurality does not satisfy Condorcet winner criterion

Ventura won, but

- 63% of voters liked him least!
- Coleman wins pairwise comparisons
  - 55% prefer Coleman to Humphrey
  - 63% prefer Coleman to Ventura

  - Coleman was the Condorcet winner
  - Ventura won

Condorcet winner criterion

Whenever there is a Condorcet winner,
that candidate is the unique winner of the election.

- Borda count does not satisfy Condorcet winner criterion
  - Exampel

Preference list	No. of occu.
A B C	3
B C A	2

  - Condorcet winner is A
  - Borda count
    - ⋆ A: $(2 \times 3) + (0 \times 1) = 6$
    - ⋆ B: $(1 \times 3) + (2 \times 2) = 7$
    - ⋆ C: $(0 \times 3) + (2 \times 1) = 2$

    $\Rightarrow$ **B is winner**

- Btw: Instant runoff does not either

Condorcet winner criterion

Whenever there is a Condorcet winner,
that candidate is the unique winner of the election.

- Plurality does not satisfy Condorcet winner criterion
- Borda count does not satisfy Condorcet winner criterion
- Instant runoff does not ...

- Are there any?
- Yes, there are!
- However, all of them run into other problems

What about other criteria?

Condorcet winner criterion example

- Choose winner based on head-to-head contests

⇒ Make sure Condorcet winner criterion is satisfied

- Example: Sequential pairwise voting
  - fix an (arbitrary) order of candidates
  - rounds of head-to-head contests between candidates following fixed order
  - winner of contest between the first two goes up against third candidate ...
  - until one candidate survives

- Satisfies the Condorcet winner criterion
  - Condorcet winner will beat everyone else on the list

Condorcet winner criterion example

Preference list	Number of occurrences
A B C D	12
B C D A	7
C D A B	5
D C B A	3
Other preferences	0

- Fixed ordering A B C D
  - 17 voters prefer A to B, only 10 B to A
  - ⇒ A beats B 17:10
  - C beats A 15:12
  - C beats D 15:12

  ⇒ C is the winner
- Fixed ordering A C B D ⇒ B is the winner
- Fixed ordering B C A D ⇒ D is the winner
- Fixed ordering B C D A ⇒ A is the winner

What?!? How ... ?!? Not good !!! 😱

Some formalism required

- Set $N = 1, 2, ..., n$ of individuals $(n \geqslant 2)$
- Set of social alternatives $X = x, y, z, ...$
- Each individual $i \in N$ has a preference ordering R_i over alternatives:
  - $\Rightarrow$ complete and transitive relation on X
- For any $x, y \in X$: xR_iy means that individual i prefers x to y
- xP_iy if xR_iy and not yR_ix ('individual i strictly prefers x to y')
- Profile

$$< R_1, R_2, ..., R_n >$$

combination of preference orderings across individuals

Some more formalism required

- Preference aggregation rule F
  - function that assigns to each profile a social preference relation $R = F(R_1, R_2, ..., R_n)$ on X

$$F : < R_1, R_2, ..., R_n > \; \rightarrow \; R = F(R_1, R_2, ..., R_n)$$

- Example: pairwise majority voting (Condorcet)
  - For any profile $< R_1, R_2, ..., R_n >$ and any $x, y \in X$:

    xRy if and only if at least as many individuals have xR_iy as have yR_ix

    or

$$|i \in N : xR_iy| \geqslant |i \in N : yR_ix|$$

Alternative criteria

- **Independence of irrelevant alternatives**

  Description

  – Candidate A is ranked higher than candidate B
  – Some voters change their preference lists, but no voter changes their preference between A and B
  ⇒ A should remain ranked higher than B

Societal preference between two candidates should depend only on the voters' preferences between A and B

  Mathematical formulation:

  – For any two profiles $< R_1, R_2, ..., R_n >$ and $< R_1^*, R_2^*, ..., R_n^* >$
  – For any $x, y \in X$
  – if for all $i \in N$:
  R_i's ranking between x and y coincides with R_i^*'s ranking between x and y

$$xRy \text{ if and only if } xR_y^*.$$

Alternative criteria

- **Independence of irrelevant alternatives**

Example: 1995 Women's Figure Skating World Championship

Ranking before last skater:

1. Chen Lu (China)
2. Nicole Bobek (US)
3. Suraya Bonaly (France)

last skater:
Michelle Kwan (US), who became 4.

Ranking after last skater:

1. Chen Lu (China)
2. Suraya Bonaly (France)
3. Nicole Bobek (US)
4. Michelle Kwan (US)

- Note: Nicole Bobek (US) and Suraya Bonaly (France) (2nd and 3rd before last skater) changed places!

Alternative criteria

- **Monotonicity**

  Description

  - Some voters move candidate A up in their preference lists
  - No voters move A down

  $\Rightarrow$ A cannot move down in the final ranking

  Mathematical formulation

  - For any profile $< R_1, R_2, ..., R_n >$ in the domain of F
  - Social preference relation R is complete and transitive

Restaurant type example

- 17 conference attandancees
- 4 suggestions for dinner restaurant type
- Selected method: Instant runoff

Preference list	Number of occurrences
Thai Chinese Italian German	6
Chinese Thai Italian German	5
Italian German Chinese Thai	4
German Italian Thai Chinese	2
Other preferences	0

- Which type of restaurant to choose for dinner?

And the winner is ...

Preference list	Number of occurrences
Thai Chinese Italian German	6
Chinese Thai Italian German	5
Italian German Chinese Thai	4
German Italian Thai Chinese	2
Other preferences	0

- Instant runoff:
  - German eliminated first
  - Chinese eliminated second
  - Italian eliminated last

  ⇒ Thai is the winner!

- Right before leaving, two voters from last row changed their mind
  - German Italian Thai Chinese
  - replaced by
  - German Thai Italian Chinese

And the winner is ...

Preference list	Number of occurrences
Thai Chinese Italian German	6
Chinese Thai Italian German	5
Italian German Chinese Thai	4
German Thai Italian Chinese	2
Other preferences	0

- Instant runoff:
  - German eliminated first
  - Italian eliminated second
  - Thai eliminated last

  ⇒ Chinese is the winner!

What?!?

Thai moved up in some preferences and went from winning to losing!

Arrow's list of conditions

- **Universal domain**
  - Voters can choose any possible preference order
  - The domain of F is the set of all logically possible profiles of complete and transitive individual preference orderings.
- **Ordering**
  This is monotonicity or ordering as discussed above
- **Independence of irrelevant alternatives**
  As discussed above

Arrow's list of conditions

- **Weak Pareto principle**
  - If all voters prefer x over y, this should hold for final ranking
  - For any profile $< R_1, R_2, ..., R_n >$ in the domain of F
  - If for all $i \in N$: xP_iy then xPy
- **Nondictatorship**
  - There should not be a dictator
  - One voter whose preference list determines the societal ranking completely.
  - There does not exist an individual $i \in N$ such that
    - ⋆ for all $< R_1, R_2, ..., R_n >$ in the domain of F
    - ⋆ for all $x, y \in X$
  - xP_iy implies xPy.

Main Result

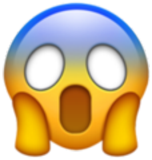
- Kenneth Joseph Arrow (1921 - 2017)
- American economist, mathematician, writer, and political theorist
- 1972 joint winner of the Nobel Memorial Prize in Economic Sciences (with John Hicks)
- Many of his former graduate students won the Nobel Memorial Prize themselves
- Most significnt contribution:

Arrow's impossibility theorem (1951)

If there are more than two candidates, then any social choice method cannot satisfy all of Arrow's five conditions.

Consequences and Implications

- All social choice methods have flaws
- Even most that are used for politcal elections throughout the world
- Also holds for most ranking methods
  - in sports
  - in games
  - etc



Consequences and Implications

- Weakening or relaxing conditions

  Works with different conditions and corresponding methods

- Example: independence of irrelevant alternatives
  - Intensity of voters' preference between two candidates
  - Number of other candidates listed between the two candidates
  - Intensity of binary independence criterion:
    - ⋆ If some voters change their preference lists
    - ⋆ No voter changes their preference between candidates A and B or the intensity of their preference

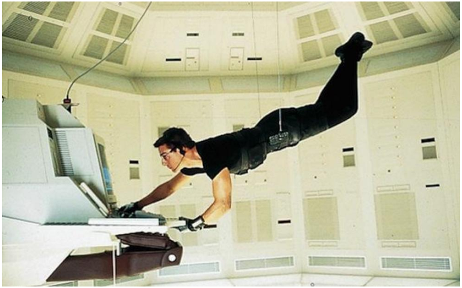
    $\Rightarrow$ Ranking of A and B in the social choice should not change

Borda count satisfies the conditions of Arrow's theorem with

- independence of irrelevant alternatives replaced by
- intensity of binary independence

Challenges for Ranking in Games

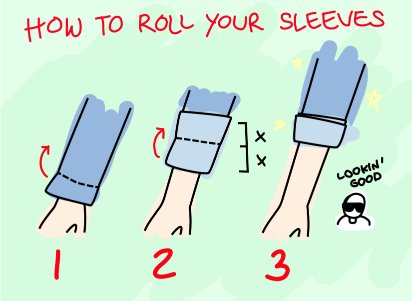
- Arrow's impossibility theorem
- Statistical Comparisons
- Choice of Fitness Functions
- Choice of Test Cases



Practical Recommendations

How to decide on a ranking method

- Rarity of criteria
- Perceived fairness
- Simplicity for transparency

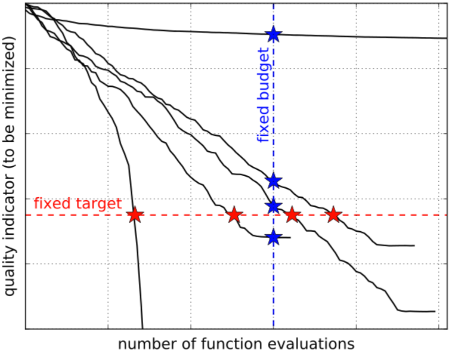


Overview of existing methods
`https://en.wikipedia.org/wiki/Comparison_of_electoral_systems`

Alternative: Mechanism Design

Lessons from EC Benchmarking

- Relevancy
- Fixed targets vs. fixed runtime
- Characteristics of problems (ELA)
- Evaluation Robustness (instances)
- Expected runtime measure
- Easy comparisons
- Info on optima?



Open Problems

- Selection of Ranking Mechanism (Which criteria can be relaxed?)
- Characterisation of problems (How do we guarantee completeness?)
- Long-term Ranking (How does ELO fit in?)
- Appropriate and practical noise handling
- Game Evaluation Measures

Games Benchmark

References

SCT2013 Ch. List, *Social Choice Theory* in Stanford Encyclopedia of Philosophy, 2013, https://plato.stanford.edu/entries/social-choice/ (last accessed 13.08.2018)

Pow2015 V. Powers, *How to choose a winner: the mathematics of social choice* in Snapshots of modern mathematics No9/2015 from Oberwolfach No 9, 2015, https://publications.mfo.de/handle/mfo/452 (last accessed 13.08.2018)

Gro2005 Ralf Grötker, *Kaputte Wahlen*, 2005 http://www.heise.de/-3402700 (last accessed 13.08.2018)