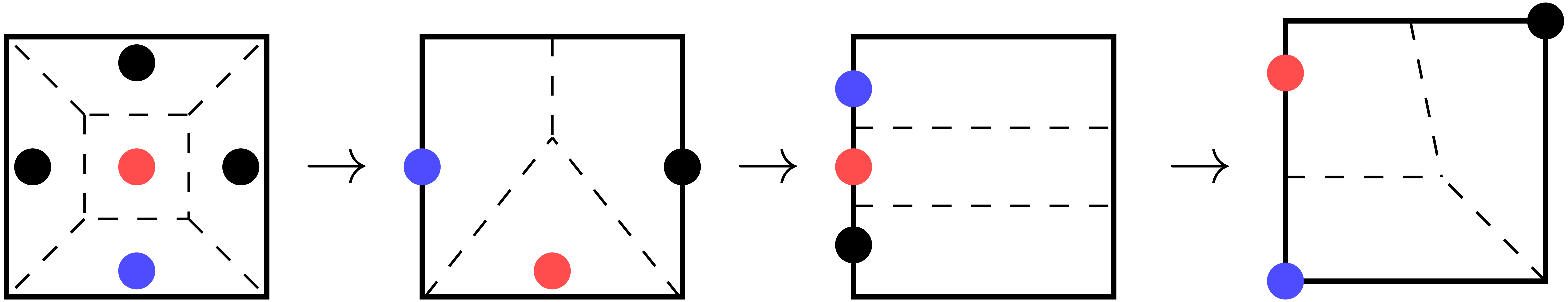
AAAI 2026



Exclusion Zones of Instant Runoff Voting

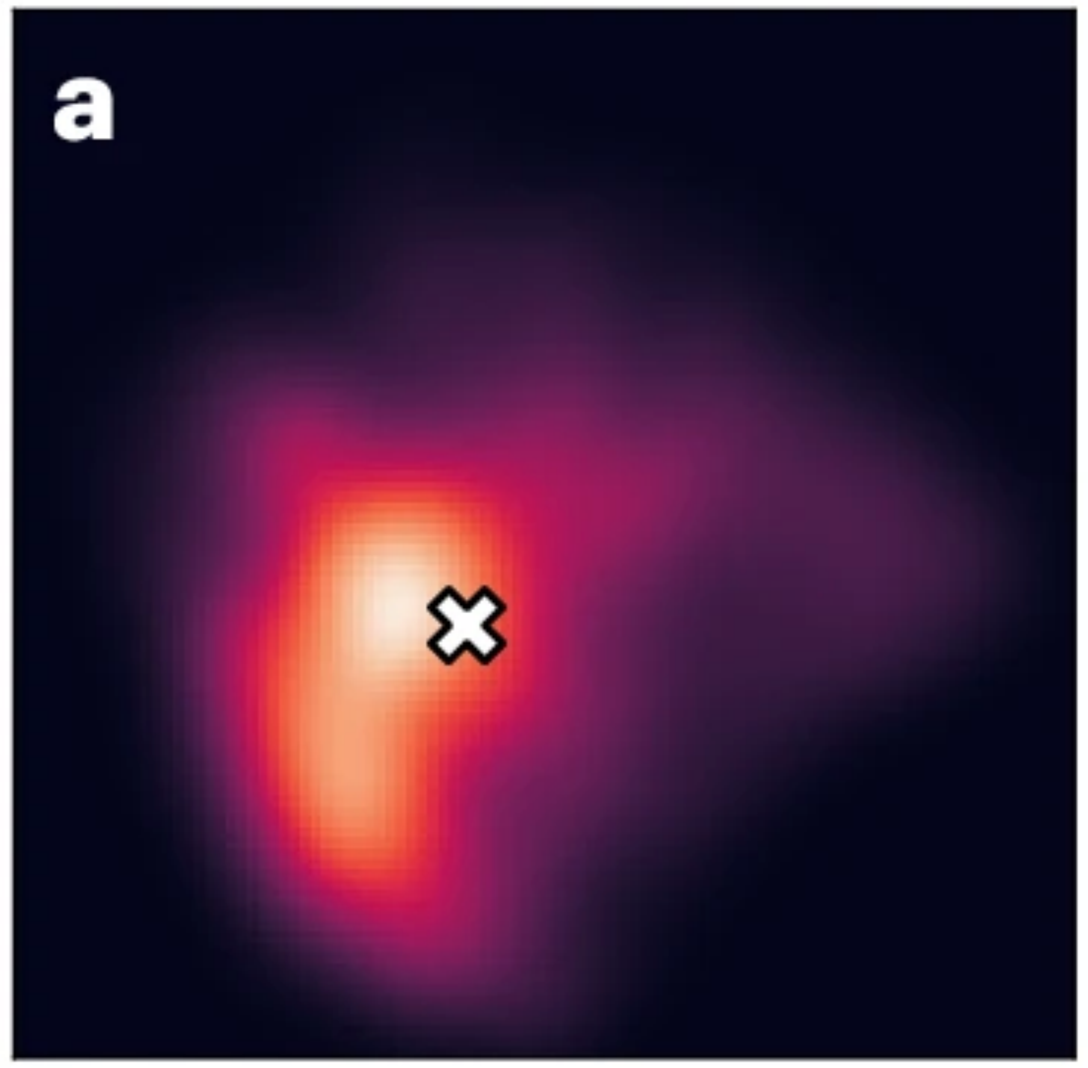
Kiran Tomlinson
Microsoft Research

Johan Ugander
Yale University

Jon Kleinberg
Cornell University

Given a distribution of voters in a metric space, what regions in the space does a voting algorithm favor?

Pew Research Center, 2004



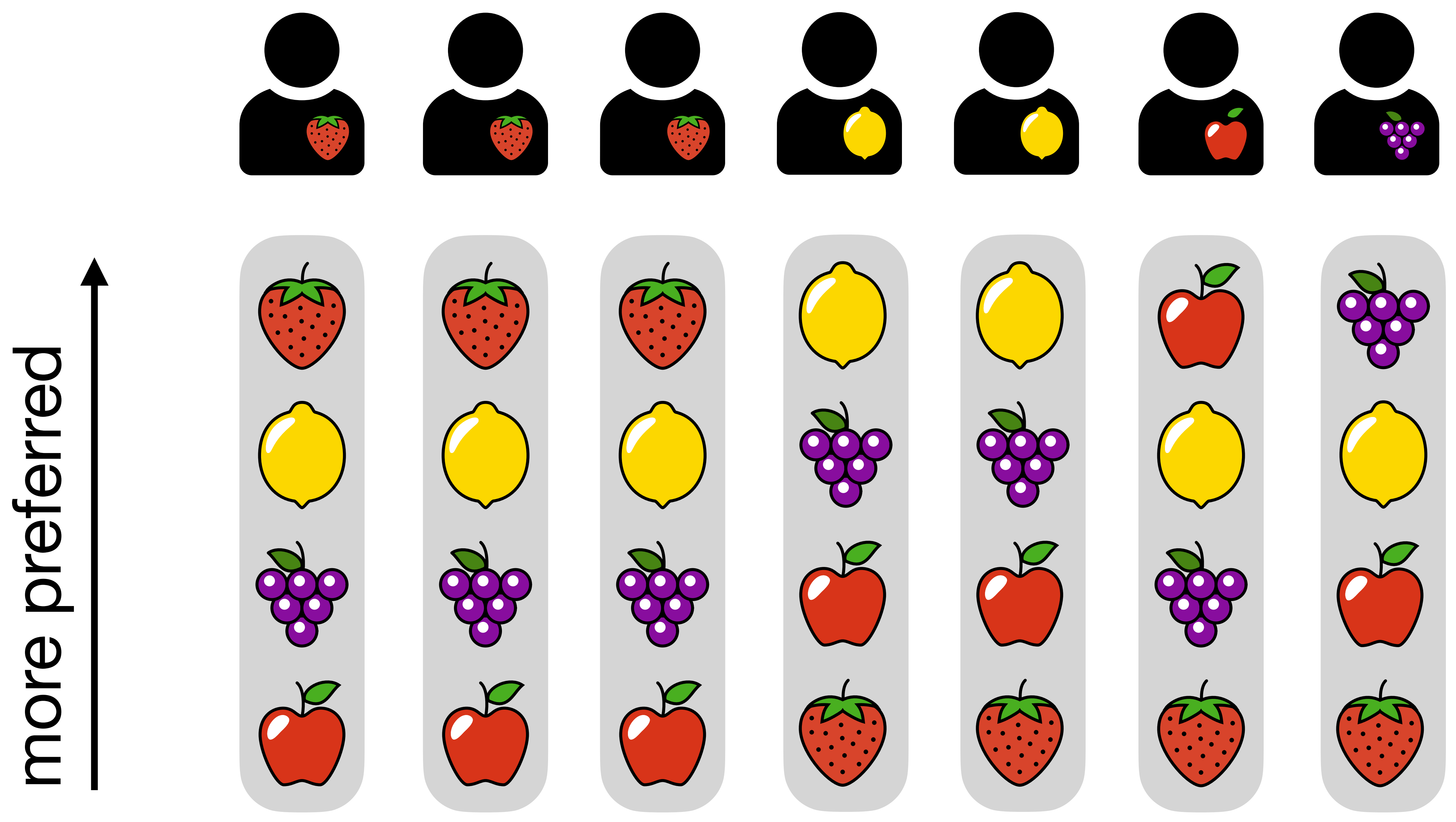
Ojer et al, *Nature Human Behavior*, 2025

E.g., will a given voting algorithm tend to elect moderates?

Instant runoff voting (IRV)

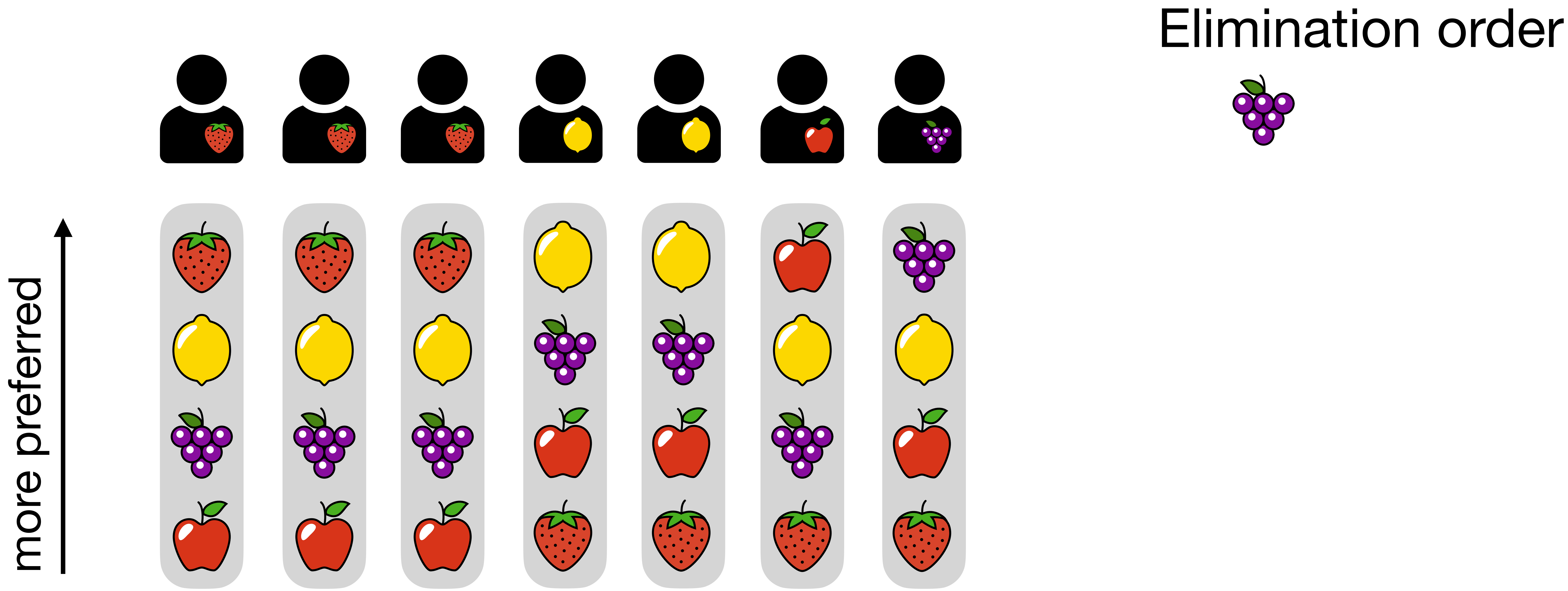
repeatedly eliminate the candidate with fewest first-place votes

Elimination order



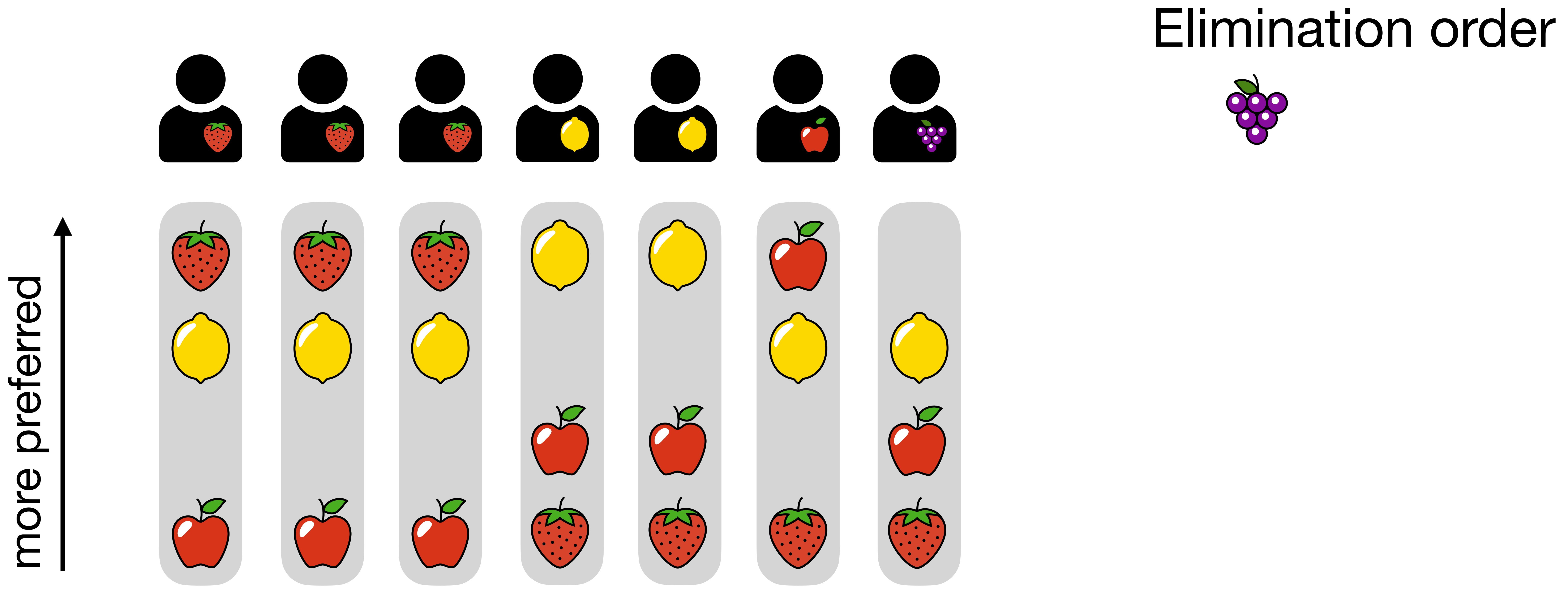
Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes



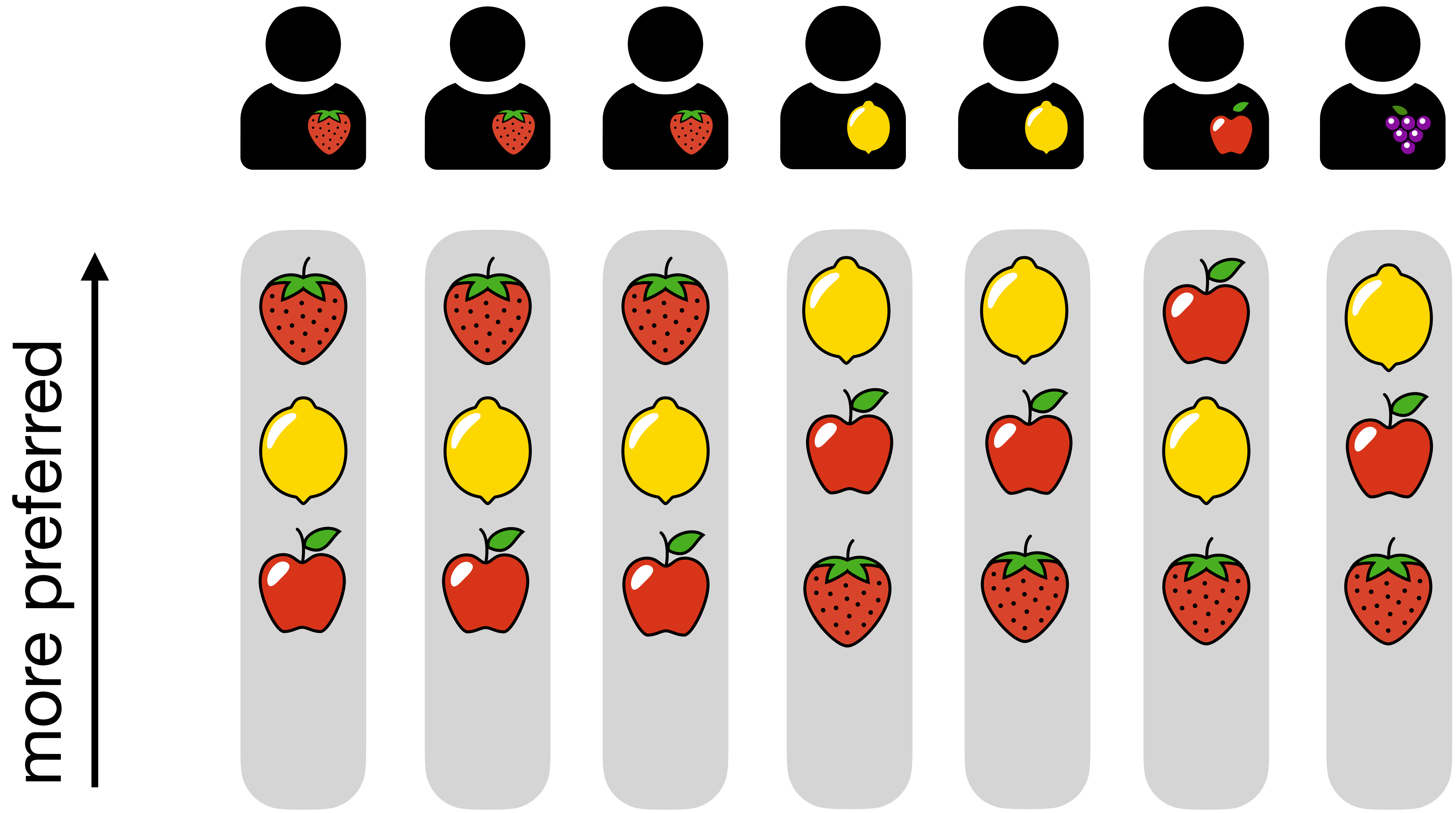
Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

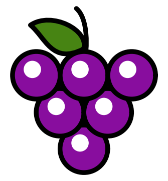


Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

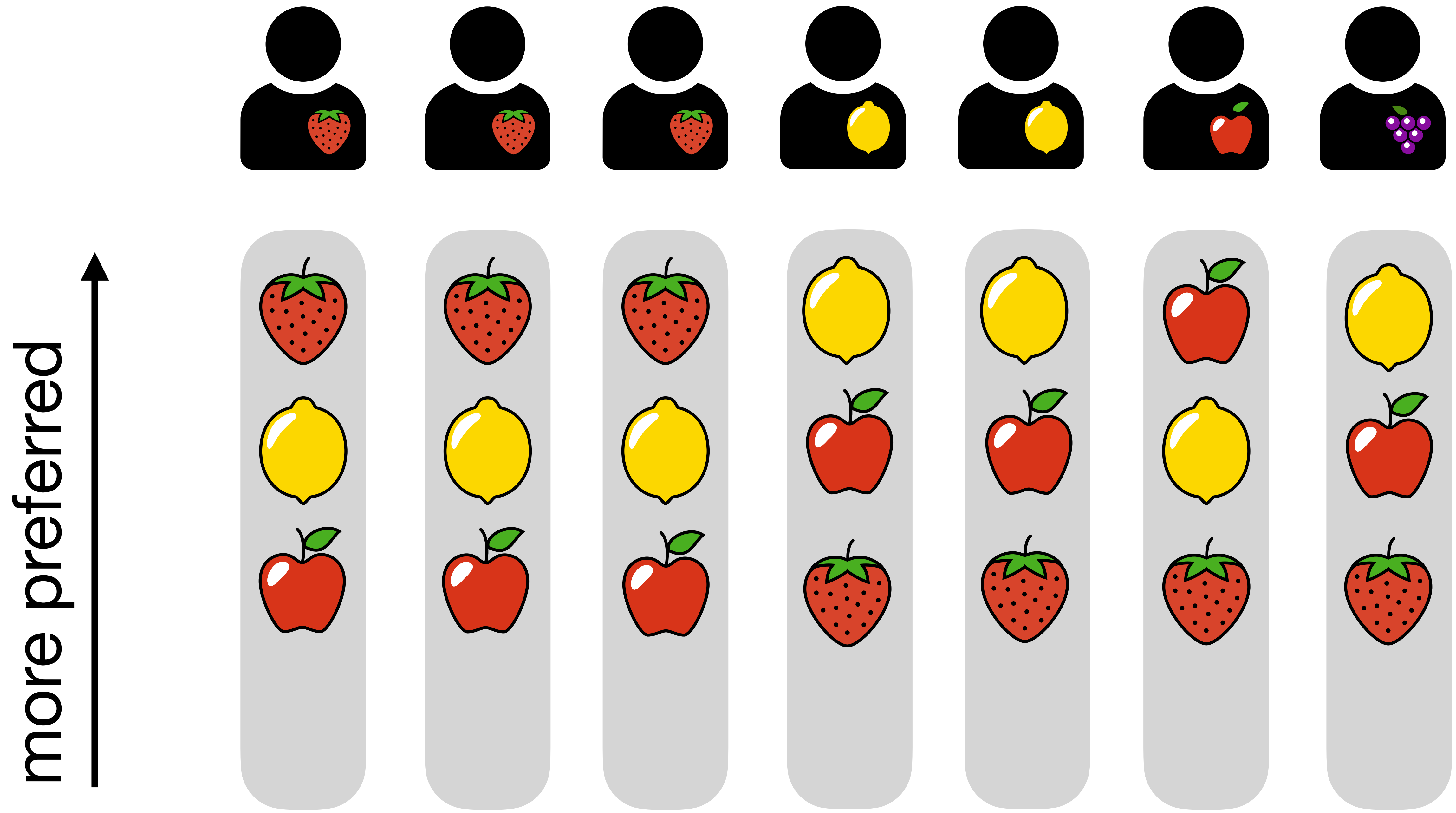


Elimination order

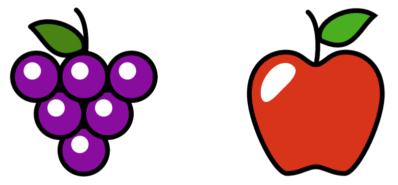


Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

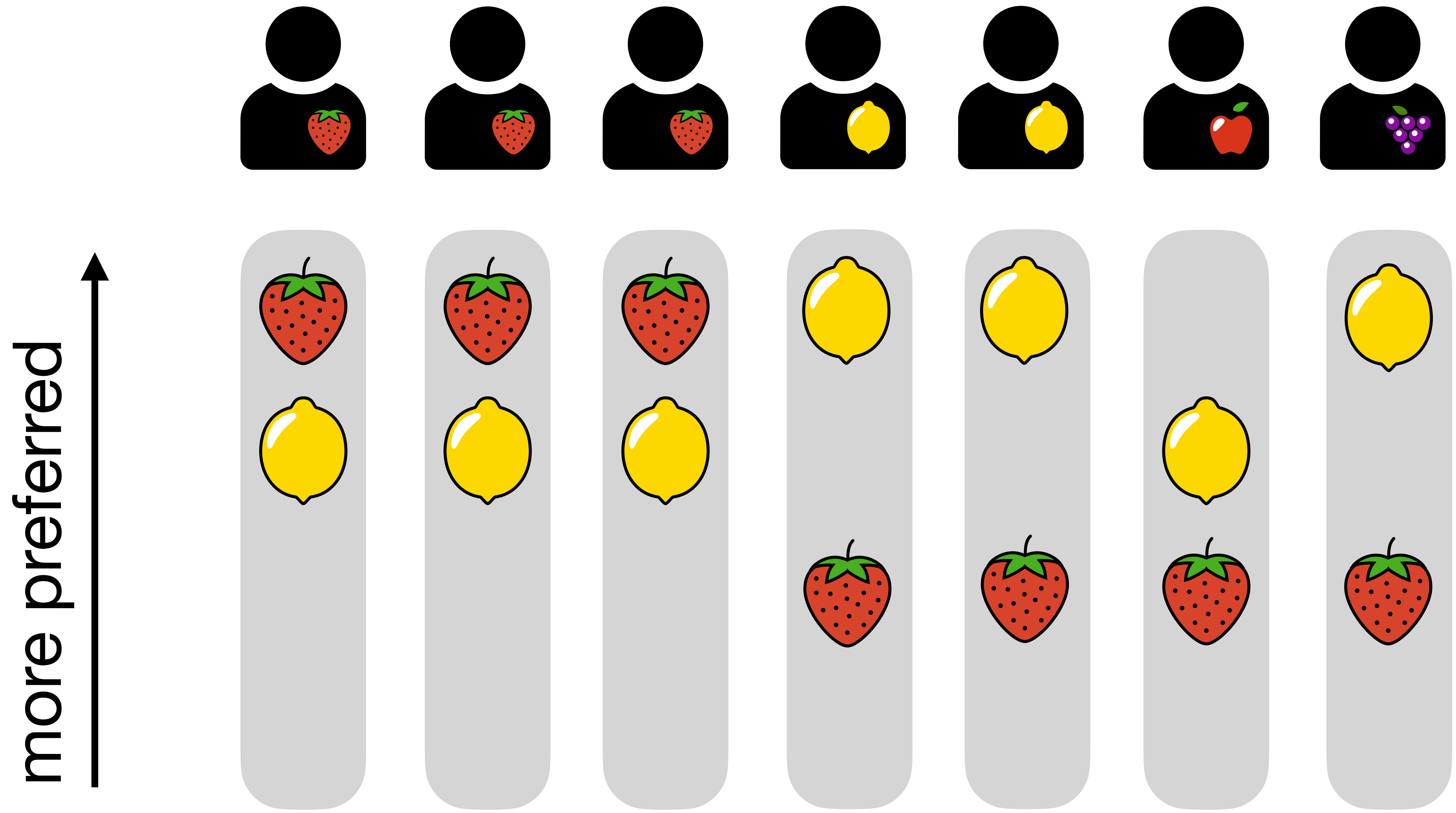


Elimination order

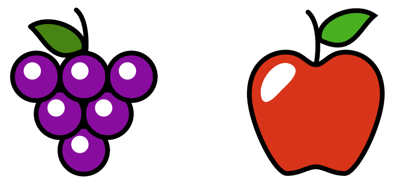


Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

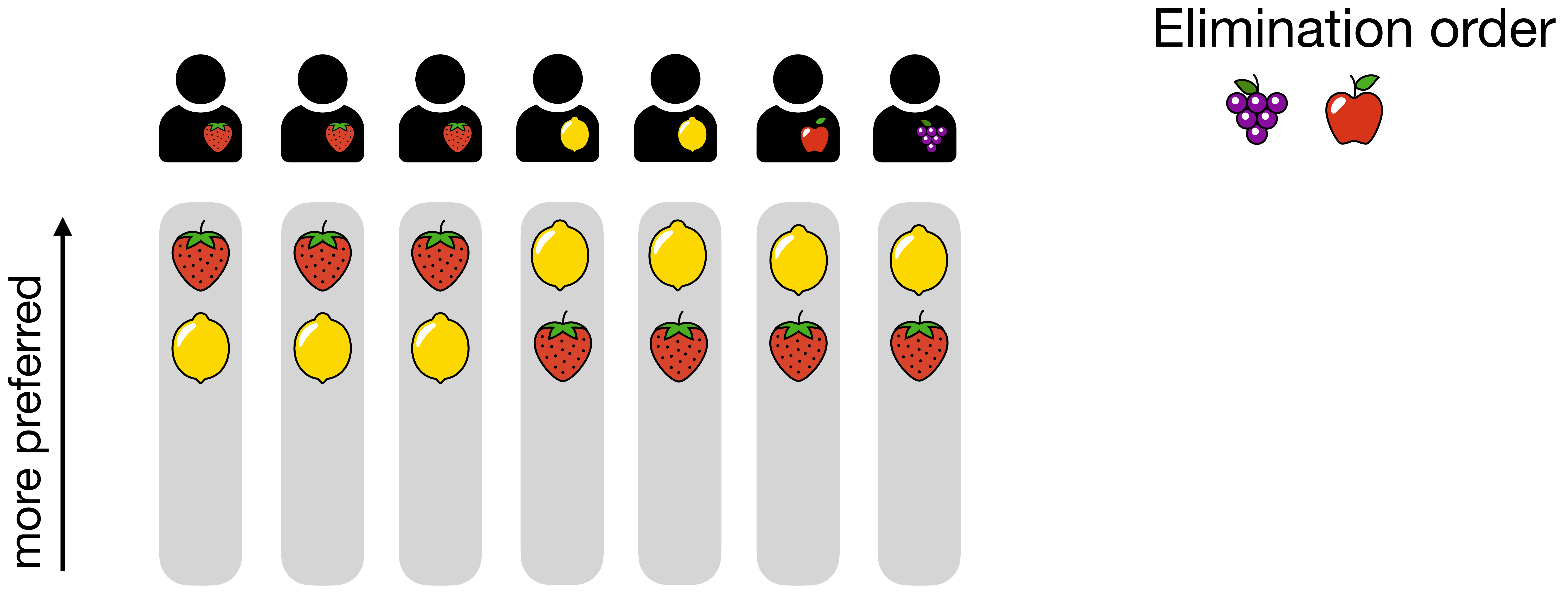


Elimination order



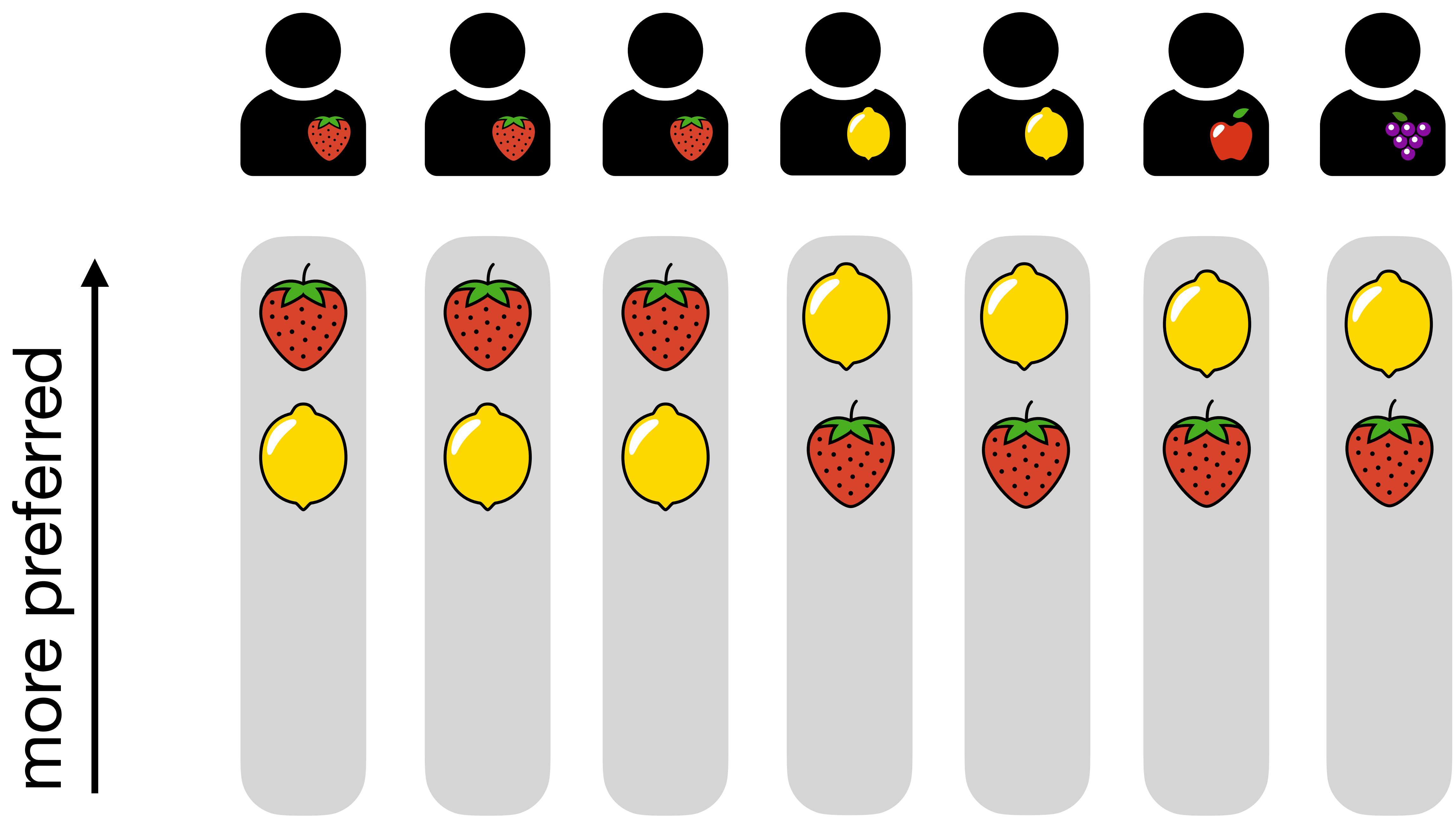
Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

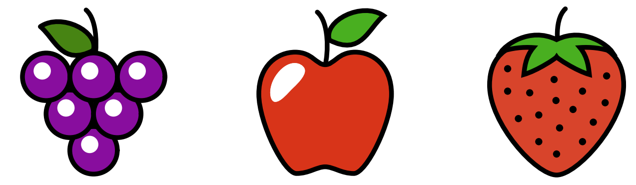


Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

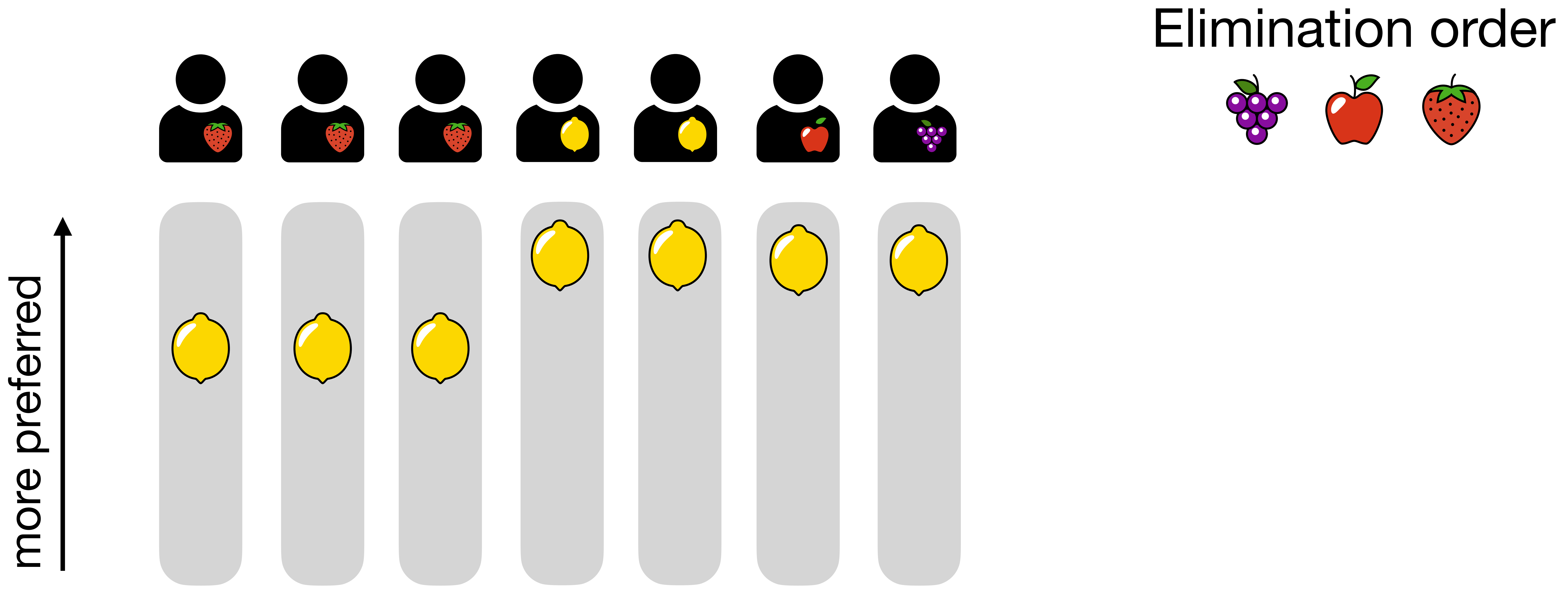


Elimination order



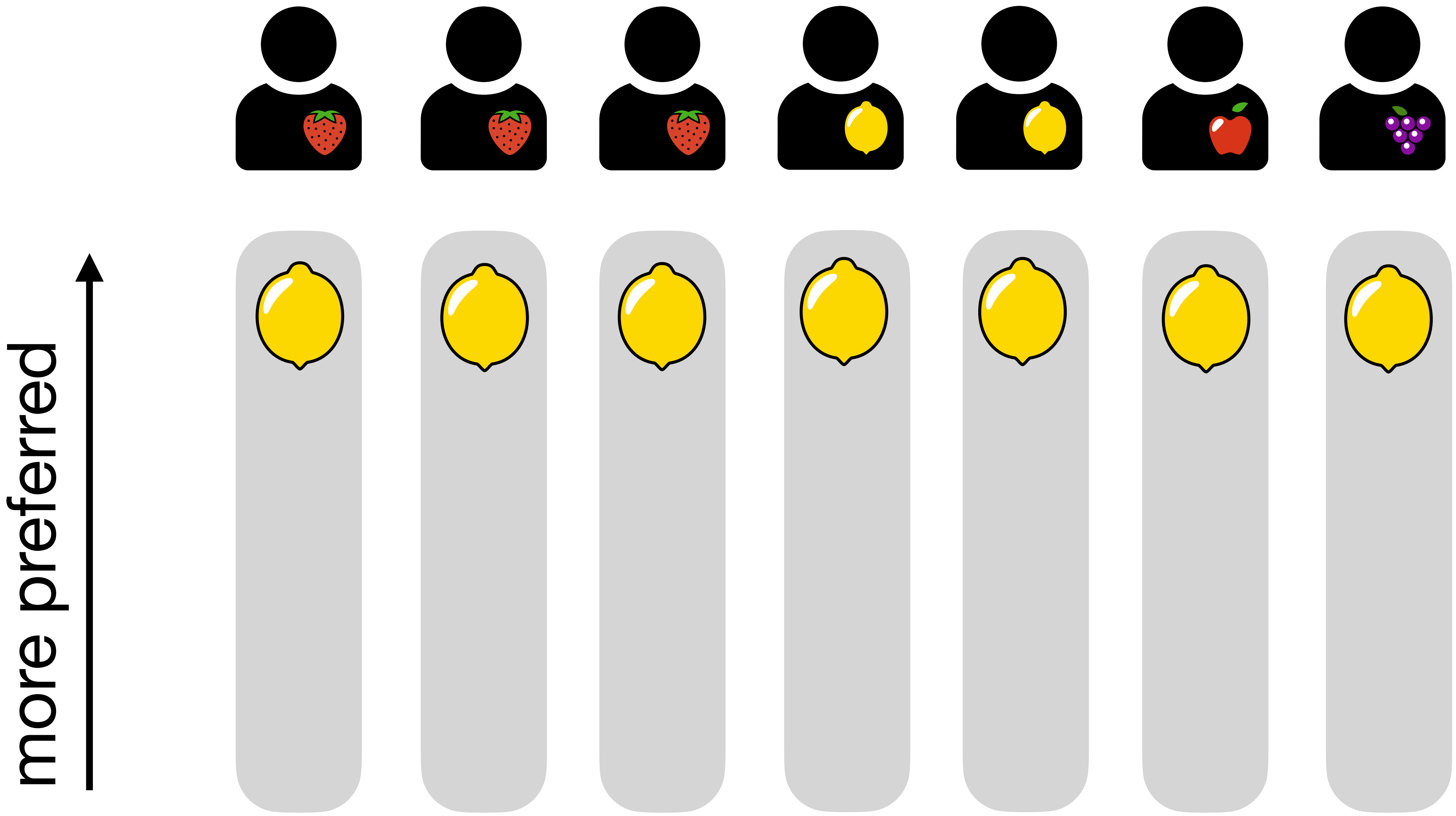
Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

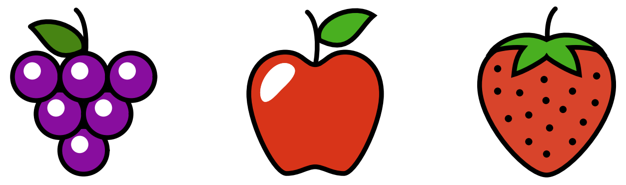


Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes

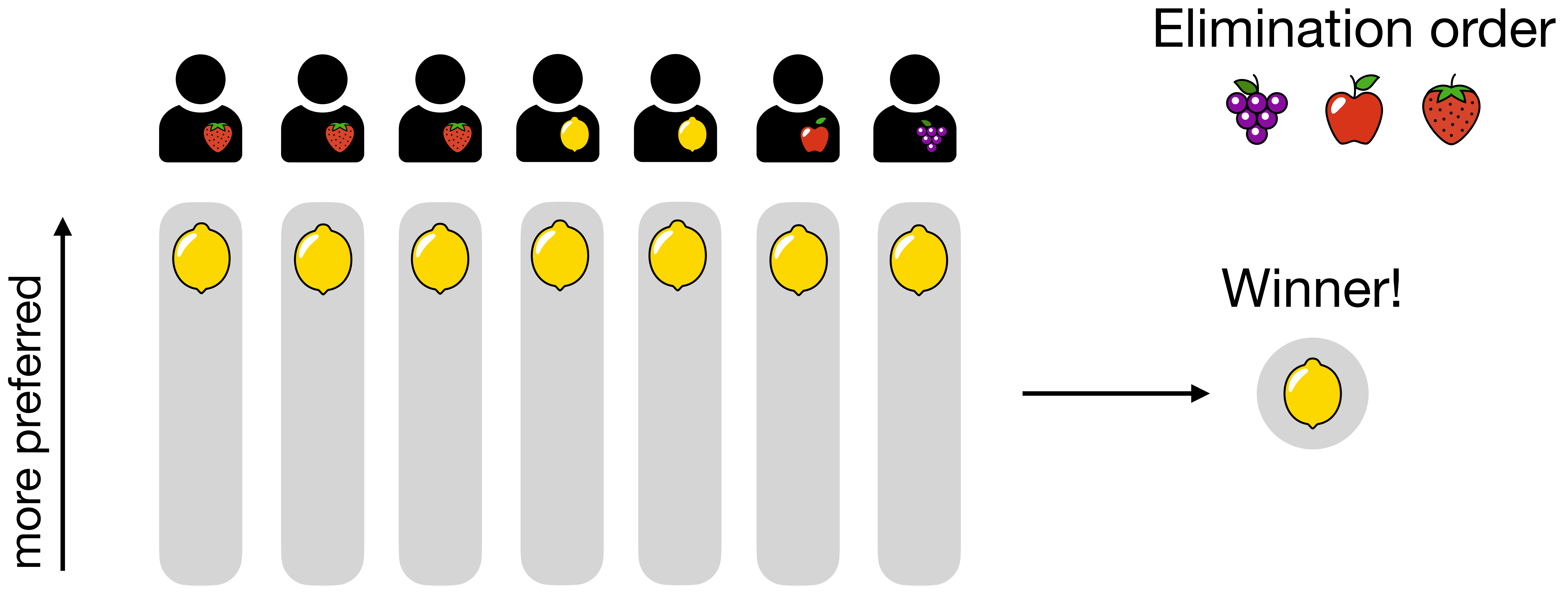


Elimination order



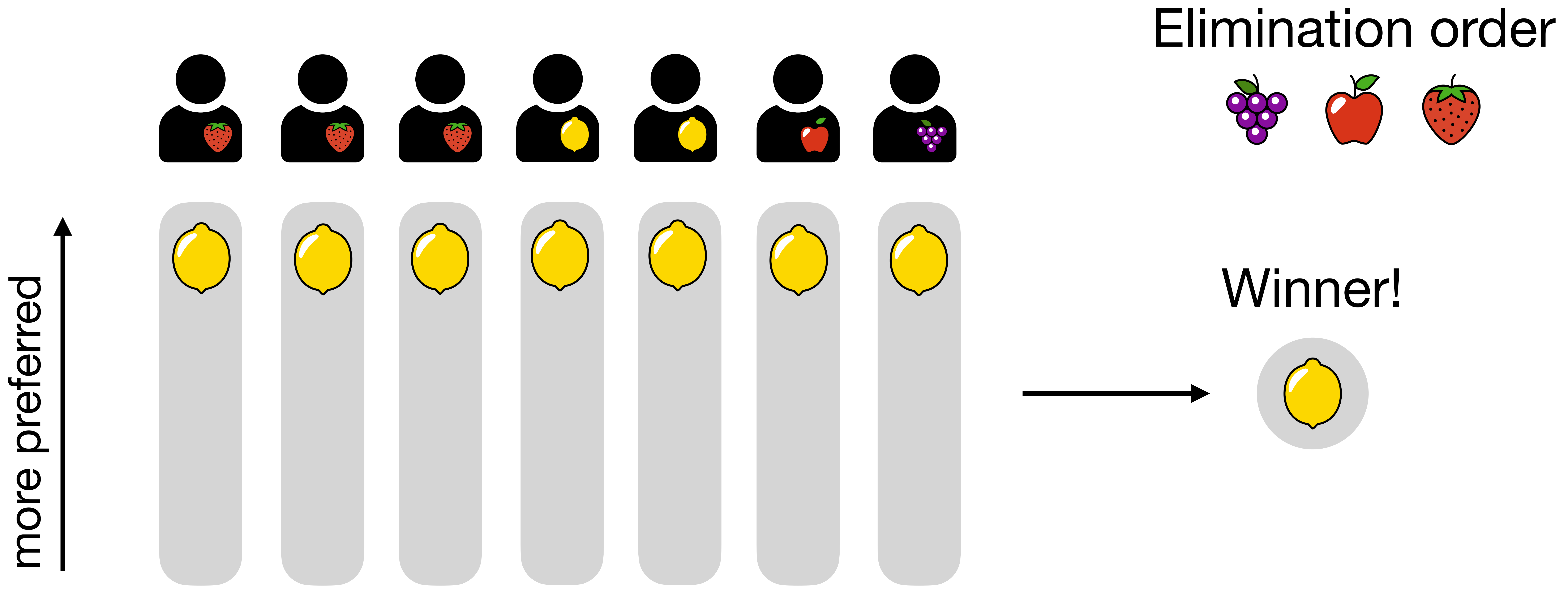
Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes



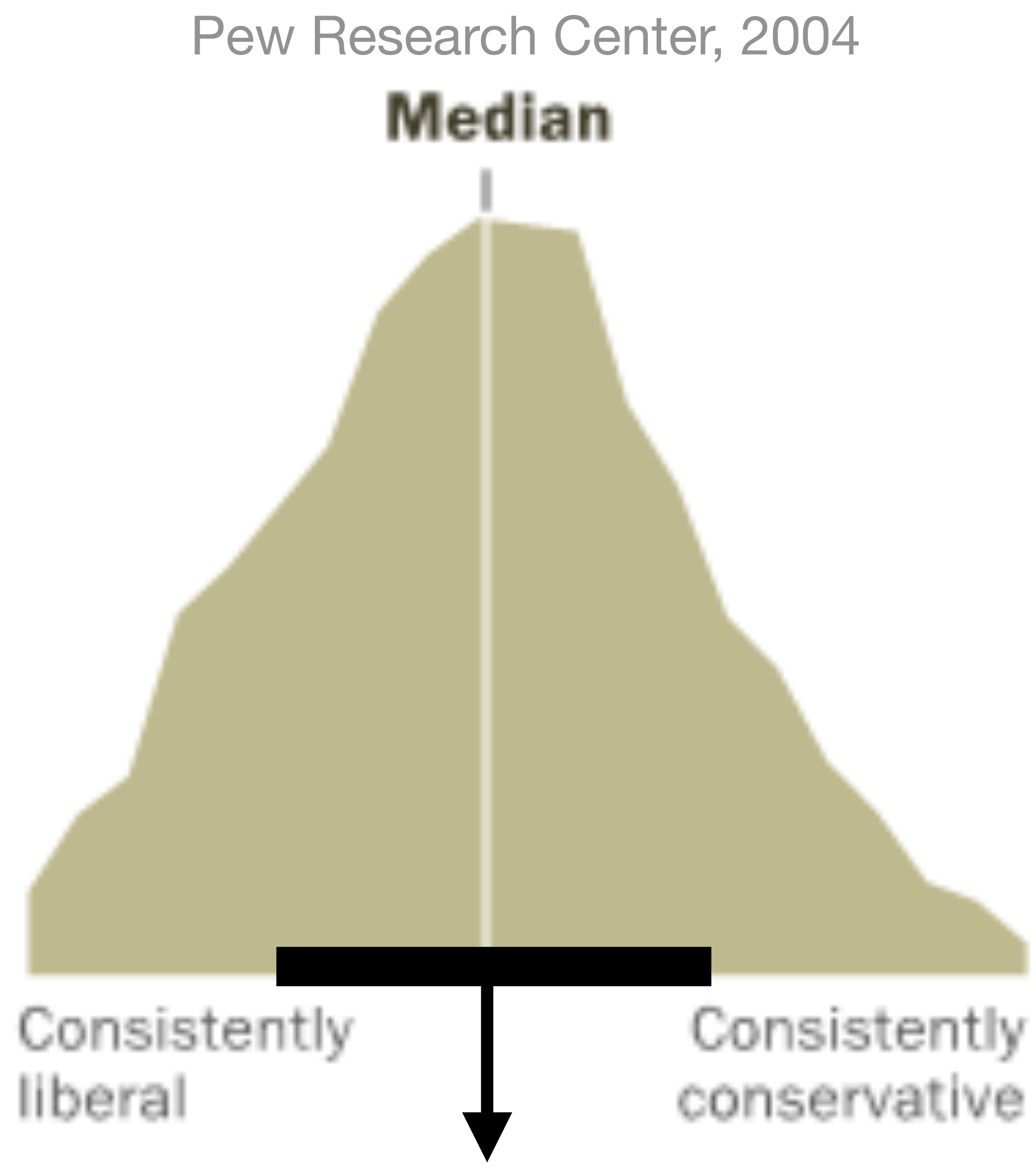
Instant runoff voting (IRV)

repeatedly eliminate the candidate with fewest first-place votes



a.k.a. RCV, STV, AV, Hare method, preferential voting

Two years ago….



IRV always favors candidates from here

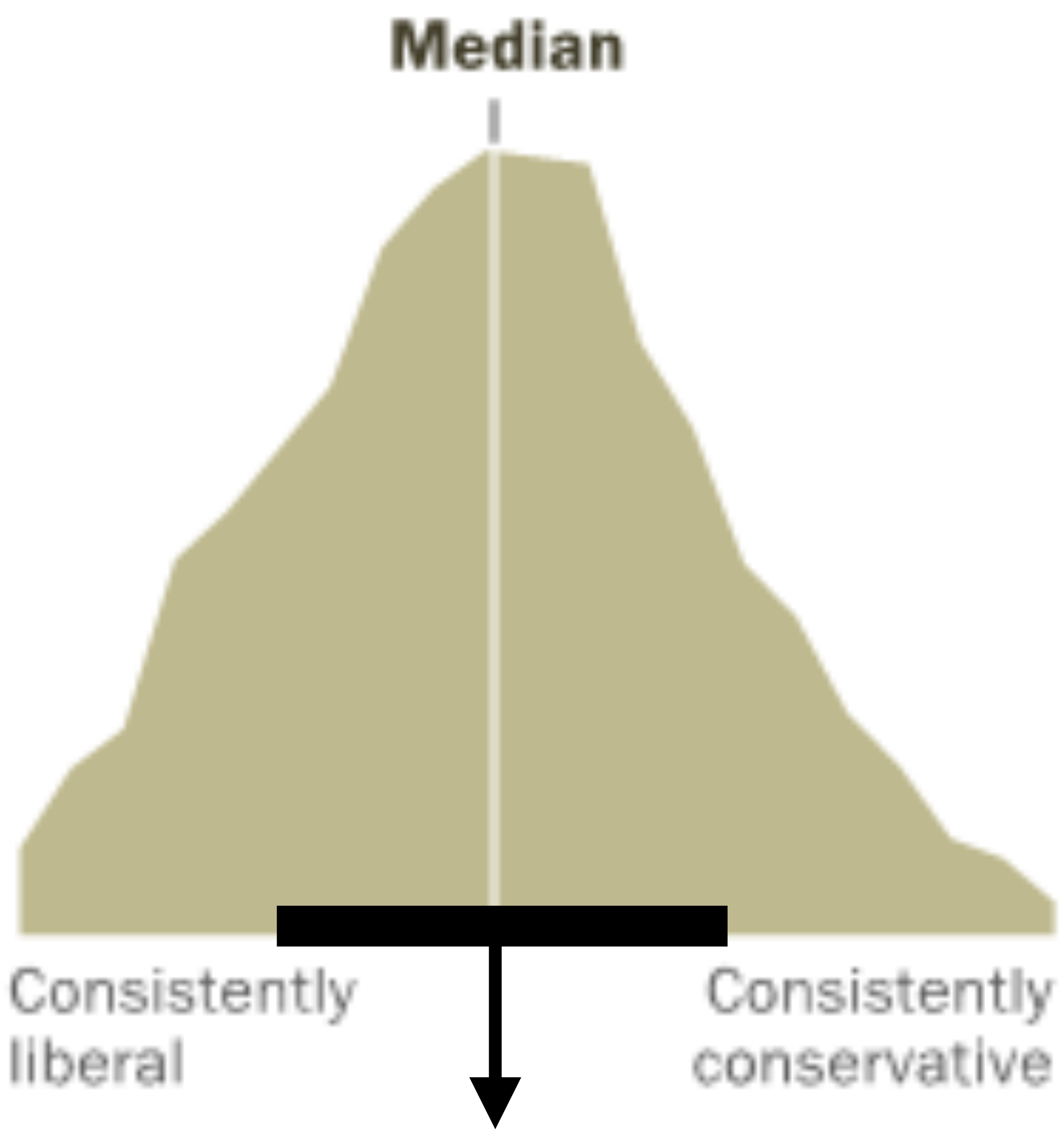
The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

The Moderating Effect of Instant Runoff Voting

Kiran Tomlinson[1], Johan Ugander[2], Jon Kleinberg[1]

Two years ago….

Pew Research Center, 2004



IRV always favors candidates from here

The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

The Moderating Effect of Instant Runoff Voting

Kiran Tomlinson[1], Johan Ugander[2], Jon Kleinberg[1]

Definition

Given a distribution of voters in a metric space M, a set $S \subseteq M$ is an **exclusion zone** of a voting system if the winner is guaranteed to come from S (unless no candidates come from S).

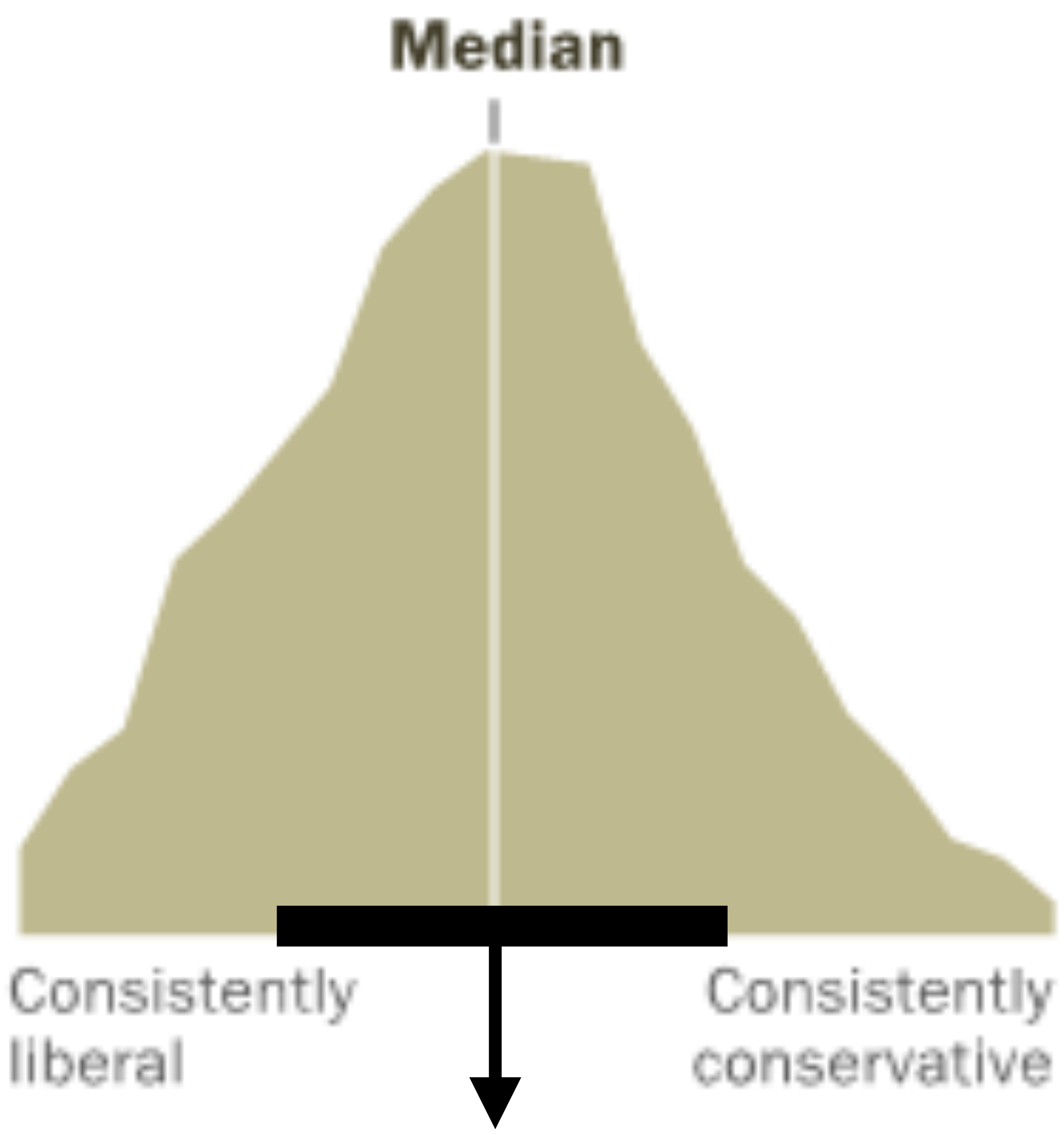
Two years ago….

The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

The Moderating Effect of Instant Runoff Voting

Kiran Tomlinson[1], Johan Ugander[2], Jon Kleinberg[1]

Pew Research Center, 2004



IRV always favors candidates from here

Definition
Given a distribution of voters in a metric space M, a set $S \subseteq M$ is an **exclusion zone** of a voting system if the winner is guaranteed to come from S (unless no candidates come from S).

Theorem [TUK ’24]
With uniform 1-Euclidean voters, $[1/6, 5/6]$ is an exclusion zone of IRV (and the smallest one).

symmetric unimodal distributions: $[F^{-1}(1/6), 1 - F^{-1}(1/6)]$

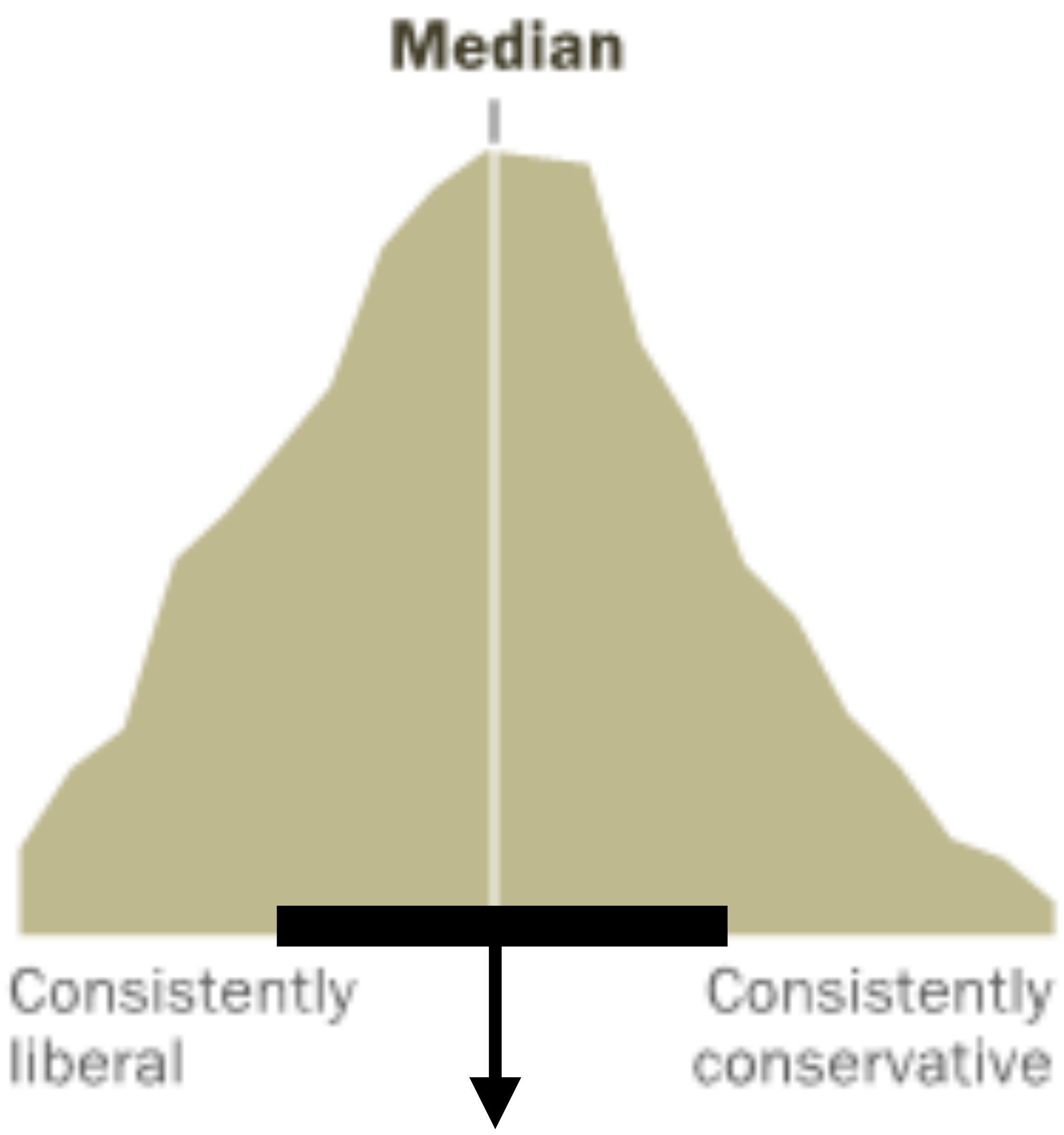
Two years ago….

The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

The Moderating Effect of Instant Runoff Voting

Kiran Tomlinson[1], Johan Ugander[2], Jon Kleinberg[1]

Pew Research Center, 2004



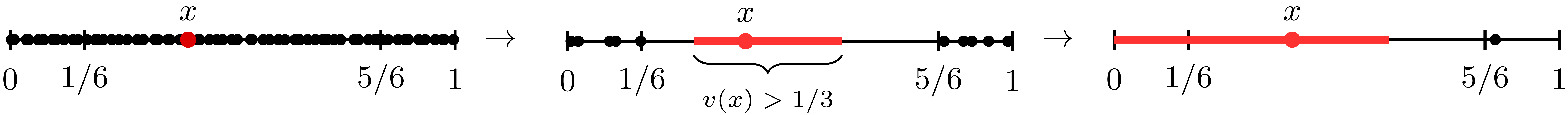
IRV always favors candidates from here

Definition
Given a distribution of voters in a metric space M, a set $S \subseteq M$ is an **exclusion zone** of a voting system if the winner is guaranteed to come from S (unless no candidates come from S).

Theorem [TUK '24]
With uniform 1-Euclidean voters, $[1/6, 5/6]$ is an exclusion zone of IRV (and the smallest one).

symmetric unimodal distributions: $[F^{-1}(1/6), 1 - F^{-1}(1/6)]$

Proof.



The space of exclusion zones in general

The space of exclusion zones in general

Proposition
Let S, T be two exclusion zones. Either $S \subset T$ or $T \subset S$.

The space of exclusion zones in general

Proposition
Let S, T be two exclusion zones. Either $S \subset T$ or $T \subset S$.

Definition
The unique **minimal exclusion zone** is given by the intersection of all exclusion zones.

The space of exclusion zones in general

Proposition
Let S, T be two exclusion zones. Either $S \subset T$ or $T \subset S$.

Definition
The unique **minimal exclusion zone** is given by the intersection of all exclusion zones.

trivial exclusion zone (all positions)

minimal exclusion zone

The space of all Condorcet method exclusion zones in 1d

Median Voter Theorem (Black, 1948)
With 1-Euclidean preferences, the candidate closest to the median voter is the Condorcet winner.

The space of all Condorcet method exclusion zones in 1d

Median Voter Theorem (Black, 1948)
With 1-Euclidean preferences, the candidate closest to the median voter is the Condorcet winner.

↓

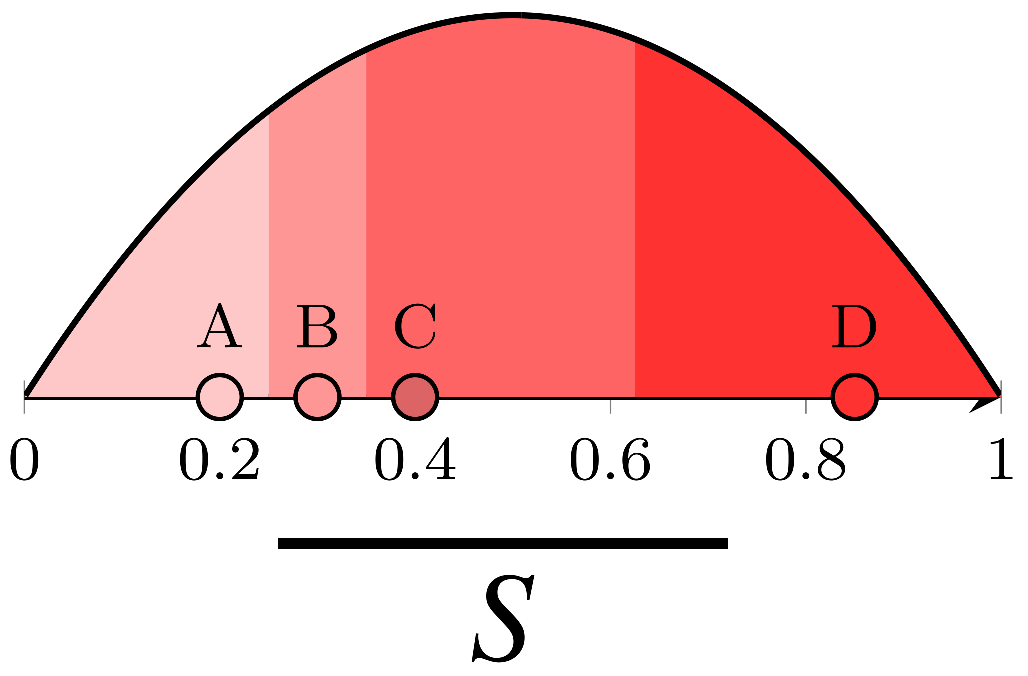
Proposition
Condorcet methods with symmetric 1-Euclidean voters have the exclusion zones:
$S = [c, 1 - c]$ for any $c \in [0, 1/2)$,
$S = (c, 1 - c)$ for any $c \in [0, 1/2)$,
$S = \{1/2\}$.

The space of all Condorcet method exclusion zones in 1d

Median Voter Theorem (Black, 1948)
With 1-Euclidean preferences, the candidate closest to the median voter is the Condorcet winner.

Proposition
Condorcet methods with symmetric 1-Euclidean voters have the exclusion zones:
$S = [c, 1-c]$ for any $c \in [0, 1/2)$,
$S = (c, 1-c)$ for any $c \in [0, 1/2)$,
$S = \{1/2\}$.

e.g.,



Does the 1-Euclidean moderating effect for IRV extend to *d*-Euclidean space?

Does the 1-Euclidean moderating effect for IRV extend to *d*-Euclidean space?

Proposition

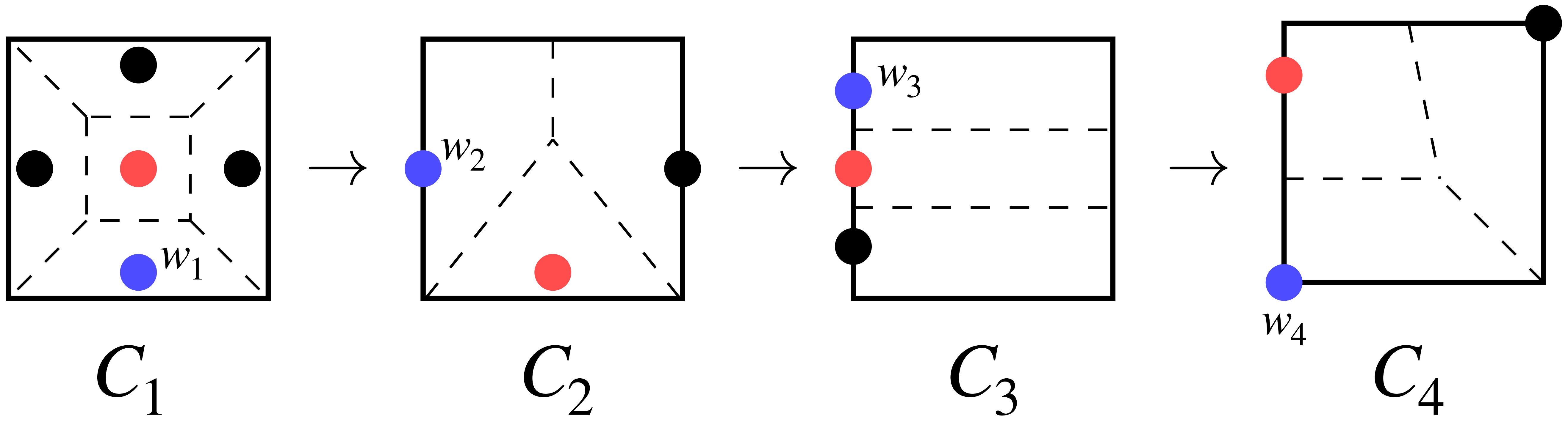
The square with uniform L_2 voters has no nontrivial IRV exclusion zone.

Does the 1-Euclidean moderating effect for IRV extend to *d*-Euclidean space?

Proposition

The square with uniform L_2 voters has no nontrivial IRV exclusion zone.

Proof sketch. "Condorcet chain"

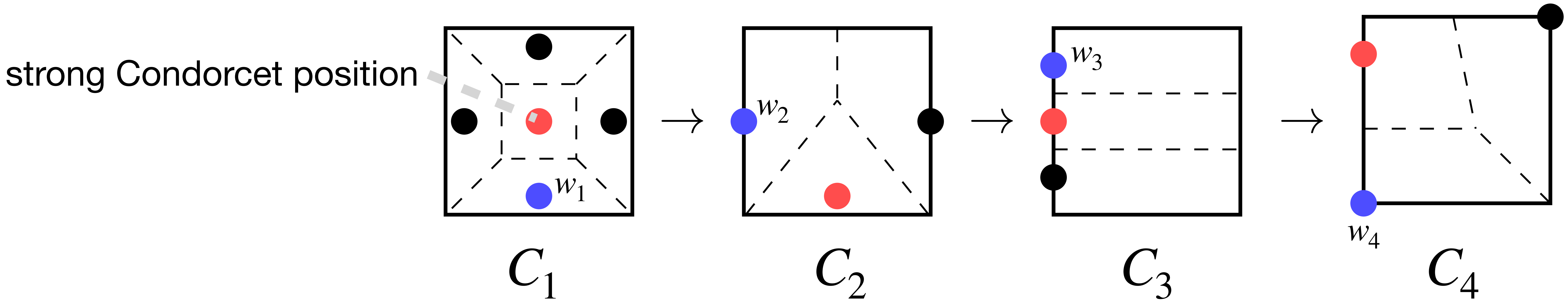


Does the 1-Euclidean moderating effect for IRV extend to *d*-Euclidean space?

Proposition

The square with uniform L_2 voters has no nontrivial IRV exclusion zone.

Proof sketch. “Condorcet chain”

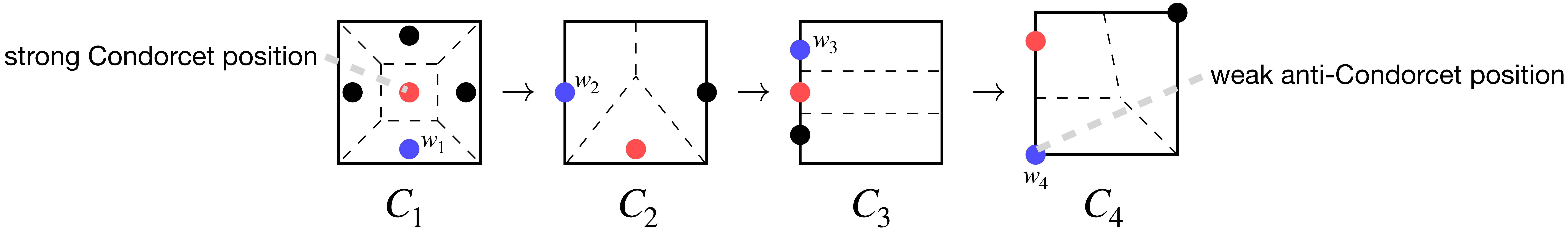


Does the 1-Euclidean moderating effect for IRV extend to *d*-Euclidean space?

Proposition

The square with uniform L_2 voters has no nontrivial IRV exclusion zone.

Proof sketch. "Condorcet chain"

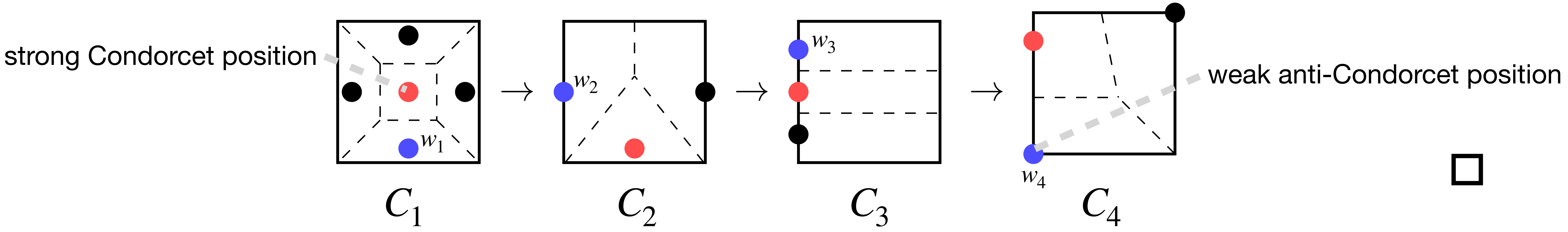


Does the 1-Euclidean moderating effect for IRV extend to *d*-Euclidean space?

Proposition

The square with uniform L_2 voters has no nontrivial IRV exclusion zone.

Proof sketch. "Condorcet chain"



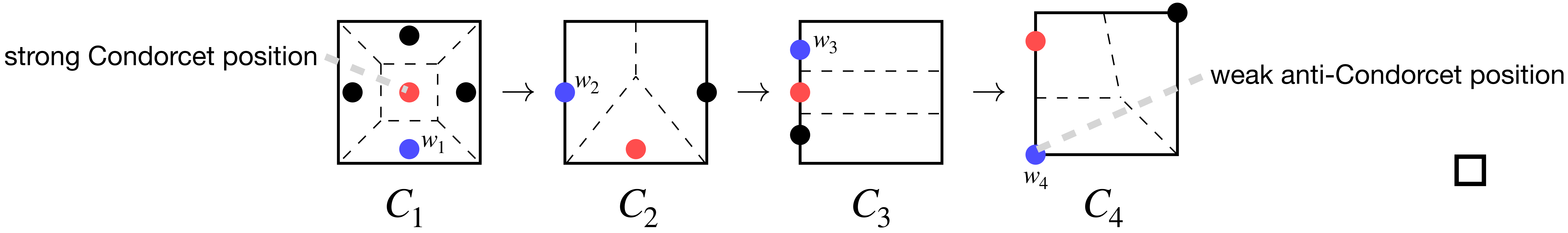
□

Does the 1-Euclidean moderating effect for IRV extend to *d*-Euclidean space?

Proposition

The square with uniform L_2 voters has no nontrivial IRV exclusion zone.

Proof sketch. "Condorcet chain"



$\square$

Theorem

Every d-dimensional hyperrectangle ($d \geq 2$) with uniform L_1 or L_2 voters has no nontrivial IRV exclusion zone.

… So does IRV only have exclusion zones in one dimension?

… So does IRV only have exclusion zones in one dimension? *No!*

… So does IRV only have exclusion zones in one dimension? *No!*

The shaded region is an IRV exclusion zone with uniform L_1 voters over this shape:

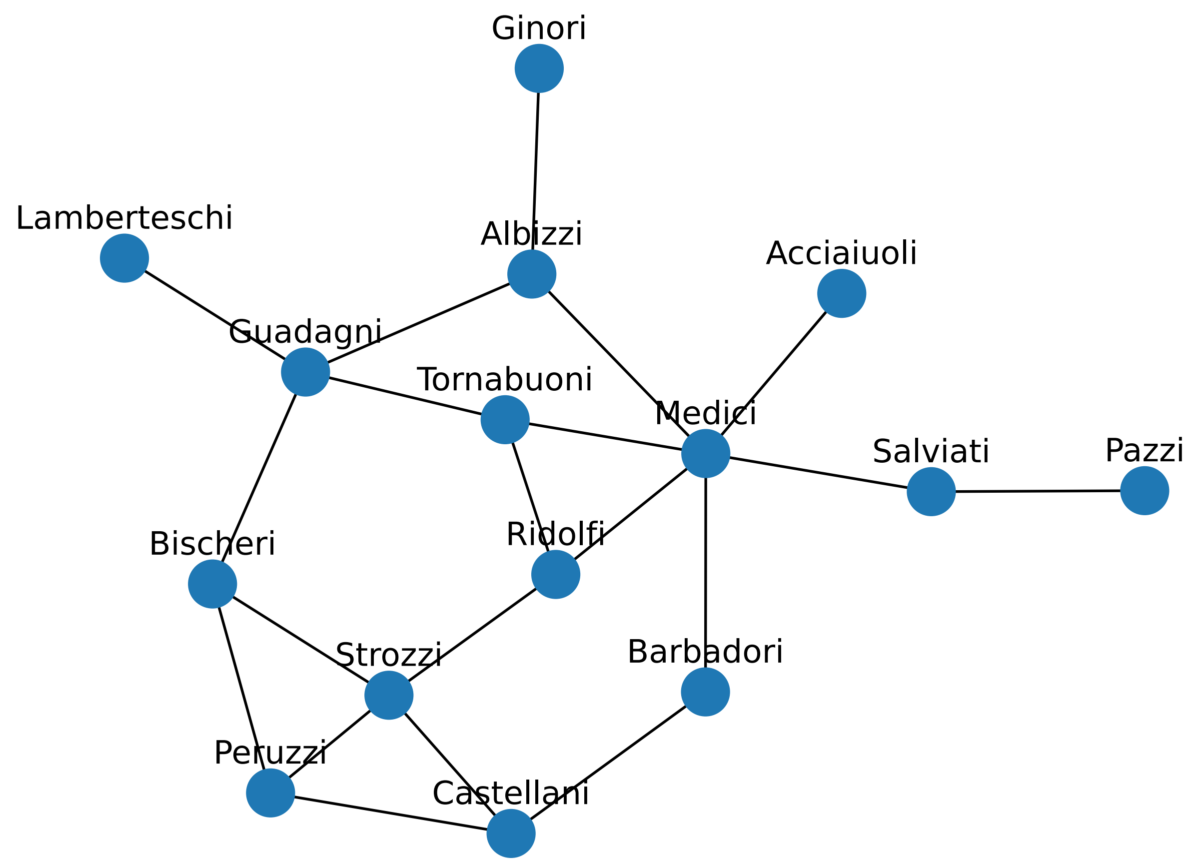
… So does IRV only have exclusion zones in one dimension? *No!*

The shaded region is an IRV exclusion zone with uniform L_1 voters over this shape:

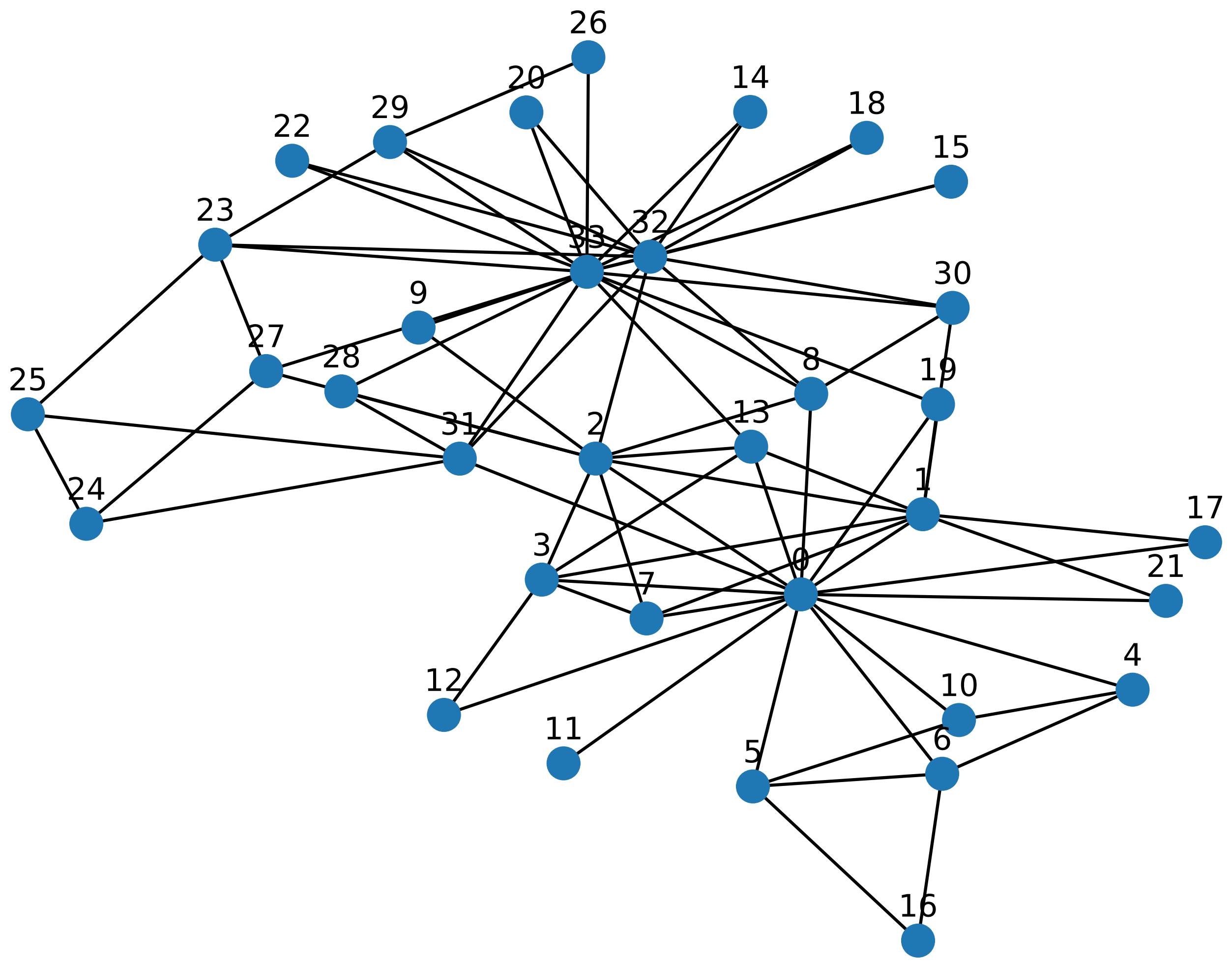
hyperrectangles have too many symmetries

Voting with the graph metric

- Nodes = voters
- Some subset of voters run for office
- Voters prefer closer candidates
- Resolve ties with Split-IRV (vote share 1 evenly split among equidistant candidates)



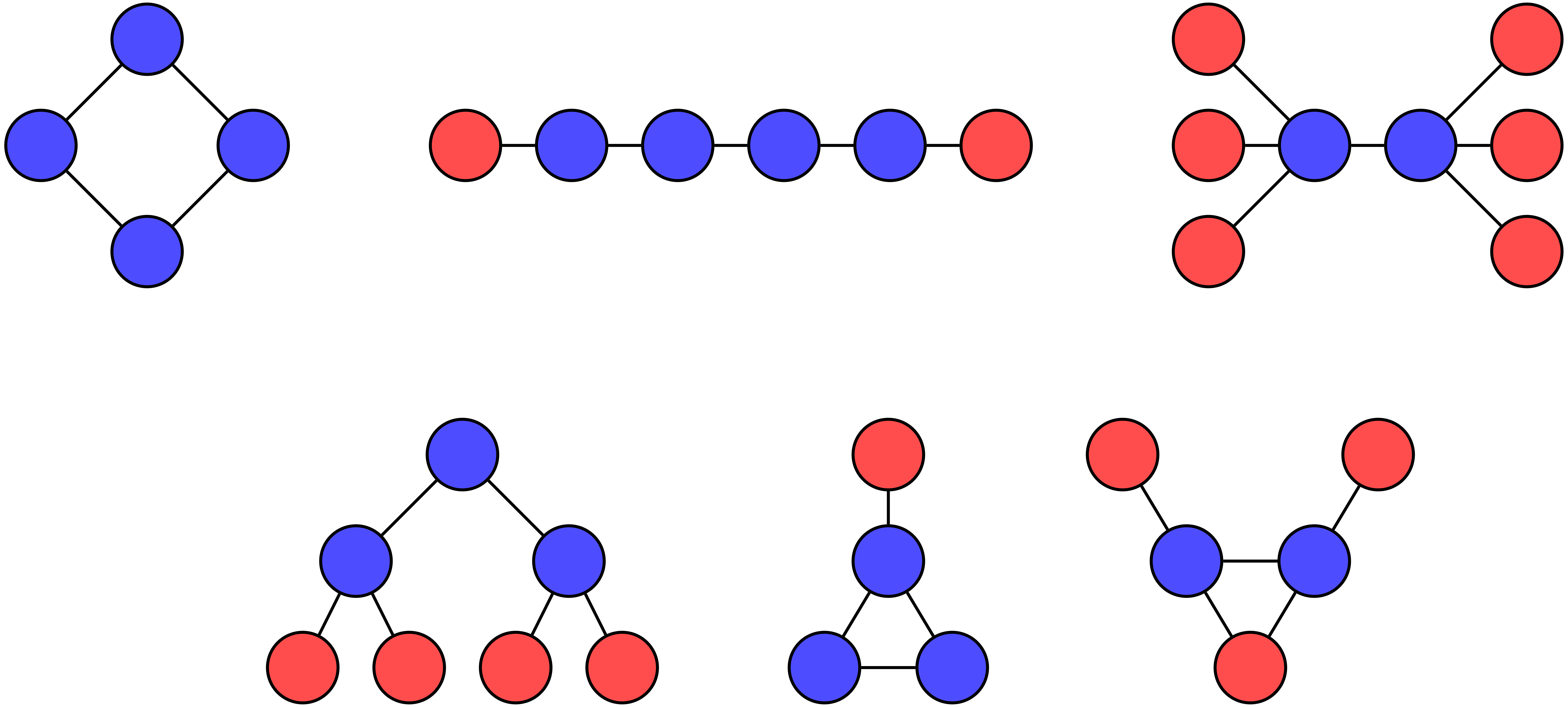
15th century Florentine marriages



Zachary's karate club

IRV exclusion zones in graphs

minimal exclusion zone



Finding IRV exclusion zones in graphs

IRV-Exclusion
Given a graph *G* and a set of nodes *S*, is *S* an IRV exclusion zone of *G*?

Finding IRV exclusion zones in graphs

IRV-Exclusion
Given a graph G and a set of nodes S, is S an IRV exclusion zone of G?

Theorem
IRV-Exclusion is co-NP-complete.

Finding IRV exclusion zones in graphs

IRV-Exclusion
Given a graph *G* and a set of nodes *S*, is *S* an IRV exclusion zone of *G*?

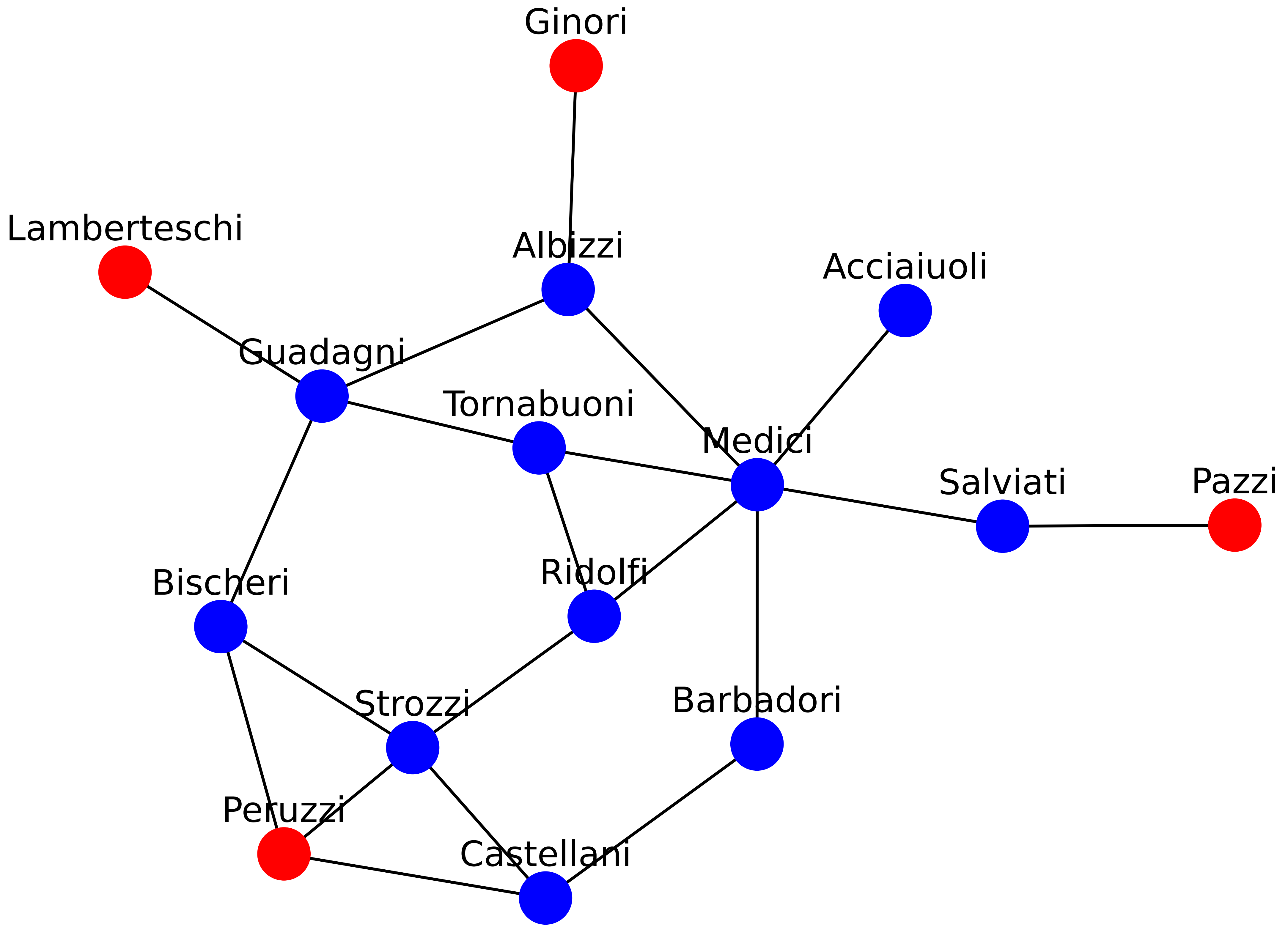
Theorem
IRV-Exclusion is co-NP-complete.

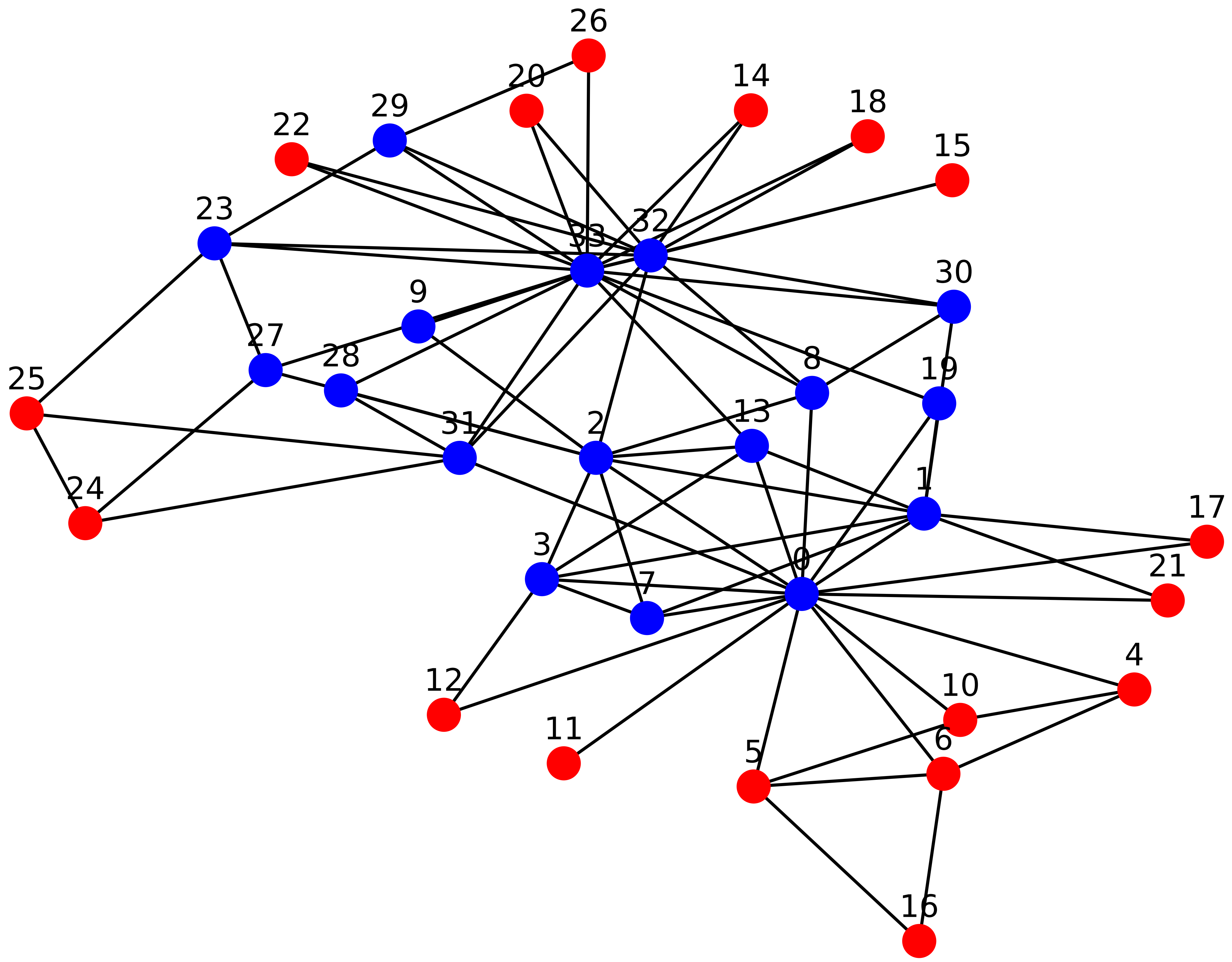
Theorem
Let *G* be a graph with *n* nodes and *m* edges. For any $\epsilon, \delta \in (0,1)$, there is a randomized algorithm returning a set *S* in time $O((n^3 + n^2m)\log(1/\delta)/e^2)$ s.t.

1. *S* is a subset of the minimal IRV exclusion zone of *G* and
2. *S* is a $(1-\epsilon)$-approximate IRV exlusion zone of *G* w.p. at least $1-\delta$.

IRV exclusion zones in graphs

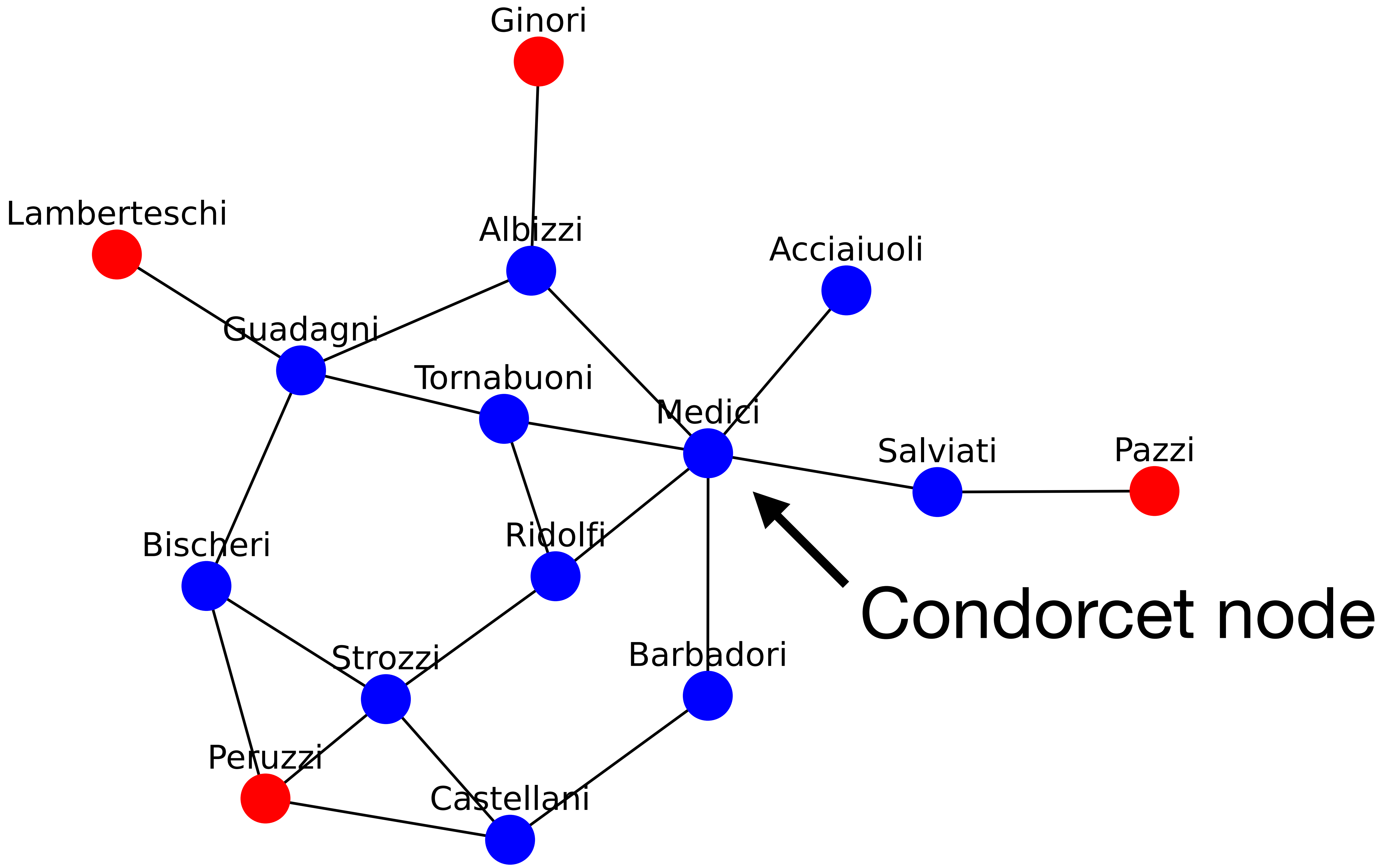
minimal exclusion zone

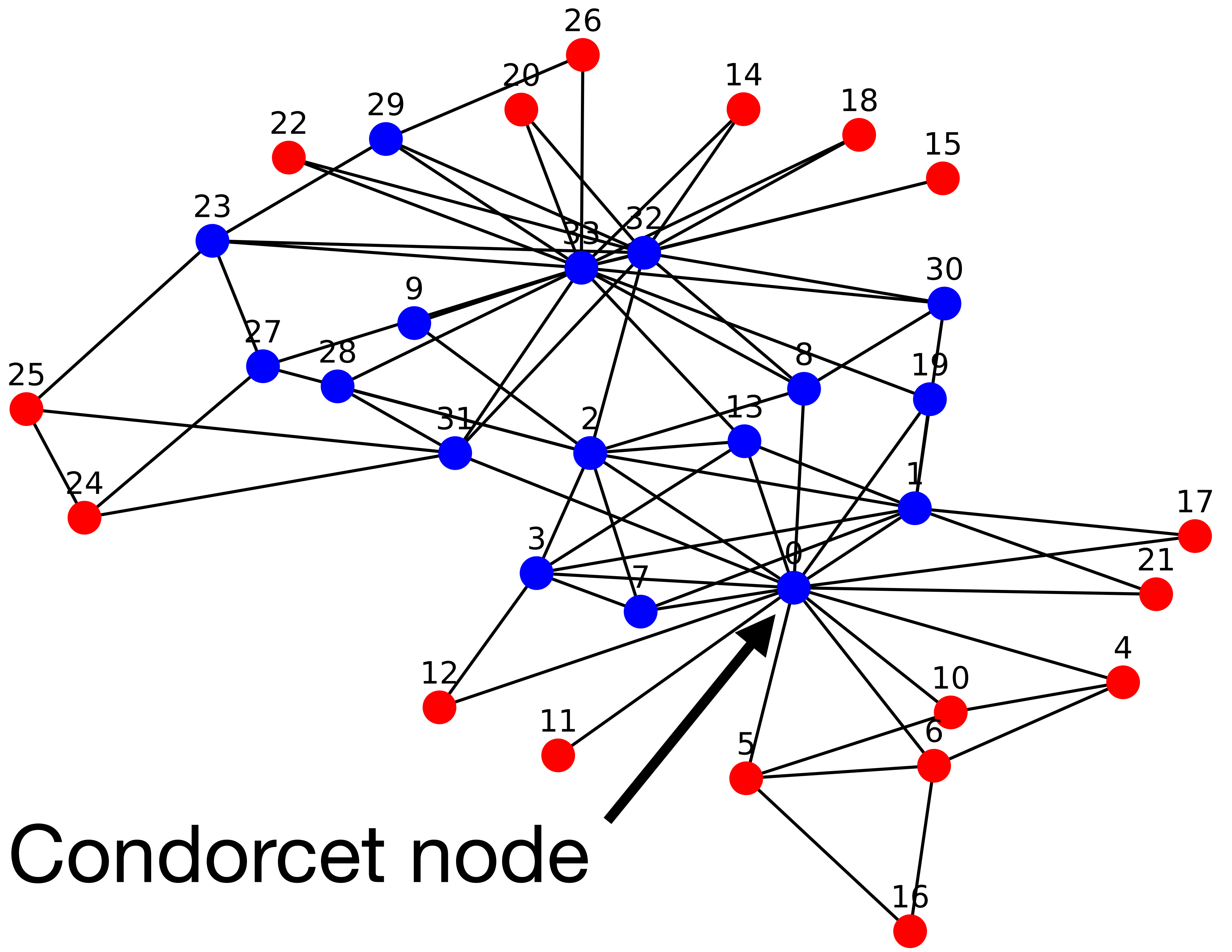




IRV exclusion zones in graphs

minimal exclusion zone





Other results

- Properties of exclusion zones
- Minimal IRV exclusion zones of paths, stars, bistars, and perfect binary trees
- Experiments: approximation algorithm on real-world networks

Other results

- Properties of exclusion zones
- Minimal IRV exclusion zones of paths, stars, bistars, and perfect binary trees
- Experiments: approximation algorithm on real-world networks

Open questions

- Do other voting rules have nontrivial exclusion zones with 1-Euclidean preferences? d-Euclidean? Graphs?
- In higher dimensions, are there “natural” voter distributions with nontrivial IRV exclusion zones?

Thank you!

Supported in part by:


Kiran Tomlinson
Microsoft Research
kitomlinson@microsoft.com


Johan Ugander
Yale University


Jon Kleinberg
Cornell University

Exclusion Zones of Instant Runoff Voting, arxiv.org/abs/2502.16719

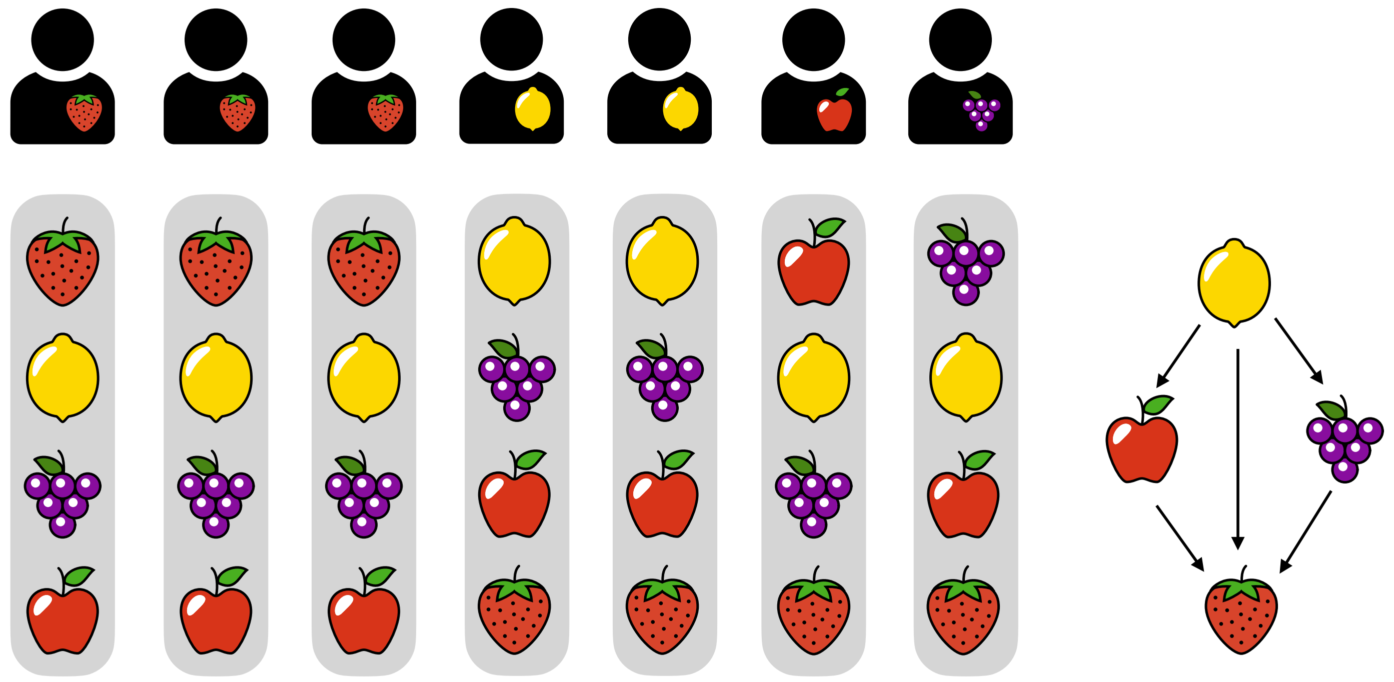
Open questions

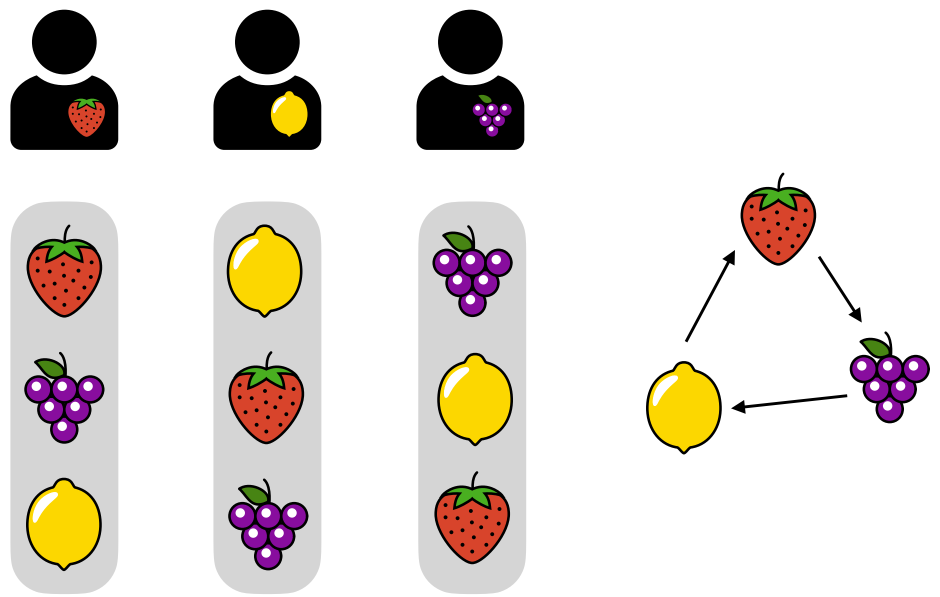
- Do other voting rules have nontrivial exclusion zones with 1-Euclidean preferences? d-Euclidean? Graphs?
- In higher dimensions, are there “natural” voter distributions with nontrivial IRV exclusion zones?

Background: Condorcet winners

Definition
A winner of every pairwise contest is a **Condorcet winner.**
A **Condorcet method** elects the Condorcet winner when one exists.

Nicolas de Condorcet
(1743 - 1794)

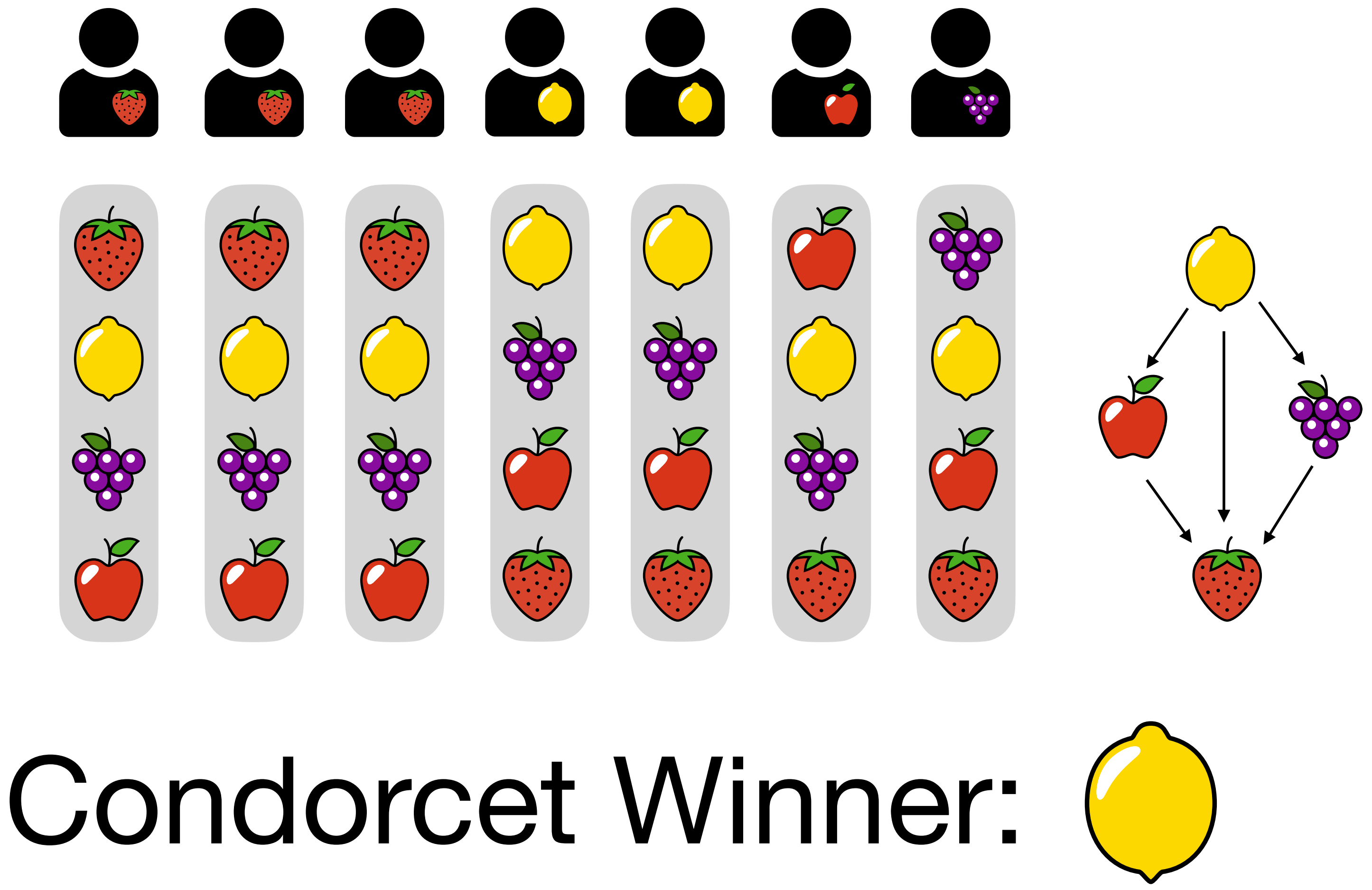




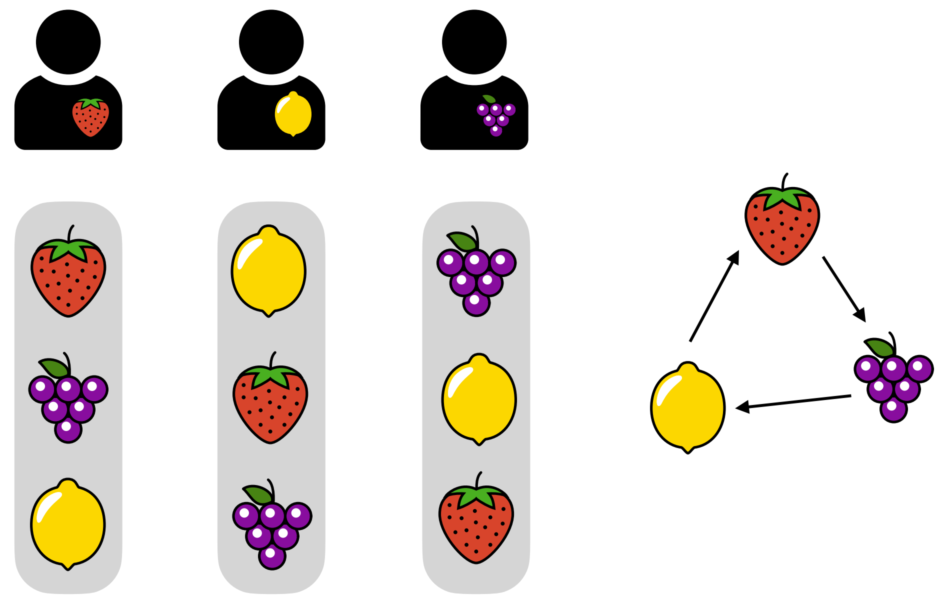
Background: Condorcet winners

Nicolas de Condorcet
(1743 - 1794)

Definition
A winner of every pairwise contest is a **Condorcet winner.**
A **Condorcet method** elects the Condorcet winner when one exists.



Condorcet Winner:

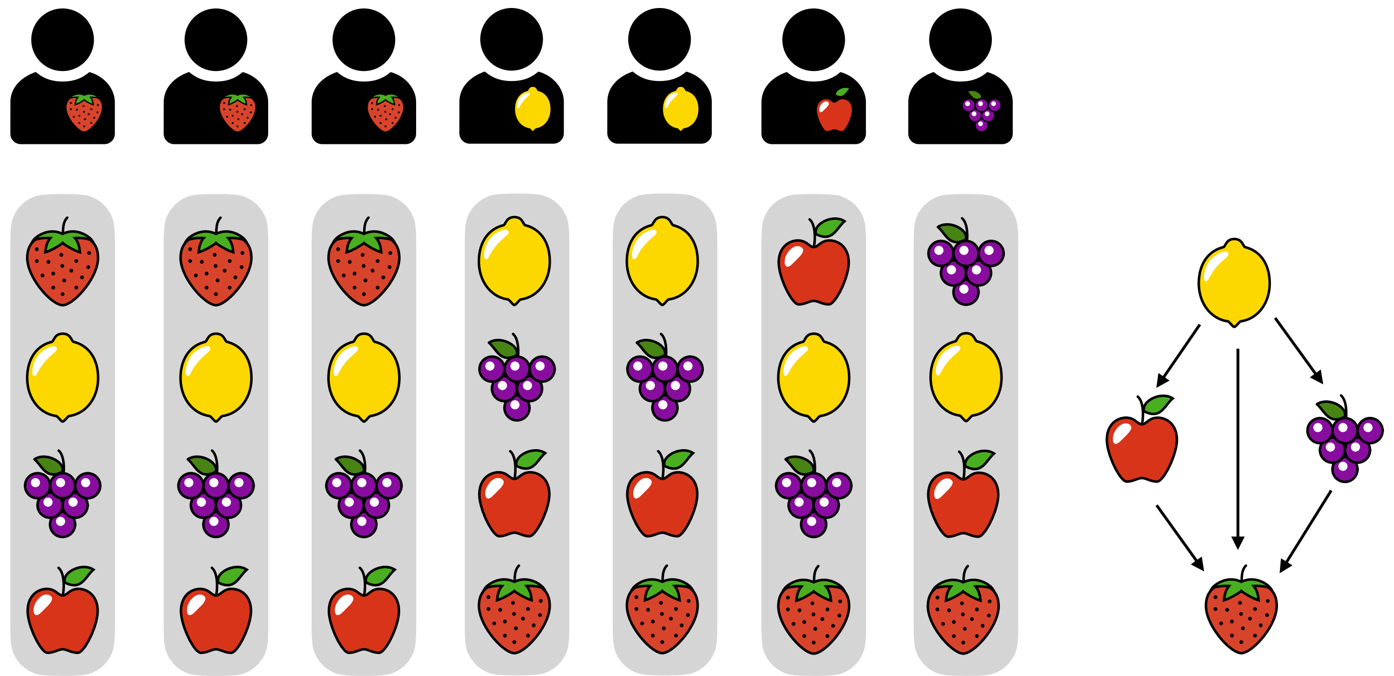


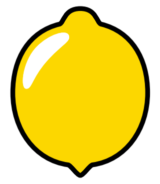
No Condorcet Winner

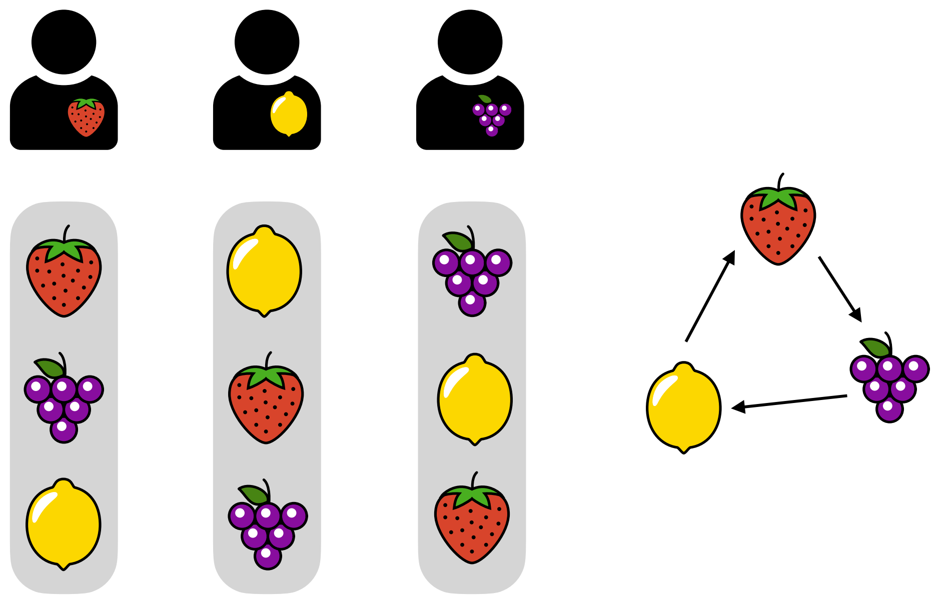
Background: Condorcet winners

Nicolas de Condorcet
(1743 - 1794)

Definition
A winner of every pairwise contest is a **Condorcet winner.**
A **Condorcet method** elects the Condorcet winner when one exists.



Condorcet Winner: 



No Condorcet Winner

Median Voter Theorem (Black, 1948)
With 1-Euclidean preferences, the candidate closest to the median voter is the Condorcet winner.

A recipe for proving the minimal exclusion zone is trivial

Proposition
For any exclusion zone $S \subseteq M$, if there is some election including $x \in S$ where y wins, then $y \in S$.

A recipe for proving the minimal exclusion zone is trivial

Proposition
For any exclusion zone $S \subseteq M$, if there is some election including $x \in S$ where y wins, then $y \in S$.

Condorcet Chain Lemma
Given an election setting, if there exist elections $C_1, \ldots, C_n$ with candidates $w_1 \in C_1, \ldots, w_1 \in C_n$ such that:

1. C_1 includes a weak Condorcet position, but a different candidate w_1 wins
2. each C_{i+1} includes w_i, but some other candidate w_{i+1} wins
3. w_n is a weak anti-Condorcet position,

then the election setting has no nontrivial exclusion zones.

Condorcet and anti-Condorcet positions

Definition
Given a metric space M and a voter distribution:

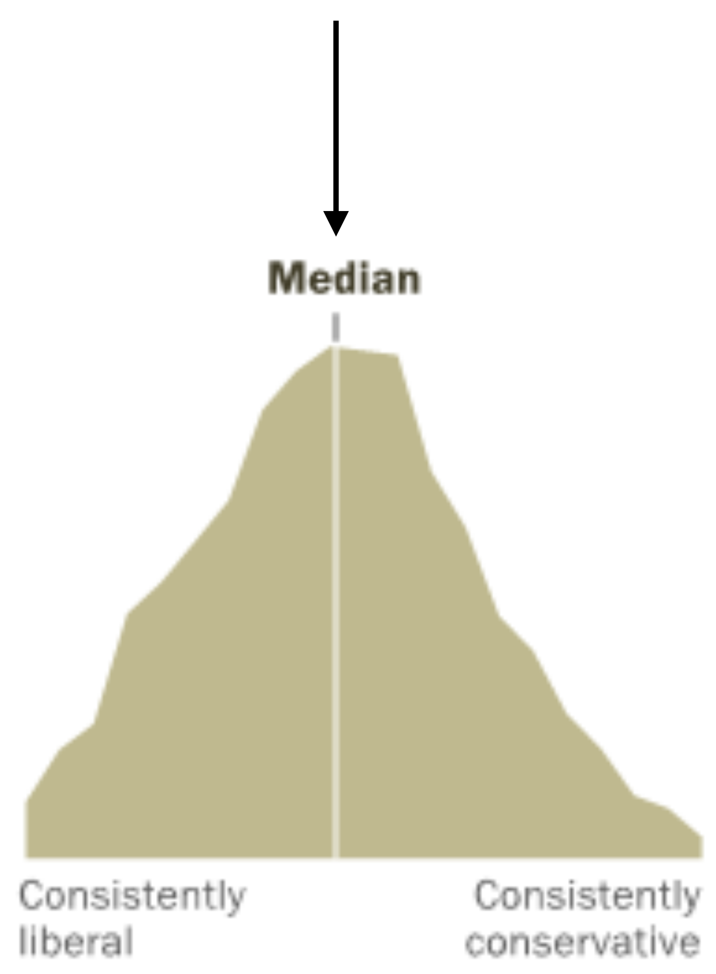
Condorcet and anti-Condorcet positions

Definition

Given a metric space M and a voter distribution:

$x \in M$ is a **weak Condorcet position** if for any other $y \in M$, at least half of the voters are closer to x than to y. (aka the **core**; strong with $>$ half).

strong Condorcet position



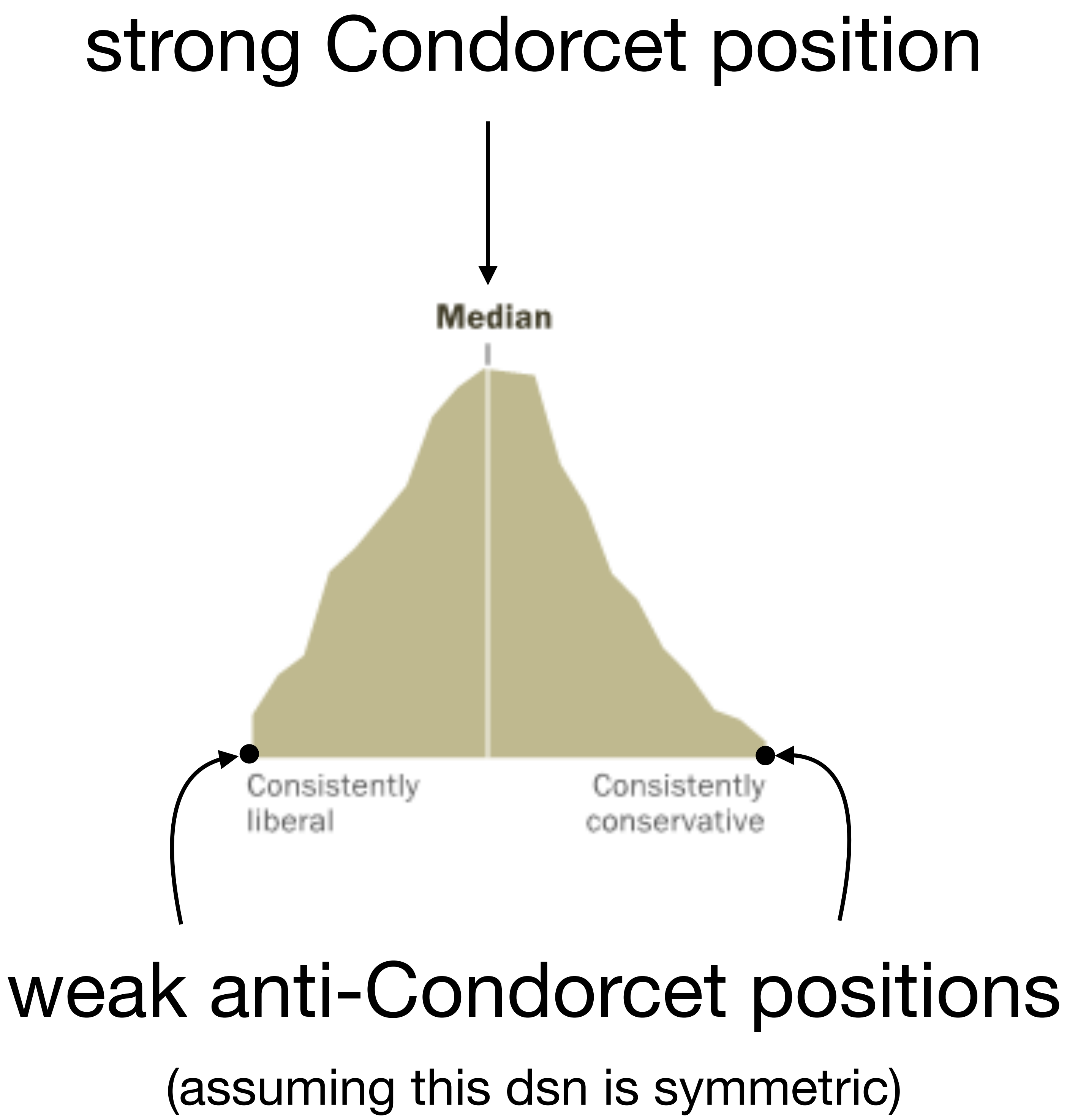
Condorcet and anti-Condorcet positions

Definition

Given a metric space M and a voter distribution:

$x \in M$ is a **weak Condorcet position** if for any other $y \in M$, at least half of the voters are closer to x than to y. (aka the **core**; strong with $>$ half).

$x \in M$ is a **weak anti-Condorcet position** if for any other $y \in M$, at least half of the voters are closer to y than to x



Condorcet and anti-Condorcet positions

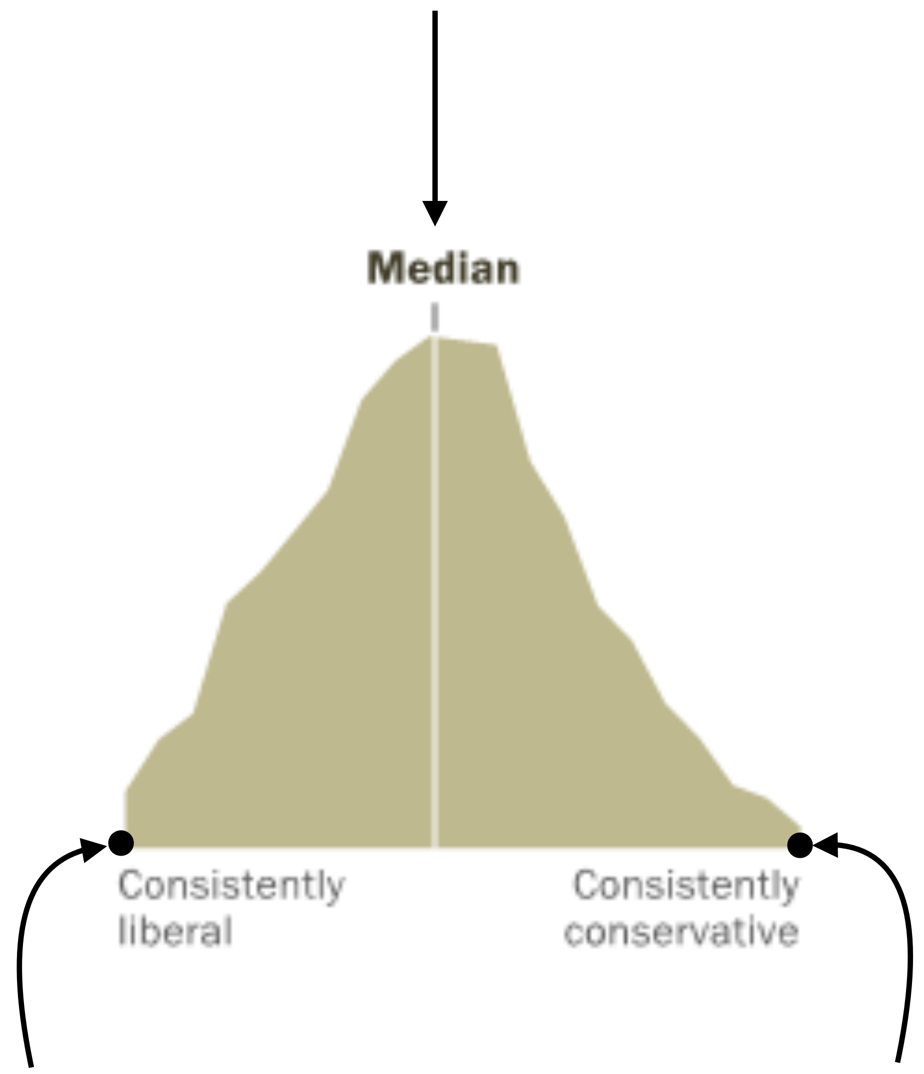
Definition

Given a metric space M and a voter distribution:

$x \in M$ is a **weak Condorcet position** if for any other $y \in M$, at least half of the voters are closer to x than to y. (aka the **core**; strong with $>$ half).

$x \in M$ is a **weak anti-Condorcet position** if for any other $y \in M$, at least half of the voters are closer to y than to x

strong Condorcet position



weak anti-Condorcet positions

(assuming this dsn is symmetric)

Proposition

For any reasonable* voting rule:

1. Any weak Condorcet position is in the minimal exclusion zone.
2. The only exclusion zone containing weak anti-Condorcet positions is trivial.

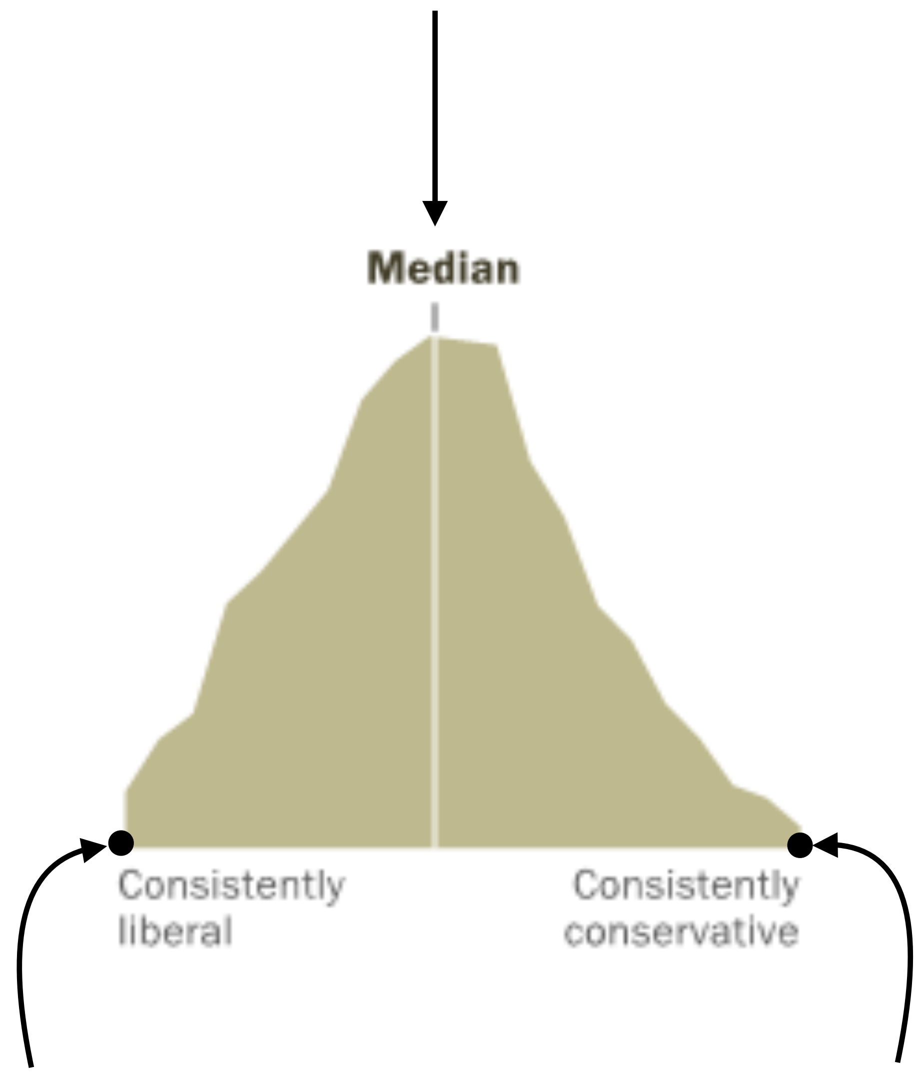
Condorcet and anti-Condorcet positions

Definition
Given a metric space M and a voter distribution:

$x \in M$ is a **weak Condorcet position** if for any other $y \in M$, at least half of the voters are closer to x than to y. (aka the **core**; strong with $>$ half).

$x \in M$ is a **weak anti-Condorcet position** if for any other $y \in M$, at least half of the voters are closer to y than to x

strong Condorcet position



weak anti-Condorcet positions
(assuming this dsn is symmetric)

Proposition
For any reasonable* voting rule:
1. Any weak Condorcet position is in the minimal exclusion zone.
2. The only exclusion zone containing weak anti-Condorcet positions is trivial.

*satisfies majority criterion in two-candidate elections