Rationality in the Full-Information Model

Ronen Gradwohl*

Abstract

We study rationality in protocol design for the full-information model, a model characterized by computationally unbounded adversaries, no private communication, and no simultaneity within rounds. Assuming that players derive some utility from the outcomes of an interaction, we wish to design protocols that are faithful: following the protocol should be an optimal strategy for every player, for various definitions of "optimal" and under various assumptions about the behavior of others and the presence, size, and incentives of coalitions. We first focus on leader election for players who only care about whether or not they are elected. We seek protocols that are both faithful and resilient, and for some notions of faithfulness we provide protocols, whereas for others we prove impossibility results. We then proceed to random sampling, in which the aim is for the players to jointly sample from a set of m items with a distribution that is a function of players' preferences over them. We construct protocols for $m \geq 3$ that are faithful and resilient when players are single-minded. We also show that there are no such protocols for 2 items or for complex preferences.

Keywords: rational cryptography, leader election, random sampling, incentive compatibility.

*Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. E-mail: r-gradwohl@kellogg.northwestern.edu. Much of this research was done while the author was a graduate student at the Weizmann Institute of Science and supported by US-Israel Binational Science Foundation Grant 2002246 and ISF Grant 334/08.

1 Introduction

The full-information model of Ben-Or and Linial [8] is one of the classically-studied settings for protocol design. In this model there are no computational limits on the adversary, there is no private communication, and there is no guarantee of simultaneity within rounds of a protocol. Three famous problems are collective coin-flipping, leader election, and random sampling. In the first, players jointly flip a coin; in the second, they jointly select a random player; and in the third, they jointly select a random element from some universe of m items. In general, the goal is to design protocols that are resilient: the outcome should be random even in the presence of an adversary who corrupts and coordinates the behavior of a fraction of the players.

In this paper we explore the role of preferences in the design of such protocols. While preferences are not explicitly considered in the well-studied formulations of the problems, they are implicitly present. For example, leader election has a fairness criterion, which requires each player to be elected with roughly equal probability (presumably because everybody wants to be the leader). A leader election protocol is resilient if an adversary can not force the elected leader to be a member of his coalition (or at least will fail to do so with constant probability). Again, the adversary *wants* a coalition-member to be elected. For collective coin-flipping and random sampling, resilience is also measured as a bound on the probability that an adversary succeeds at something. It is implicitly assumed that the adversary wants to do this, and that the honest (non-adversarial) players do not wish him to achieve his goal.

The study of preferences in the design of protocols is primarily the domain of mechanism design. In mechanism design a planner wishes to implement some function of players' private information. His goal is to design a mechanism and provide incentives for the players so that their optimal strategy is to truthfully reveal their private information, and more generally to adhere to the mechanism. The optimality of players' strategies is measured via some solution concept: following the mechanism should be in some equilibrium, most commonly Nash, ex post Nash, or dominant strategy. In this paper we take a similar approach – we define new solution concepts appropriate for the full-information model, and seek protocols that are *faithful*: following them is optimal for players with respect to these solution concepts (in addition to the usual resilience guarantees).

For any problem of protocol design, making the structure of preferences explicit has two potential benefits, both of which we achieve in this paper. First, it can result in better protocols – protocols are arguably of little use if players have no incentives to follow them. If one can obtain faithful protocols without harming the original guarantees of the protocol, then one has only gained. Second, it may be possible to sidestep some impossibility results of the original problem, since often these impossibility results are based on arbitrary play by the adversary. If players do not play arbitrarily but rather obey some preference structure, then many of these results no longer hold.

The model In the full-information model all communication is by broadcast. In each round, some of the players send a message, which may depend on messages sent in previous rounds. The main difficulty is that adversarial players are allowed to "rush" – to wait until all messages have been sent within a round, and only then to send their own messages.

This paper We are largely motivated by recent work in rational cryptography, in which the aim is to design cryptographic protocols that participants *want* to follow. Two of the main difficulties

encountered when attempting a game theoretic analysis of cryptographic protocols are computational limits and potentially adversarial timing. In this paper we focus solely on the latter issue by considering a model in which (adversarial) players may be computationally unbounded, and the guaranteed security (i.e. resilience) is information-theoretic. We highlight the various challenges and subtleties caused by a combination of rational and adversarial players, particularly in the presence of adversarial timing. We also draw a possibility-impossibility border for various problems and requirements in this setting. Finally, we believe that this paper is an illuminating stepping-stone towards a game theoretic analysis of more general cryptographic protocols.

1.1 Our Results and Organization

Definitions (Section 2 and Appendix A) The initial difficulty encountered when considering preferences in the full-information model is to precisely formulate a notion of equilibrium. The first notion to consider is Nash equilibrium (NE), in which each player's strategy must be expected utility maximizing assuming others also follow their Nash strategies. If the protocol is such that only one player sends a message in each round, then this suffices. One such protocol is Baton Passing [31], a protocol that is resilient and in fact satisfies our weaker solution concept[1]. However, state-of-the-art protocols are often round-efficient, and allow multiple players to broadcast within a round. Because of the lack of synchrony within rounds, however, NE does not suffice. In the Lightest Bin protocol [14], for example, a player may increase his chance of winning from $1/n$ (where n is the number of players) to a constant by deviating. To deal with asynchrony, we will require that for *any ordering* of the players within each round, the protocol is in a NE. In Section 2 we formalize this and other notions of what it means to be faithful and faithful in the presence of adversarial players. An alternative but equivalent formulation of these definitions appears in Appendix A.

Impossibility with complex preferences (Section 3) In Section 3 we encounter our first impossibility result. Theorem 3.1 states that no random selection protocol can satisfy even our weakest solution concept if players have a full preference order over the outcomes of the protocol. One implication of this impossibility result is that collective coin-flipping is impossible with players who have some preference about the outcome. For leader election and random sampling, this result forces us to concentrate on more restricted preferences for players. For the former, we assume that players care only about whether or not they are elected, and are indifferent otherwise. For the latter, we assume players are single-minded: each prefers one of the items, and is indifferent about the others.

Faithfulness with resilience (Section 4.2) The standard aim of selection protocols in the full-information model is resilience: if an adversary corrupts and coordinates the actions of a fraction of the players, he still fails to force his desired outcome with non-negligible probability. In Section 4.2 we construct optimal protocols that both satisfy a notion of equilibrium and are resilient. Players wish to faithfully adhere to the protocol if the others also do, and there is a resilience guarantee in the presence of an adversary.

[1] More specifically, it is in a full-information ex post NE – see Definition 2.6. It is not, however, in a full-information dominant strategy equilibrium (Definition 2.3).

Faithfulness in the face of an adversary (Section 4.3) In Section 4.3 we consider the problem of constructing leader election protocols that are in equilibrium even when not all others follow the protocol. We show that it is impossible to construct such protocols in the presence of a malicious adversary, even if the adversary has his own objective of maximizing the probability that a coalition-member wins. However, if the adversary maximizes this probability, but also only deviates from the protocol if he strictly gains from doing so (i.e. if deviating is costly), then we do design a resilient protocol.

Resilience to rational coalitions (Section 4.4) A different form of resilience against adversarial play is when there is no controlling adversary, but instead players may form "rational coalitions" to benefit all members. In Section 4.4 we give an impossibility result for one notion of a "rational coalition", but for a weaker notion provide a protocol that is resilient against all such coalitions of size at most $n - 2$.[2]

Random sampling (Section 5) Our final set of results concerns random sampling. Each player has some preferences over a universe of m items, and the goal is to design a protocol in which an item is sampled with a probability distribution that is a function of those preferences. We design protocols that are simultaneously in a full-information ex post Nash equilibrium (in which truthful revelation of one's preferences is optimal) and resilient against adversarial coalitions.

1.2 Related Work

This paper draws from three different literatures – protocol design in cryptography and distributed computing, and algorithmic mechanism design. The extensive literature on collective coin-flipping, random sampling, and leader election in the full-information model includes [31, 16, 26, 3, 11, 12, 17, 30, 14, 13, 32, 6]. The paper most closely related to ours is that of Antonakopoulos [6], who also considers 1-round protocols in which individual players have no incentive to deviate. However, his protocols all attain either faithfulness or resilience, but never both. Similarly, Ben-Or and Linial [8] have a 2-round protocol that is faithful but not resilient to larger coalitions. The paper most closely related to ours from the mechanism design literature is that of Altman and Tennenholtz [5], who construct 1-round protocols that are faithful (but also not resilient). Their goal is to attain arbitrary distributions over the players. Also related is the literature on ranking games [10], in which players have preferences about their rankings in some game.

While we believe that we are the first to study notions of rationality tailored specifically for the full-information model, such notions have been studied in other settings for distributed computing. For example, Monderer and Tennenholtz [24] consider an implementation problem in a distributed network. Shneidman and Parkes [33] introduce the idea that protocols should be faithful. Additionally, the field of Distributed Algorithmic Mechanism Design (DAMD) focuses on implementing mechanisms for various problems in a distributed setting. In a general "mission statement" for DAMD, Feigenbaum and Shenker [15] argue that it would be desirable to incorporate various fault models into the DAMD framework. Also, Halpern [20, 21] has expressed the need to incorporate faulty or malicious behavior into distributed settings with rational players. Some papers that address this issue are Aiyer et al. [2], Abraham et al. [1], and Gradwohl [18].

[2]Compare this with the fact that there are *no* protocols that are resilient against an adversary of size $n/2$ [31].

Finally, as mentioned in the introduction, this work is closely related to the growing literature on rational cryptography (see, for example, Katz [23] and the references therein). Many works in this literature study rational behavior in a cryptographic setting, for which the full-information model is a special case. However, due to computational issues, the definitions in the general setting are messier (and often also weaker). We note that the way we model rushing is closely related to an idea of Ong et al. [28], who adopt the methods of Kalai [22] to a protocol design setting. The idea of considering rational coalitions was also explored in this context by Ong et al. [27].

Our notions of stability of coalitions are related to similar notions in the game theory literature, such as the strong Nash equilibrium of Aumann [7] and the coalition-proof equilibrium notions of Bernheim et al. [9], Moreno and Wooders [25], and Abraham et al. [1].

2 Protocols and Games

For any vector $X = (X_1, \ldots, X_n)$ and $S \subset [n]$, we denote by $X_S = \{X_i\}_{i \in S}$ and by $X_{-S} = \{X_i\}_{i \notin S}$.

2.1 Resilient Protocols

We are interested in protocols involving many players and the incentives of players in following these protocols. Thus, we will assume that players have preferences over possible outcomes, as well as other private information. As in the game theory literature, all this information is collectively called a player's type. Player i's type is denoted by t_i, and the vector $t = (t_1, \ldots, t_n)$ is called the type profile. The space of possible types of player i is T_i.

Definition 2.1 (selection protocol) *An n-player* selection protocol $\mathcal{P}$ *is specified by a function f, a natural number q, and, for each of the n players, a set of q randomized functions $\left\{S_i^1, \ldots, S_i^q\right\}_{i \in [n]}$. The protocol proceeds as follows:*

- *At round j, the i'th player broadcasts a random message M_i^j obtained by applying the randomized function S_i^j to all previous messages sent, namely $\{M_k^l : k \in [n], l < j\}$, as well as player i's type t_i. The randomness of the function comes from the player's independent coins.*
- *After q rounds, the players output $f(\{M_k^l : k \in [n], l \in [q]\})$ which is an element of $[m]$ in an m-*item random sampling protocol *and an element of $[n]$ in a* leader election protocol*. If all players follow the protocol then the output is a uniformly random element (unless stated otherwise).*

In any round j, a player i's *legal* messages are those in $\bigcup_{t_i \in T_i} supp(S_i^j(t_i, \{M_k^l : k \in [n], l < j\}))$[3]. We assume that if a player noticeably deviates from the protocol (by broadcasting a message that is not legal), then his message is changed to some default legal value.

Players may not legally base their messages in round j on the messages of other players in round j. However, since we can not guarantee simultaneity within a round, we allow the dishonest players to *rush*: they may base their messages on the messages of other players from the same round (but not from later rounds). A leader election protocol is *ε-resilient* to coalitions of size t if the following holds: If at most t players are playing a coordinated rushing strategy, then the probability that the

[3]Note that a player's legal actions include messages in the support of S_i^j for all types, not just the true one. This is so because the other players do not know i's true type.

elected leader is a cheating player is at most $1-\varepsilon$. Often we will implicitly be referring to a family of protocols, one for each value of n. In this case, we say a protocol is resilient if there exists some $\varepsilon > 0$ such that all protocols in the family with enough players are ε-resilient.

A protocol is *oblivious* if players' messages are based only on their internal coin tosses. A protocol is *explicit* if players' messages and the function f are computable in probabilistic polynomial time (in the number of players and $\log(m)$).

2.2 Extensive-Form Games and Protocols

An n-player *extensive-form game* is specified by a game tree in which every node is owned by a player and outgoing edges are labelled by actions. The game begins at the root node and proceeds down the tree – at every node following the edge labelled by the action played by the node's owner. Payoffs for players are specified at the leaves.

Definition 2.2 (Nash equilibrium (NE)) *A* Nash equilibrium (NE) *in an extensive-form game is a mixed strategy for every player at every node that he owns, such that: if all players play their NE strategy, then no player obtains a higher expected payoff by deviating at any of his nodes.*

We note that in the games we consider, the NE will be completely mixed strategies (i.e. players will play every action with positive probability). Such Nash equilibria are in fact subgame perfect (see [29]).

Consider a selection protocol, where each player i derives some utility $u_i : T_i \times [m] \mapsto \mathbb{R}$ from outputs of the protocol. u_i is such that for $o \neq o' \in [m]$, we have that $u_i(t_i, o) > u_i(t_i, o')$ if and only if player i of type t_i strictly prefers o to o'. Then any protocol in which only one player sends a message in each round can be viewed as an extensive-form game[4]: if after $j-1$ rounds and messages $M_1, \ldots, M_{j-1}$ player i plays in round j, i owns the node at level j in the game tree reached by the game path $M_1, \ldots, M_{j-1}$. Player i's payoff from an instance of play resulting in o is $u_i(t_i, o)$. Such a selection protocol is in a NE if, in the associated game, it is a NE for every player i to play according to strategy S_i^j if any of his nodes at level j is reached (we say that i follows strategy S_i).

2.3 Rationality in Selection Protocols – Definitions

We now define notions of what it means for a protocol to be faithful, i.e. in which it is in players' best interests to follow the protocol specification. Because there is no synchrony within rounds of a selection protocol, we may view the possibility of rushing as a strategy for players. That is, a player may choose to wait until others have played, and only then submit his message. Thus, NE does not suffice as a solution concept for such games. However, if only one player plays in each round, then this does not matter (since rushing is only allowed *within* rounds), and so for such protocols NE is a reasonable solution concept. For general protocols, we would like the protocol to be optimal for players *regardless of the order of play within a round*. This motivates the following.

For any q-round protocol $\mathcal{P}$, we can construct protocol $\mathcal{P}'$ with at most qn rounds, and such that only one player sends a message in each round. We say that $\mathcal{P}'$ is a *linearization* of $\mathcal{P}$, and it is constructed as follows: Let $\pi : [n] \times [q] \mapsto [nq]$ be some bijective map. Then $\mathcal{P}'$ is such that in round ℓ, if $(i, k) = \pi^{-1}(\ell)$ then player i sends a message sampled from S_i^k. This is well-defined for

[4]While it is possible to model simultaneous play as an extensive-form game with *imperfect* information, the ability to rush and the lack of synchrony are more difficult to incorporate into this framework.

oblivious protocols, and essentially means players play in an arbitrary order, but only one player per round. For non-oblivious protocols, we require π to be *round-respecting*: $\pi(i,k) = \ell$ if and only if for all $j \in [n]$ and $k' < k$ it holds that $\pi(j,k') < \ell$. That is, here the arbitrary ordering is only within rounds.

We note that the idea of considering all linearizations appears also in Ong et al. [28].

Our first solution concept for selection protocols in the full-information model is a full-information dominant strategy equilibrium, which essentially means that for any player i, regardless of the messages sent by others in all rounds, i can never strictly increase his utility by deviating from the protocol. The following generalizes the definition of [5] to multi-round protocols.

Definition 2.3 (full-information dominant strategy equilibrium) *An oblivious selection protocol $\mathcal{P}$ is in a* dominant strategy equilibrium *if for all type profiles, all linearizations of $\mathcal{P}$ are in a NE.*

An alternative, more direct but equivalent formulation is the following:

Definition 2.4 (full-information dominant strategy equilibrium – alternative formulation) *An oblivious n-player, q-round, m-item selection protocol is in a* full-information dominant strategy equilibrium *if for all $i \in [n]$ and messages $M_{-i} = \{M_k^l : k \in [n] \setminus \{i\}, l \in [q]\}$ sent by all other players in all rounds, it holds that $u_i(t_i, f(M_{-i}, M_i)) = u_i(t_i, f(M_{-i}, M_i'))$, where $M_i, M_i' \in supp(S_i^1) \times \ldots \times supp(S_i^q)$.*

Remark 2.5 The reason we have equality above, as opposed to an inequality, is that the actions in the support of S_i^j are *all* dominant. That is, all these actions are best-responses, even conditioned on the actions of others. It can thus not be that one such action is better than the other, for then the other would not be dominant.

The definition of a full-information dominant strategy equilibrium is rather strong, but still achievable (for example, Theorems 4.1, 4.3, and 4.4 below). We note that our impossibility result, Theorem 3.1, applies even to our weaker solution concepts.

In a full-information ex post NE the requirement is a bit relaxed: a player i can not strictly increase his expected utility in any round j by deviating, regardless of the messages of players in all rounds up to *and including* round j. That is, regardless of the order of play within the current round, i has no incentive to deviate (on expectation over play in future rounds). The following definition is new:

Definition 2.6 (full-information ex post Nash equilibrium) *A selection protocol $\mathcal{P}$ is in an* ex post NE *if for all type profiles, all round-respecting linearizations of $\mathcal{P}$ are in a NE.*

The alternative, more direct but equivalent formulation for this solution concept is a bit more involved, and appears in Appendix A.1.

2.4 Rationality in the Face of an Adversary – Definitions

A protocol that satisfies the definitions of Section 2.3 is an optimal strategy for players assuming all others also follow the protocol. If some of the players are adversarial, however, then this may not hold. In this case, we actually want a stronger guarantee. To this end, we need the following definition, first defined by [1] (for normal-form games):

Definition 2.7 (v-tolerant NE) *A* v-tolerant NE *in an extensive-form game is a mixed strategy for every player at every node that he owns, such that the following holds: for any* $V \subset [n]$ *of size at most* v*, if all players in* $[n] \setminus \{V\}$ *play their NE strategy, then none of them can obtain a higher expected payoff by deviating from the NE at any of their nodes* regardless of the actions of players in V.

The ideal faithfulness guarantee that we would like for selection protocols is roughly the following: no player should be able to strictly improve his expected payoff by deviating, assuming most players follow the protocol, some play arbitrarily, and the order within any round is also arbitrary.

Definition 2.8 (full-information v-tolerant ex post NE) *A leader election protocol* $\mathcal{P}$ *is in a* full-information v-tolerant ex post NE *if all round-respecting linearizations of* $\mathcal{P}$ *are in a* v*-tolerant NE.*

One possible weakening of this definition is to consider an adversary who does not act arbitrarily, but also has his own utility function u_A. Suppose an adversary corrupts a set V of players. Then we say he is playing a *coalition-optimal strategy* with respect to strategies $S = (S_1, \ldots, S_n)$ if, when the players not in V follow strategies S_{-V}, the members of V play a coordinated strategy that maximizes the expectation of u_A. We say he is playing a *strictly coalition-optimal strategy* with respect to strategies S if the above holds, and if, at every node owned by some $i \in V$, i follows S_i if his part of the coordinated deviation does not strictly increase the expectation of u_A. (A more formal definition appears in Appendix A.3).

Definition 2.9 (v-tolerant NE with (strictly) self-interested adversary) *A* v-tolerant NE with self-interested adversary *in an extensive-form game is a mixed strategy* S_j *for every player* j *for every node that he owns, such that the following holds: for any* $V \subset [n]$ *of size at most* v *and any player* $i \notin V$*, if the players in* V *play any coalition-optimal strategy and the others play their* S_j *strategy, then* i *can not increase his expected utility by deviating from* S_i*. If this holds only when the players* V *play a strictly coalition-optimal strategy, then the equilibrium is a* v-tolerant NE with strictly self-interested adversary.

Definition 2.10 (full-information v-tolerant ex post NE with (strictly) self-interested adversary) *A leader election protocol* $\mathcal{P}$ *is in a* full-information v-tolerant ex post NE with a (strictly) self-interested adversary *if all round-respecting linearizations of* $\mathcal{P}$ *are in a* v*-tolerant NE with a (strictly) self-interested adversary.*

2.5 Resilience to Rational Coalitions – Definitions

In Section 2.4 the adversarial coalition could act arbitrarily, or by maximizing some joint utility function u_A. In this section we define notions of rational coalitions – i.e. coalitions that rational players might reasonably want to form. In the following definitions, we assume there is some prescribed protocol $\mathcal{P}$ for the players. When we say players are "at least as well off" or "strictly gain", this is with respect to following the prescribed protocol.

Definition 2.11 (Pareto coalition) *A coalition* V *is a* Pareto coalition *if there exists a coordinated rushing strategy* S_V^* *for the players in* V *such that all players in* V *are at least as well off when playing* S_V^**, and one player strictly gains.*

Definition 2.12 (strong coalition) *A coalition V is* strong *if there exists a coordinated rushing strategy S^*_V for players V such that the expected utility of every $i \in V$ strictly increases when playing S^*_V.*

Definition 2.13 (stable coalition) *A coalition V is* stable *if there exists a coordinated rushing strategy S^*_V for players V such that the expected utility of every $i \in V$ strictly increases when playing S^*_V, and, in addition, for all sub-coalitions $V' \subset V$ and any coordinated rushing strategy $S^*_{V'}$, playing $S^*_{V'}$ does not increase the expected utility of all players in V' when players $V \setminus V'$ play S^*_V.*

3 Impossibility with Complex Preferences

A player in a selection protocol has *complex preferences* if for any two outcomes $o \neq o'$ he strictly prefers one over the other. We now show that there are no faithful selection protocols for players with such preferences.

Theorem 3.1 *No selection protocol can be in a full-information ex post NE for players with complex preferences.*

Proof Sketch: Suppose there exists an *oblivious* selection protocol $\mathcal{P}$ in an ex post Nash equilibrium, and fix some round-respecting linearization of $\mathcal{P}$. Let T be the corresponding game tree, where some player i owns a node u (that is reached with positive probability) at the lowest non-leaf level ℓ. Suppose the protocol specification is for i to play mixed strategy S_i at level ℓ. Now, if different actions in $supp(S_i)$ result in leaves with different outcomes, then i prefers one outcome over the others (due to complex preferences). However, due to the full-information ex post NE this can not be the case: a player's different actions should not affect his expected utility, for otherwise he would have a beneficial deviation. We conclude that player i's actions do not influence the final choice of item. Hence, u can safely be omitted, resulting in a new, smaller tree. We continue shrinking the tree in this manner, yielding a deterministic selection protocol (a contradiction). The extension to non-oblivious protocols appears in Appendix B. ■

4 Rational Leader Election Protocols

4.1 Basic Faithful Leader Election Protocols

Because of Theorem 3.1, we must limit the preferences in order to obtain protocols. One natural setting for leader election is that of *self-interested* players: players care only about whether or not they are elected (they either want to win or want to not win), but are indifferent otherwise. Note that if a leader election protocol is in a full-information dominant strategy equilibrium for self-interested players, then the messages sent by others determine whether a player is elected or not (because the equilibrium holds for all type profiles). That player can only determine who is elected if he is not. The same holds for leader election protocols in a full-information ex post NE, but on expectation over messages in future rounds.

There are some basic protocols that we will use in our constructions. The first is a 1-round leader election protocol that is in a full-information dominant strategy equilibrium (but is not resilient). This protocol was given by Antonakopoulos [6] for the uniform distribution, and then generalized by Altman and Tennenholtz [5].

Theorem 4.1 ([5]) *For any $n \geq 4$ and any distribution $\mathcal{D}$ over $[n]$ there exists a 1-round, n-player leader election protocol $\mathcal{P}_{AT}$ in a full-information dominant strategy equilibrium, and in which each player i is elected with probability $\mathcal{D}(i)$.*

[5] also showed that there is no faithful 1-round leader election protocol for 3 players. The following protocol, which we will use in our constructions, does work for 3 players, albeit at the cost of having 2 rounds[5]. Fix any natural number $k \geq 3$, and denote $i_+ = (i \mod (k-1)) + 1$. Then for any positive $p_1, \ldots, p_k$ with $p_1 + \ldots + p_k = 1$ define

Protocol $\mathcal{P}_k$:

1. Player k chooses one player $i \neq k$, each with probability $\frac{p_{i_+}}{1-p_k}$.
2. For each $j \in \{1, \ldots, k-1\}$, if player j is chosen in round 1, he elects player k with probability p_k and player j_+ with probability $1 - p_k$ as leader.

Proposition 4.2 *$\mathcal{P}_k$ is a leader election protocol in a full-information ex post NE that elects each player i with probability p_i.*

4.2 Combining Rationality and Resilience in Leader Election

Neither of the protocols of Section 4.1 is resilient for any $t > 1$. The following theorem can be combined with resilient leader election protocols to obtain protocols that are both resilient and in full-information dominant strategy equilibria.

Theorem 4.3 *For any $n \geq 4$, $k = \Omega(\sqrt{n})$, and any explicit, oblivious $r(n)$-round leader election protocol $\mathcal{P}$ there exists an explicit protocol $\mathcal{P}'$ in a dominant strategy equilibrium that has $r(\lceil n/4 \rceil)$ rounds. If $\mathcal{P}$ is resilient to $t(n)$ faults, then $\mathcal{P}'$ is resilient to $t(\lfloor n/4 \rfloor) - k$ faults.*

Proof: In the protocol below and the rest of the proof, indices are cyclical. We will prove the theorem for n a multiple of 4. The general case follows similar lines. The players are partitioned into 4 disjoint sets C_1, C_2, C_3, C_4, where $i \in C_j$ if $\lceil 4i/n \rceil = j$. The following is done in **parallel**:

1. Each set C_i runs protocol $\mathcal{P}$ to select a representative R_i.
2. For each i, R_i chooses a random player from C_{i+1}, say L_{i+1}, and outputs a random message b_i to $\mathcal{P}_{AT}$ (i.e. b_i is a random element of B_i, where $\mathcal{P}_{AT}$ takes inputs from $B_1 \times \ldots \times B_4$).
3. The winner is L_j, where j is the winner of $\mathcal{P}_{AT}$ with inputs b_1, b_2, b_3, b_4.

Since the 3 steps are done in parallel, all players choose a random player and a random input in step 2., but the output depends only on the choices of the R_i's.

Fix some player x, and suppose $x \in C_i$. x is chosen as the leader only if R_{i-1} chooses x. The probability that this occurs does not change regardless of the actions of x. Additionally, for x to win, i must be the winner of $\mathcal{P}_4$. However, since $\mathcal{P}_{AT}$ is in an ex post NE, no player in C_i can influence the probability that this occurs. Hence, from x's perspective, it does not matter who is chosen as L_i.

[5]Note that the case of 2 players is impossible by the lower bound of [31], regardless of the number of rounds.

Now consider some cheating coalition of t players. In order for a member of the coalition to win, at least one member of the coalition must be chosen as L_i for some i. In order for this to occur, either R_{i-1} must be a member of the coalition (and then he can choose a fellow member in C_i), or R_{i-1} is an honest player who chooses a member of the coalition. Suppose there are c_1 faulty players in C_{i-1} and c_2 faulty players in C_i, where $c_1 + c_2 \leq t$. Then the probability that there are more than $c_2 + k$ honest players in C_{i-1} who choose a faulty player is at most a constant $e = \exp(-2k^2/(n/4 - c_1)) < \exp(-16k^2/n) < 1$ by a multiplicative Chernoff bound, and using the fact that $c_1 \leq t(n/4) < n/8$ (since no leader election protocol can be resilient to more than half the players). Thus, with probability at least $1 - e$, there are at most $c_1 + c_2 + k \leq t(n/4)$ players in C_{i-1} who choose a coalition member in C_i. The maximal probability that one of them wins and becomes R_{i-1} is at most a constant $\varepsilon < 1$ (since we can view the honest players who chose a coalition-member as additional faulty players). The probability that a coalition member becomes L_i for any i is thus at most $1 - (1 - \varepsilon)^4 \cdot (1 - e)^4$, which is some constant < 1. ■

In Theorem 4.3 is that the size of the coalition shrinks by about a factor of 4, and so we can not use it to get a faithful protocol with resilience close to the optimal $n/2$. The following protocol has optimal resilience, is in a full-information dominant strategy equilibrium, and has $\log^*(n) + O(1)$ rounds (same as in the state-of-the-art leader election protocols [30, 14]). The proof is in Appendix C.

Theorem 4.4 *For every constant $\delta > 0$ and $n \geq 4$ there exists an explicit $(\log^* n + O(1))$-round leader election protocol resilient against $n(1/2 - \delta)$ faults that is in a full-information dominant strategy equilibrium.*

Extensions and Further Results Theorem 4.3 can actually be generalized to obtain any distribution over the players (Appendix D.1). If we plug a 1-round leader election protocol into Theorem 4.3 with any distribution, we get a 1-round protocol that implements any distribution and is in a full-information dominant strategy equilibrium. This confirms a conjecture of Altman and Tennenholtz [4] about the existence of such protocols in which all players influence the outcome of the protocol in some instance. We can also construct protocols that satisfy a stronger notion of resilience against adversarial coalitions – namely, they have bounded cheaters' edge [6] – that are in a full-information ex post NE (Appendix D.2). Finally, our protocols can also be used to construct leader election protocols in which a player is elected at random *from the set of players who want to be elected* (Appendix D.3).

4.3 Rationality in the Face of an Adversary

While the protocols of Section 4.2 are resilient against adversarial behavior, they are in equilibrium only if all players follow the protocol. What if this is not the case? Can an honest player's protocol specification be optimal even when some others play adversarially? The main difficulty here is that a player's actions may now also influence the actions of adversarial players in future rounds. Even if the protocol is oblivious, an adversary's strategy might not be. Definition 2.8 defines a the concept of an full-information v-tolerant ex post NE to deal precisely with this issue.

Unfortunately, Theorem 4.5 below implies that no leader election protocol can be in a v-tolerant ex post NE, and so we must look for some relaxation. For Definition 2.8 we make no assumptions about the adversary. If we assume that the adversary also has some preferences, then we may

be able to weaken this restriction. We will assume here that the adversary's goal is to maximize the probability that some member of his coalition gets elected (the standard assumption for leader election) – that is, we consider Definition 2.10, where u_A is the probability that a member of the coalition gets elected. Theorem 4.5 also shows that this relaxation does not suffice:

Theorem 4.5 *There does not exist a leader election protocol in a v-tolerant ex post Nash equilibrium with self-interested adversary for any $v \geq 1$.*

Proof: Fix some protocol $\mathcal{P}$ in an ex post Nash equilibrium and a round-respecting linearization of $\mathcal{P}$. Suppose i is the first player who has a mixed strategy in the game, where two possible messages in i's support are I_1, I_2. Because i is eventually chosen by $\mathcal{P}$ with some probability that i can not influence himself (he wins with the same expected probability whether he plays I_1 or I_2), there must exist some other player whose choice of messages does influence this probability. In the subtree rooted at the node following i choosing I_1 there must exist some player j who has a strategy S_1 that increases the probability of i getting elected, and some other strategy S_2 that decreases this probability. Because $\mathcal{P}$ is in an ex post Nash equilibrium, these choices of player j do not harm his own chance of getting elected. A valid (adversarial) strategy for player j is to play S_1 whenever i plays I_1. Alternatively, j can play S_2 whenever i plays I_1. Because i does not know which strategy j is using (since j is adversarial), and in either case one of I_1 or I_2 is strictly better than the other, no single strategy of player i can be optimal in both cases. ■

If we limit the adversary to be strictly self-interested we can get an explicit protocol (Appendix E).

Theorem 4.6 *For any positive k and $n = 3^k$ there exists an explicit n-player $2\log_3(n)$-round leader election protocol $\mathcal{P}$ that is in a full-information n-tolerant ex post Nash equilibrium with a strictly self-interested adversary. Furthermore, $\mathcal{P}$ is resilient against $n^{\log_3(2)}/2$ faults.*

To get an idea for the proof, we show that $\mathcal{P}_k$ with $k = 3$ and the uniform distribution is in a full-information 3-tolerant ex post NE with a strictly self-interested adversary. If none or all of the players are adversarial, then all non-adversarial players should follow the protocol (since it is in a full-information ex post NE). If two players are adversarial, then they can always force a win, and so the third player may as well follow the protocol. Finally, if only one player is adversarial, then he can not increase his chance of winning (by the full-information ex post NE), and since the adversary is strictly self-interested he will not deviate. Hence, it is also a full-information ex post NE for the others to follow the protocol.

To generalize this to more players, we divide the players into sets of 3, each running $\mathcal{P}_3$. We then repeat this on the winners, until only one is left. Appendix E contains further details and an analysis of the resilience of this protocol.

4.4 Resilience to Rational Coalitions

In Sections 4.2 and 4.3 the adversary corrupts some set of v players, and coordinates their actions. Here we let players form a "rational coalition" to benefit all members – namely, we consider the definitions of Section 2.5. For the following theorems (whose proofs appear in Appendix F), we restrict ourselves to the case in which players are self-interested, and all *want* to be elected. First, we show that it is impossible to have resilience against our weakest notion of a rational coalition.

Theorem 4.7 *Every leader election protocol in a full-information ex post NE has a Pareto coalition of two players.*

For a stronger notion, however, we can get a protocol that side-steps the impossibility of leader election with adversarial coalitions of size $n/2$:

Theorem 4.8 *There exists an explicit 2-round leader election protocol in a full-information ex post Nash equilibrium with only 1 stable coalition. The coalition is of size $n-1$.*

The protocol that achieves this is $\mathcal{P}_k$ with $k = n$ and the uniform distribution (see Section 4.1). We also have the following theorem, as a weak illustration that we gained something by weakening our requirement from strong to stable coalitions.

Theorem 4.9 *For any n-player leader election protocol in a full-information ex post Nash equilibrium, all coalitions of size $n-1$ are strong.*

5 Rational Random Sampling Protocols

We consider some universe of m items, and will construct protocols that output each item with probability proportional to the number of players who (claim to) like that item most. In Appendix H we discuss generalizations to other distributions. Due to Theorem 3.1, we restrict ourselves to *single-minded* players – each i's type $t_i \in [m]$ is the item he prefers, and he is indifferent about the others. Theorem 3.1 also implies that no random sampling is possible with $m = 2$. If $m > 2$ but players prefer only one of two items we are sampling from two items. So we must limit the type profiles. We do this by considering balanced profiles: a profile $(t_1, \ldots, t_n)$ for m items is z-balanced if each type occurs between $n/m + z$ and $n/m - z$ times.

Theorem 5.1 *For any $n \geq 66$ and explicit $r(n)$-round leader election protocol resilient to $t(n)$ faults in a full-information ex post NE, there exists an explicit $(r(n)+3)$-round random selection protocol for a universe of size $m \geq 66$ that is in a full-information ex post NE for all $(n(1/66 - 1/m))$-balanced type profiles. For such profiles, the random selection protocol is resilient to $t(\lfloor n/3 - n/66 \rfloor)$ faults.*

Proof: The protocol is the following:

1. Each player announces his preferred item. Players are split into 3 categories C_1, C_2, C_3 as follows: all players with the same announced type are in the same category, and the categories are "roughly" balanced: sets of players with the same declared type are greedily assigned to the smallest C_i. Fix $c_i = |C_i|$, and note that $|c_i - c_j| \leq d$ for $d = n/66$ (assuming at most one player lies about his preferred item).

2. For each i, players in C_i run the leader election protocol $\mathcal{P}$ to elect a representative R_i.

3. For each i, the players in $C_{i+1} \cup C_{i+2}$ run the leader election protocol $\mathcal{P}$, and the winner chooses a uniformly random player L_i from C_i.

4. R_1, R_2, and R_3 run $\mathcal{P}_3$. The protocol is run so that players are elected with probabilities $\frac{c_1}{n}$, $\frac{c_2}{n}$, and $\frac{c_3}{n}$ respectively.

5. The protocol's output is the announced item of player L_j, where j is the winner of $\mathcal{P}_3$ in the last round.

If a player $i \in C_j$ truthfully announces his type, then he can no longer change the probability of his type getting chosen: he only affects which of the other types are potential winners (via his choice of L_k for $k \neq j$) or which player from C_i participates in $\mathcal{P}_3$. However, since $\mathcal{P}_3$ is in a full-information ex post NE, this does not matter either.

It remains to show that it is optimal for i to truthfully reveal his type. Suppose i's preferred item is B, the fraction of other players who announce B is β, and they all get placed in C_j. Suppose i lies about his type and gets placed in $C_k \neq C_j$. How can i cheat? i wins the leader election protocol of step (2) with probability $1/c_k$ and the leader election protocol of step (3) (choosing L_j) with probability $1/(c_k + c_\ell)$ for $\ell \neq j, k$. If i is elected in both leader election protocols, he can force the winner to be a player who wants B with probability at most 1. If he wins only the leader election protocol of step (2), he can cause j to win in $\mathcal{P}_3$ with probability $1 - c_k/n$. If he wins only the leader election protocol of step (3), he can force L_j to be a player who wants B (but that player wins $\mathcal{P}_3$ with probability c_j/n. The probability that B is the chosen type given that i is cheating is

$$\begin{aligned}\Pr[B \text{ wins}] &< \left(1 - \frac{1}{c_k + c_\ell} - \frac{1}{c_k} + \frac{1}{c_k} \cdot \frac{1}{c_k + c_\ell}\right) \cdot \beta + \frac{1}{c_k + c_\ell} \cdot \frac{c_j}{n} + \frac{1}{c_k} \cdot \frac{\beta n}{c_j}\left(1 - \frac{c_k}{n}\right) + \frac{1}{c_k} \cdot \frac{1}{c_k + c_\ell} \\ &= \beta - \frac{\beta}{c_k + c_\ell} - \frac{\beta}{c_k} + \frac{c_j}{(c_k + c_\ell)n} + \frac{\beta n}{c_k \cdot c_j} - \frac{\beta}{c_j} + \frac{1}{c_k} \cdot \frac{1}{c_k + c_\ell}.\end{aligned}$$

By our balancedness assumption, we know that $n/3 - d \leq c_1, c_2, c_3 \leq n/3 + d$, and that $\beta \leq d/n$. Plugging in these values (and performing some manipulations) yields

$$\Pr[B \text{ wins}] < \beta + \left(\frac{n + 3d}{2n - 6d}\right)\frac{1}{n} + \frac{9d + 18}{\left(\frac{n}{3} - d\right)^2}.$$

It can be verified that when $d \leq n/66$ and $n \geq 66$, we get that $\Pr[B \text{ wins}] < \beta + 1/n$. Now, if player i were to bid truthfully, then the probability that B wins would be $\beta + 1/n$ (since i's vote adds to B's chance of winning). Thus, it is an optimal strategy for i to bid truthfully.

What about resilience? Suppose there is an adversary of size at most $t(\lfloor n/3 - n/66 \rfloor)$ faults. In order to force an outcome in some predefined set, the adversary must win at least one of the 6 runs of the leader election protocol $\mathcal{P}$, and each runs on a set of at least $\lfloor n/3 - n/66 \rfloor$ players. Since $\mathcal{P}$ is resilient for this number of adversaries, the probability that the adversary loses all of them is at least ε^6 for some constant $\varepsilon > 0$. ■

The following works for smaller m, and is proved in Appendix G.

Theorem 5.2 *For $n \geq 3$, any explicit $r(n)$-round leader election protocol resilient up to $t(n)$ faults in a full-information ex post NE, and any constant natural number $m \geq 3$, there exists an explicit $(r(n) + 4)$-round random selection protocol for a universe of size m that is in a full-information ex post NE for all z-balanced profiles, where $z = n/10m^2$. For such profiles, the random selection protocol is resilient up to $t(\lfloor n/m - z \rfloor)$ faults.*

6 Conclusion and Open Problems

Perhaps the main insight of this paper is that the full-information model is a setting that allows for a relatively clean examination of the interplay between rationality and adversarial behavior in the presence of asynchronous communication. While we have explored numerous aspects of this interplay, we are now faced with many more open questions.

The first set of questions consists of direct extensions of the results presented here. For example, can one generalize the types of preferences for which there are faithful and resilient protocols? For random selection protocols, for example, one might consider a setting in which each players likes some set of items, and dislikes the others. Are there random sampling protocols with weaker balancedness assumptions? How about such protocols that are rational in the face of an adversary, or resilient to rational coalitions? Also, are there protocols with few strong coalitions? Finally, one may consider approximate solution concepts: for example, one may desire all linearizations of a protocol to be in an ε-Nash equilibrium for a small but positive ε. Note that in this case our impossibility result of Theorem 3.1 no longer applies.

The second set of questions is more open-ended. What can one say about rationality for more general protocol problems in the full-information model? And are there other tractable models for the study of the interplay between rationality and adversarial behavior?

Acknowledgements I would like to thank Ran Canetti, Moni Naor, and Omer Reingold for helpful conversations. I am also grateful to the anonymous referees for their comments.

References

[1] I. Abraham, D. Dolev, R. Gonen, and J. Halpern. Distributed computing meets game theory: Robust mechanisms for rational secret sharing and multiparty computation. *Proceedings of 25th Annual ACM Symposium on Principles of Distributed Computing*, pages 53–62, 2006.

[2] A. Aiyer, L. Alvisi, A. Clement, M. Dahlin, J.-P. Martin, and C. Porth. Bar fault tolerance for cooperative services. *Proceedings of 20th ACM Symposium on Operating Systems Principles*, pages 45–58, 2005.

[3] N. Alon and M. Naor. Coin-flipping games immune against linear-sized coalitions. *SIAM Journal of Computing*, 22(2):403–417, 1993.

[4] A. Altman and M. Tennenholtz. Selection games and deterministic lotteries. `http://iew3.technion.ac.il/~moshet/selection-lottery.pdf`, 2008.

[5] A. Altman and M. Tennenholtz. Strategyproof deterministic lotteries under broadcast communication. *Proceedings of AAMAS*, 2008.

[6] S. Antonakopoulos. Fast leader-election protocols with bounded cheaters' edge. *Proceedings of STOC*, pages 187–196, 2006.

[7] R. J. Aumann. Acceptable points in general cooperative n-person games. *Contributions to the Theory of Games, Annals of Mathematical Studies*, IV:287–324, 1959.

[8] M. Ben-Or and N. Linial. Collective coin fliping. *Advances in Computing Research*, 5:91–115, 1989.

[9] B. D. Bernheim, B. Peleg, and M. Whinston. Coalition proof nash equilibrium: Concepts. *Journal of Economic Theory*, 42(1):1–12, 1989.

[10] F. Brandt, F. Fischer, and Y. Shoham. On strictly competitive multi-player games. *Proceedings of AAAI*, 2006.

[11] I. B. Damgård. Interactive hashing can simplify zero-knowledge protocol design without computational assumptions. *Proceedings of CRYPTO*, 1993.

[12] I. B. Damgård, O. Goldreich, and A. Wigderson. Hashing functions can simplify zero-knowledge protocol design (too). Technical Report TR RS-94-39, BRICS, 1994.

[13] Y. Z. Ding, D. Harnik, A. Rosen, and R. Shaltiel. Constant-round oblivious transfer in the bounded storage model. *Proceedings of STOC*, 2004.

[14] U. Feige. Noncryptographic selection protocols. *Proceedings of 40th Annual Symposium on Foundations of Computer Science*, pages 142–152, 1999.

[15] J. Feigenbaum and S. Shenker. Distributed algorithmic mechanism design: Recent results and future directions. *Proceedings of the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pages 1–13, 2002.

[16] O. Goldreich, S. Goldwasser, and N. Linial. Fault-tolerant computation in the full information model. *SIAM J. Computing*, 27(2), 1998.

[17] O. Goldreich, A. Sahai, and S. Vadhan. Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge. *Proceedings of 30th STOC*, 1998.

[18] R. Gradwohl. Fault tolerance in distributed mechanism design. *Proceedings of the 4th Workshop on Internet and Network Economics*, 2008.

[19] R. Gradwohl, S. Vadhan, and D. Zuckerman. Random selection with an adversarial majority. *Proceedings of CRYPTO*, 2006.

[20] J. Halpern. A computer scientist looks at game theory. *Games and Economic Behavior*, 45(1):114–131, 2003.

[21] J. Halpern. Computer science and game theory: A brief survey. *The New Palgrave Dictionary of Economics*, 2008.

[22] E. Kalai. Large robust games. *Econometrica*, 72(6):1631–1665, 2004.

[23] J. Katz. Bridging game theory and cryptography: Recent results and future directions. *Proceedings of TCC*, 2008.

[24] D. Monderer and M. Tennenholtz. Distributed games: from mechanisms to protocols. *Proceedings of the 16th National Conference on Artificial Intelligence*, pages 32–37, 1999.

[25] D. Moreno and J. Wooders. Coalition-proof equilibrium. *Games and Economic Behavior*, 17(1):80–112, 1996.

[26] M. Naor, R. Ostrovsky, R. Venkatesan, and M. Yung. Perfect zero-knowledge arguments for NP can be based on general complexity assumptions. *J. Cryptology*, 11, 1998.

[27] S. J. Ong, D. Parkes, A. Rosen, and S. Vadhan. Fairness with an honest minority and a rational majority. Preliminary version, October 2007.

[28] S. J. Ong, D. Parkes, A. Rosen, and S. Vadhan. Fairness with an honest minority and a rational majority. *Proceedings of the Fourth Theory of Cryptography Conference*, pages 36–53, 2009.

[29] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, MA, 1994.

[30] A. Russell and D. Zuckerman. Perfect-information leader election in $\log^* n + O(1)$ rounds. *Journal of Computer and System Sciences*, 63:612–626, 2001.

[31] M. Saks. A robust noncryptographic protocol for collective coin flipping. *SIAM Journal on Discrete Mathematics*, 2(2):240–244, 1989.

[32] S. Sanghvi and S. Vadhan. The round complexity of two-party random selection. *SIAM Journal of Computing*, 32(2):523–550, 2008.

[33] J. Shneidman and D. C. Parkes. Specification faithfulness in networks with rational nodes. *Proceedings of PODC*, 2004.

A Definitions – Alternative Formulation

In this section we give definitions of various solution concepts for the full-information model. The definitions are more general than those given in Sections 2.3 to 4.3, and their formulation is also a bit different and more direct.

A.1 Rationality in the Full-Information Model

Recall our strongest notion of equilibria for oblivious selection protocols.

Definition A.1 (full-information dominant strategy equilibrium) *An oblivious n-player, q-round, m-item selection protocol is in a* full-information dominant strategy equilibrium *if for all $i \in [n]$ and messages $M_{-i} = \{M_k^l : k \in [n] \setminus \{i\}, l \in [q]\}$ sent by all other players in all rounds, it holds that $u_i\left(t_i, f(M_{-i}, M_i)\right) = u_i\left(t_i, f(M_{-i}, M_i')\right)$, where $M_i, M_i' \in supp(S_i^1) \times \ldots \times supp(S_i^q)$.*

If a player's strategy is in a dominant strategy equilibrium, then regardless of the actions he plays, he can not gain or lose. Our protocols will require players to play a mixture of such actions, and such a mixture can thus not harm them.

Before defining our next notion of equilibria, we formalize the notion of a rushing strategy.

Definition A.2 (rushing strategy) *In a q-round, n-player selection protocol, a* rushing strategy *for player i is a set of q random functions $\widetilde{S_i^1}, \ldots, \widetilde{S_i^q}$, where each $\widetilde{S_i^j}$ is a function of messages sent in all rounds up to* and including *round j, namely $\{M_k^l : k \in [n], l < j\} \cup \{M_k^j : k \in [n] \setminus \{i\}\}$, as well as player i's type t_i.*

Definition A.3 (full-information ex post Nash equilibrium) *A q-round selection protocol is in an* full-information ex post Nash equilibrium *if*

1. *for all $i \in [n]$, player types $t_1, \ldots, t_n$, and rounds $j \leq q$,*
2. *for all messages $M^{<j} = \{M_k^l : k \in [n], l < j\}$ sent by all players in rounds prior to j that satisfy $M_k^l \in supp\left(S_k^l\left(\{M_k^{l'} : k \in [n], l' < l\}, t_k\right)\right)$,*
3. *for all messages $M_{-i}^j = \{M_k^j : k \in [n] \setminus \{i\}\}$ sent by all players other than i in round j that satisfy $M_k^j \in supp\left(S_k^j\left(\{M_k^{l'} : k \in [n] \setminus \{i\}, l' < j\}, t_k\right)\right)$,*
4. *for all rushing strategies $\widetilde{S_i^{>j}} = \{\widetilde{S_i^{j+1}}, \ldots, \widetilde{S_i^q}\}$ of player i,*
5. *for all messages $M_i^j \in supp\left(S_i^j\left(M^{<j}, t_i\right)\right)$ of player i that are in the support of his strategy given his true type, and*
6. *for all messages $\overline{M_i^j} \in \bigcup_{t_i' \in T_i} supp\left(S_i^j\left(M^{<j}, t_i'\right)\right)$ of player i that are in the support of his strategy given any type,*

it holds that $\mathrm{E}\left[u_i\left(t_i, f\left(S_{-i}^{>j}, \widetilde{S_i^{>j}}, M^{<j}, M_{-i}^j, M_i^j\right)\right)\right] \geq \mathrm{E}\left[u_i\left(t_i, f\left(S_{-i}^{>j}, \widetilde{S_i^{>j}}, M^{<j}, M_{-i}^j, \overline{M_i^j}\right)\right)\right]$.
The expectation is over the randomness of players' strategies in rounds $j+1$ to q.

The above definition captures player i's ability to rush by requiring that any pair of actions M_i^j and $\overline{M_i^j}$ in the support of player i's strategy yield the same payoff, conditioned on all messages sent in previous rounds, the messages of other players in the current round, and on expectation over future rounds. If $\overline{M_i^j}$ is a legal message not in the support given i's true type, then it may be strictly better for player i to play M_i^j.

A.2 Adversarial Coalitions

The usual guarantee studied in the full-information model is a form of resilience against a rushing adversary. Before the protocol begins, the adversary chooses some set of t players to corrupt, and assigns them coordinated rushing strategies. For any n-player leader election protocol $\mathcal{P}$, denote by $\mathrm{fail}_{\mathcal{P}}(n,t)$ the maximal probability that the chosen leader is one of the faulty players, assuming the non-faulty players follow the protocol (and the t corrupted players follow a coordinated rushing strategy that maximizes this probability). Then a protocol $\mathcal{P}$ is *resilient* for $t = t(n)$ if there exists some $\varepsilon > 0$ such that $\mathrm{fail}_{\mathcal{P}}(n, t(n)) \leq 1 - \varepsilon$ for all $n \geq n_0$. A protocol has *bounded edge* up to $t(n)$ faults if there exist constants c and n_0 such $n/t \cdot \mathrm{fail}_{\mathcal{P}}(n,t) - 1 \leq c$ for all $n \geq n_0$ and $t \leq t(n)$. In words – an adversary of size t can force the output to lie be a member with probability at most $O(t/n)$.

For random sampling protocols, we use similar notation as leader election: Consider a set of m items, and suppose the adversary wishes to force an element of a subset $S \subseteq [m]$ of size $|S| = s$. Then $\text{fail}_{\mathcal{P}}(n, t, m, s)$ is the maximal probability that he succeeds after corrupting t players and assigning them coordinated rushing strategies. In this case, our notions of resilience and cheaters' edge are the following:

Definition A.4 (resilience for random selection protocol) *A random sampling protocol $\mathcal{P}$ is* ε-resilient *for $t = t(n)$ if there exists some n_0 such that*

$$\mathit{fail}_{\mathcal{P}}(n, t(n), m, s) \leq 1 - \varepsilon$$

for all $n \geq n_0$ and all $s < m$. A protocol is resilient *if it is ε-resilient for a constant $\varepsilon > 0$.*

Definition A.5 (edge for random selection protocol) *For $t \geq 1$, the* cheaters' edge *of a random sampling protocol $\mathcal{P}$ is*

$$\mathit{edge}_{\mathcal{P}}(n, t, m, s) = \frac{m}{s} \cdot \mathit{fail}_{\mathcal{P}}(n, t, m, s) - 1.$$

$\mathcal{P}$ has bounded edge *if there exist constants c and n_0 such $\mathit{edge}_{\mathcal{P}}(n, t(n), m, s) \leq c$ for all $n \geq n_0$ and all natural numbers $s \leq m$.*

A.3 Rationality in the Face of Adversarial Coalitions

This section addresses the question of how to get both resilience and faithfulness together. The simplest combination requires a protocol to separately be in an equilibrium and be resilient. For example, an oblivious leader election protocol $\mathcal{P}$ is in a dominant strategy equilibrium and has bounded edge if it satisfies Definition A.1 and has bounded edge.

The next ways of combining equilibria with resilience use the following definition.

Definition A.6 (V-rushing strategy) *Fix a player i, a set $V \subseteq [n] \setminus \{i\}$, and a q-round, n-player selection protocol. Then a V-*rushing strategy *for player i is a set of q random functions $\widetilde{S_i^1}, \ldots, \widetilde{S_i^q}$, where each $\widetilde{S_i^j}$ is a function of (1) messages sent by all players in all rounds up to and including round $j-1$, namely $\{M_k^l : k \in [n], l < j\}$, (2) messages sent by players $[n] \setminus (V \cup \{i\})$ in round j, namely $\{M_k^j : k \in [n] \setminus (V \cup \{i\})\}$, and (3) player i's type t_i.*

Definition A.7 (full-information v-tolerant ex post Nash equilibrium) *A q-round selection protocol is in a* full-information v-tolerant ex post Nash equilibrium *if for all items 1–2 and items 5–6 of Definition A.3, and*

(a) for all $V \subseteq [n] \setminus \{i\}$ of size $|V| \leq v$,

(b) for all messages $M^j_{-(V \cup \{i\})} = \{M_k^j : k \in [n] \setminus (V \cup \{i\})\}$ sent by honest players in round j and satisfying $M_k^j \in \mathit{supp}\left(S_k^j\left(\{M_k^{l'} : k \in [n], l' < j\}, t_k\right)\right)$,

(c) for all V-rushing strategies $\widetilde{S_i^{\geq j}} = \{\widetilde{S_i^{j+1}}, \ldots, \widetilde{S_i^q}\}$ of player i, and

(d) for all coordinated rushing strategies $S_V^{*\geq j} = \{S_V^{j*}, \ldots, S_V^{q*}\}$ of players in V,

it holds that:

$$\mathrm{E}\left[u_i\left(f\left(S^{>j}_{-(V\cup\{i\})}, S^{*\geq j}_{V}, \widetilde{S^{>j}_{i}}, M^{<j}, M^{j}_{-(V\cup\{i\})}, M^{j}_{i}\right)\right)\right]$$
$$\geq \mathrm{E}\left[u_i\left(f\left(S^{>j}_{-(V\cup\{i\})}, S^{*\geq j}_{V}, \widetilde{S^{>j}_{i}}, M^{<j}, M^{j}_{-(V\cup\{i\})}, \overline{M^{j}_{i}}\right)\right)\right].$$

The expectation is over the randomness of players' strategies.

For Definition A.7 we make no assumptions about the adversary. In item (d), we allow the adversary to play any coordinated rushing strategy, and in particular this includes strategies that may actually be harmful to the adversary. If we assume that the adversary also has some utility function, say u_A, then we may be able to weaken this restriction. The following two definitions are two ways of doing this: the first only allows strategies that do not harm the adversary, and the second allows only strategies that strictly benefit the adversary. Note, however, that we do not require the adversary to optimally choose which players to corrupt – the equilibrium must hold for any such choice.

Definition A.8 (full-information v-tolerant ex post Nash equilibrium with self-interested adversary
A q-round selection protocol is in a full-information v-tolerant ex post Nash equilibrium with a self-interested adversary *if it satisfies Defintion A.7, but with item (d) replaced by:*

(d) for all coordinated rushing strategies $S^*_V = \left(S^*_{V_1}, \ldots, S^*_{V_v}\right)$, where $S^*_{V_k} = \left(S^{1*}_{V_k}, \ldots, S^{q*}_{V_k}\right)$ for all $k \in [v]$, satisfying:

- *for any other coordinated rushing strategies $S^{*'}_V$ for players V, $\mathrm{E}\left[u_A\left(f\left(S_{-V}, S^*_V\right)\right)\right] \geq \mathrm{E}\left[u_A\left(f\left(S_{-V}, S^{*'}_V\right)\right)\right]$, and*
- *for all $k \in [v]$ and $j \in [q]$ with $S^j_{V_k} \not\equiv S^{j*}_{V_k}$,*
  $\mathrm{E}\left[u_A\left(f\left(S_{-V}, S^*_V\right)\right)\right] \geq \mathrm{E}\left[u_A\left(f\left(S_{-V}, S^*_{(V\setminus\{V_k\})}, S^{1*}_{V_k}, \ldots, S^j_{V_k}, \ldots, S^{q*}_{V_k}\right)\right)\right]$.

Definition A.9 (full-information v-tolerant ex post Nash equilibrium with strictly self-interested ad
A q-round selection protocol is in a full-information v-tolerant ex post Nash equilibrium with a strictly self-interested adversary *if it satisfies Defintion A.8 with the final inequality of (d) replaced by a strict inequality.*

B Proof of Theorem 3.1

We first restate the theorem.

Theorem B.1 (Theorem 3.1 restated) *No selection protocol can be in a full-information ex post NE for players with complex preferences.*

Proof: Suppose there exists a selection protocol $\mathcal{P}$ in a full-information ex post Nash equilibrium. If $\mathcal{P}$ is not an oblivious protocol, then the proof sketch of Section 3 does not immediately work for the following reason: In non-oblivious protocols a player's message may also be a function of his type. Thus, one can not immediately reduce the communication, because the protocol does not

"know" the players' types. However, it is possible to get around this difficulty by considering a different protocol $\mathcal{P}'$.

Let $\mathcal{P}'$ be the following protocol: in round 1, all players broadcast their types. After this, the players follow the protocol $\mathcal{P}$ with one modification: for each player i and round j, i's legal moves are those in $supp(S_i^j(t_i, \{M_k^\ell : k \in [n], \ell < j\}))$, where t_i is the type i declared in round 1.

$\mathcal{P}'$ is in a full-information ex post NE because $\mathcal{P}$ is in a full-information ex post NE *for all type profiles*. That is, even if all the players know each other's types, they will have no incentive to deviate. Thus, all players may as well reveal their types to one another. Furthermore, because $\mathcal{P}$ is in a full-information ex post NE, no player would have any reason to send a message that is illegal in $\mathcal{P}'$.

From this point on, the proof continues as in the proof sketch in Section 3, but with $\mathcal{P}'$ instead of $\mathcal{P}$. ■

C Proof of Theorem 4.4

In this section we prove Theorem 4.4. We begin with a definition of resilience for random selection protocols, construct such a protocol (as a variant of a protocol of Gradwohl et al. [19]), and then use this protocol in Theorem 4.4. We say a selection protocol is $(\beta, \mu, \varepsilon)$-*resilient* if when at most a β fraction of players are cheating and V is any subset of $[m]$ of density at most μ, the probability that the output lands in S is at most ε. More formally,

Definition C.1 *An m-item n-player selection protocol $\mathcal{P}$ is called* $(\beta, \mu, \varepsilon)$-resilient *if $fail_{\mathcal{P}}(n, \beta \cdot n, m, s) \leq \varepsilon$ for all $s \leq \mu \cdot m$.*

A slight variation on a theorem of [19] yields the following protocol:

Theorem C.2 *For any constant $\delta > 0$, there exists a constant $\varepsilon < 1$ and an m-item n-player selection protocol with the following properties:*

(i) The protocol has $\max(\log^* n, \log^* m) + O(1)$ *rounds.*

(ii) The protocol is $(1 - \alpha + \delta, \alpha - \delta, \varepsilon)$-resilient for all $0 \leq \alpha \leq 1$.

The proof uses the following as a sub-protocol:

Definition C.3 *A* $[(n, m) \mapsto (n', m')]$-universe+player reduction protocol *is an n-player protocol whose output is a sequence $(s_1, \ldots, s_{m'})$ of elements of $[m]$ and a sequence $(t_1, \ldots, t_{n'})$ of elements of $[n]$. Such a protocol is* $[(\beta, \mu) \overset{\gamma}{\mapsto} (\beta', \mu')]$-resilient *if when at most a β fraction of players are cheating and S is any subset of $[m]$ of density at most μ, the probability that at most a β' fraction of the first output sequence are cheating players and at most a μ' fraction of the second output sequence is in S is at least γ.*

The following is a theorem of [19]:

Theorem C.4 (many-round universe+player reduction) *For every $m, n \in \mathbb{N}$ and every $\beta, \theta, \varepsilon > 0$, there exists a $[(n, m) \mapsto (n', m')]$-universe+player reduction protocol that is $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} (\beta + \theta, \mu + \theta)]$-resilient for every $\mu > 0$, with*

$$\begin{aligned} m' &= \text{poly}(\log(1/\varepsilon), 1/\theta) \\ n' &= \text{poly}(\log(1/\varepsilon), 1/\theta). \end{aligned}$$

Moreover, the number of rounds is $t = \max\{\log^* m, \log^* n\} - \log^* m' + O(1)$.

We now prove Theorem C.2.

Proof: The protocol proceeds in two steps:

1. The players run the $[(n, m) \mapsto (n', m')]$-universe+player reduction protocol of Theorem C.4 with $\beta = 1 - \alpha$, $\theta = \delta/2$, and $\varepsilon > 0$ an arbitrary constant. This yields a constant number n' players and a universe of size m' (which is also a constant). With probability at least $1 - \varepsilon$, the fraction of honest players in n' is larger than the fraction of elements from the target set in m'.

2. Replicate each of the m' items n' times. The players now run the elimination protocol: players sequentially eliminate one of the $m' \cdot n'$ items at random, until one remains. Note that with constant probability, every honest player eliminates an element from the target set at every turn. Since the fraction of honest players is larger, this means that with some constant nonzero probability the remaining item is not in the target set.

■

We can finally prove Theorem 4.4.

Theorem C.5 (Theorem 4.4 restated) *For every constant* $\delta > 0$ *and* $n \geq 4$ *there exists an explicit* $(\log^* n + O(1))$*-round leader election protocol resilient against* $n(1/2 - \delta)$ *faults that is in a dominant strategy equilibrium.*

Proof: Let $\mathcal{P}'$ be an $(n/2)$-item $(n/2)$-player random selection protocol guaranteed by Theorem C.2. Partition the players into two sets $C_1 = \{1, \ldots, \lfloor n/2 \rfloor\}$ and $C_2 = \{\lfloor n/2 \rfloor + 1, \ldots, n\}$. The players in C_1 run $\mathcal{P}'$ twice simultaneously to select two distinct players from the other side – call the lexicographically smaller one L_1 and the lexicographically larger one L_2. The players in C_2 simultaneously runs $\mathcal{P}'$ twice to select two distinct players from C_1– call the lexicographically smaller one L_3 and the lexicographically larger one L_4. Next, players 1, $\lfloor n/2 \rfloor$,$\lfloor n/2 \rfloor + 1$, and n run $\mathcal{P}_{AT}$ with the uniform distribution. The winner is L_i if the i'th of the four players won in $\mathcal{P}_3$.

Note that this protocol is in a dominant strategy equilibrium: a player only influences the other candidates for $\mathcal{P}_{AT}$, and his own candidacy is not in his control. The four players who play $\mathcal{P}_{AT}$ do not have any incentive to deviate, since either they are not candidates, or they are but can not increase their chances of getting elected (since $\mathcal{P}_3$ is also in a dominant strategy equilibrium).

Finally, the protocol is resilient: The total number of adversarial players is $n(1/2 - \delta)$, so if the fraction of adversarial players in C_1 is α, the fraction of adversarial players in CS_2 is at most $1 - \alpha - \delta$. Thus, the probability that an adversarial player is selected in any run of the random selection protocol is at most $1 - \varepsilon$ for a constant $\varepsilon > 0$. With probability at least $\varepsilon^4 > 0$, no adversarial player is selected, and the output of the protocol is an honest player. ■

D Extensions to Rational Leader Election Protocols

D.1 Theorem 4.3 and D.2 with Any Distribution

Theorem D.1 *Theorem 4.3 and D.2 can attain any distribution* $(p_1, \ldots, p_n)$ *over the players with support size at least 4.*

Proof: For both theorems, choose every set C_i of players so that there is at least one player $j \in C_i$ with $p_j \neq 0$. Now, whenever a player R_j chooses an element k from C_i, he does so with probability $\Pr[p_k] / (\sum_{j \in C_i} p_j)$. When players in Theorem 4.3 run $\mathcal{P}_{AT}$, they do so with probabilities q_1, q_2, q_3, q_4, where $q_i = \Pr[C_i \text{ contains the winning player}]$. In Theorem D.2, the player who chooses between the two final candidates flips a biased coin, where the probability of each candidate is the probability that the winning player is in his respective set. ■

D.2 Bounded Cheaters' Edge

Recall Definition A.5 on bounded edge: a leader election protocol has bounded cheaters' edge up to $t(n)$ faults if for any $t \leq t(n)$, the adversary can force the output to be a coalition member with probability at most $O(t/n)$.

Theorem D.2 *Let $\mathcal{P}$ be an explicit $r(n)$-round leader election protocol with bounded cheaters' edge up to $t(n)$ faults. Then there exists an explicit leader election protocol $\mathcal{P}'$ in a full-information ex post Nash equilibrium with bounded cheaters' edge up to $t(\lfloor n/2 \rfloor)$ faults and $r(\lceil n/2 \rceil) + 2$ rounds.*

Proof: Partition the n players into two sets $C_1 = \{1, \ldots, \lfloor n/2 \rfloor\}$ and $C_2 = \{\lfloor n/2 \rfloor + 1, \ldots, n\}$. The following are then done sequentially:

1. The players in C_1 use $\mathcal{P}$ to elect a representative R_1, and simultaneously the players in C_2 use $\mathcal{P}$ to elect a representative R_2.
2. R_1 chooses a uniformly random player L_2 from C_2, and R_2 chooses a uniformly random player L_1 from C_1.
3. The lexicographically first player other than L_1 flips a coin to choose between L_1 and L_2, where each L_i is chosen with probability $|C_i| / n$.

Note that no player has any incentive to deviate from the protocol. The probability that a faulty leader is elected is at most the probability that one of L_1 of L_2 are faulty. L_1 is faulty if R_2 is faulty (which occurs with probability $O(t/n)$ due to bounded cheaters' edge) or if a R_2 is non-faulty, but chooses a faulty player as L_1 (which occurs with probability at most t/n). Thus, the probability that either L_1 or L_2 is faulty is at most $O(t/n)$, and so $\mathcal{P}'$ has bounded cheaters' edge. ■

Note that the protocol above is in a full-information ex post NE but not in a full-information dominant strategy equilibrium. The reason is that if a player i in C_2 knows that player 1 flipped a 0, but player 2 flipped a 1 (in step 3 of the protocol), then if i is elected as R_2, he will not choose player 2 as L_1, in order to increase the chance that the deciding player chooses a winner from C_2 (which helps i if he is also elected as L_2).

D.3 Selecting a Leader from Those Who Want to Win

A type profile for players that are self-interested is a list of players who want to be elected, and a list of players who want to **not** be elected. Instead of electing a leader uniformly at random, we may wish to elect a leader at random from those who want to be elected. We assume that if nobody wants to be elected, then we elect a leader uniformly at random from all players. We have the following theorem:

Theorem D.3 *Let $\mathcal{P}$ be an n-player, $r(n)$-round leader election protocol in a full-information ex post NE that is resilient to $t(n)$ faults that elects a player uniformly at random. Then there exists a leader election protocol $\mathcal{P}'$ that elects a uniformly random player from those who want to be elected. $\mathcal{P}'$ is also in a full-information ex post NE, it has $r(n)+1$ rounds, and it is resilient to $t(\lfloor n/2 \rfloor)$ faults.*

Proof: The protocol is the following:

1. Players announce their type (i.e. whether or not they want to be elected). Denote by C_1 the set of players who want to be elected, and by C_2 the set of players who do not want to be elected.
2. If $|C_1| \geq |C_2|$, the players in C_1 run the leader election protocol $\mathcal{P}$ and output the winner.
3. If $|C_1| < |C_2|$, the players in C_2 run the leader election protocol $\mathcal{P}$ to choose a representative R. If $C_1 = \emptyset$, R is the elected leader. Otherwise, R chooses a random player from C_1 to be the elected leader.

Note that no player can gain by lying about his type, so the protocol is in a full-information ex post NE. In order for the adversary to force a win, he must win in the run of the leader election protocol $\mathcal{P}$. However, $\mathcal{P}$ is always run on at least $\lfloor n/2 \rfloor$ players, so he succeeds with probability at most some constant $\varepsilon < 1$. ■

E Proof of Theorem 4.6

Theorem E.1 (Theorem 4.6 restated) *For any positive k and $n = 3^k$ there exists an explicit n-player $2\log_3(n)$-round leader election protocol $\mathcal{P}$ that is in a full-information n-tolerant ex post Nash equilibrium with a strictly self-interested adversary. Furthermore, $\mathcal{P}$ is resilient against $n^{\log_3(2)}/2$ faults.*

Proof: Recall the protocol $\mathcal{P}_k$ with $k = 3$ and the uniform distribution. This protocol $\mathcal{P}_3$ is in a full-information ex post Nash equilibrium with a strictly self-interested adversary. To see this, note the two possible cases – either two of the players are adversarial, or one of them is (the cases of zero or three adversarial players are trivial). In the former, a strictly self-interested adversary will always win regardless of the actions of the honest player, and so the honest player may as well stick to the protocol. In the latter, the adversarial player can not increase his chances of winning regardless of his actions, and so by the assumption that the adversary only deviates if he can gain we get that he does not deviate. Since he does not deviate, it is as if all players are honest, and so the protocol is in equilibrium.

We now use $\mathcal{P}_3$ with the uniform distribution $p_1 = p_2 = p_3 = 1/3$ as a building block in our protocol. The protocol $\mathcal{P}$ is the following:

1. In the beginning all players are active.
2. In iteration j, the active players are tripled-off, and each triplet (simultaneously) follows $\mathcal{P}_3$. The winner in each triplet remains active, and the other two are deactivated.
3. The elected leader is the last active player.

$\mathcal{P}$ has $\log_3(n)$ iterations, and thus $2\log_3(n)$ rounds. Note also that $\mathcal{P}$ is in a full-information ex post NE: the only way player i can win is if he wins at every iteration, and he can not increase the probability of this occurring regardless of his strategy. Finally, $\mathcal{P}$ is in a full-information ex post Nash equilibrium with a strictly self-interested adversary: For any triplet and any iteration of $\mathcal{P}$, there can be anywhere between zero and three players of the triplet who are adversarial. However, regardless of which of these is the case, no honest player can increase his chances of winning by deviating (by the argument above).

We now turn to the resilience properties of $\mathcal{P}$. Let c be the number of adversarial players in some instance. After an iteration of the protocol, how many of these players will survive to the next iteration? Note that if a pair of adversarial players are in the same triplet, then exactly one of them will survive with probability 1. If some adversarial player is alone in a triplet, he will survive with probability 1/3. Finally, the adversary has no reason to place three such players in the same triplet.

The adversary has to choose in the beginning which players to corrupt. However, for the analysis of this protocol we actually give him more power: in each iteration i, suppose c_i adversarial players are still alive. Then we allow him to corrupt *any* set of c_i players in this round.

In some iteration i the adversary may choose to place c_i^2 of the players doubled-up in the same triplet, and c_i^1 of the players as lone adversarial players in a triplet, where $c_i = c_i^2 + c_i^1$. Now, for some adversarial player j who survives to the last round, we can trace his route through the protocol: sometimes he was doubled-up with another adversarial player, and sometimes he was a lone player in a triplet. Our idea is to duplicate the player, and allow one duplicate to be doubled-up and the other to be a lone player in each round. We note that the probability that the player survives is at most the probability that one of the duplicates survives.

Since the total number of adversarial players is c, and we duplicate each player, we now have $2c$ adversarial players. Let c^2 be the players who are doubled-up in every round, and c^1 be the lone players in every round, where $c_1^2 = c$ and $c_1^2 = c$. Additionally, let us assume that c^2 is a power of 2, and if not add more adversarial players until this holds (yielding $c^2 < 2c$).

Now, the number of players from c^2 who survive to the last iteration is

$$\frac{c^2}{2^{\log_3 n}} < \frac{2c}{2^{\log_3 n}} = 1$$

when $c < 2^{\log_3(n)-1}$. The probability that any player from c^1 survives to the last round is at most

$$\varepsilon = \frac{1}{3}^{\log_3(n)} \cdot c^1 < \frac{1}{3}^{\log_3(n)} \cdot 2^{\log_3(n)-1}$$

when $c < 2^{\log_3(n)-1}$, where ε is smaller than any positive constant.

Overall, with probability $1-\varepsilon$ only one adversarial player survives to the last round (from c^2), and so the adversary wins with probability at most $(1-\varepsilon)/3$, which is a constant < 1. Thus, the protocol is resilient for $c < 2^{\log_3(n)-1} = n^{\log_3(2)}/2$. ■

F Proofs – Leader Election with Few Rational Coalitions

We begin with the following theorem.

Theorem F.1 (Theorem 4.7 restated) *Every leader election protocol in a full-information ex post NE has a Pareto coalition of two players.*

Proof: The proof is essentially the same as that of Theorem 4.5. Let i be the first with a mixed strategy, let j be the player who can improve i's chance of winning (the adversary in Theorem 4.5). Since j can help i without harming himself, the coalition $\{i, j\}$ is a Pareto coalition. ■

We now prove Theorem 4.8.

Theorem F.2 (Theorem 4.8 restated) *There exists an explicit 2-round leader election protocol in a full-information ex post Nash equilibrium with only 1 stable coalition. The coalition is of size* $n - 1$.

Proof: The protocol is $\mathcal{P}_k$ with $k = n$ and the uniform distribution.

Let $V \subseteq [n]$ be some stable coalition of players. $|V| < n$, since the coalition of all players can not be strong (someone must lose if all others gain). Additionally, $\mathcal{P}_n$ is in a full-information ex post NE, and so $|V| > 1$. Now consider the case in which $n \in V$. Then there exists some player $i \notin V$ such that $(i \mod n-1)+1 \in V$. Since V is stable it must also be strong, and so player n must gain from joining the coalition. The only way he can gain is if at least some player in V chooses n with probability $> 1/n$ in step 2 of $\mathcal{P}_n$. Now observe that if the stable deviation S^*_V of the players requires player n to choose player i with some positive probability in step 1, then player n can gain by deviating and never choosing player i (since player i only chooses n with probability $1/n$). However, if S^*_V does not require player n to choose i with positive probability, then player $(i \mod n-1)+1$ never gets elected, and so the coalition is not strong.

Next consider the case in which $n \notin V$, and suppose $|V| < n-1$. Then again there exists some player $i \in \{1, \ldots, n-1\}$ such that $i \notin V$ and in addition $(i \mod n-1)+1 \in V$. However, since $n, i \notin V$, the probability that $i+1 \mod n-1$ is elected is exactly $1/n$. Thus, V is again not strong.

The only stable coalition is the coalition $V^* = \{1, \ldots, n-1\}$, which has size $n-1$. The stable strategy for this coalition is for all players to never choose player n in step 2 of the protocol. ■

We also have the following theorem, as a weak illustration that we "gained" something by weakening our requirement from strong to stable coalitions.

Theorem F.3 (Theorem 4.9 restated) *For any n-player leader election protocol in a full-information ex post Nash equilibrium, all coalitions of size $n-1$ are strong.*

Proof: We will actually prove something a bit stronger – in any leader election protocol in a full-information ex post Nash equilibrium that obtains any distribution $(p_1, \ldots, p_n)$ over the players, the following holds: for any i such that $p_i \neq 0$, the coalition $V = [n] \setminus \{i\}$ has a coordinated non-rushing strategy such that for every $j \in V$ with $p_j \neq 0$, the coordinated strategy elects j with probability strictly greater than p_j.

The proof is by induction on the number of rounds q of the protocol. For the base case, consider some 1-round n-player leader election protocol $\mathcal{P}$ in a full-information ex post Nash equilibrium, and fix some player i with $p_i \neq 0$ and the coalition $V = [n] \setminus \{i\}$. The properties of the equilibrium imply that regardless of the actions of player i, whether or not he is elected is entirely determined by the players in V. So instead of playing strategies $\bigotimes_{k \in V} S^1_k$, the coalition members can play the coordinated (non-rushing) strategy $\big((S^1_1, \ldots, S^1_{i-1}, S^1_{i+1}, \ldots, S^1_n) | f(\{S^1_k : k \in V\}) \neq i\big)$. Note that the condition $(f(\{S^1_k : k \in V\}) \neq i)$ is well-defined since the players in V determine whether or not $f(\cdot) = i$ or not. If the players follow this strategy, then every player $j \in V$ gets elected with probability $p_j/(1-p_i)$.

Now suppose the claim holds for any $(q-1)$-round leader election protocol, and consider some q-round leader election protocol $\mathcal{P}$. Again fix some player i with $p_i \neq 0$ and the coalition $V = [n] \setminus \{i\}$. Consider some set of messages sent by all players in the first round, say $M = \{M_k^1 : k \in [n]\}$. We can view the leader election protocol $\mathcal{P}$ conditioned on having seen M in the first round as a new $(q-1)$-round leader election protocol $\mathcal{P}^M$, whose distribution over chosen leaders is the initial distribution of $\mathcal{P}$, conditioned on M. By the inductive hypothesis, if i wins with positive probability in $\mathcal{P}^M$, then the players V have a strategy by which they all gain. If i wins with positive probability in $\mathcal{P}^M$ for all $M \in \bigotimes_{k\in[n]} supp(S_k^1(t_k))$, then we are done: players in V play according to the regular strategy in round 1, and according to the strong strategy in every sub-protocol $\mathcal{P}^M$.

Some subtleties arise when for some M, player i wins with probability 0 in $\mathcal{P}^M$. This is because in such a sub-protocol, playing the coordinated strategy does not increase the winning probability of players in V (at it remains $p_j/(1-p_i) = p_j$ for each $j \in V$). It is possible that some player $j \in V$ only has a nonzero chance of winning in sub-protocols in which i has a zero chance of winning. In this case, following the coordinated strategy from round 2 onwards will not increase the chance of j winning, and so is not strong.

For each M and each $j \in V$, suppose that following the coordinated strategy in $\mathcal{P}^M$ increases j's chance of winning by ε_j^M. Let $\varepsilon = \min_{j,M:\varepsilon_j^M \neq 0} \varepsilon_j^M$, and note that ε is the smallest nonzero amount by which any player in V gains in any sub-protocol $\mathcal{P}^M$ by following the coordinated strategy.

Now, recall that the ex post guarantee of $\mathcal{P}$ implies that the probability of player i winning in $\mathcal{P}^M$, say p_i^M, is entirely determined by the players in V. That is, $p_i^M = p_i^{M_V}$, where M_V is the set of messages sent by players V. So we set the coordinated strategy for the players in V in round 1 to be the following. Suppose $M_{V,1}, \ldots, M_{V,\ell}$ are such that $p_i^{M_{V,w}} = 0$, for $w \in [\ell]$. Then players in V choose each $M_{V,w}$ with probability

$$\frac{(1+\delta) \cdot \Pr[M_{V,w}]}{\sum_{y\in[\ell]} \Pr[M_{V,y}]},$$

for

$$\delta = \frac{\varepsilon \cdot \min_{M:\Pr[M]\neq 0} \Pr[M]}{2}.$$

Thus, the players who only win with positive probability when $p_i = 0$ now gain probability δ. The other players in V lose at most δ, which is smaller than half of what they gain, and so they still gain by following this coordinated strategy. ■

G Proof of Theorem 5.2

A useful property of $\mathcal{P}_k$ is the following:

Lemma G.1 *In $\mathcal{P}_k$, the maximal probabilities that a cheating player i can force a player $j \neq i$ to win are the following:*

1. *k can force $i \neq k$ with probability at most $1 - p_k$.*
2. *$i \neq k$ can force i_+ with probability at most $\frac{p_{i_+}}{1-p_k}$.*
3. *$i \neq k$ can force k with probability at most $\frac{p_{i_+}}{1-p_k} + \left(1 - \frac{p_{i_+}}{1-p_k}\right) \cdot p_k$.*

4. $i \neq k$ can force $j \notin \{k, i_+\}$ with probability at most p_j.

Proof:

1. k can always choose i, who then chooses i_+ with probability $1 - p_k$.
2. If i is chosen by k, he can force i_+ in his next choice. This happens with probability $\frac{p_{i_+}}{1-p_k}$. If i is not chosen by k, he can do nothing.
3. If i is chosen by k, he can force k in his next choice. This happens with probability $\frac{p_{i_+}}{1-p_k}$. If $j \neq i$ is chosen by k (with probability $\frac{p_{j_+}}{1-p_k}$), k is chosen by j with probability p_k.

■

Lemma G.2 *Let $A_1, \ldots, A_m$ be m players that are supposed to win with probabilities $a_1, \ldots, a_m$ respectively (where $a_1 + \ldots + a_m = 1$ and all are positive). Suppose players run protocol $\mathcal{P}_m$ where the identities of players in the protocol are chosen at random. Then the maximal probability with which some player $B \in (A_1, \ldots, A_m)$ wins given that $C \in (A_1, \ldots, A_m)$ is cheating and the others are honest is*

$$\max_{p_{c_+}, p_k \in \{a_1, \ldots, a_m\}} \left(\frac{1 - p_a + p_{a_+} + (m-2) \cdot p_b}{m} - \frac{p_b \cdot p_{a_+}}{(1-p_b) \cdot m} + \frac{p_{a_+}}{(1-p_1) \cdot m} \right).$$

Proof: The pair $(C, B) = (A_i, A_j)$ above with probability $m(m-1)$ for each pair $i \neq j \in [m]$. For each such pair, Lemma G.1 gives the maximal probability that C can force B. For $(C, B) = (A_i, A_j)$, denote by $p_C = \Pr[C] = a_i$ and $p_B = \Pr[C] = a_i$. Also, $b_+ = i_+$ and $c_+ = j_+$. To calculate the maximal probability that B wins given a cheating C we just need to average over all $m(m-1)$ cases:

$$\begin{aligned}
\Pr[B \text{ wins}] &\leq \max_{p_{c_+}, p_k} \Bigg(\frac{1}{m(m-1)} \cdot \Bigg((m-1)(1-p_c) + (m-1) \left(\frac{p_{c_+}}{1-p_b} + \left(1 - \frac{p_{c_+}}{1-p_b} \right) \cdot p_b \right) \\
&\quad + (m-1) \frac{p_{c_+}}{1-p_k} + (m-3)(m-1) \cdot p_b \Bigg) \Bigg) \\
&\leq \max_{p_{c_+}, p_k} \left(\frac{1 - p_c + p_{c_+} + (m-2) \cdot p_b}{m} - \frac{p_b \cdot p_{c_+}}{(1-p_b) \cdot m} + \frac{p_{c_+}}{(1-p_k) \cdot m} \right).
\end{aligned}$$

■

We now prove Theorem 5.2.

Proof: The protocol is the following:

1. Each player announces his preferred item. Players are split into m categories $T_1, \ldots, T_m$ according to their announced type.
2. For each i, T_i runs the leader election protocol $\mathcal{P}$ to elect a leader L_i. T_1 runs another leader election protocol $\mathcal{P}$ (in parallel) to elect a second leader L_1'.

3. L'_1 chooses a uniformly random mapping of $(1,\ldots,m)$ to $(A_1,\ldots,A_m)$.

4. $L_1\ldots,L_m$ run $\mathcal{P}_m$, with the ordering $(A_1,\ldots,A_m)$ as chosen above. The protocol is run so that players are elected with probabilities $\frac{|T_i|}{n}$ for each i.

5. The protocol's output is the announced type of the winning player.

We first show that this protocol is in a full-information ex post Nash equilibrium. Fix some player i. Suppose first that in step (1) of the protocol, i announces his type truthfully. Since in the following steps player i can never increase the chance of his preferred type getting chosen, the rest of the protocol is easily seen to be in a full-information ex post Nash equilibrium.

Now consider the case in which player i lies about his type in step (1). Suppose i's true type is B, and he claims it to be C. Then with probability $1-1/|T_C|$, i is not chosen as leader L_C, and so can no longer influence the final choice. In this case, B is chosen with probability $|T_B|/n$. With probability $1/|T_C|$ player i is chosen as L_C, and now wishes to force B in step (4). Lemma G.2 gives a bound on the maximal probability with which he succeeds.

Then overall, the maximal probability that B wins with a cheating i of claimed type C is:

$$\Pr[B \text{ wins}] \leq \max_{p_{c_+},p_k}\left(\left(1-\frac{1}{p_a n}\right)\cdot b+\frac{1}{p_a n}\cdot\left(\frac{1-p_a+p_{a_+}+(m-2)\cdot p_b}{m}-\frac{p_b\cdot p_{a_+}}{(1-p_b)\cdot m}+\frac{p_{a_+}}{(1-p_k)\cdot m}\right)\right)$$

by Lemma G.2. Now, it is an optimal strategy for i to bid truthfully if the above is at most $b+\frac{1}{n}$, since truthful bidding adds $1/n$ to the probability with which B is elected. It is readily verified that this holds as long as each $p_i \in [1/m-\delta, 1/m+\delta]$ for $\delta = 1/10m^2$.

What about resilience? Suppose there is an adversary of size at most $t(\lfloor n(1/m+\delta)\rfloor)$. In order to force an outcome in some predefined set, the adversary must win at least one of the $2m$ runs of the leader election protocol $\mathcal{P}$. Since $\mathcal{P}$ is resilient, the probability that the adversary loses all of them is at least ε^{2m} for some constant $\varepsilon > 0$. Since m is a constant, the adversary loses with constant positive probability. ■

H Random Selection with Other Distributions

In Section 5 we construct protocols in which each item is selected with probability proportional to the fraction of players who (claim to) prefer that item the most. A natural question is whether or not this can be generalized to other distributions. We offer some discussion here about a partial answer to this question. Since this discussion is somewhat informal, we limit ourselves to a generalized version of Theorem 5.2, and specifically to the case of $m = 3$ items.

Suppose the players reveal their types, and suppose the fraction of each type i is α_i. We wish to design a protocol in which type i is selected with probability $p(\alpha_i)$ for each i.

Consider the following protocol. In the last step of the protocol of Theorem 5.2, when the players run $\mathcal{P}_3$, the do so with a slight modification. Instead of running the protocol with probabilities $\alpha_1,\alpha_2,\alpha_3$, they do so with probabilities $p(\alpha_1),p(\alpha_2),p(\alpha_3)$. This selects each item with the desired probability. Furthermore, this protocol has the same resilience as the original protocol of Theorem 5.2.

Is the protocol in a full-information ex post Nash equilibrium? First observe that if a player truthfully reveals his type, then he can no longer influence the probability that his preferred item is selected. It remains to show that truthfully revealing one's type is expected utility maximizing. This is the part that does not always hold.

In Theorem 5.2, truthful revelation of one's type is expected utility maximizing if the potential gain from lying is at most $1/n$, which is the gain in truthfulness. The potential gain from lying is given as a function of α_1, α_2, and α_3. When sampling from a distribution $p(\alpha_i)$, the potential gain from lying is now a function of α_1, α_2, α_3, $p(\alpha_1)$, $p(\alpha_2)$, and $p(\alpha_3)$. Additionally, the gain from truthfulness is no longer $1/n$, but rather $p(\alpha_i + 1/n) - p(\alpha_i)$.

It remains to check that for the desired distribution, the gain from lying is bounded above by the gain from truthfulness. We conjecture that this holds for all functions p satisfying

$$p(\alpha_i) = \frac{g(\alpha_i)}{\sum_j g(\alpha_j)},$$

where g is any concave function.