TCC 2018 (Goa)

Game Theoretic Notions of Fairness in Multi-Party Coin Toss

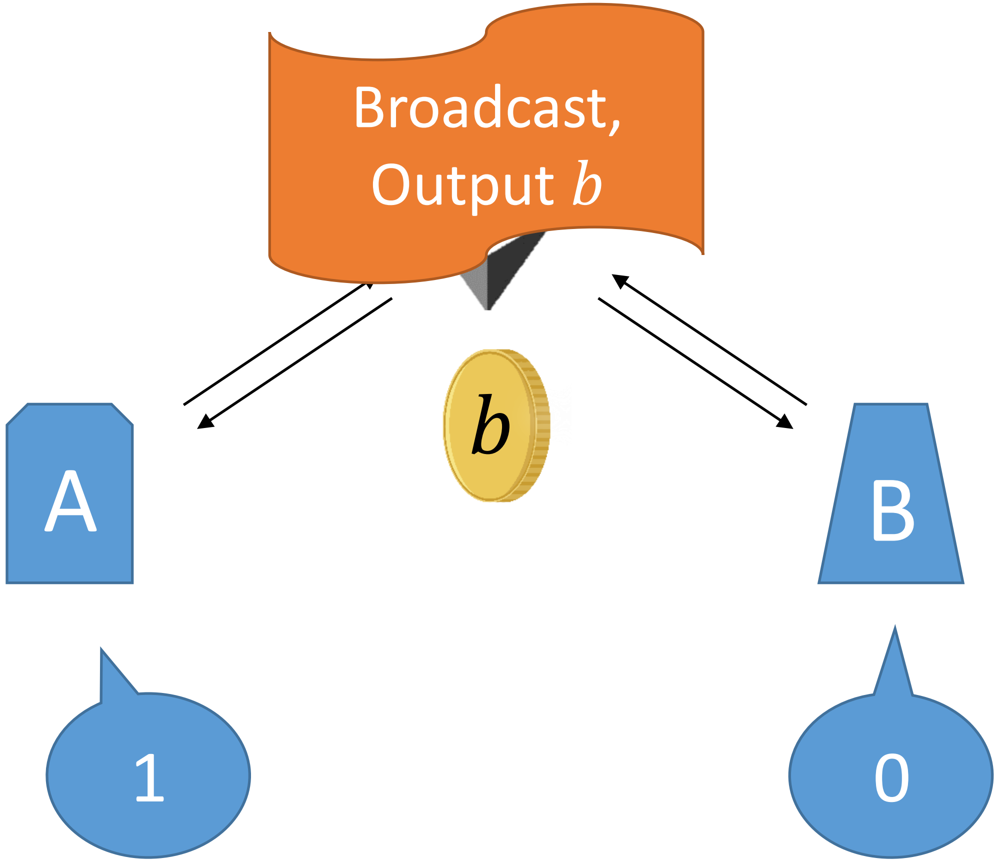
Kai-Min Chung, Yue Guo, **Wei-Kai Lin**, Rafael Pass, and Elaine Shi

Nov 13, 2018

中央研究院 ACADEMIA SINICA
Cornell University
CORNELL TECH | HOME OF THE JACOBS TECHNION-CORNELL INSTITUTE

Who Gets to TCC in Goa?

- Soft merge of A and B
- Only one gets to present



		A	B
Preference		1	0
Payoff	$b = 0$	0	1
	$b = 1$	1	0

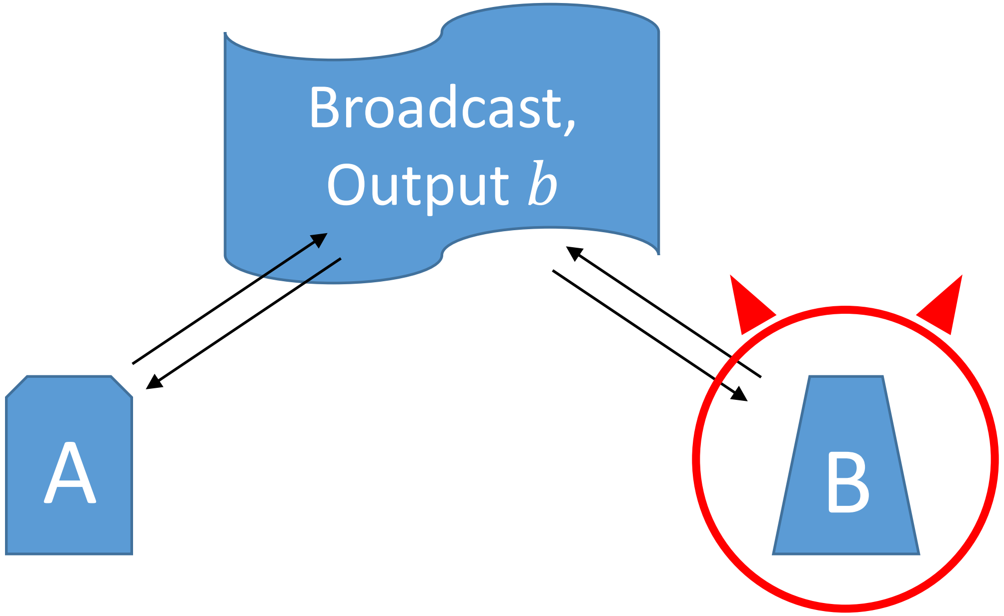
Strong Fairness of Coin Toss

Expected **output** of honest $= 0.5$

Corrupt majority, aborts early

[Cleve'86] Any n-party, $n \geq 2$,
Impossible even adversary is
comp-bounded and fail-stop

fail-stop:
aborts early,
otherwise honest



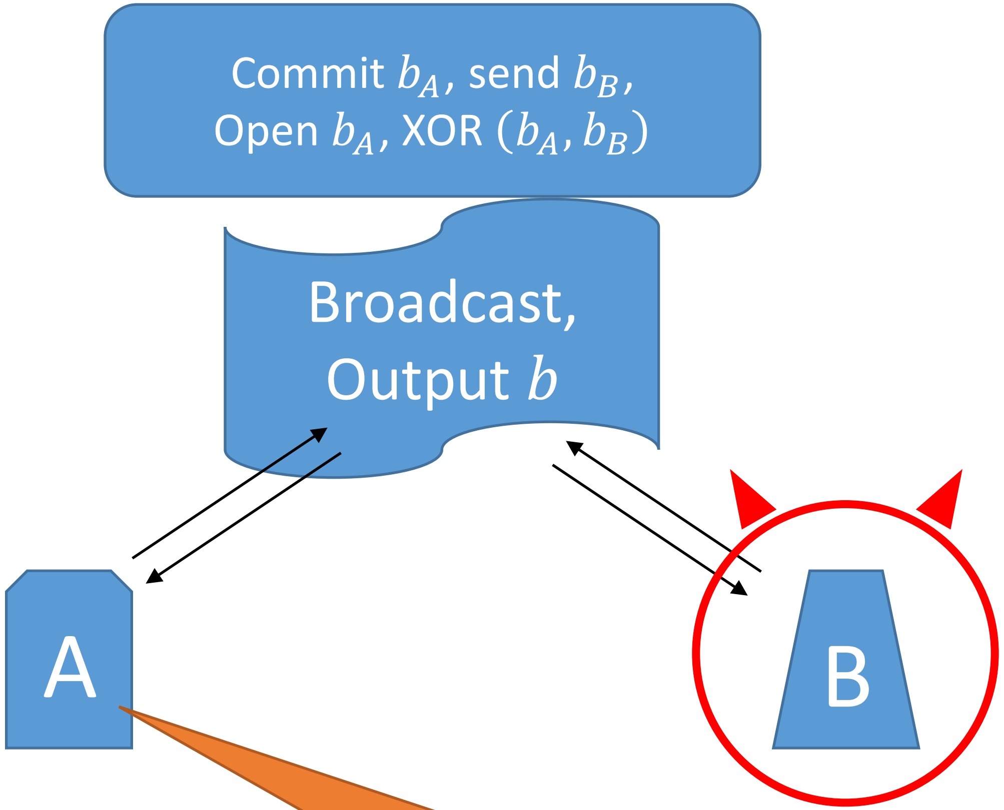
		A	B
Preference		1	0
Payoff	$b = 0$	0	1
	$b = 1$	1	0

Blum's Coin Toss

Intuition: no harm to honest

Expected **payoff** of honest $\geq \mathbf{0.5}$

[Blum'81]
2-party protocol from crypto commitments



If B aborts early, then A outputs 1

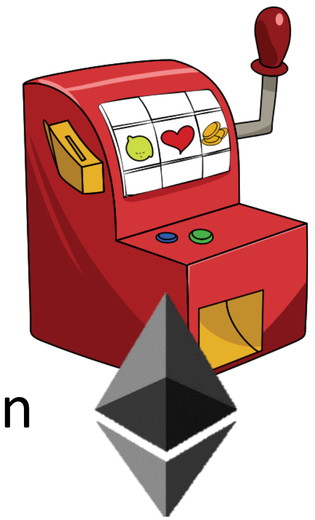
		A	B
Preference		1	0
Payoff	$b = 0$	0	1
	$b = 1$	1	0

Definition of 3-Party Weak Fairness?

Public-identifiable abort

Broadcast, Output b

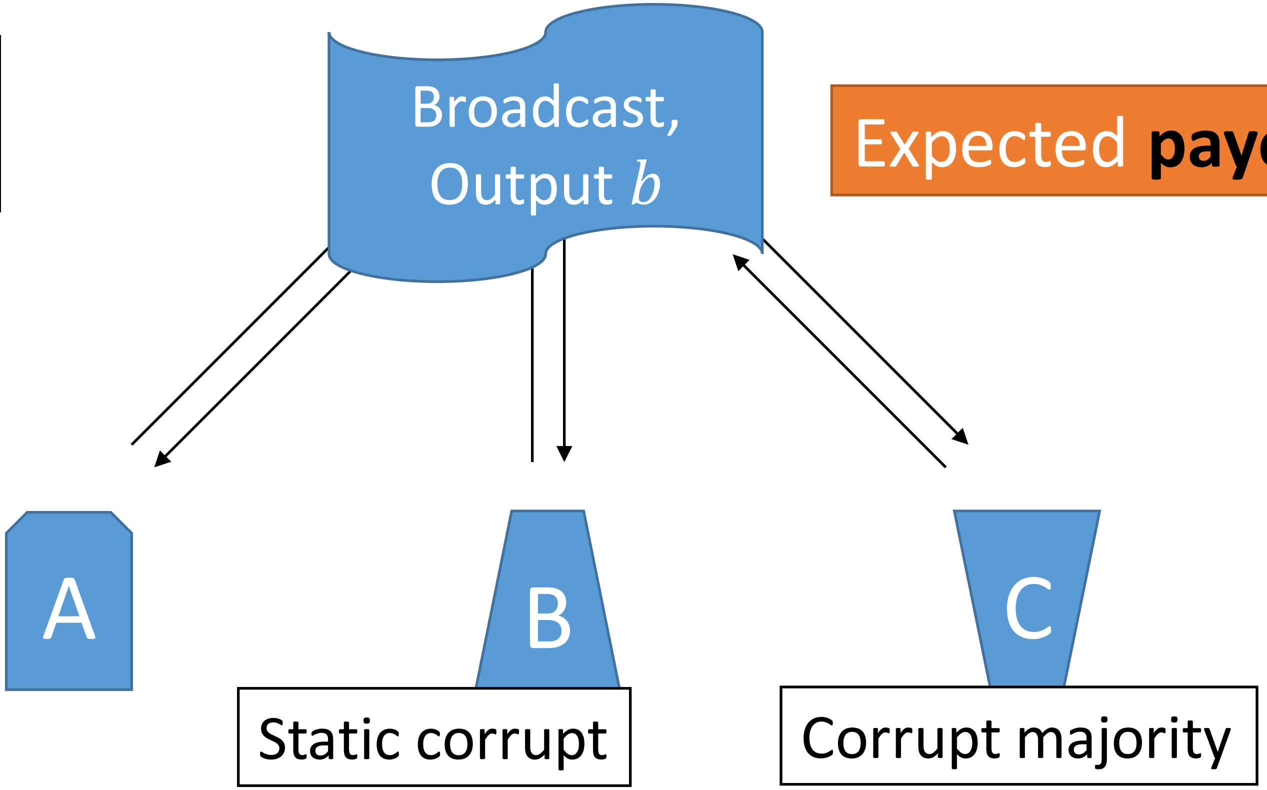
E.g. Gamble on blockchain



		A	B	C
			Static corrupt	Corrupt majority
Public Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

Definition of Maximin Fairness

Public-identifiable abort



Expected **payoff** of honest $\geq \mathbf{0.5}$

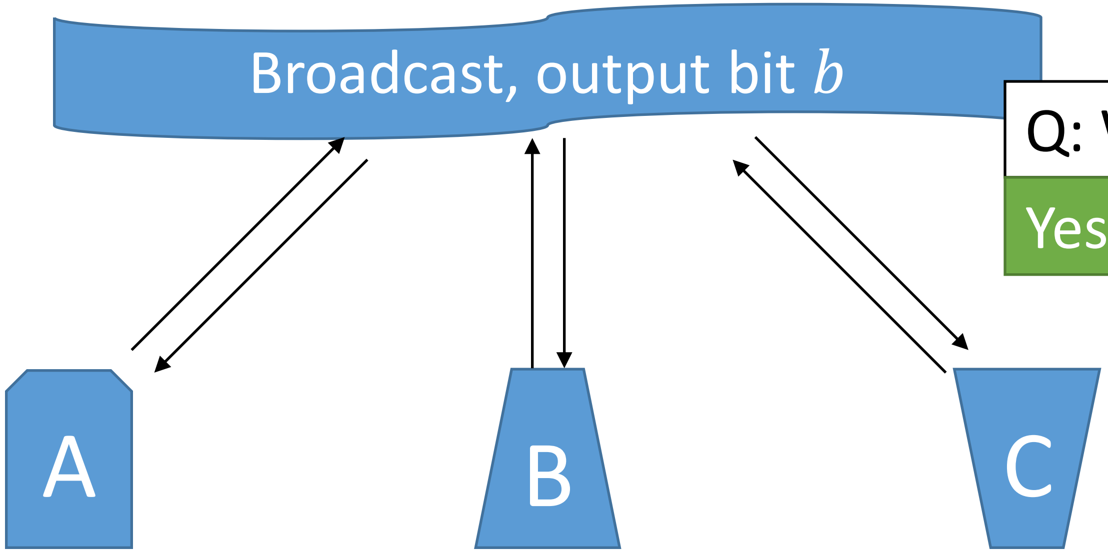
No harm to honest payoff

There are several "natural extensions"

Public

		A	B	C
Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

Maximin Fairness of 3-Party, Unanimous



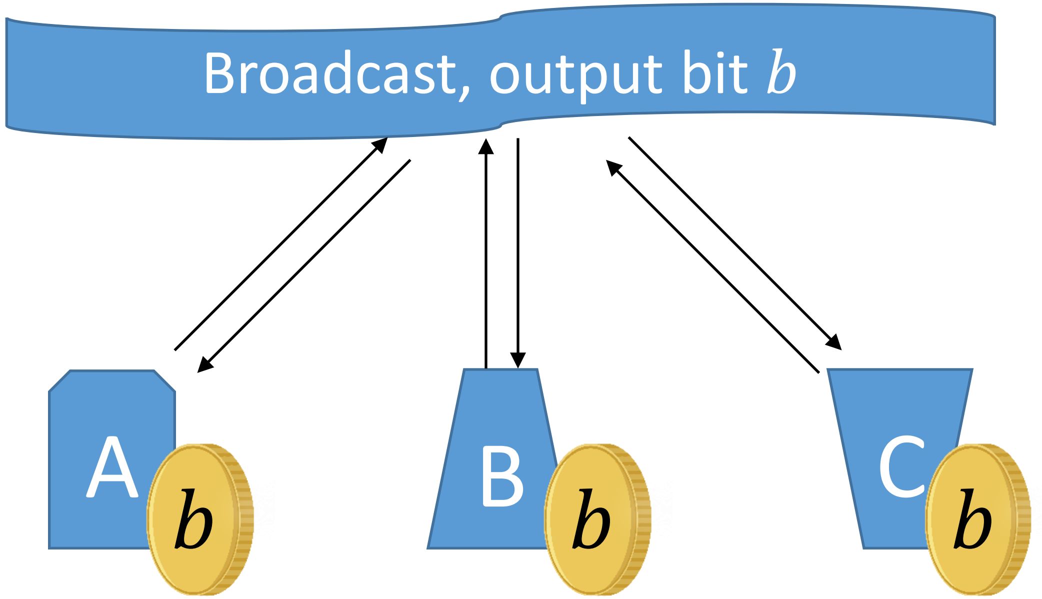
Q: Weak fairness?

Yes, Just output preference

		A	B	C
Public Preference		1	1	1
Payoff	$b = 0$	0	0	0
	$b = 1$	1	1	1

Maximin Fairness of 3-Party, Fail-Stop

abort early,
otherwise honest



		A	B	C
Public Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

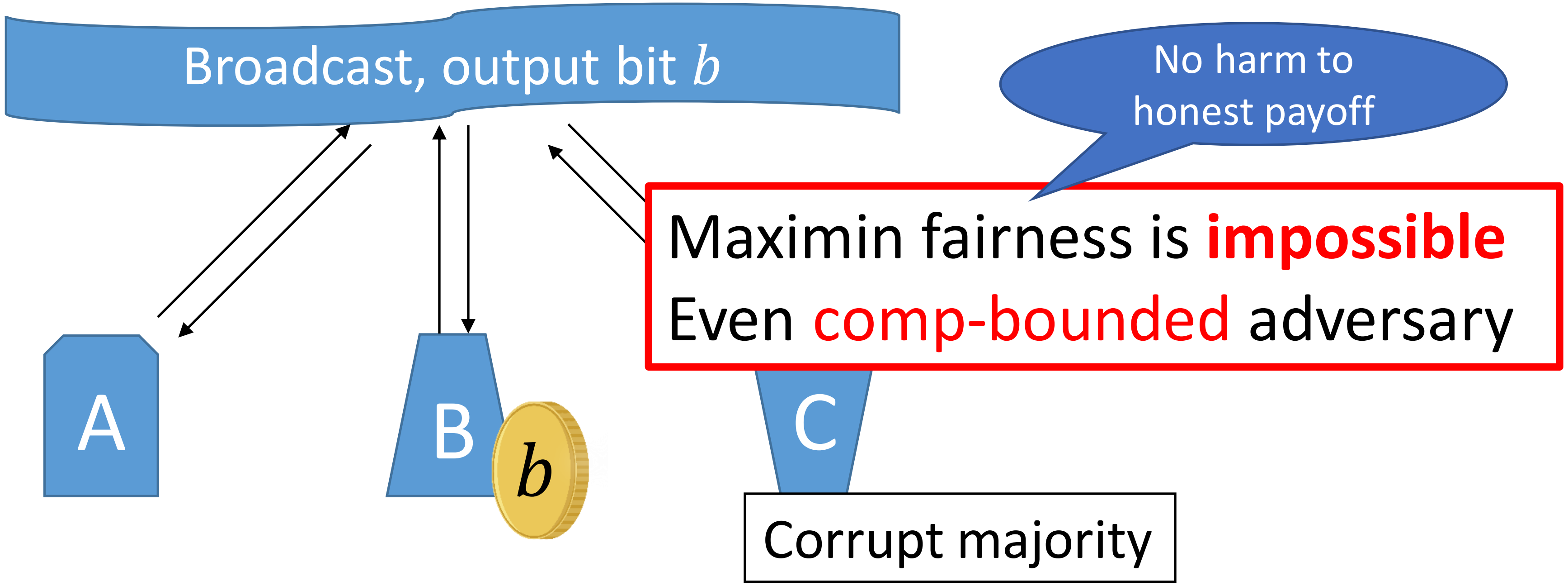
Q: Weak fairness?

Yes:

1. B sample bit b, sends b to A, C
2. A, C output b if received, output 1 if not received; B output b

Maximin Fairness of 3-Party, **Malicious?**

abort early & tamper random tape



		A	B	C
Public Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

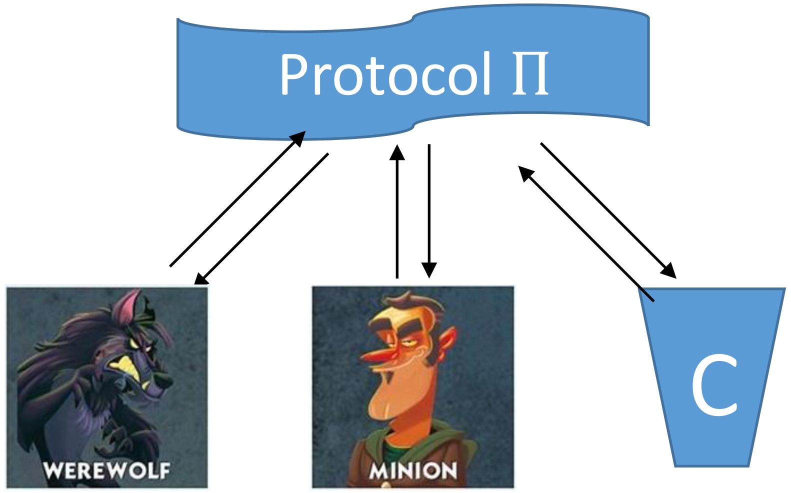
Proof of Impossibility

Impossible even comp-bounded adversary

No harm to honest payoff

Proof roadmap:

1. [Lone-wolf] Single corrupt A (or C)
2. [Lone-minion] Single corrupt B
3. [Wolf-minion] Corrupt A+B (or C+B)

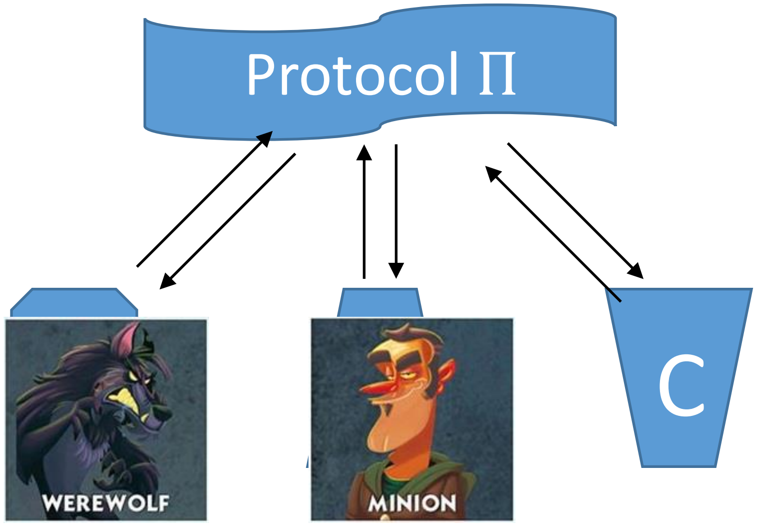


Public Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

Proof of Impossibility

Impossible even comp-bounded adversary

No harm to honest payoff



Public Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

Proof roadmap:

1. [Lone-wolf] Single corrupt A (or C)
2. [Lone-minion] Single corrupt B
3. [Wolf-minion] Corrupt A+B (or C+B)

Cleve's Attackers

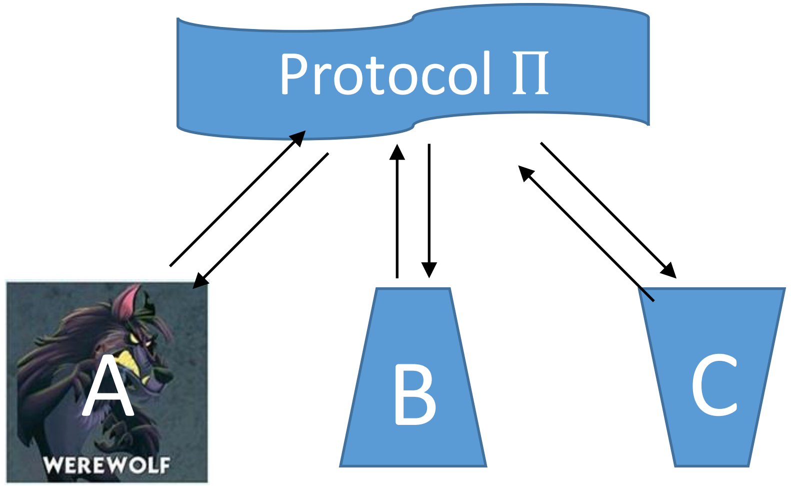
Lone-Wolf Condition

Claim:
Single-corrupt lone-wolf A (or C) cannot make any bias

$E[b] = 0.5$

Proof.
By fairness, cannot harm honest B and C.

No harm to honest payoff



		A	B	C
Public Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

Lone-Minion Condition

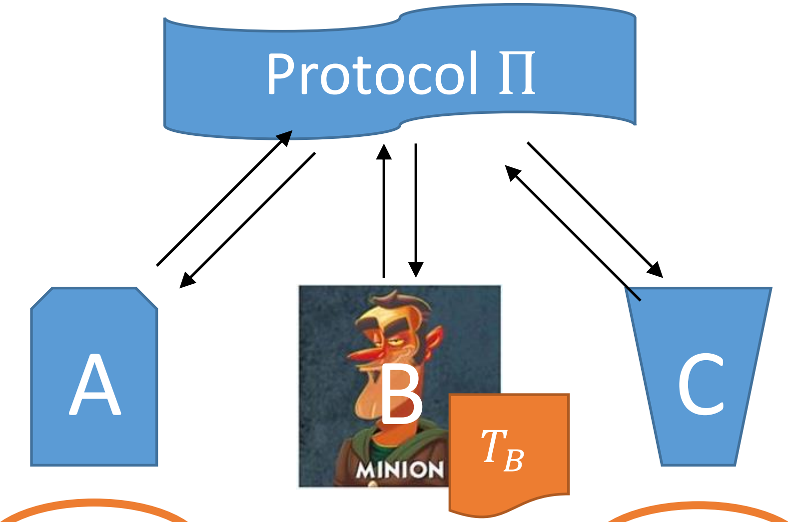
Claim:
Almost all random tapes T_B of B are equal

$E[b \mid T_B] = 0.5$

Proof.

- If not, then some T_B bias toward 1 by fairness
- But, average over all T_B is 0.5
- Then, exists some T_B bias toward 0 not fair to A and C

No harm to honest payoff

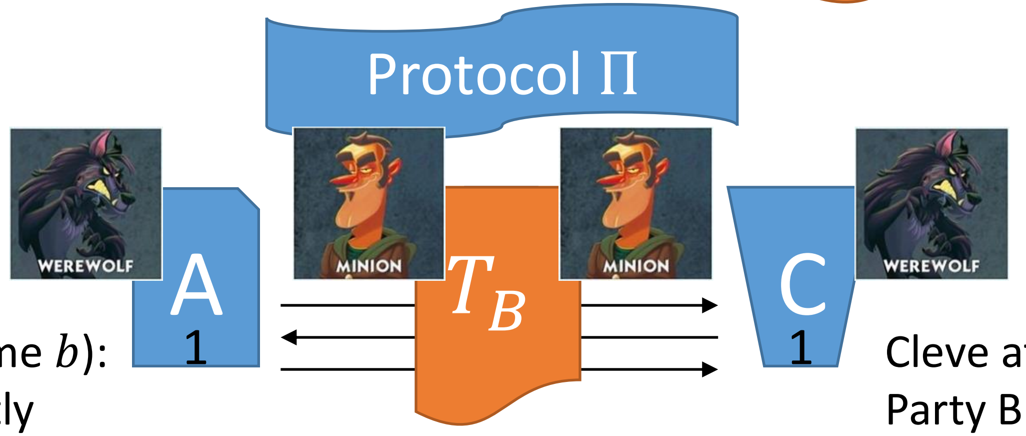


		A	B	C
Public Preference		1	0	1
Payoff	$b = 0$	0	1	0
	$b = 1$	1	0	1

Fixed = Public

Cleve Attackers, Fixed Equal T_B

$4R$ attackers
R: # of rounds



Cleve attacker $\mathcal{A}_i^b$ (round i, outcome b):
Party B: always follow Π, T_B honestly
Party A:

1. Follow Π until round i
2. Given transcript τ_i, Π-outcome α_i
3. $\alpha_i = b$, abort after i-th msg;
   $\alpha_i \neq b$, abort (no i-th msg)

Cleve attacker $\mathcal{C}_i^b$ (round i, outcome b):
Party B: always follow Π, T_B honestly
Party C:

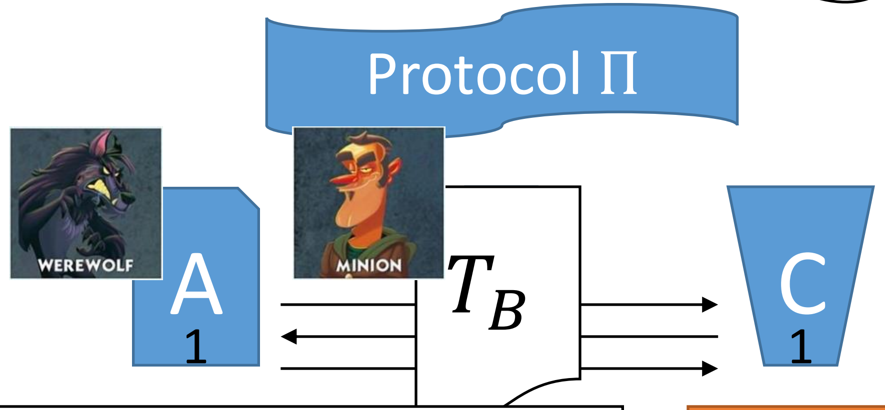
1. Follow Π until round i
2. Given transcript τ_i, Π-outcome β_i
3. $\beta_i = b$, abort after i-th msg;
   $\beta_i \neq b$, abort (no i-th msg)

[Cleve'86]:
Average bias of attackers $\left(\mathcal{A}_i^b, \mathcal{C}_i^b\right)$ is $\Omega\left(\frac{1}{4R}\right)$

Fixed = Public

Cleve Attackers, Fixed Good T_B

$4R$ attackers
R: # of rounds



[Cleve'86]:
Average bias of attackers $\left(\mathcal{A}_i^b, \mathcal{C}_i^b\right)$ is $\Omega\left(\frac{1}{4R}\right)$

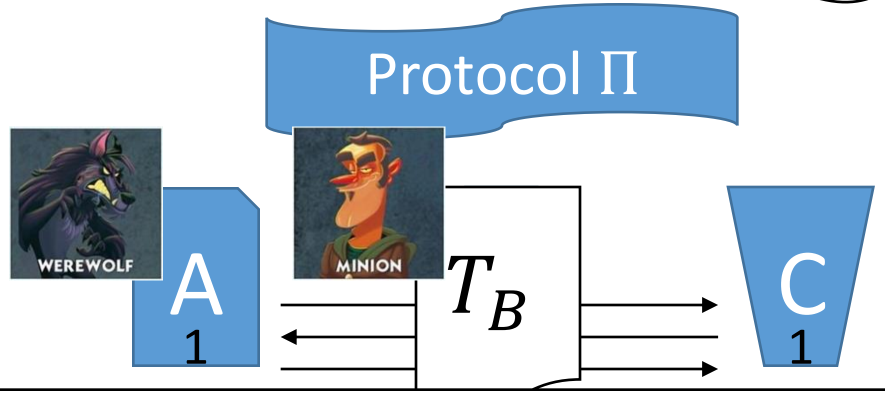
Maximin fair (no harm to 1)
$\Rightarrow$ Exist $\mathcal{A}dv_{T_B} \in \left(\mathcal{A}_i^1, \mathcal{C}_i^1\right)$ toward 1

Almost all T_B

Let such T_B be Good

Cleve Attackers, Uniform Rand T_B

$4R$ attackers
R: # of rounds



Weak fair (no harm to 1) ⇒ For each Good T_B, Exist $\mathcal{Adv}_{T_B} \in (\mathcal{A}_i^1, \mathcal{C}_i^1)$ toward 1

Almost all

"Benign"

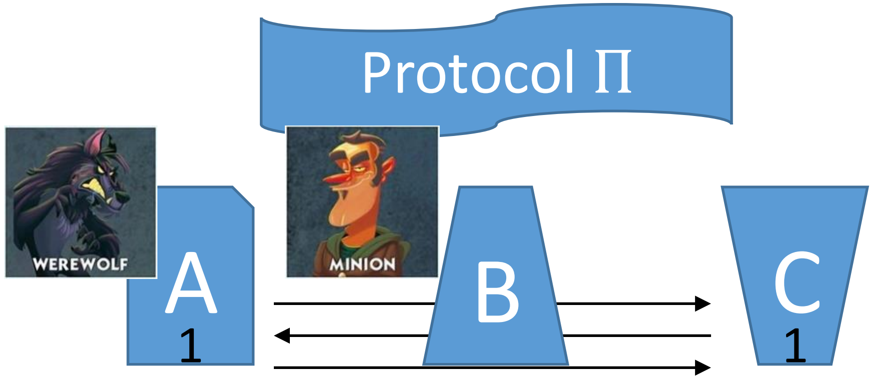
$\mathcal{Adv}$ (some round i):
Party B: always follow Π Unif. Rand. T_B
Party A:

1. Follow Π until round i
2. Given transcript τ_i, Π-outcome α_i
3. $\alpha_i = 1$, abort after i-th msg;
   $\alpha_i \neq 1$, abort (no i-th msg)

Averaging over all T_B
⇒ Exist $\mathcal{Adv}$ toward 1

"Benign"

Wolf-Minion Attackers



"Benign" $\mathcal{Adv}$ toward 1

$\mathcal{Adv}$ (some round i):
Party B: always follow Π, Unif. Rand. T_B
Party A:
1. Follow Π until round i
2. Given transcript τ_i, Π-outcome α_i
3. $\alpha_i = 1$, abort after i-th msg;
   $\alpha_i \neq 1$, abort (no i-th msg)

$\overline{\mathcal{Adv}}$ (some round i):
Party B: always follow Π, Unif. Rand. T_B
Party A:
1. Follow Π until round i
2. Given transcript τ_i, Π-outcome α_i
3. $\alpha_i = 1$, abort (no i-th msg)
   $\alpha_i \neq 1$, abort after i-th msg

Expected outcome:
$E[\mathcal{Adv}] + E[\overline{\mathcal{Adv}}]$
$= 0.5 + 0.5$ (by lone-wolf condition)

$\Rightarrow \overline{\mathcal{Adv}}$ toward 0

Π is not <u>maximin fair</u>

No harm to honest payoff

Summary of Maximin Fairness, $n \geq 3$

<table>
<tr><th></th><th>Fail-Stop</th><th>Malicious</th></tr>
<tr><td>Unanimous
Preference (1, 1, 1, …)</td><td colspan="2">Yes</td></tr>
<tr><td>Almost Unanimous
Preference (0, 1, 1, …)</td><td>Yes</td><td>Impossible
reduce to 3-party</td></tr>
<tr><td>Other
Preference (0, 0, 1, …)</td><td colspan="2">Impossible
reduce to 2-party [Cleve’86]</td></tr>
</table>

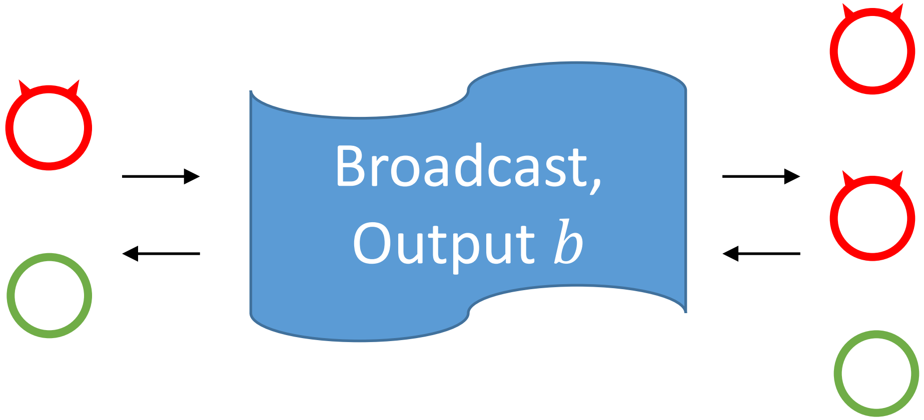
Strong-Nash-Equilibrium (SNE) Fairness

Public-identifiable abort

Maximin:
No harm to **honest payoff**

SNE:
No adversary **increases every corrupt** expected payoff significantly

No incentive to deviate



Public

Preference		1	0
Payoff	$b = 0$	0	1
	$b = 1$	1	0

Equivalent in Blums' 2-party

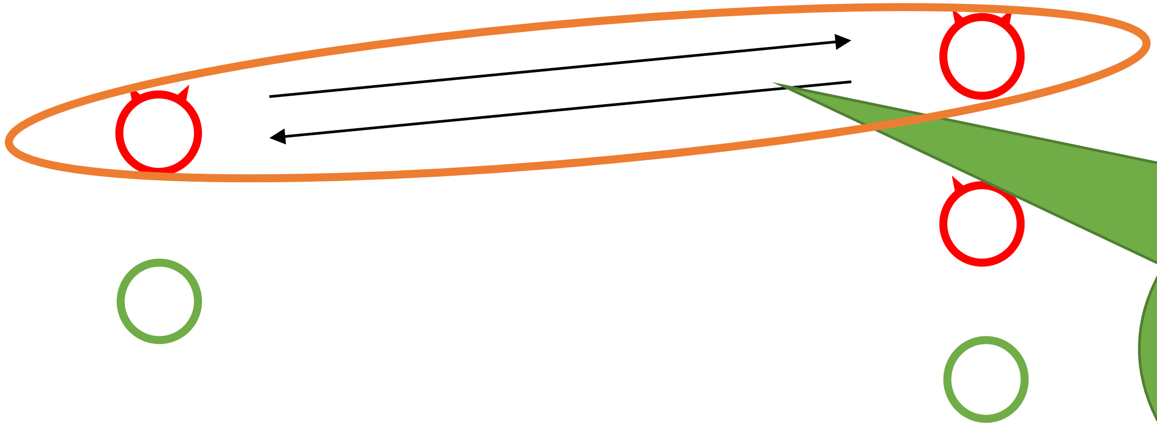
Feasibility of SNE Fairness

No adversary **increases every corrupt** expected payoff significantly

Public-identifiable abort

Commit b_A, send b_B,
Open b_A, XOR (b_A, b_B)

No incentive to deviate



Pick any two opposites,
Run Blum’s 2-party

Public

Preference		1	0
Payoff	$b = 0$	0	1
	$b = 1$	1	0

Fairness Notions of Coin Toss

Maximin

Impossible (except for simple cases)

Group Maximin

Total loss/gain of honest/corrupt

Coalition-Strategy-Proof (CSP)

Strong Nash Equilibrium (SNE)

Fair protocol against malicious adv.

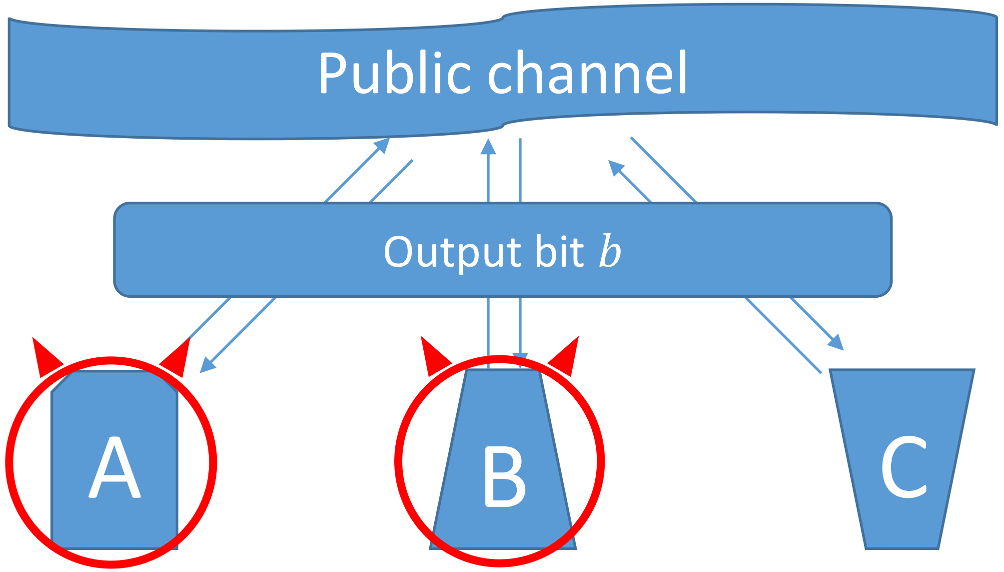
All are equivalent in 2-party (Blum)

More Settings/Problems

- More game-theoretic notions (e.g. self-enforcing)
- Private preference, non-public abort, adaptive adversary
- Gap between upper & lower bounds
- Payoff functions (e.g. zero-sum)
- Other functionalities:
  - Finite random variable
  - Functions imply coin toss
  - …
- Composition of functionalities

Thank you!

Private Preference



Harder to achieve fairness

Impossibility follows

		A	B	C
Preference		1	0	[redacted]
Payoff	$b = 0$	0	1	[redacted]
	$b = 1$	1	0	[redacted]