Proper Scoring Rules

~~Wagering Mechanisms~~: From Forecaster Selection to Fair Division

Rupert Freeman
University of Virginia, Darden School of Business

Based on joint work with Andreas Krause, David Pennock, Chara Podimata, Jennifer Wortman Vaughan, and Jens Witkowski

Eliciting Truthful Forecasts with Scoring Rules

- A central entity wants to predict whether the number of COVID-19 cases will increase the next day.

Increase	Decrease	Payment:
60%	40%	$s(0.6,0) = 0.64$
30%	70%	$s(0.3,0) = 0.91$
p_i	$1 - p_i$	$s(p_i, x) = 1 - (x - p_i)^2$

Quadratic score [Brier 1950]

Strictly proper (incentive compatible): Forecaster strictly maximizes their expected score by truthfully reporting p_i

Eliciting Truthful Forecasts with Scoring Rules

- A central entity wants to predict whether the number of COVID-19 cases will increase the next day.

	Day 1:		Day 2:		Day 3:	
	Up	Down	Up	Down	Up	Down
	60%	40%	40%	60%	45%	55%
	30%	70%	55%	45%	50%	50%
	30%	70%	55%	45%	40%	60%

. . .

Proper Scoring Rules – Quick Summary

- Scoring rule: Function that assigns a score/payment to a forecaster based on their report p_i and the event outcome x
  - Quadratic/Brier scoring rule very popular in practice

- Scoring rule is (strictly) proper if the forecaster (strictly) maximizes their expected score by truthfully reporting their subjective probability

- (Informal) More accurate prediction = higher expected score

#	Δ2d	Team Name	Score	Entries	Last Submission UTC (Best – Last Submission)
1	↑1	**Miroslaw Horbal**	**0.57421**	**34**	**Fri, 06 Nov 2015 04:10:10**
2	↓1	NxGTR	0.59159	47	Fri, 06 Nov 2015 02:45:39
3	↑8	Branden Murray	0.59890	17	Fri, 06 Nov 2015 03:57:12
4	↑3	(͡° ͜ʖ ͡°)	0.60761	3	Thu, 05 Nov 2015 20:59:52
5	↓2	Siddha	0.60838	14	Thu, 05 Nov 2015 17:01:32 (-2.6d)
6	↓2	Jordan Goblet	0.61620	23	Fri, 06 Nov 2015 08:36:56 (-34.3h)
7	↑75	KW Wu	0.62250	9	Fri, 06 Nov 2015 07:49:57 (-25.2h)
8	↑7	Keiku	0.62470	7	Fri, 06 Nov 2015 08:27:57
9	↓3	Hui Hu	0.62915	26	Fri, 06 Nov 2015 05:37:06
10	↓5	Eric	0.63030	28	Wed, 04 Nov 2015 11:17:47 (-4.1h)

GOOD JUDGMENT

Season leaderboard

Entire season

RANK	NAME	POINTS	PERCENTILE
1	Griffin Colaizzi	+1,126.2	99th
2	Joseph Ewbank	+1,100.6	99th
3	Peter Keith	+1,057.9	99th
4	Jan Hájek	+1,052.5	99th
5	Chandrasekhar Cidambi	+1,052.4	99th
6	Maxime Turgeon	+1,037.6	99th
7	Jeff Rolfes	+1,024.1	99th
8	Caleb Heartbird	+1,022.5	99th
9	Trevor Horton	+1,015.1	99th
11	Jack Overby	+1,008.2	99th
12	Jonathan Markowitz	+1,003.3	99th
13	Andrew Kastelman	+1,001.0	99th

FiveThirtyEight

Search for a pollster

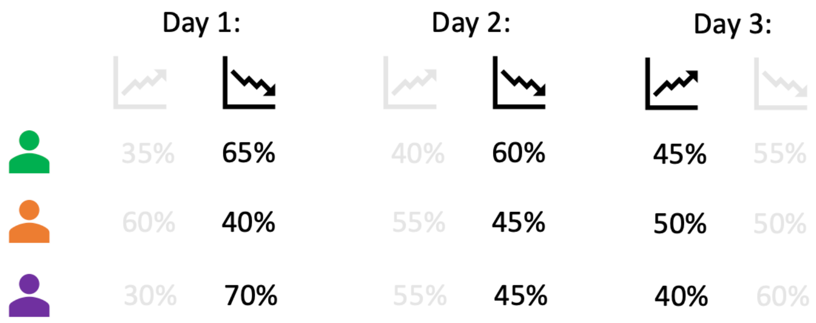
POLLSTER	METHOD	LIVE CALLER WITH CELLPHONES	NCPP/ AAPOR/ ROPER	POLLS ANALYZED	ADVANCED +/-	PREDICTIVE +/-	538 GRADE	BANNED BY 538	MEAN-REVERTED BIAS
SurveyUSA	IVR/ online/ live		●	777	-1.1	**-0.9**	A		D+0.1
Rasmussen Reports/Pulse Opinion Research	IVR/ online			711	+0.2	**+0.6**	C+		R+1.5
Zogby Interactive/ JZ Analytics	Online			464	+0.6	**+1.0**	C		R+0.9
Mason-Dixon Polling & Research Inc.	Live	●		420	-0.5	**-0.3**	B+		R+0.7
Public Policy Polling	IVR/ online			411	-0.4	**0.0**	B		D+0.3
YouGov	Online			375	-0.4	**+0.1**	B		D+0.3

Forecasting Competitions

Incentive Compatible Forecasting Competitions. Jens Witkowski, Rupert Freeman, Jennifer Wortman Vaughan, David Pennock, Andreas Krause. Management Science 2022.

Forecasting Competitions

- Forecasting Competition: Given a sequence of predictions and outcomes, select a single forecaster
  - Forecasters derive positive utility from being selected, zero otherwise



- In practice: Forecasters are scored by quadratic score, highest score wins

- Not incentive compatible… [Lichtendahl and Winkler 2007]

- Theorem: No deterministic mechanism is (strictly) incentive compatible.

The Single Event Case

- First attempt: Select each forecaster with probability proportional to their quadratic score
  - Not incentive compatible
- Instead: Borrow a trick from Kilgour and Gerchak [2004]
- Event Lotteries Forecaster selection (ELF): Select forecaster i with probability

$$\frac{1}{n} + \frac{1}{n}\left(\boxed{s(p_i, x)} - \boxed{\frac{1}{n-1} \sum_{j \neq i} s(p_j, x)} \right)$$

Score of agent i

Average score of other agents

Accuracy of ELF

- Suppose that the event has an underlying true probability θ
  - Let $s(p,\theta)$ denote the expected score for reporting p when true probability is θ
  - If $s(p_i,\theta) > s(p_j,\theta)$ then we say forecaster i is more accurate than j
- ELF selects forecaster i with probability

$$\frac{1}{n} + \frac{1}{n}\left(s(p_i,\theta) - \frac{1}{n-1}\sum_{j\neq i} s(p_j,\theta)\right)$$

  - Most accurate forecaster is selected with $> \frac{1}{\mathrm{n}}$ probability
- Theorem: For two forecasters, no incentive-compatible mechanism selects the most accurate forecaster with higher probability than ELF

The Multiple Event Case

- ELF: Choose one event at random, run single-event ELF
  - Retains incentive-compatibility even if events are arbitrarily correlated
  - Doesn’t provide better accuracy guarantees than the single-event version

- I-ELF: Run single-event ELF on each event to find a winner w_k for each event k. Select the forecaster that wins the most events.
  - Is incentive compatible when events are independent*
  - Selects the most accurate forecaster with probability approaching 1 as number of events grows

Example: Predicting COVID-19 cases

- A central entity wants to predict whether the number of COVID-19 cases will increase the next day.

	Day 1:		Day 2:		Day 3:		
	35%	65%	40%	60%	45%	55%	
	55%	45%	55%	45%	50%	50%	. . .
	45%	55%	55%	45%	30%	70%	
	?		?		?		

Incentive-Compatible Online Learning

No-Regret and Incentive-Compatible Online Learning. Rupert Freeman, David Pennock, Chara Podimata, Jennifer Wortman Vaughan. ICML 2020.

The Problem

1. For each event $t \in T$:
2. Each of n experts ***strategically*** reports a probabilistic *prediction* $p_{i,t}$
3. Learner chooses *prediction* $\overline{p_t} = \sum_i \pi_{i,t}\, p_{i,t}$
4. Event is realized (e.g., )
5. Every prediction incurs quadratic loss: $(p - x)^2$
6. Learner updates *distribution* $\boldsymbol{\pi_t} \rightarrow \boldsymbol{\pi_{t+1}}$ over experts

Learner's Goal – achieve "no regret"
$Loss(algo) - Loss(best\_expert) \leq o(T)$

Expert's goal (at $t \in [T]$):
Report prediction to maximize $\pi_{i,t+1}$

The Problem

1. For each event $t \in T$:
2. Each of n experts ***strategically*** reports a probabilistic *prediction* $p_{i,t}$
3. Learner chooses *prediction* $\bar{p}_t = \sum_i \pi_{i,t}\, p_{i,t}$
4. Event is realized (e.g.,)
5. Every prediction incurs quadratic loss: $(p - X)^2$
6. Learner updates *distribution* $\boldsymbol{\pi}_t \rightarrow \boldsymbol{\pi}_{t+1}$ over experts

Learner's Goal – achieve "no regret"
$$Loss(algo) - Loss(best\_expert) \leq o(T)$$

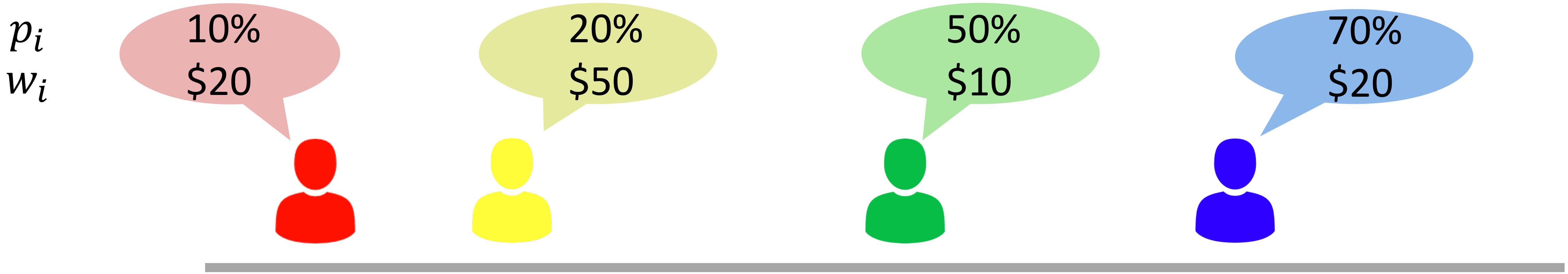
Multiplicati… …ate

$O(\sqrt{T\log(n)})$

Not incentive compatible!

Wagering Mechanisms [Lambert et al. 2008]

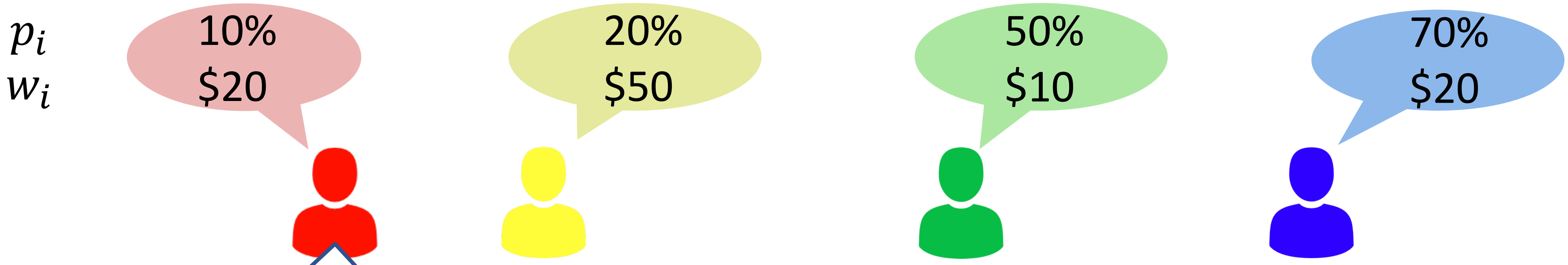
Will COVID-19 cases increase tomorrow?



	$14.30	$44.25	$12.75	$28.70
	$22.70	$55.25	$8.95	$13.10

Wagering Mechanisms [Lambert et al. 2008]

Will COVID-19 cases increase tomorrow?



$$\frac{1}{n} + \frac{1}{n}\left(s(p_i, x) - \frac{1}{n-1} \sum_{j \neq i} s(p_j, x) \right)$$

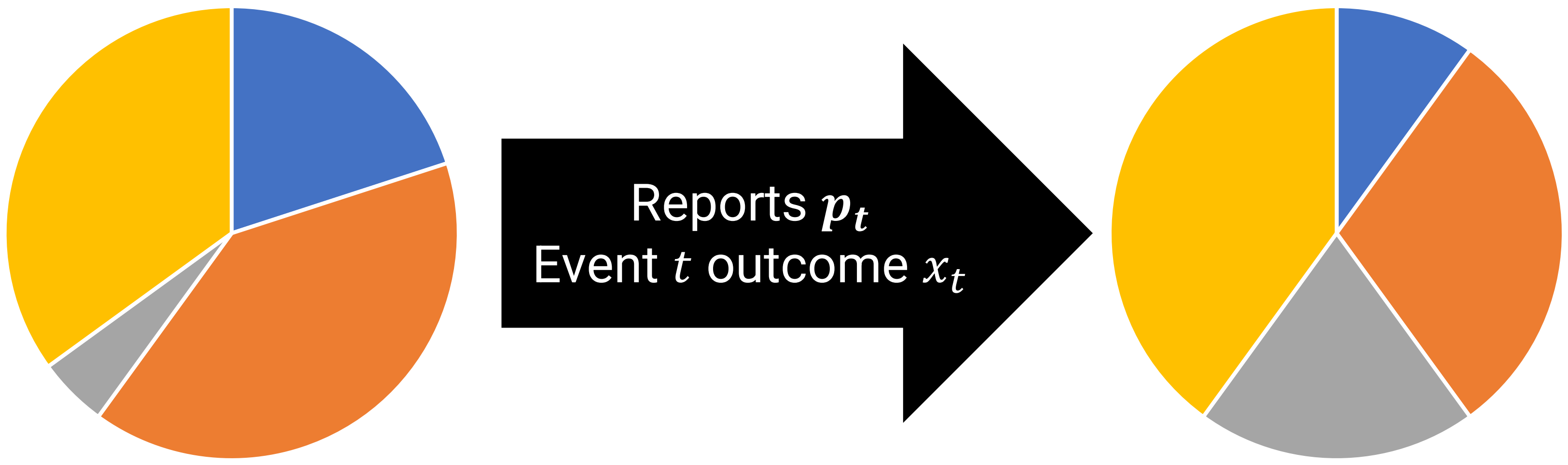
Weighted Score Wagering Mechanism

$$\mathrm{WSWM}_i(\boldsymbol{p}, \boldsymbol{w}) = w_i \left(1 + \frac{\sum_j \ell(p_j, x) w_j}{\sum_j w_j} - \ell(p_i, x) \right)$$

1) incentive compatible

2) strictly budget balanced

Online Learning and Wagering



$\boldsymbol{\pi_t}$ ✓

$\pi_{i,t}$ = wager

incentive compatible

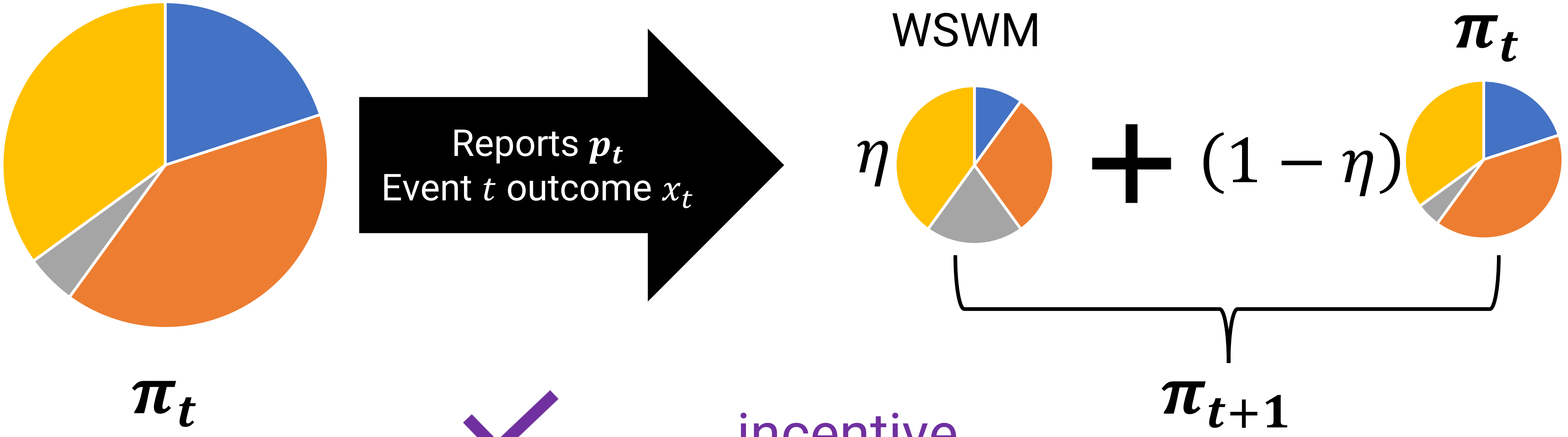
$\boldsymbol{\pi_{t+1}}$

$\pi_{i,t+1}$ = payment

Wagering Mechanisms for Online Learning

- Can learn using a wagering mechanism to update the distribution over experts
  - Takes care of incentive compatibility
  - What about regret?

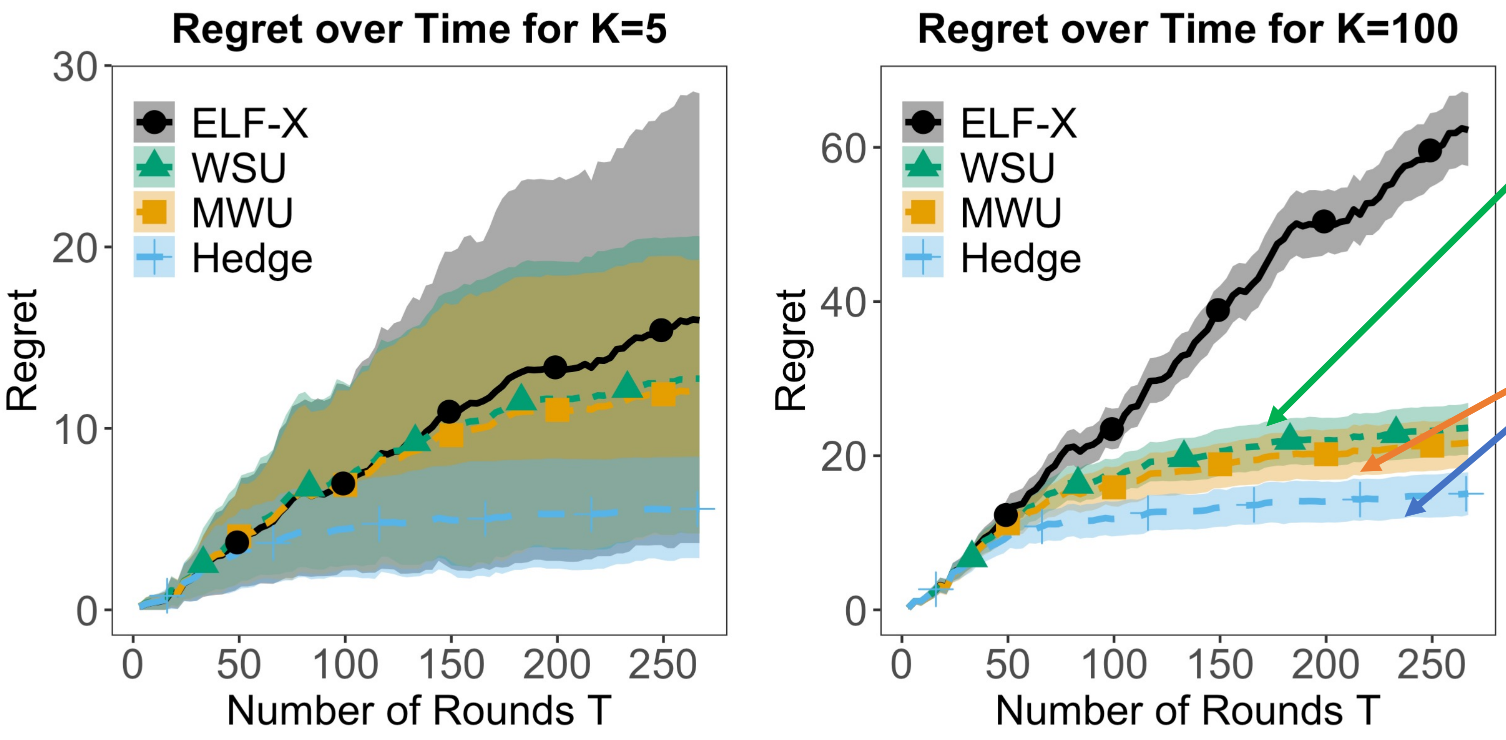
Wagering Mechanisms Made No-Regret



✓ incentive compatibility

$$Loss(algo) - Loss(best\_expert) \leq O\left(\sqrt{T\log(n)}\right)$$

Experiments on FiveThirtyEight NFL18-19 data



Our Algorithm

State-of-the-art no-regret **(but not IC)** algorithms

Fair Division

An Equivalence Between Wagering and Fair-Division Mechanisms. Rupert Freeman, David Pennock, Jennifer Wortman Vaughan. AAAI 2019.

ALGORITHMS

How to Cut Cake Fairly and Finally Eat It Too

28

Computer scientists have come up with an algorithm that can fairly divide a cake among any number of people.

spliddit

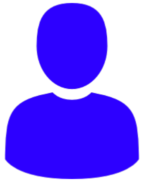
[PDF] **Dominant Resource Fairness**: Fair Allocation of Multiple **Resource** Types.

A Ghodsi, M Zaharia, B Hindman, A Konwinski… - NSDI, 2011 - static.usenix.org

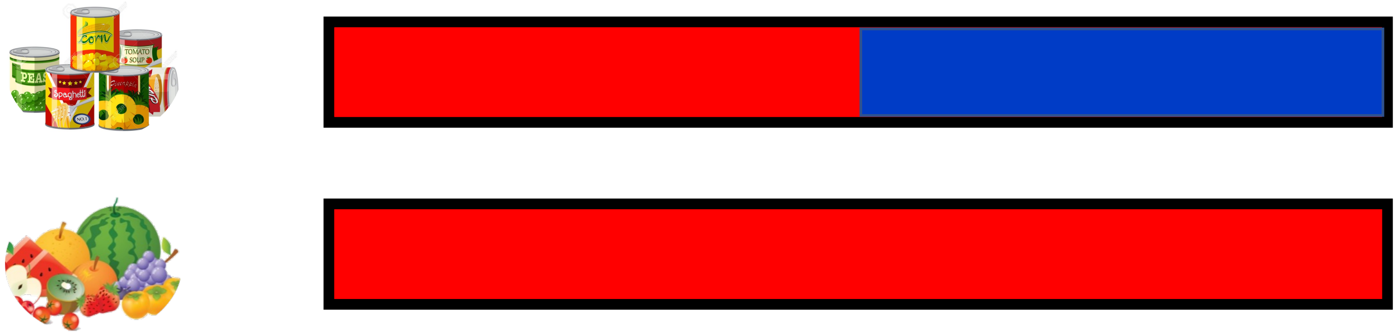
Abstract We consider the problem of fair resource allocation in a system containing different resource types, where each user may have different demands for each resource. To address this problem, we propose Dominant Resource Fairness (DRF), a generalization of max-min

Cited by 649 Related articles All 40 versions

Fair Division: Food Bank

			
$\frac{1}{2}\times 0.7 = 0.35$		0.7	0.3
$\frac{1}{2}\times 0.6 + 0.4 = 0.7$		0.6	0.4

← Red agent is indifferent between 2kg of canned food and 3kg of fresh food



Desirable Properties

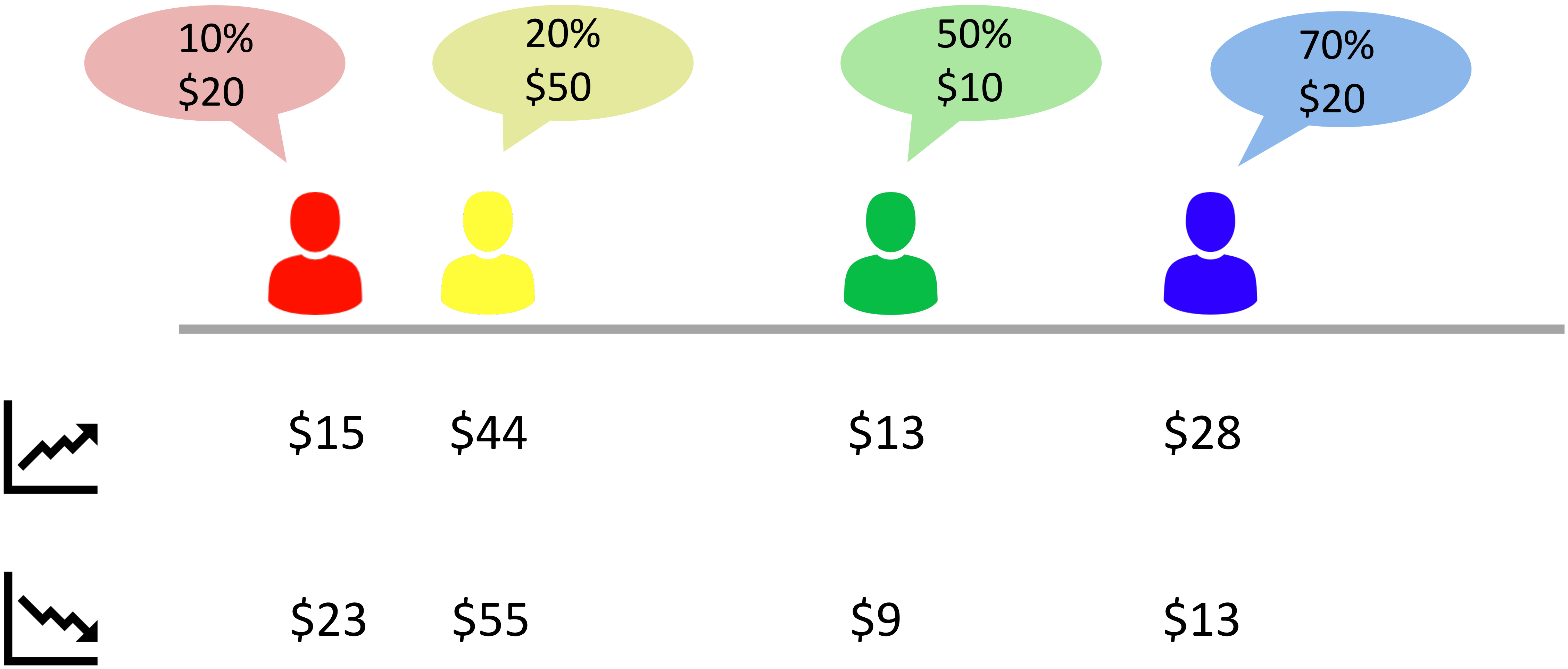
- Proportionality: Each agent receives 1/n of their value for all the goods
- Envy-freeness: No agent prefers the allocation of another agent to her own allocation
- Incentive Compatibility: An agent can never achieve higher utility by lying about their values
- Pareto Optimality: It is impossible to make some agent better off without making another agent worse off

Wagering and Allocation are Equivalent

- Theorem: There is a one-to-one correspondence between weakly budget-balanced wagering mechanisms and allocation mechanisms
- The correspondence preserves several desirable properties.

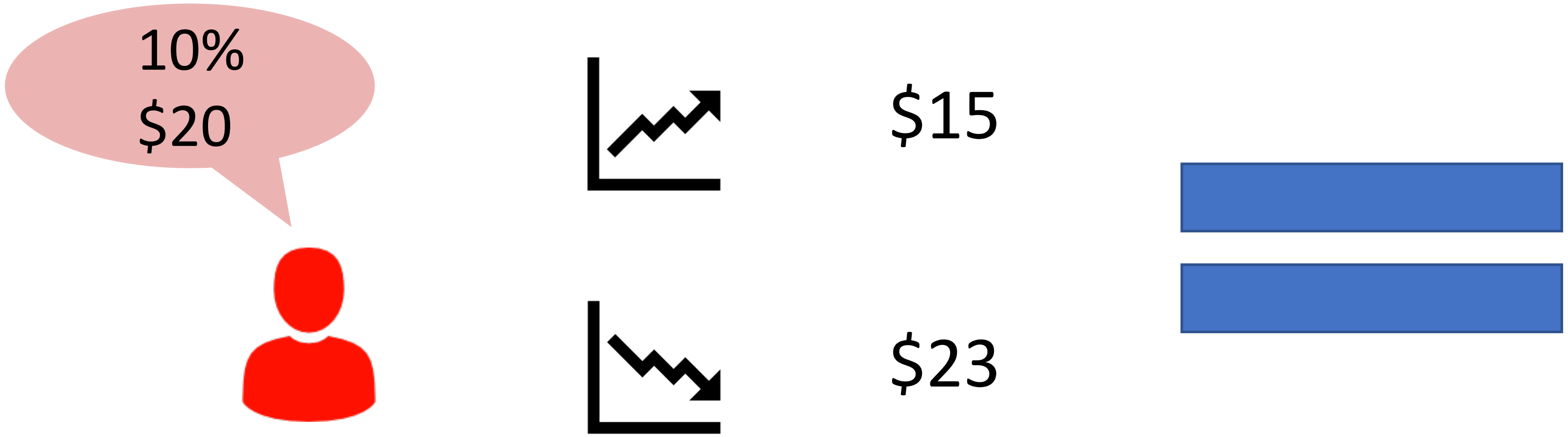
Fair Division	Wagering
Incentive Compatibility	Incentive Compatibility
Proportionality	Individual Rationality

Will COVID-19 Cases Increase Tomorrow?



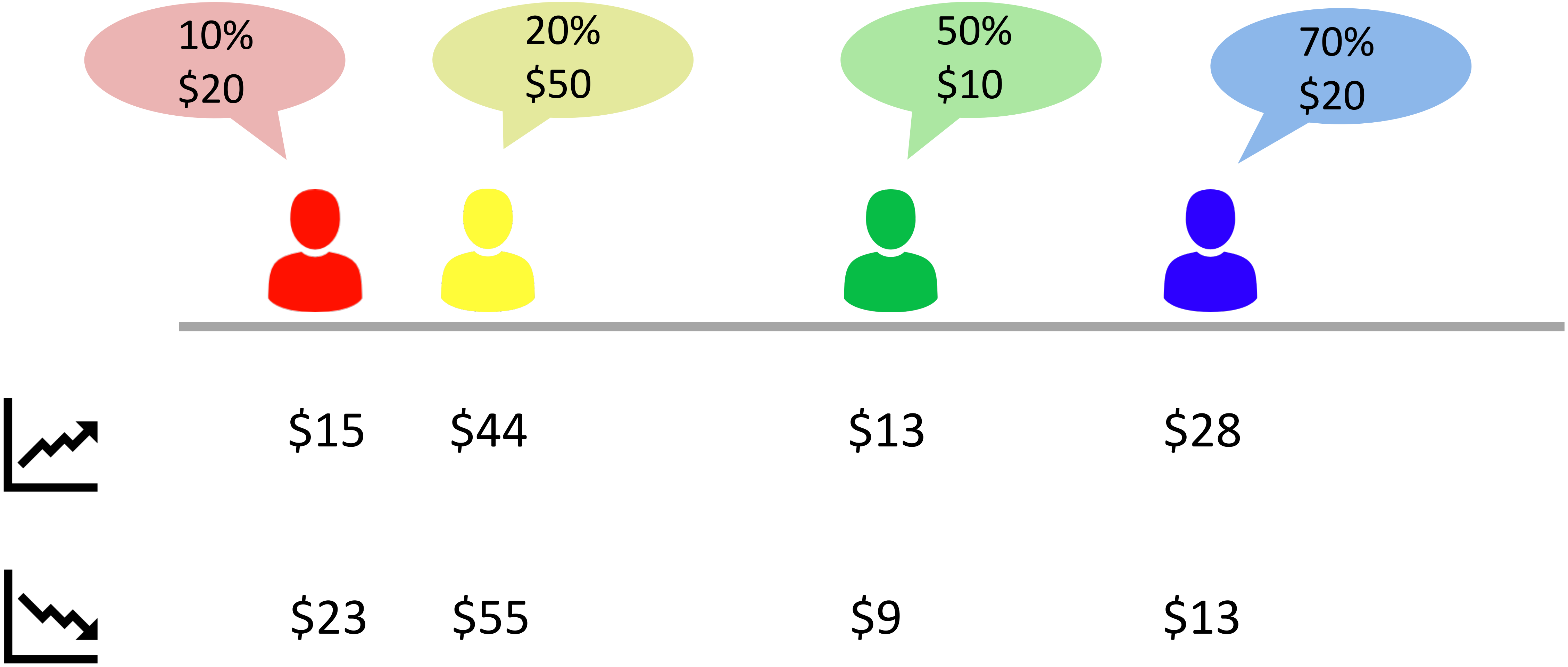
Thinking about securities

- Consider two types of securities: ‘yes’ securities which each pay out $1 if the event occurs, and ‘no’ securities which pay out $1 if it doesn’t.
  - Note: A ‘yes’/‘no’ pair is exactly equivalent to $1
  - Forecaster values ‘yes’ securities at p_i and ‘no’ securities at $1 - p_i$

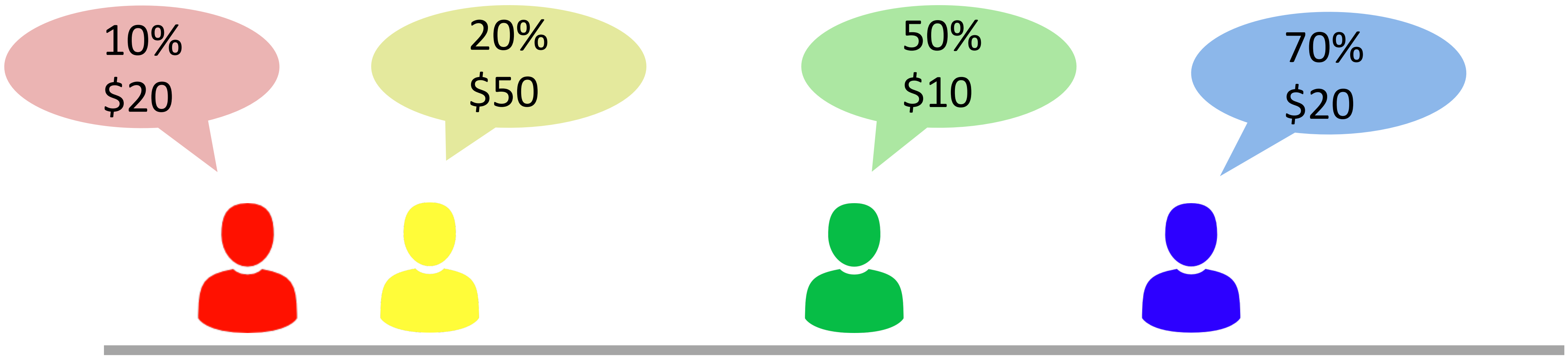


Agent is allocated 15 ‘yes’ securities and 23 ‘no’ securities

Will COVID-19 Cases Increase Tomorrow?



Will COVID-19 Cases Increase Tomorrow?



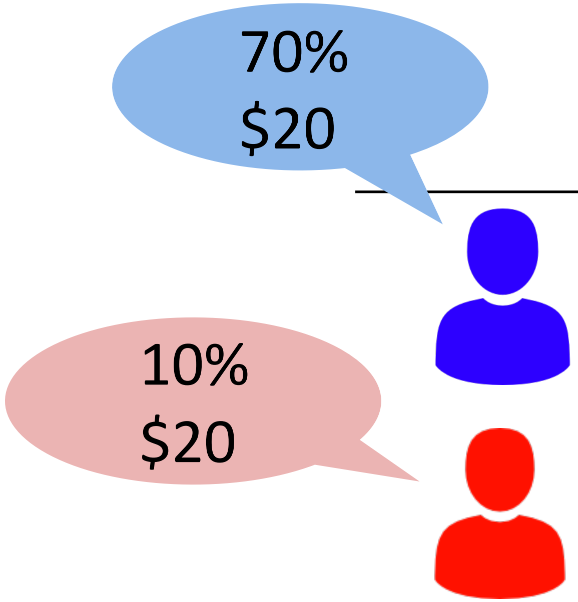
15 ‘yes’ and 23 ‘no’

44 ‘yes’ and 55 ‘no’

13 ‘yes’ and 9 ‘no’

28 ‘yes’ and 13 ‘no’

Will COVID-19 Cases Increase Tomorrow?

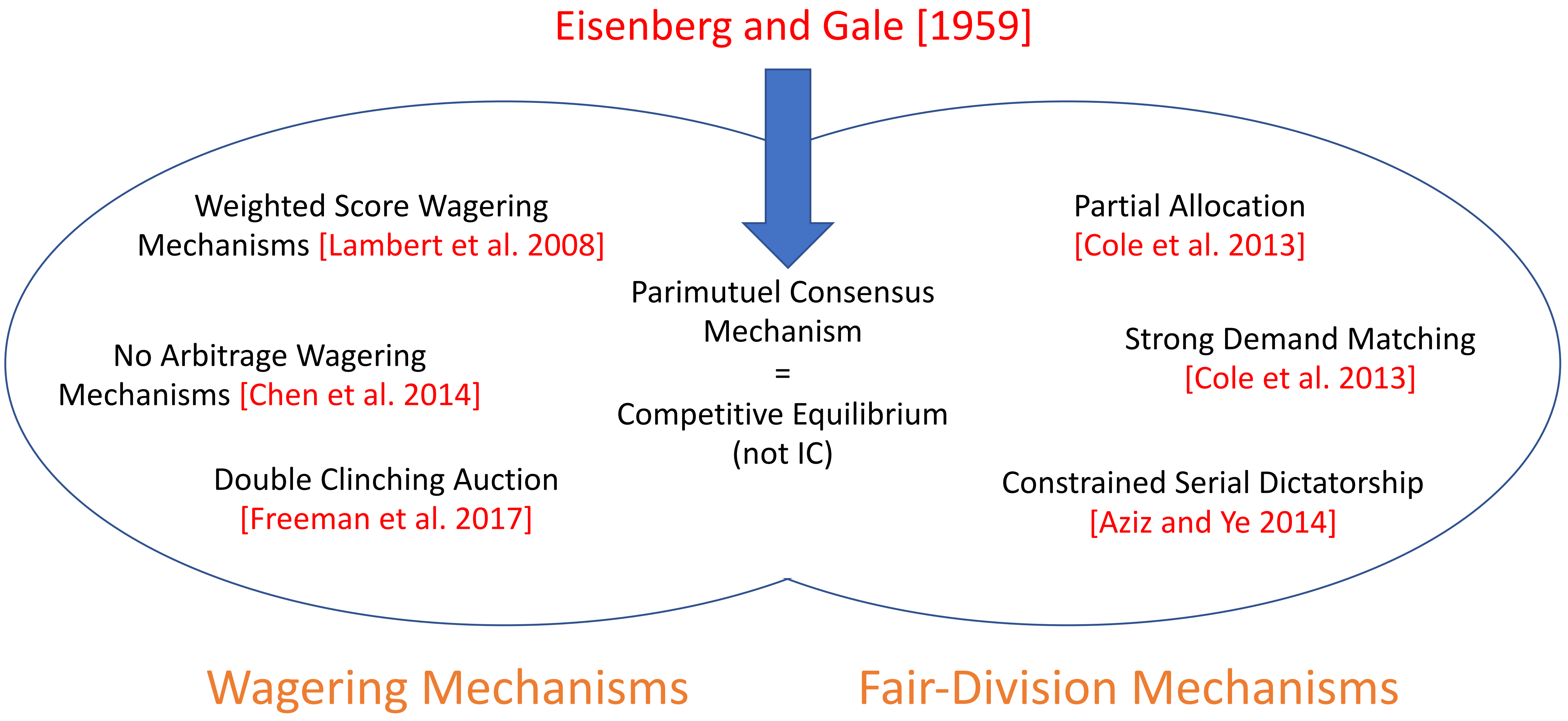


	'Yes' Securities	'No' Securities
	0.7	0.3
	0.1	0.9

Equivalence

Fair Division	**Wagering**
n agents	*n* forecasters
m items	*m* outcomes
Valuations	Probabilities
Weights	Wagers

Consequences



Consequences

- Wagering mechanisms as allocation mechanisms
  - Weighted Score Wagering Mechanism
    - First strictly incentive compatible allocation mechanisms
    - First non-trivial, incentive compatible, envy-free and proportional allocation mechanisms
- Allocation mechanisms as wagering mechanisms
  - Constrained Serial Dictatorship:
    Wagering mechanism that requires only ordinal probability judgments
  - Strong Demand Matching:
    Satisfies side-bet Pareto optimality at the expense of (minimal) individual rationality violations

Conclusion

- We have seen three (surprising?) applications of scoring rules
  - Forecasting competitions
  - No-Regret Learning
  - Fair Division
- Common technical theme: Dividing finite “resource” in incentive compatible way
- I haven’t found other applications but maybe you have one!

Thank you!