Game Theoretic Notions of Fairness in Multi-Party Coin Toss*

Kai-Min Chung[2], Yue Guo[1], Wei-Kai Lin[1], Rafael Pass[1], and Elaine Shi[1]

[1] Cornell/CornellTech, {yueguo,wklin,rafael}@cs.cornell.edu, runting@gmail.com
[2] Academia Sinica, kmchung@iis.sinica.edu.tw

Two-party coin toss protocol

- Definition.
- Fairness notions: strong and weak.
- Strong fairness ⇒ Weak fairness but not the reverse.
- Feasibility results:
  - Assuming the existence of one-way functions, a weakly fair 2-party protocol can be constructed for malicious, comp. bounded adversaries. [Blum. *Coin flipping by telephone*.]
  - There is no 2-party protocol guaranteeing strong fairness even for computationally bounded, fail-stop adversaries. [Cleve. *Limits on the security of coin flips when half the processors are faulty*.]

Results about Multiparty coin toss

- Strong fairness can be achieved in a multiparty coin toss protocol assuming honest majority and existence of one-way functions even against malicious computationally bounded adversaries. [Goldreich et al.]
- For corrupt majority, Cleve's result extends to multiple parties, i.e., strong fairness cannot be achieved even for fail-stop and computationally bounded adversaries.

- But these are only about strong fairness!!

What about weak fairness for multiparty coin toss?

- Can we achieve Blum’s weak fairness notion in multi-party coin toss protocols?
  - Note that strong fairness has been extensively studied for multiparty coin toss protocols.
- Can we overcome Cleve’s impossibility for a corrupt majority multi-party coin toss with weak fairness?
- How do we even define weak fairness in multi-party coin toss protocols?

Focus of the paper

Explore different fairness notions in case of multiparty protocols, particularly for the case of corrupt majority. (For honest majority, all these are achievable because strong fairness is achievable.)

Contents

- Preliminaries: Execution model and Corruption models (fail-stop, malicious).
- Formal definition of a multiparty coin toss protocol.
  - Defining strong fairness (for the sake of completeness).
- Maximin fairness
- Cooperative Strategy Proof (CSP) fairness
- Strong Nash Equilibrium (SNE) fairness
- The case of private preference profiles
- Conclusion

Contents

- Preliminaries: Execution model and Corruption models (fail-stop, malicious).
- Formal definition of a multiparty coin toss protocol.
  - Defining strong fairness (for the sake of completeness).
- Maximin fairness
- Cooperative Strategy Proof (CSP) fairness
- Strong Nash Equilibrium (SNE) fairness
- The case of private preference profiles
- Conclusion

Execution Model

- Parties modelled by ITMs
- All the corrupt parties are controlled by an adversary $\mathcal{A}$
- Synchronous broadcast medium. *(Messages sent by honest parties in round r will be delivered to all honest parties at the beginning of round r + 1).*
- Identifiable abort. *(If a party i aborts the protocol in round r without sending any message, then all honest parties can detect such abort by detecting the absence of i s message at the beginning of round r + 1.)*

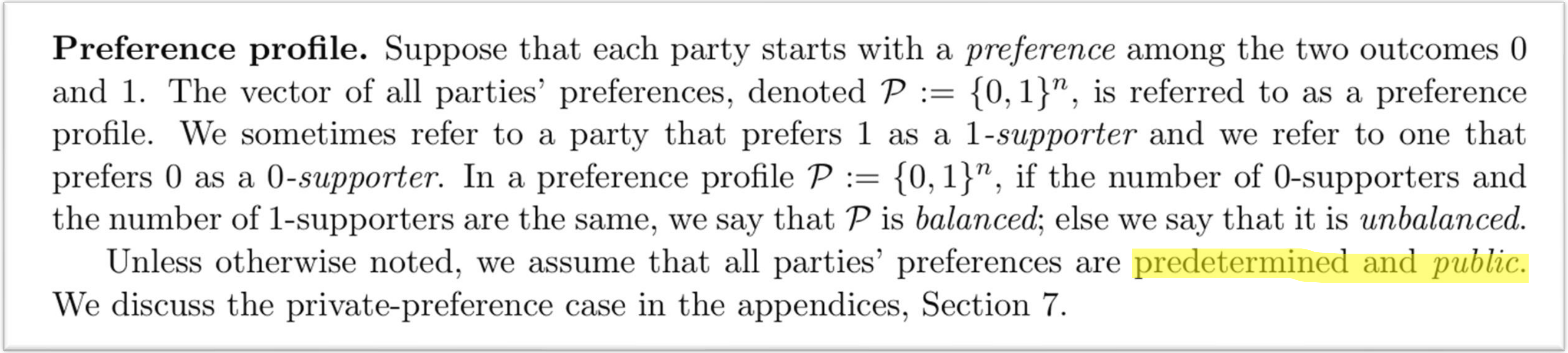
Corruption Model

- For any fixed adversary algorithm $\boldsymbol{\mathcal{A}}$, the set of parties it wants to corrupt is deterministically encoded in the description of $\boldsymbol{\mathcal{A}}$ *(i.e., for any fixed adversary $\boldsymbol{\mathcal{A}}$, there is no randomness in the choice of the corrupt coalition).*
- $\boldsymbol{\mathcal{A}}$ can be fail-stop or malicious:
  - *Fail-stop:* Corrupt nodes always follow the honest protocol but may abort in the middle of the protocol. The decision to abort (or not) can depend on the corrupt parties' view in the protocol so far.
  - *Malicious:* The adversary can make corrupt parties deviate arbitrarily from the prescribed protocol, including sending arbitrary messages, choosing randomness arbitrarily, and aborting prematurely.

Contents

- Preliminaries: Execution model and Corruption models (fail-stop, malicious).
- Formal definition of a multiparty coin toss protocol.
  - Defining strong fairness (for the sake of completeness).
- Maximin fairness
- Cooperative Strategy Proof (CSP) fairness
- Strong Nash Equilibrium (SNE) fairness
- The case of private preference profiles
- Conclusion

Preference profile (Def.)



Preference profile. Suppose that each party starts with a *preference* among the two outcomes 0 and 1. The vector of all parties' preferences, denoted $\mathcal{P} := \{0,1\}^n$, is referred to as a preference profile. We sometimes refer to a party that prefers 1 as a 1-*supporter* and we refer to one that prefers 0 as a 0-*supporter*. In a preference profile $\mathcal{P} := \{0,1\}^n$, if the number of 0-supporters and the number of 1-supporters are the same, we say that $\mathcal{P}$ is *balanced*; else we say that it is *unbalanced*.

Unless otherwise noted, we assume that all parties' preferences are predetermined and *public*. We discuss the private-preference case in the appendices, Section 7.

Coin toss protocol (Def.)

- A protocol Π with n parties (each with a preference of some bit) is said to be a coin toss protocol if there is a polynomial-time computable deterministic function, which, given the transcript of the protocol execution, outputs a bit $b \in \{0,1\}$, often said to be the *outcome* of the protocol.
- **Correctness**:
  - If some parties have differing preferences, in an all-honest execution (when all parties are honest), the probability that the outcome is 0 (or 1) is exactly 1/2.
  - If all parties happen to prefer the same bit $b \in \{0,1\}$, the honest execution should output the preferred bit b with probability 1.

- **Note: Payoff function.** If the protocol's outcome is b, a party who prefers b receives a reward (or payoff) of 1; else, it receives a reward (or payoff) of 0.

Preference profile (Characterization)

Types of preference profiles:

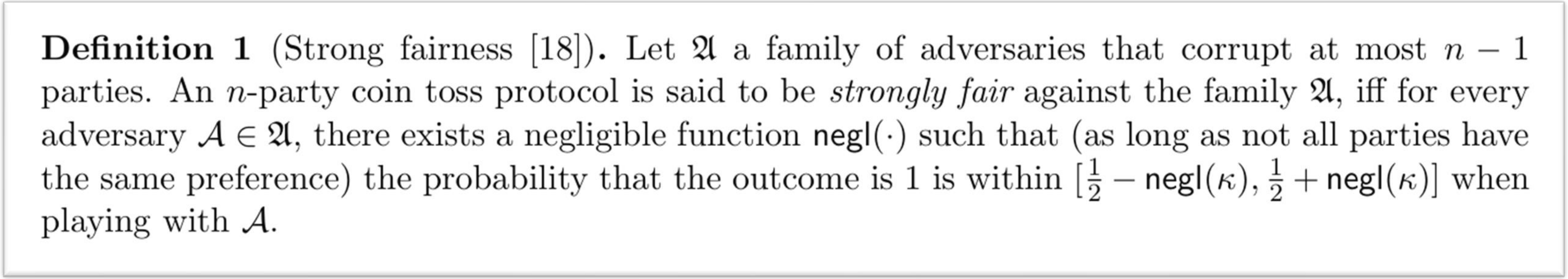
- *Unanimous*: Every party prefers the same bit *b.*
- *Almost unanimous*: All parties *but one* prefer the same bit *b.*
- *Amply divided*: ≥2 parties prefer the bit *0,* and ≥2 parties prefer the bit *1*.

- **Observation**. The above types are disjoint and exhaustive.

Trivial case: Unanimous preference profile

- In this case, it is not required that an honest execution produce an unbiased coin, since it makes sense for the outcome to be the bit that is globally preferred.
- In the remainder of the paper, for the case of public preference, if everyone prefers the same bit $b \in \{0,1\}$, it is assumed that the protocol simply fixes the outcome to be the universally preferred bit b regardless of how parties act.
- Everyone obtains a payoff of 1, and no deviation from the protocol can influence the outcome.
- Therefore, all game-theoretic fairness notions considered now are trivially satisfied when the preference profile is unanimous. So, this case is ignored for the rest of the presentation.

Strong fairness



Definition 1 (Strong fairness [18]). Let $\mathfrak{A}$ a family of adversaries that corrupt at most $n-1$ parties. An n-party coin toss protocol is said to be *strongly fair* against the family $\mathfrak{A}$, iff for every adversary $\mathcal{A} \in \mathfrak{A}$, there exists a negligible function $\mathsf{negl}(\cdot)$ such that (as long as not all parties have the same preference) the probability that the outcome is 1 is within $[\frac{1}{2} - \mathsf{negl}(\kappa), \frac{1}{2} + \mathsf{negl}(\kappa)]$ when playing with $\mathcal{A}$.

Adversarial power and strength of fairness

- Adversarial power*:*

Fail-stop < Computationally bounded malicious

< Computationally unbounded malicious

- Strength of fairness:

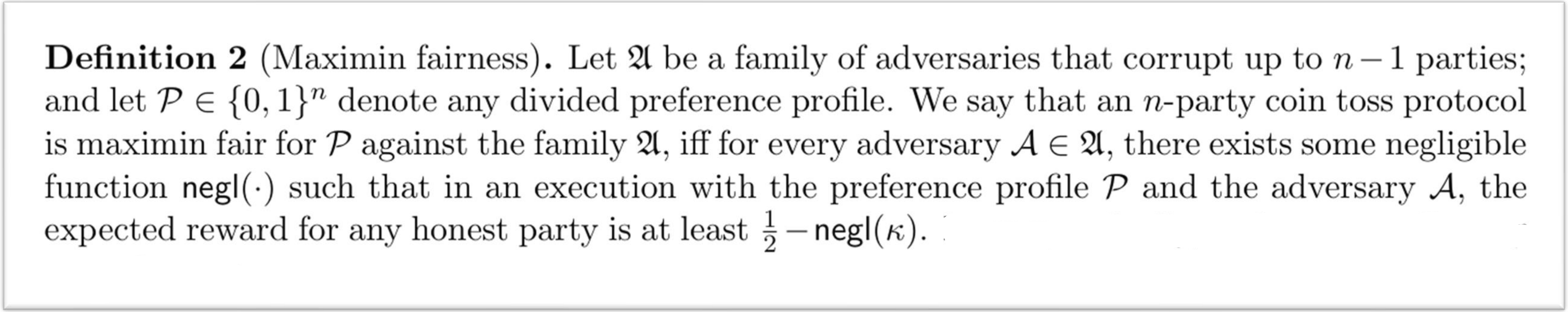
Computational fairness < *Statistical fairness* < *Perfect fairness*

(fair against p.p.t. *adversaries) (against unbounded adversaries) (perfectly fair)*

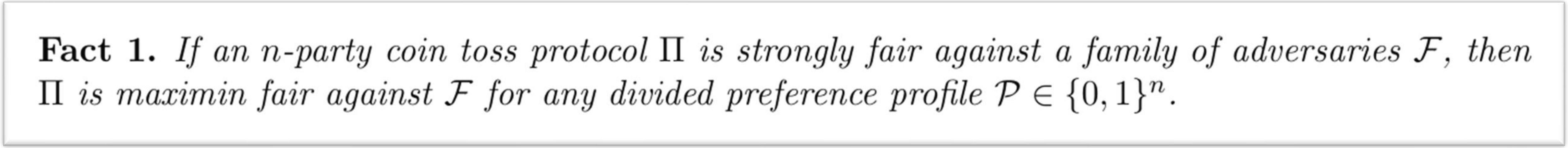
Contents

- Preliminaries: Execution model and Corruption models (fail-stop, malicious).
- Formal definition of a multiparty coin toss protocol.
  - Defining strong fairness (for the sake of completeness).
- Maximin fairness
- Cooperative Strategy Proof (CSP) fairness
- Strong Nash Equilibrium (SNE) fairness
- The case of private preference profiles
- Conclusion

Maximin Fairness (Def.)

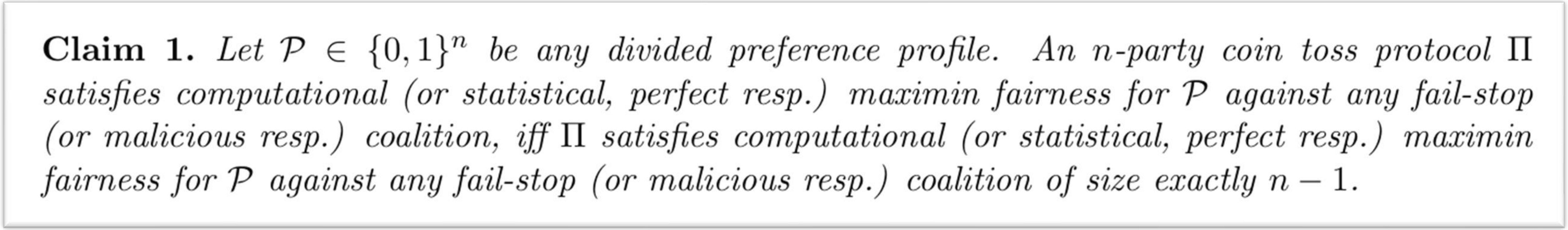


Definition 2 (Maximin fairness)**.** Let $\mathfrak{A}$ be a family of adversaries that corrupt up to $n-1$ parties; and let $\mathcal{P} \in \{0,1\}^n$ denote any divided preference profile. We say that an n-party coin toss protocol is maximin fair for $\mathcal{P}$ against the family $\mathfrak{A}$, iff for every adversary $\mathcal{A} \in \mathfrak{A}$, there exists some negligible function $\mathsf{negl}(\cdot)$ such that in an execution with the preference profile $\mathcal{P}$ and the adversary $\mathcal{A}$, the expected reward for any honest party is at least $\frac{1}{2} - \mathsf{negl}(\kappa)$.



Fact 1. *If an n-party coin toss protocol Π is strongly fair against a family of adversaries $\mathcal{F}$, then Π is maximin fair against $\mathcal{F}$ for any divided preference profile $\mathcal{P} \in \{0,1\}^n$.*

Maximin Fairness (at most = exactly)



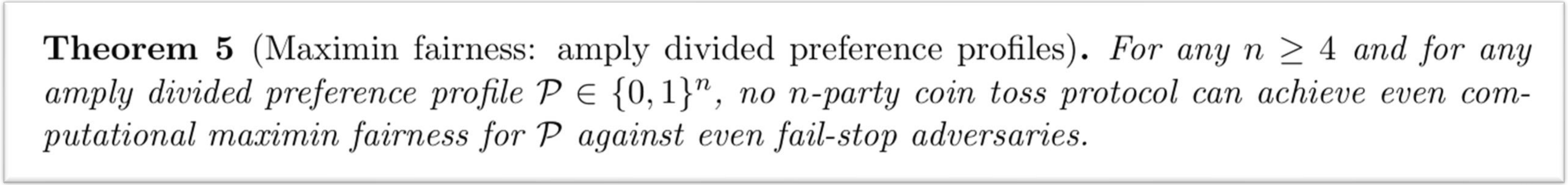
Claim 1. *Let $\mathcal{P} \in \{0,1\}^n$ be any divided preference profile. An n-party coin toss protocol Π satisfies computational (or statistical, perfect resp.) maximin fairness for $\mathcal{P}$ against any fail-stop (or malicious resp.) coalition, iff Π satisfies computational (or statistical, perfect resp.) maximin fairness for $\mathcal{P}$ against any fail-stop (or malicious resp.) coalition of size exactly $n-1$.*

Proof. Consider an adversary that controls a coalition C of size at most $n-1$. So there is at least one honest party P_i. View this execution as a coalition C' that consists of all the parties except P_i. The parties in C' but not in C are just assumed to be adversaries following the honest protocol.

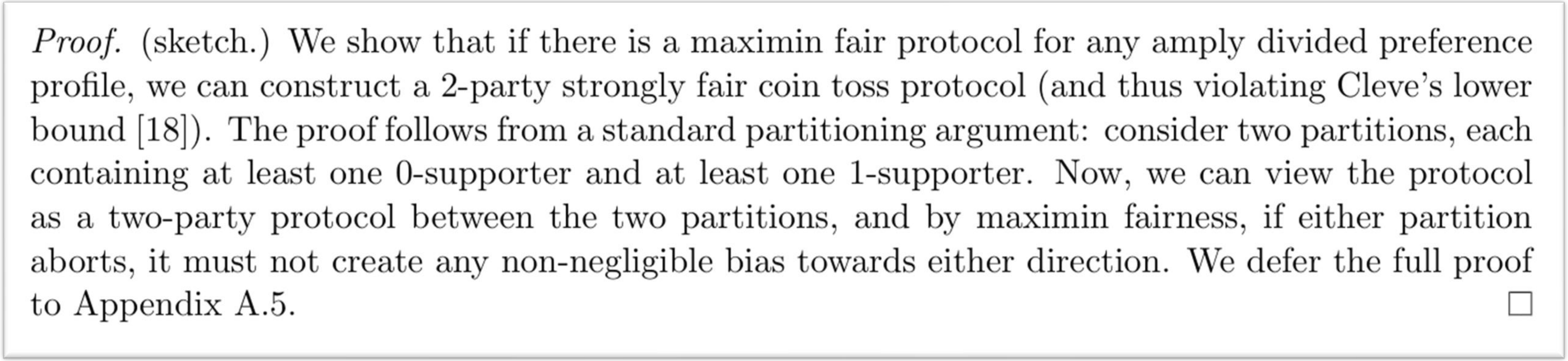
Maximin Fairness (game theoretic interpretation)

- If a coin-toss protocol is maximin fair, then the following hold:
  - The honest strategy maximizes a player’s worst-case expected payoff (even when everyone else is colluding against the player); this explains the name maximin fairness.
  - When playing the honest strategy, a player’s worst-case payoff is what it would have gained in an all-honest execution – note that a player’s worst-case (expected) payoff obviously cannot be more than its payoff in an all-honest execution.

Maximin Fairness (Amply divided preference profiles)



Theorem 5 (Maximin fairness: amply divided preference profiles). *For any $n \geq 4$ and for any amply divided preference profile $\mathcal{P} \in \{0,1\}^n$, no n-party coin toss protocol can achieve even computational maximin fairness for $\mathcal{P}$ against even fail-stop adversaries.*



Proof. (sketch.) We show that if there is a maximin fair protocol for any amply divided preference profile, we can construct a 2-party strongly fair coin toss protocol (and thus violating Cleve's lower bound [18]). The proof follows from a standard partitioning argument: consider two partitions, each containing at least one 0-supporter and at least one 1-supporter. Now, we can view the protocol as a two-party protocol between the two partitions, and by maximin fairness, if either partition aborts, it must not create any non-negligible bias towards either direction. We defer the full proof to Appendix A.5. □

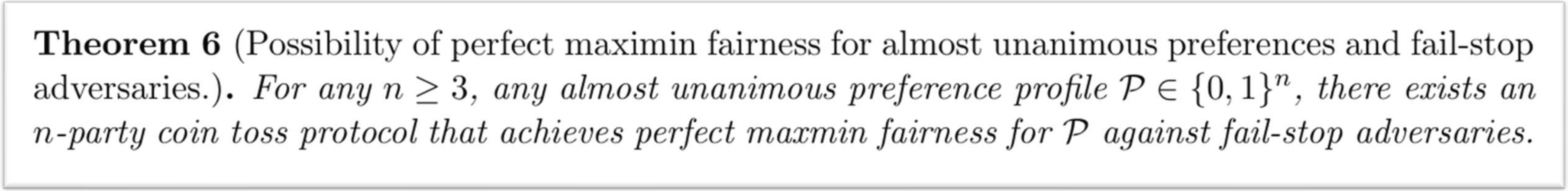
Maximin Fairness (Characterization)

- Amply divided preference profile:

	Computational fairness	*Statistical fairness*	*Perfect fairness*
Fail-stop adversary	No	No	No
Comp. bounded malicious adversary	No	No	No
Comp. unbounded malicious adversary	No	No	No

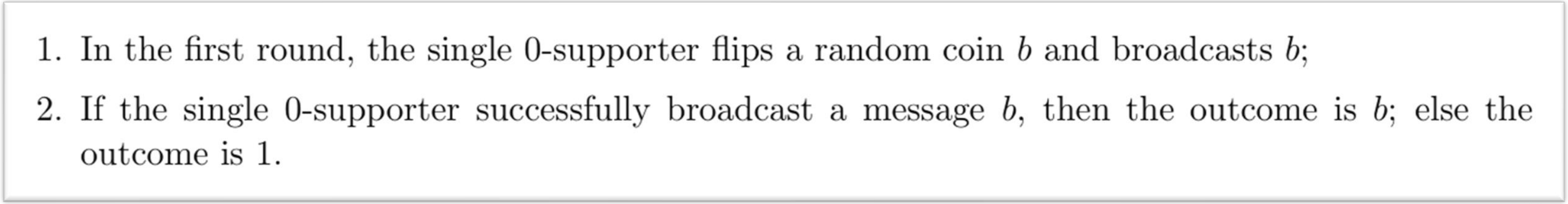
Maximin Fairness (Almost unanimous preference profiles)

- Fail-stop adversaries:



Theorem 6 (Possibility of perfect maximin fairness for almost unanimous preferences and fail-stop adversaries.). *For any $n \geq 3$, any almost unanimous preference profile $\mathcal{P} \in \{0,1\}^n$, there exists an n-party coin toss protocol that achieves perfect maxmin fairness for $\mathcal{P}$ against fail-stop adversaries.*

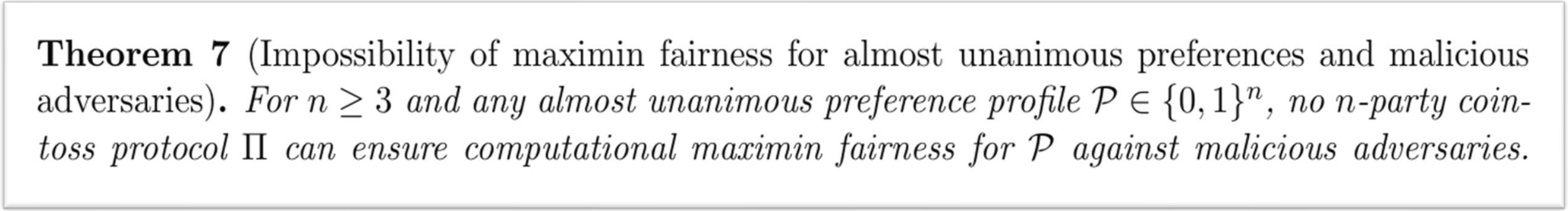
- The protocol:



1. In the first round, the single 0-supporter flips a random coin b and broadcasts b;
2. If the single 0-supporter successfully broadcast a message b, then the outcome is b; else the outcome is 1.

Maximin Fairness (Almost unanimous preference profiles)

- Malicious adversaries:



Theorem 7 (Impossibility of maximin fairness for almost unanimous preferences and malicious adversaries)**.** *For $n \geq 3$ and any almost unanimous preference profile $\mathcal{P} \in \{0,1\}^n$, no n-party coin-toss protocol Π can ensure computational maximin fairness for $\mathcal{P}$ against malicious adversaries.*

- So, to sum it up, for the Maximin fairness notion for an almost unanimous preference profile, even perfect fairness is possible against fail-stop adversaries, but not even computational fairness is possible against even computationally bound malicious adversaries.

Maximin Fairness (Characterization)

- Almost unanimous preference profile:

	Computational fairness	*Statistical fairness*	*Perfect fairness*
Fail-stop adversary	Yes	Yes	Yes
Comp. bounded malicious adversary	No	No	No
Comp. unbounded malicious adversary	No	No	No

Contents

- Preliminaries: Execution model and Corruption models (fail-stop, malicious).
- Formal definition of a multiparty coin toss protocol.
  - Defining strong fairness (for the sake of completeness).
- Maximin fairness
- Cooperative Strategy Proof (CSP) fairness
- Strong Nash Equilibrium (SNE) fairness
- The case of private preference profiles
- Conclusion