Junta Distributions and the Average Case Complexity of Manipulating Elections

A. D. Procaccia & J. S. Rosenschein

Lecture outline

- Introduction to social choice theory and voting
- A few examples, a few intuitions, a few axioms – Arrow, Gibbard-Satterthwaite
- Manipulation
- Scoring protocols

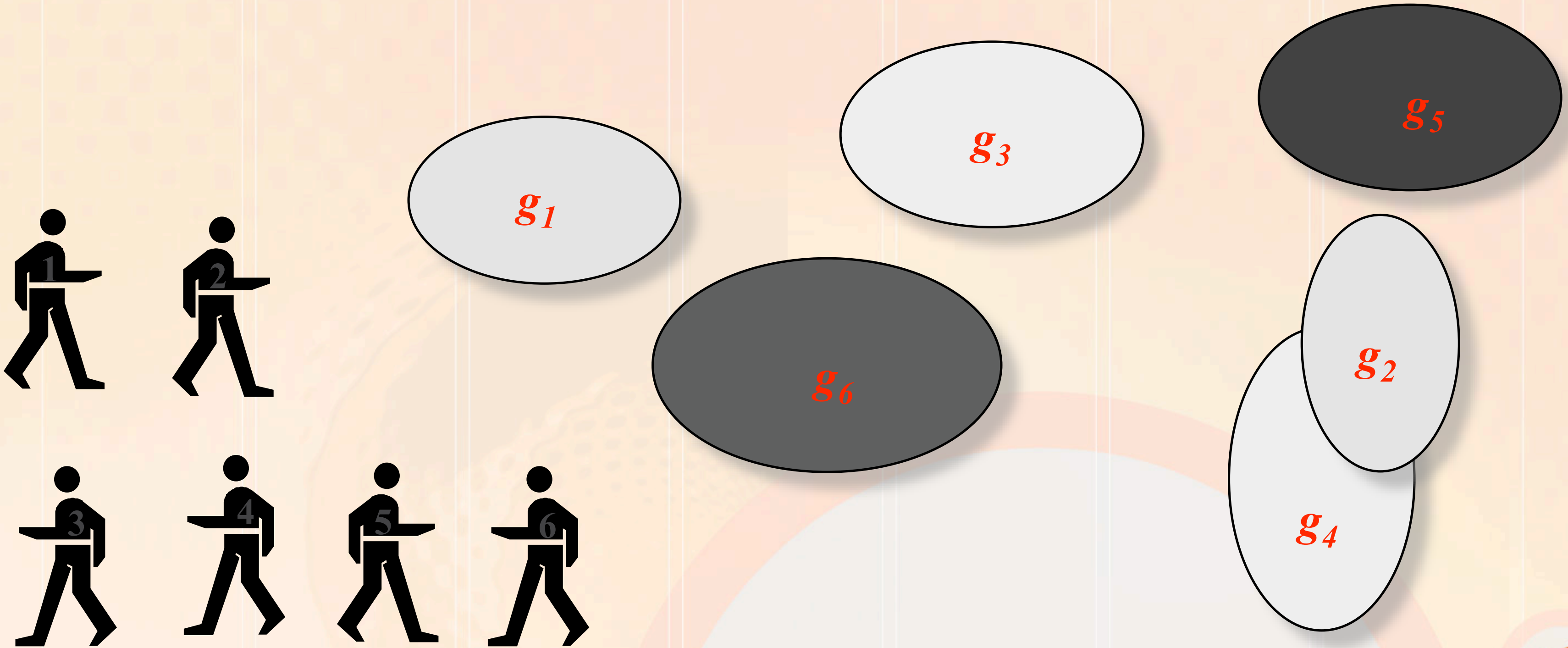
Background

- Our average case analysis
  - Junta distributions
- Manipulating scoring protocols is NP-hard, but easy in the average-case
- Conclusions

Recent Research

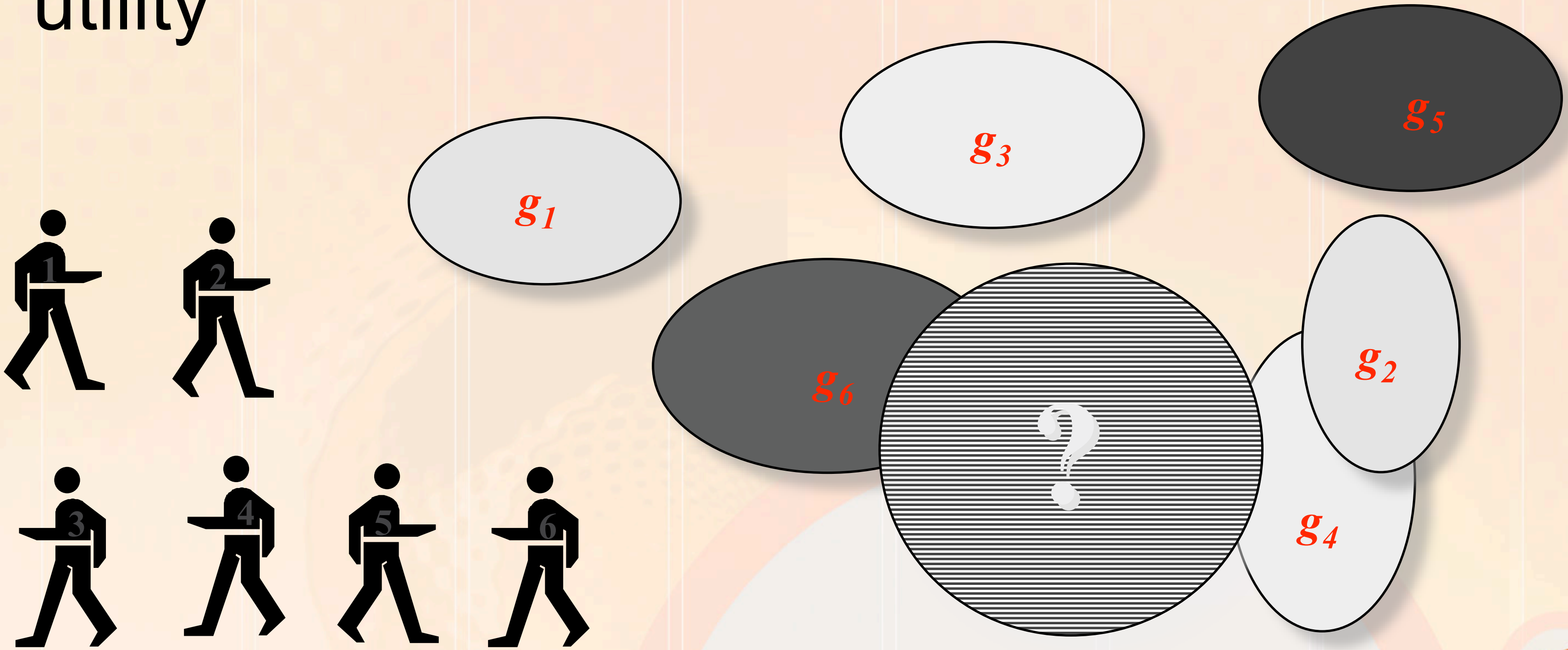
One Motivation

Agents need to reach consensus about what to do next in shared environment



Aim of the Process

Search for a joint plan that brings agents to the consensus state that optimizes global utility

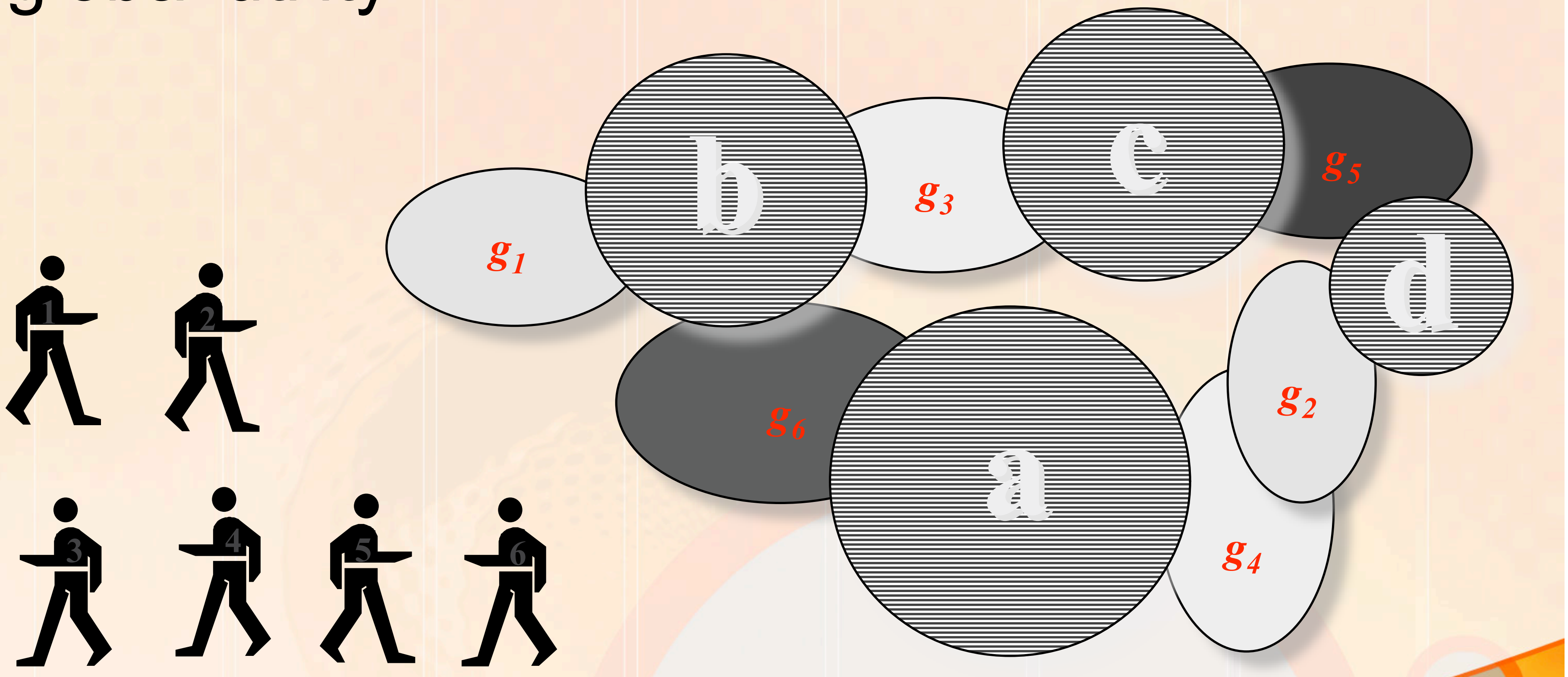


Alternative Routes to Consensus

- Centralized planning
- (Generalized) Partial Global Plans
- Negotiation (in many forms)
- Market mechanisms
- Synchronization of pre-existing plans
- Voting, a means of preference aggregation
  - Agents reveal preferences by ranking candidates
  - Winner determined by a voting protocol

Ordering Candidate States

Choosing the consensus state that optimizes global utility



Social Choice Theory

- Studies how decisions are made among a collection of alternatives, when there are voters with separate opinions
- Group choice should reflect the individual voters’ desires (by some definition, as much as possible…)

Ordinal Voting Methods

- Group of voters with ranked ordinal preference over more than two alternatives, decide on an *ordering*, or on a *choice*
- Example (order of preferences over candidates a, b, c, d):

1 voter	1 voter	1 voter
a	c	b
b	a	d
d	b	c
c	d	a

Arrow's Impossibility Theorem

- **Universality**: should create a deterministic, complete social preference order from every possible set of individual preference orders
- **Citizen sovereignty**: every possible order should be achievable by some set of individual preference orders
- **Non-dictatorship**: the social welfare function should be sensitive to more than the wishes of a single voter
- **Monotonicity**: change favorable to candidate *x* does not hurt *x*
- **Independence of irrelevant alternatives**: if we restrict attention to a subset of options and apply the social welfare function only to those, then the result should be compatible with the outcome for the whole set of options
- **No system meets all these criteria when there are two or more voters, and three or more choices**

Gibbard–Satterthwaite Theorem

(Regarding systems that choose a single winner)

- For three or more candidates, one of the following three things must hold for every voting rule:
  - The rule is dictatorial; or
  - There is some candidate who cannot win, under the rule, in any circumstances; or
  - The rule is manipulable

Manipulations

- Voters may prefer to reveal their intentions untruthfully

Manipulations

- Voters may prefer to reveal their intentions untruthfully
- Can happen in the full knowledge case (where a manipulating voter knows others' votes), or strategically (heuristically) without full knowledge of others' votes
- This is undesirable, since the outcome may be one that does not maximize social welfare

A Few Examples, A Few Criteria

- Sequential Pairwise Voting
- Pareto Criterion
- Plurality Voting
- Condorcet Winner Criterion
- Plurality with Run-off
- Monotonicity Criterion
- The Borda Count
- **Scoring Protocols**

Ordinal Voting Methods

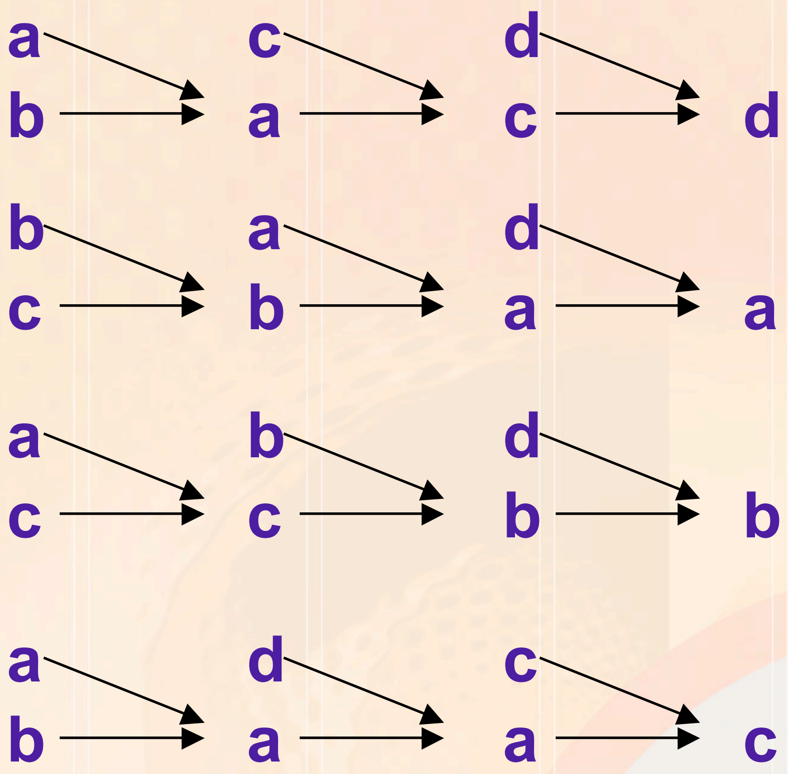
- A group of voters, with ranked ordinal preferences over more than two alternatives, have to decide on a choice
- Example:

1 voter	1 voter	1 voter
a	c	b
b	a	d
d	b	c
c	d	a

Sequential Pairwise Voting

Different possible agendas

a	c	b
b	a	d
d	b	c
c	d	a



a b c a d c d

Agenda i

b c a b d a a

Agenda ii

a c b c d b b

Agenda iii

a b d a c a c

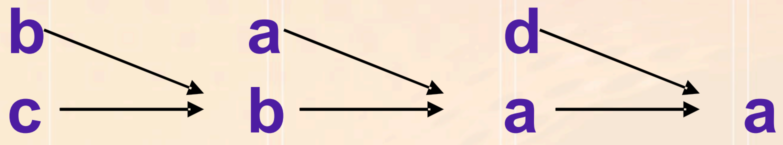
Agenda iv

In this example, anyone can be a winner!

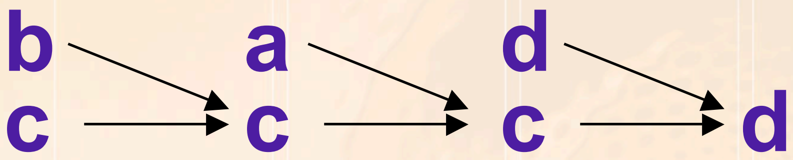
Rule of thumb: bring up your favorite as late as possible

Manipulation

- The vote, of course, is also susceptible to insincere, manipulative voting
- Example: third voter votes insincerely for c instead of for b (just in first election):



Agenda ii



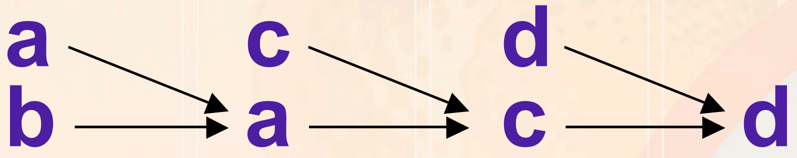
Agenda ii (insincere third voter)

a	c	b
b	a	d
d	b	c
c	d	a

Third voter gets second choice (d) instead of last choice (a), by lying

Pareto Criterion

- *“If every voter prefers an alternative x to an alternative y, a voting rule should not produce y as a winner”*
- Sequential pairwise voting violates this criterion; for example, in Agenda i, d wins, even though everyone prefers b to d



Agenda i

a	c	b
b	a	d
d	b	c
c	d	a

Plurality Voting

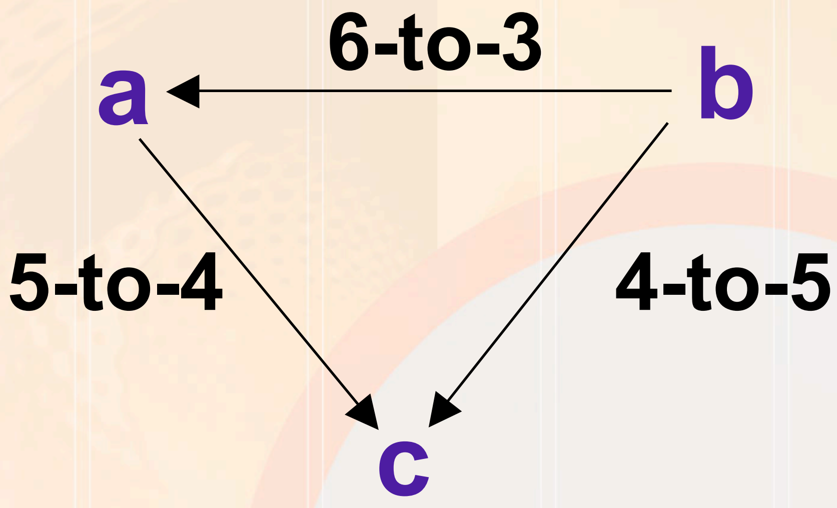
- Each voter votes for one alternative; the one with the most votes wins
- Example (9 voters):

3 voters	2 voters	4 voters
a	b	c
b	a	b
c	c	a

- Plurality voting has c winning, even though 5-4 majority rate c last

Even More Disturbing…

- In pairwise decisions, b, which came in last in the plurality vote, would beat both c and a, and c, which won the plurality vote, would have lost all pairwise contests:



3	2	4
a	b	c
b	a	b
c	c	a

Condorcet Winner Criterion

- *“If there is an alternative x that would win in pairwise contests against every other alternative, a voting rule should choose x as the winner”*
- If such an *x* exists, it is unique and is called the Condorcet winner
- Often there is no Condorcet winner
- Sequential pairwise voting, despite its other faults, does satisfy the Condorcet Winner Criterion

Plurality with Run-off

- Example (17 voters):

6 voters	5 voters	4 voters	2 voters	
a	c	b	b	a
b	a	c	a	b
c	b	a	c	c

- Plurality voting, a and b are top two, and a beats b in run-off by 11-to-6
- But, if last two voters changed their minds in favor of a, i.e., a b c instead of b a c, then a and c are top two, and c beats a by 9-to-8
- a gets more first place votes, and loses an election it would have won!

Monotonicity Criterion

- *"If x is a winner under a voting rule, and one or more voters change their preferences in a way favorable to x (without the changing the order in which they prefer any other alternatives), then x should still be the winner"*
- Straight plurality voting satisfies monotonicity
- Plurality with a run-off violates it
- They of course both tempt voters to vote insincerely

The Borda Count

- Each voter submits preferences over the *n* alternatives
- Each alternative receives
  - no points for being ranked last
  - 1 point for being ranked second-to-last
  - …
  - up to *n*-1 points for being ranked first
- Points for each alternative are summed across all voters, and the alternative with the highest total is the winner

Borda Count Example

1 voter	1 voter	1 voter	
a	c	b	3 points
b	a	d	2 points
d	b	c	1 point
c	d	a	0 points

- With Borda count, a gets 3 points from first voter, 2 points from the second, and 0 from the third
- Final Borda count totals: a:5, b:6, c:4, d:3
- b is the Borda winner

Advantages of the Borda Count

- Uses information from entire preference rankings of the voters (not just first or last rankings)
- Chooses the alternative that occupies the highest position on the average in the voters' preference rankings
  - *x*'s Borda count, divided by the number of voters, is the average number of alternatives ranked below *x*
- The Borda winner should be "broadly acceptable"

Advantages of the Borda Count

- The Borda count equals the number of votes an alternative would get in pairwise contests with the other alternatives (if all voters have strict preference orderings), and also equals the sum of items ranked below it across all voters
- The Borda count satisfies the Pareto condition, and the Monotonicity condition (and others)
- The Borda count does not satisfy the Condorcet winner criterion…

Violate Condorcet Winner Criterion

3 voters	2 voters	
a	b	2 points
b	c	1 point
c	a	0 points

- Borda counts, a:6, b:7, c:2
- b wins, but a is the Condorcet winner
- Even worse: a has an absolute majority of first place votes!
- The existence of c allows the 2 voters to weight b over a more heavily than the 3 voters choose to weight a over b, enabling b to win the Borda count

Scoring Protocols

- $\alpha = <\alpha_1, \dots, \alpha_m>$ where $\alpha_i \geq \alpha_{i+1}$. Candidate receives α_i points for each voter that ranks it in *i*'th place
- Examples:
  - Plurality: <1, 0, …, 0>
  - Veto: <1, …, 1, 0>
  - Borda: <*m*-1, *m*-2, …, 0>
- **Sensitive** scoring protocols are scoring protocols where $\alpha_{m-1} > \alpha_m=0$
  - Including Veto and Borda

Lecture outline

- Introduction to social choice theory and voting
- A few examples, a few intuitions, a few axioms – Arrow, Gibbard-Satterthwaite
- Manipulation
- Scoring protocols

Background

- Our average case analysis
    - Junta distributions
- Manipulating scoring protocols is NP-hard, but easy in the average-case
- Conclusions

Recent Research

Coalitional Manipulation

- Voting protocol is non-dictatorial implies there are elections where an agent is better off voting untruthfully
- Coalitional Manipulation: Given a set S of weighted votes (i.e., other voters' choices are known), a set T of manipulators' weights, and a candidate *p*. Can the votes in T be cast so that *p* wins?
- Manipulation is (presumably) undesirable
- Bounded rationality comes to the rescue!

Complexity as Scourge or Savior

- Computational complexity can be an obstacle to desirable computations
- Computational complexity can be an obstacle to *undesirable* computations
- Example: RSA encryption
- Manipulation is (basically) always possible, but if it’s too hard to calculate, perhaps a voting process can be manipulation-resistant

Some Previous Results

- [Bartholdi and Orlin 1991] There are voting protocols that are NP-hard for a single voter to manipulate
- [Conitzer and Sandholm 2002, 2003a] Some manipulations of common voting protocols are NP-hard, even for a small number of candidates
- [Conitzer and Sandholm 2003b] Adding a pre-round to some voting protocols can make manipulation hard (even PSPACE-hard in some cases)

NP-hard manipulations

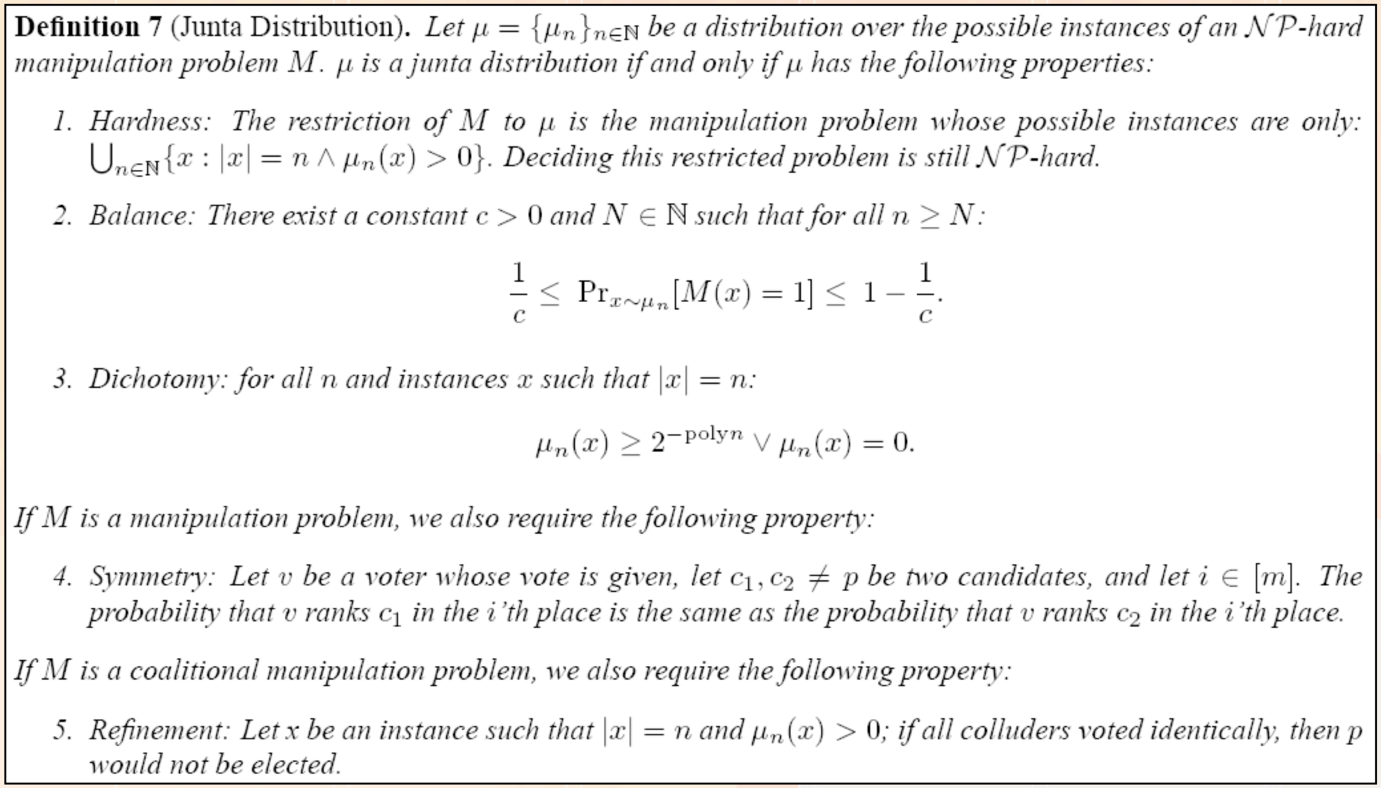
- Individual manipulation of some protocols is NP-hard when the number of candidates *m* is large
- We proved coalitional manipulation of sensitive scoring protocols is NP-hard, even when *m*=3 (generalization of Conitzer/Sandholm result)
- But…*this may be a weak guarantee of resistance to manipulation*
- Given a reasonable distribution, how hard is it to manipulate?

Average Case Analysis

- Traditional average case complexity theory seems inappropriate for our purposes
- **Distributional problem** = <M,μ>; M is a decision (manipulation) problem, μ is a distribution over the possible inputs
- Algorithm A is a **heuristic polynomial time algorithm** for <M,μ> if A runs in polynomial-time, and ∃*p* s.t. ∀*x* of size *n*: $\Pr_{\mu}[A(x) \neq M(x)] \leq 1/p(n)$

Junta Distributions

- If an algorithm succeeds in deciding instances drawn from a junta distribution, it will also succeed with most reasonable distributions



Definition 7 (Junta Distribution). *Let $\mu = \{\mu_n\}_{n \in \mathbb{N}}$ be a distribution over the possible instances of an $\mathcal{NP}$-hard manipulation problem M. μ is a junta distribution if and only if μ has the following properties:*

1. *Hardness: The restriction of M to μ is the manipulation problem whose possible instances are only: $\bigcup_{n \in \mathbb{N}} \{x : |x| = n \wedge \mu_n(x) > 0\}$. Deciding this restricted problem is still $\mathcal{NP}$-hard.*
2. *Balance: There exist a constant $c > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$:*

$$\frac{1}{c} \leq \Pr_{x \sim \mu_n}[M(x) = 1] \leq 1 - \frac{1}{c}.$$

3. *Dichotomy: for all n and instances x such that $|x| = n$:*

$$\mu_n(x) \geq 2^{-\mathrm{poly} n} \vee \mu_n(x) = 0.$$

If M is a manipulation problem, we also require the following property:

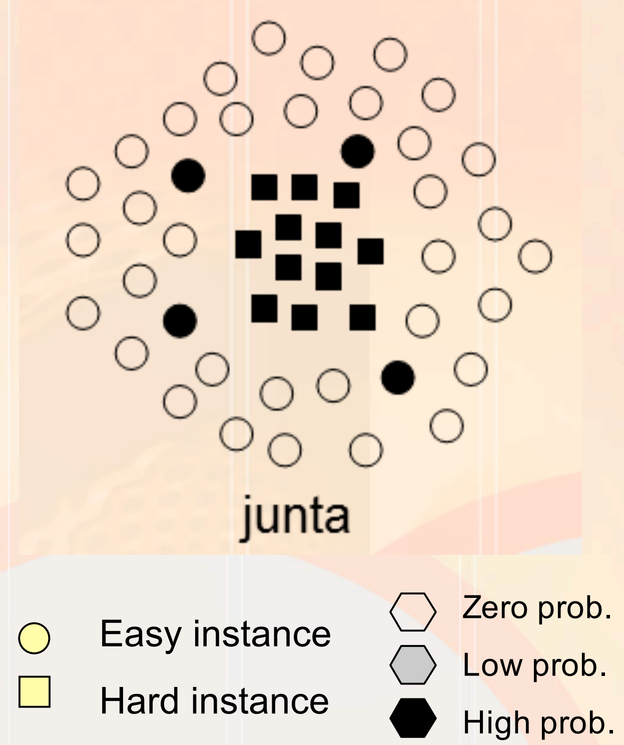
4. *Symmetry: Let v be a voter whose vote is given, let $c_1, c_2 \neq p$ be two candidates, and let $i \in [m]$. The probability that v ranks c_1 in the i'th place is the same as the probability that v ranks c_2 in the i'th place.*

If M is a coalitional manipulation problem, we also require the following property:

5. *Refinement: Let x be an instance such that $|x| = n$ and $\mu_n(x) > 0$; if all colluders voted identically, then p would not be elected.*

Junta Distributions

- If an algorithm succeeds in deciding instances drawn from a junta distribution, it will also succeed with most reasonable distributions



○ Easy instance

□ Hard instance

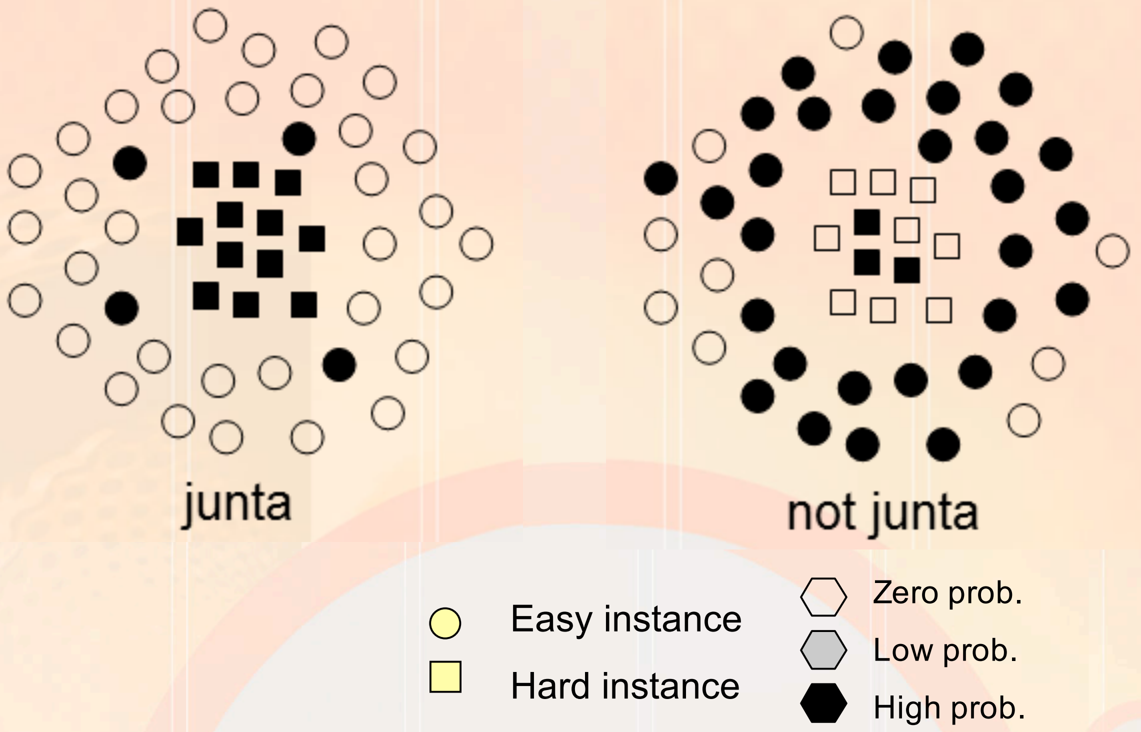
⬡ Zero prob.

⬡ Low prob.

⬢ High prob.

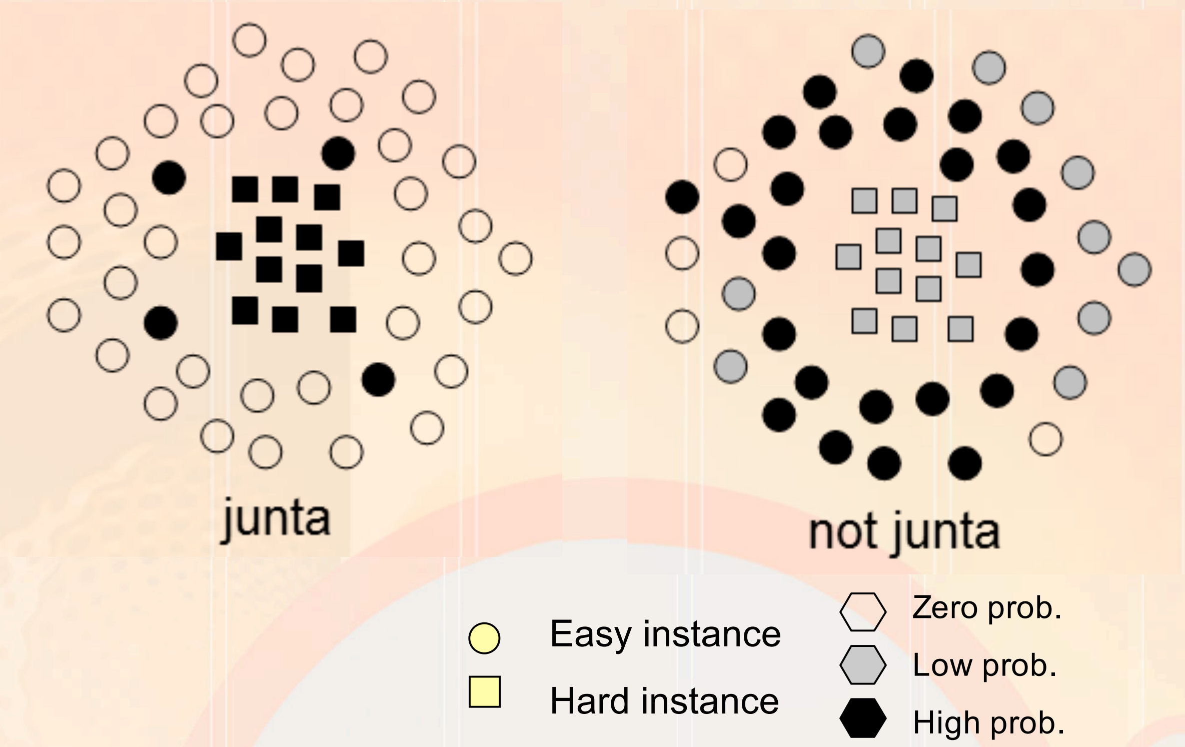
Junta Distributions

- If an algorithm succeeds in deciding instances drawn from a junta distribution, it will also succeed with most reasonable distributions
- Properties:
  - Hardness: Still enough hard instances



Junta Distributions

- If an algorithm succeeds in deciding instances drawn from a junta distribution, it will also succeed with most reasonable distributions
- Properties:
  - Hardness: Still enough hard instances
  - Dichotomy: Instances are either probable or impossible



Junta Distributions

Additional Properties:

- Balance: Can't answer the decision problem correctly by always saying "yes", or always saying "no"
- Symmetry: A voter is as likely to vote for one candidate as for another
- Refinement: Manipulation fails if all colluders vote identically

Susceptibility to manipulation

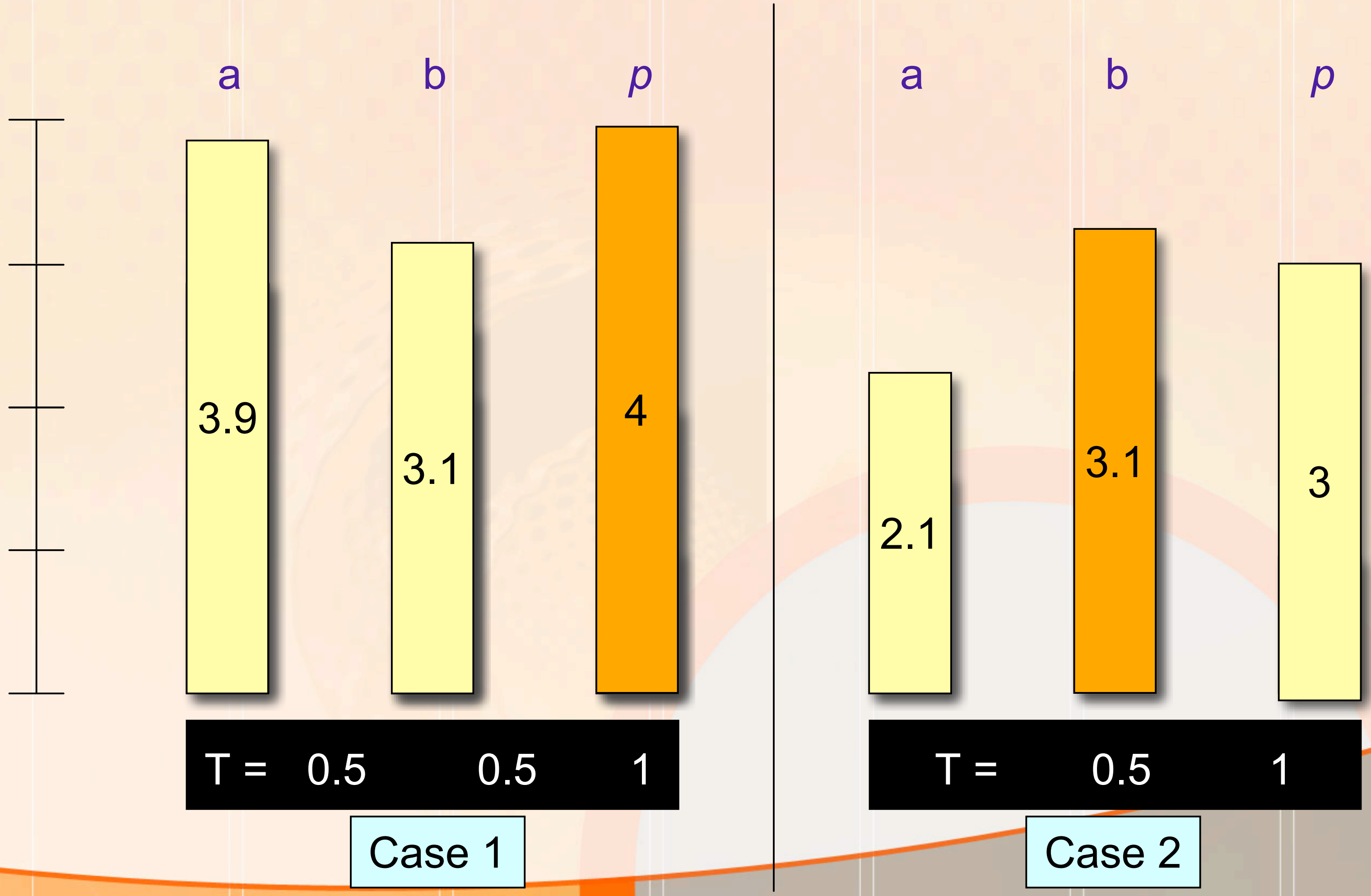
- A mechanism is **susceptible** to a manipulation M if there exists a junta distribution μ, s.t. there exists a heuristic polynomial-time algorithm for <M, μ>
- **Theorem:** Let P be a sensitive scoring protocol. Then P is susceptible to coalitional manipulation when the number of candidates is a constant.

A junta distribution

- Sampling algorithm for μ^*:
  - All *v* in T: randomly choose w(*v*) in [0,1]. Total weight is then called W.
  - All candidates ≠ *p*: randomly choose initial score in $[(\alpha_1-\alpha_2)W, \alpha_1 W]$.
- μ^* is a junta distribution
- μ^* is intuitively appealing

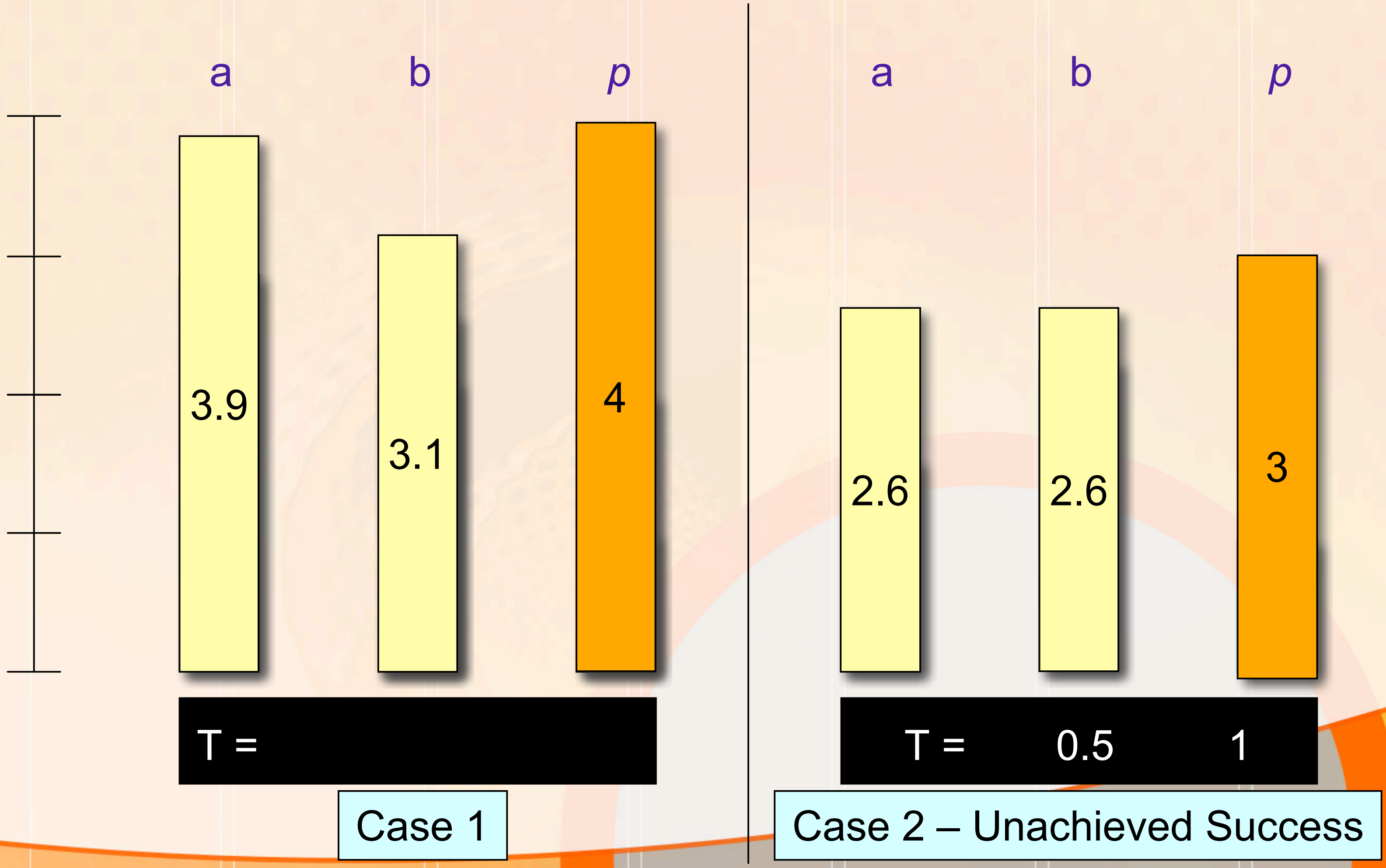
A heuristic polynomial time alg

- Greedy algorithm: each voter in T ranks *p* first, and the other candidates in an order inversely proportional to their current score.



A heuristic polynomial time alg

- Greedy algorithm: each voter in T ranks *p* first, and the other candidates in an order inversely proportional to their current score.



Case 1

Case 2 – Unachieved Success

Proof idea

- If there is no manipulation, the algorithm is surely correct. The algorithm might err if there is a manipulation.
- The alg errs only if there is subset of candidates with high initial scores, since this requires a careful distribution of points among these candidates.
- Formally, the algorithm errs only if there is d in {2,…,m} and a subset of candidates of size d {c_{j1},…,c_{jd}} such that:

$$\underbrace{W\sum_{i=1}^{d}\alpha_1 - W\sum_{i=1}^{d}\alpha_{m+2-i}}_{} - \underbrace{\frac{d(d-1)}{2}\alpha_2}_{} \le \underbrace{\sum_{i=1}^{d} S[c_{j_i}]}_{} \le \underbrace{W\sum_{i=1}^{d}\alpha_1 - W\sum_{i=1}^{d}\alpha_{m+2-i}}_{}.$$

- This only happens with polynomially small probability.

Additional Result

- [Uncertain Votes Weighted Manipulation (UVWM) problem] Given: a weight for each voter, a distribution over all the votes, a candidate *p*, and a number *r* in the range 0 to 1; can the manipulator cast its vote so that *p* wins with probability greater than *r* ?
- Let P be a voting protocol such that there exists a junta distribution over the instances of UVWM in P, with the following property: *r* is uniformly distributed in the range 0 to 1. Then P is susceptible to UVWM.

Conclusions

- Starting point for studying average case complexity of manipulating different protocols and mechanisms
- Introduced tools for showing that mechanism manipulation is *easy* in the average case
- Sensitive scoring protocols are susceptible to such manipulation if number of candidates is constant
- Which protocols are average-case *hard* to manipulate?