Fault Tolerance in Large Games

Ronen Gradwohl * Omer Reingold †

Abstract

A Nash equilibrium is an optimal strategy for each player under the assumption that others play according to their respective Nash strategies, but it provides no guarantees in the presence of irrational players or coalitions of colluding players. In fact, no such guarantees exist in general. However, in this paper we show that *large* games are innately fault tolerant. We quantify the ways in which two subclasses of large games – λ-continuous games and anonymous games – are resilient against Byzantine faults (i.e. irrational behavior), coalitions, and asynchronous play. We also show that general large games have some non-trivial resilience against faults.

JEL codes: C72

Keywords: Nash equilibrium, large games, fault tolerance

1 Introduction

Game theory provides a means of studying an interaction between many agents by modeling it as a game. One of the simplest models is a normal-form game, in which all agents play simultaneously and only once. Given such a game, a solution concept – most commonly a Nash equilibrium – is hypothesized to explain or predict the behavior of the agents.

Nowadays, with the advent of the Internet, many of the interactions we wish to model are taking place online, in a distributed setting. Such interactions are characterized by unpredictably asynchronous communication, the occasional presence of faults, and perhaps some exogenous communication between the players. The simple model of a normal-form game is far from accurate in describing this setting. Furthermore, even if it were somehow sufficient, the concept of a Nash equilibrium would provide no guarantees about the optimality of strategies in this fault-ridden environment, and hence would not suffice.

One approach to correcting this deficiency is to incorporate the characteristics of the distributed setting into the model. This, however, requires precise knowledge of the setting to be modeled, and may make the model much more complicated. A second approach that is taken in some recent literature is to quantify the damage caused by the presence of faulty players (see the section Related Work below).

*Corresponding author. Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. E-mail: r-gradwohl@kellogg.northwestern.edu.

†Microsoft Research Silicon Valley and The Weizmann Institute of Science, 1065 La Avenida Mountain View, CA, 94043. E-mail: omer.reingold@microsoft.com.

In this paper we take a different approach, and show that sometimes the model of a normal-form game and the concept of a Nash equilibrium do suffice. More specifically, we show that sometimes the strategies of a Nash equilibrium in a normal-form game remain close to best-responses for each player even if the game is modified to permit asynchronous communication, faulty behavior, and the possibility of collusion.

We study large games – games that involve many agents – because in such games it is particularly relevant to examine the robustness of equilibria against faults. Additionally, in large games it seems plausible that the sheer size of the game renders it robust, whereas in small games such an approach is easily seen to fail. Of course, a large game could be comprised of a small game embedded in a setting with many players, in which case our approach would not work. But then again, such a game is not really a "large" game. Thus, we consider various properties of the game that are reasonable characterizations of large games, and show how these properties relate to the fault tolerance of the game.

Faulty Behavior The guarantee provided by a Nash equilibrium is that each player's strategy is optimal, assuming all others play their designated strategies. One type of faulty behavior consists of Byzantine faults[1], wherein some of the players do not play according to their equilibrium strategies. Perhaps the actions of these faulty players are altered due to an error in communication, perhaps the players are irrational, or perhaps they have some unknown utility. In any case, their actions can be arbitrary or even adversarial.

Two types of resilience we wish to have against Byzantine faults are immunity and tolerance. Immunity means that even if some players fault, the utility of the non-faulty players is not affected by much. Such a property is useful in that it provides the (non-faulty) players a guaranteed utility under the equilibrium strategies, even if some others are faulty. More formally, we say that an equilibrium is (ε, t)-immune if players' expected utilities do not decrease by more than ε when any t other players deviate arbitrarily. Tolerance means that even if some players fault, the original strategies of non-faulty players are still optimal (even conditioned on the actions of the faulty players), although their payoffs may be different from the case in which no players fault. More formally, an equilibrium is (ε, t)-tolerant if players' equilibrium strategies remain best responses (within ε) even if t other players deviate arbitrarily.

Another type of faulty behavior is the potential strategizing enabled by asynchrony. Nash equilibrium strategies are optimal if all players play simultaneously, but suppose some player was delayed, and was then exposed to the chosen actions of some of the other players. This player may then use the information he obtained from the exposed actions to get a better payoff for himself. In this case, his original Nash strategy may no longer be optimal. Resilience to asynchrony means that this does not occur – that a player can **not** improve his payoff by too much, even after he is exposed to the actions of some other players. More formally, a strategy profile is an (ε, p, t)-ex post Nash equilibrium if with probability $1-p$ any player's original strategy is still nearly optimal (up to an additive ε), even after he sees the chosen actions of a set of t players.

A final type of faulty behavior consists of collusion among players. Nash equilibria only guarantee that no player can benefit by a unilateral deviation from his strategy, but imply nothing about deviations by a colluding set of players. Resilience against collusion means that coalitions of players can not improve their payoffs even by a coordinated deviation. More formally, an equilibrium is (ε, t)-coalitional if no player can

[1] Byzantine faults are named after the Byzantine generals' problem of Pease et al. (1982).

gain more than ε even with a joint deviation by himself and $t-1$ other players.

Our Results The first subclass of games we examine is one that has received much attention in the economics literature (see, for example, Kalai (2004), Azrieli and Shmaya (2010), and the references therein). This subclass consists of λ-continuous games, and it turns out that it is particularly well-suited to our demands of fault tolerance. Roughly, in such games no player is influenced too much by the other players. More precisely, if a coalition consisting of a δ-fraction of the players changes their actions, then this can change the utility of a non-member by at most $\lambda \cdot \delta$. For games that are λ-continuous, we observe that all Nash equilibria (ε, t)-tolerant, (ε, t)-immune, and (ε, t)-coalitional for $t = O(n)$ (assuming λ and ε are arbitrary constants). They are also resilient to asynchronous exposure to a large number of players – always to at least $O(n)$ players, and sometimes to all $n-1$ other players.

λ-continuous games have very strong fault tolerance, but perhaps the restriction on the game is too strong. More precisely, in these games a player's utility given **any** profile of the others' actions changes little if a small fraction of players changes their action. For fault tolerance, however, we are really only interested in what happens when the actions of (non-faulty) players come from specific distributions – namely, the strategies of a Nash equilibrium.

The next natural subclass of games we study is that of anonymous games, in which the utility of each player is a function of his own action and the empirical distribution of other players' actions. Thus, in anonymous games a player does not care which player performed each action, but only how many performed each action. Anonymous games are widely studied in economics. Note that anonymous games can be very far from being λ-continuous: For example, if a player's utility is determined by the majority of the other players' binary actions, then a change of even 1 player's action can flip the utility from 1 to 0.

We show that every anonymous game has an equilibrium that is (ε, t)-tolerant, (ε, t)-immune, and (ε, t)-coalitional, for $t = O(\sqrt{n})$. Additionally, the equilibrium is an (ε, p, t')-ex post Nash equilibrium for $t' = O(n)$. In order to prove this theorem we actually show that all equilibria that are "mixed enough" – in which players play every action with at least some minimal probability – have these fault tolerance guarantees. Section 5.2 contains some discussion about the applicability and naturalness of this assumption, as well examples demonstrating its necessity.

Finally, for general large games, we give an example in which there can be no fault tolerance, even to one Byzantine fault (see Section 6). However, not all is lost. We show that there exists an equilibrium for which the following holds: If the faulting players are chosen randomly (but their actions remain adversarial), then the strategies provided by the equilibrium remain near best responses for most players. We give a reduction from the case of arbitrary games with randomly-chosen faulting players to anonymous games with arbitrarily-chosen faulting players. Using this technique we show that the number of faulting players can be as large as the number of faulting players in any anonymous game – that is, $t = O(\sqrt{n})$ for immunity and tolerance and $t' = O(n)$ for asynchronous exposure.

Our results are summarized in Table 1, together with some previously known, related results (see the Related Work section for further details). We note that our results are optimal in the following sense: The restrictions placed on the games, the equilibria, and the number of faulting players are pretty much necessary.

Table 1: Summary of Results – Maximal t for Constant $\varepsilon > 0$

Class of Games	Immunity	Tolerance	Coalition Size	Asynchronous Exposure (Constant $p > 0$)
λ-Continuous	$O(n)$	$O(n)$	$O(n)$	$O(n)$ $n-1$ if also anonymous: Kalai (2004)
Anonymous	$O(\sqrt{n})$	$O(\sqrt{n})$	$O(\sqrt{n})$	$O(n)$
General, but randomly chosen t	$O(\sqrt{n})$	$O(\sqrt{n})$	$O(\sqrt{(n)})$	$O(n)$: Gradwohl and Reingold (2010)

It is not possible to weaken them by much, and we describe some games that serve as counter-examples.

The techniques we utilize for the different classes of games are different but related. We first define a notion of stability of equilibria in games: Roughly, a strategy profile is (ε, t)-stable if for any fixed action of any player, his expected utility does not change much even if t arbitrary other players deviate. We demonstrate the usefulness of stability by showing that it implies immunity, tolerance, and resilience to collusion. To prove the results on fault tolerance of various classes of games we then typically demonstrate stability, from which the other properties follow.

Proving stability is different in each class of games. For λ-continuous games, stability follows almost directly from the definition. For anonymous games, however, stability follows on the fact that utilities depend on the number of times each action was played. Since players' strategies are independent, a law of large numbers implies that the probabilities an action was played k times or $k+t$ times are close (when t is on the order of $\sqrt{n}$). Since probabilities are close, expectations will also be close. Finally, for general games the results follow from a reduction of the game to an anonymous game, a reduction that preserves a notion of "stability on average."

Related Work The original motivation for this work arose from the recent paper of Kalai (2004), in which he shows that in large normal-form games that are anonymous and λ-continuous, the Nash equilibria are extensively robust. This means that even when the game is altered in such a way as to allow sequential play, the possibility of revision, and communication between the players, the original Nash equilibria of the game survive. In particular, the Nash equilibria of such games are ex post Nash: A player's strategy remains a best response even after he sees the chosen actions of all the other players.

In an attempt to generalize the work of Kalai, we (Gradwohl and Reingold, 2010) introduced the notion of partial exposure. In that paper we show that in **any** large game, players' Nash strategies remain nearly optimal even after they are exposed to the chosen actions of *some* of the other players.

These properties of ex post Nash and partial exposure can be seen as a form of resilience to asynchrony, since players who play later are exposed to the chosen actions of those who played earlier. The resilience holds since the properties guarantee that players are not too motivated to change their actions despite the new information. The current paper expands on these ideas, and places this resilience in the broader context of fault tolerance.

Fault tolerance is one of the major topics of study in distributed computing, and much work in this field has focused on designing multi-agent protocols that are resilient against various types of faults. Game theory, however, is relatively unconcerned with these issues. In recent survey articles, Halpern (2003, 2008) has argued that perhaps these issues should be studied more extensively in a game theoretic framework.

Notions of resilience against Byzantine faults have appeared in a few implementation frameworks. Eliaz (2002) shows how to implement a mechanism in such a way that if some players fault, the other players' strategies are still best responses; this is the notion we call tolerance. Aiyer et al. (2005) implement a protocol in an asynchronous setting that is resilient against Byzantine, altruistic, and rational behavior. Chen et al. (2010) and Chen and Micali (2012) design auctions when some players may collude. Additionally, the burgeoning field of rational cryptography (see Nielsen (2007) for a survey) is concerned with the implementation of protocols that are secure against both rational and Byzantine players. Finally, some recent work of Abraham et al. (2006, 2008) on the implementation of mediators has considered solution concepts that take into account the possibility of irrational play by some of the players.

Questions on how to deal with asynchrony have also come up in game theory. Specifically, the works of Monderer and Tennenholtz (1999a, 1999b) present a framework that incorporates game theoretic analysis in a distributed setting.

Solution concepts in game theory that deal with deviations by coalitions have a longer history. The first work is perhaps that of Aumann (1959) – in his solution concept, there can be no coalitions of players in which all players benefit from a deviation. Bernheim et al. (1989) present the notion of a *coalition-proof Nash equilibrium*, in which there can be no deviations in which all players improve and in which there are no further deviations by a subset of the coalition from the deviation. Moreno and Wooders (1996) extend this work to allow correlated strategies. However, one of the main differences between these papers and the current one is that our notion of robustness agains collusion holds with a bound on the number of colluding players. In the papers above no coalition is excluded due to its size. In their work on implementing mediators, Abraham et al. (2006, 2008) do consider solution concepts in which some (bounded subset of) players may coordinate a deviation. Their notion requires that no such deviation will benefit any player in the coalition. However, one of the problems with many of the notions of stability against collusion is that in general, a game need not have such a coalition-proof equilibrium (as opposed to a Nash equilibrium, which always exists).

In addition, several recent papers quantify the difference between a regular equilibrium and one that is resilient against some adversarial players. Specifically, Moscibroda et al. (2006) study the so-called *price of malice* in a virus inoculation game, and Babaioff et al. (2007) and Roth (2008) study a different notion with the same name in congestion games.

Finally, there is some additional recent work on large games that is tangentially related to this paper, such as Azevedo and Budish (2012), Deb and Kalai (2010), and Kash et al. (2011).

2 Definitions

For any natural number k we denote by $[k] = \{0, 1, \ldots, k\}$. We use the standard definition of a normal-form game (c.f. Osborne and Rubinstein, 1994):

Definition 2.1 (game) *A game G is described by a triple $G = (N, A, u)$ as follows:*

- *$N = \{1, \ldots, n\}$ is the set of players.*
- *A is a finite set of player actions, and without loss of generality $A = \{0, 1, \ldots, m\}$.*
- *$u = (u_1, \ldots, u_n)$ is the vector of payoff functions. For every i, the payoff of player i is $u_i : A^n \mapsto [0, 1]$. We will also slightly abuse notation, and denote by $u_i(\mathbf{A}) = \mathrm{E}[u_i(\mathbf{A})]$, where $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_n)$ is a random variable over A^n.*

We also use the following standard notation. For a subset $I \subseteq N$ and some vector $X = X_1 \times \ldots \times X_n$, let $X_I = \bigotimes_{i \in I} X_i$, and $X_{-I} = \bigotimes_{i \notin I} X_i$. Furthermore, $\mathbf{A}_{-i} : \mathbf{A}'_i = \mathbf{A}_1 \times \ldots \times \mathbf{A}_{i-1} \times \mathbf{A}'_i \times \mathbf{A}_{i+1} \times \ldots \times \mathbf{A}_n$.

The most common solution concept for such games is a Nash equilibrium.

Definition 2.2 (Nash equilibrium) *A product distribution $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ over A^n is a δ-*Nash equilibrium *if for all i and all $\mathbf{A}'_i$,*

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-i} : \mathbf{A}'_i) - \delta.$$

It is a Nash equilibrium *if this holds for $\delta = 0$.*

We now define a property of distributions.

Definition 2.3 (ε-uniform) *A distribution $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_n)$ is ε-*uniform *if for all i and all $a \in A$ it holds that $\Pr[\mathbf{A}_i = a] \geq \varepsilon$.*

Suppose $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ are the strategies of a Nash equilibrium. If $\mathbf{A}$ is ε-uniform, then no player i has a pure strategy (unless $|A| = 1$).

We now define some properties of functions. In the following, let $f : A^n \mapsto [0, 1]$.

Definition 2.4 (λ-continuous) *f is λ-*continuous *if for all $x \in A^n$ and $y \in A^n$ it holds that*

$$|f(x) - f(y)| \leq \lambda \cdot \Delta(x, y),$$

where

$$\Delta(x, y) = \frac{|\{i : x_i \neq y_i\}|}{n}.$$

For a player i and an action $\overline{a} \in A$, define $u_i^{\overline{a}} : A^{n-1} \mapsto [0, 1]$ as

$$u_i^{\overline{a}}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = u_i(x_1, \ldots, x_{i-1}, \overline{a}, x_{i+1}, \ldots, x_n).$$

Then a game G is λ-continuous if for every $i \in N$ and every $\overline{a} \in A$ the function $u_i^{\overline{a}}$ is λ-continuous.

Definition 2.5 (anonymous) *f is* anonymous *if $f(x) = f(y)$ whenever*

$$|\{i : x_i = \overline{a}\}| = |\{i : y_i = \overline{a}\}|$$

for all $\overline{a} \in A$. This occurs if and only if

$$f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

for all permutations $\sigma : N \mapsto N$. For $z \in N^m$ we often write $f(z_1, \ldots, z_m)$ instead of $f(x)$ when

$$z_j = |\{i : x_i = j\}|.$$

As a convention, we do not write the number of 0's – it can be calculated from the number of elements of all other values.

A game G is anonymous if for every i and every $\overline{a} \in A$ the function $u_i^{\overline{a}}$ is anonymous.

3 Faulty Behavior

In this section we describe the various types of faulty behavior in which we are interested, and prove some lemmas about them.

3.1 Byzantine Faults

When a Byzantine fault occurs, the player who faulted plays an arbitrary action. His action is not rational, and does not obey any known distribution. The only knowledge players have about such faults is that the number of players that fault is at most some t. There are two possible properties we may want: immunity and tolerance. Immunity means that even if some players fault, the utilities of the non-faulty players are not affected by much. Tolerance means that even if some players fault, the original strategies of non-faulty players are still optimal, although their payoffs may be different from the case in which no players fault.

Definition 3.1 (immunity) *In a game G, a strategy profile $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is (ε, t)-*immune *if for every set $S \subset N$ of size $|S| \leq t$, every $b_S \in A^{|S|}$, and every player $i \notin S$,*

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : b_S) - \varepsilon.$$

Definition 3.2 (tolerance) *In a game G, a strategy profile $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is (ε, t)-*tolerant *if for every set $S \subset N$ of size $|S| \leq t$, every $b_S \in A^{|S|}$, every player $i \notin S$, and every strategy $\mathbf{A}'_i$,*

$$u_i(\mathbf{A}_{-S} : b_S) \geq u_i(\mathbf{A}_{-(S \cup \{i\})} : b_S, \mathbf{A}'_i) - \varepsilon.$$

Note that if a profile is (ε, t)-tolerant for some $t \geq 0$ then that profile is also an ε-Nash equilibrium.

The notions of immunity and tolerance serve two purposes and are very distinct properties. However, the following definition provides a sufficient condition for both.

Definition 3.3 (stable) *In a game G, a strategy profile $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is (ε, t)-*stable *if for every set $S \subset N$ of size $|S| \leq t$, every $b_S \in A^{|S|}$, every player $i \notin S$, and every action $\overline{a} \in A$,*

$$\left| u_i(\mathbf{A}_{-i} : \overline{a}) - u_i(\mathbf{A}_{-(S \cup \{i\})} : b_S, \overline{a}) \right| \leq \varepsilon.$$

The following lemma shows that in equilibrium stability is sufficient for both immunity and tolerance:

Lemma 3.4 *Suppose that in a game G, a δ-Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is (ε, t)-stable. Then $\mathbf{A}$ is (ε, t)-immune and $(2\varepsilon + \delta, t)$-tolerant.*

Proof: Immunity follows almost directly. Now consider some alternate strategy $\mathbf{A}'_i$ for player i. Then

$$
\begin{aligned}
&u_i(\mathbf{A}_{-(S\cup\{i\})} : \mathbf{A}'_i, b_S) \\
&\quad = \sum_{\overline{a}\in A} u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \overline{a}) \cdot \Pr\left[\mathbf{A}'_i = \overline{a}\right] \\
&\quad \leq \sum_{\overline{a}\in A} \left(u_i(\mathbf{A}_{-i} : \overline{a}) + \varepsilon\right) \cdot \Pr\left[\mathbf{A}'_i = \overline{a}\right] && (1) \\
&\quad = \left(\sum_{\overline{a}\in A} u_i(\mathbf{A}_{-i} : \overline{a}) \cdot \Pr\left[\mathbf{A}'_i = \overline{a}\right]\right) + \varepsilon \\
&\quad = u_i(\mathbf{A}_{-i} : \mathbf{A}'_i) + \varepsilon \\
&\quad \leq u_i(\mathbf{A}) + \varepsilon + \delta && (2) \\
&\quad = \left(\sum_{\overline{a}\in A} u_i(\mathbf{A}_{-i} : \overline{a}) \cdot \Pr\left[\mathbf{A}_i = \overline{a}\right]\right) + \varepsilon + \delta \\
&\quad \leq \left(\sum_{\overline{a}\in A} \big(u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \overline{a}) + \varepsilon\big) \cdot \Pr\left[\mathbf{A}_i = \overline{a}\right]\right) + \varepsilon + \delta && (3) \\
&\quad = u_i(\mathbf{A}_{-S} : b_S) + 2\varepsilon + \delta,
\end{aligned}
$$

where (1) and (3) follow from (ε, t)-stability and (2) holds since $\mathbf{A}$ is a δ-Nash equilibrium (and so $\mathbf{A}_i$ is a better strategy than $\mathbf{A}'_i$). Thus, $\mathbf{A}$ is $(2\varepsilon + \delta, t)$-tolerant. ■

3.2 Coalitions

Another type of fault occurs when sets of players collude in an attempt to increase their payoffs.

Definition 3.5 (coalitional Nash equilibrium) *In a game G, a strategy profile $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is an* (ε, t)-coalitional Nash equilibrium *if for every set $S \subset N$ of size $|S| \leq t$, every strategy $\mathbf{A}'_S$, and every player $i \in S$,*

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : \mathbf{A}'_S) - \varepsilon.$$

Observe that if a strategy profile is an (ε, t)-coalitional Nash equilibrium then it is also an ε-Nash equilibrium.

There are major differences between the notions of coalitional Nash equilibrium, immunity, and tolerance: For example, in a coalitional Nash equilibrium, the player i whose utility does not change is a member of the coalition S. In the case of immunity, on the other hand, i is **not** a member of S. However, the three notions are closely related, and in fact immunity and tolerance imply resilience against coalitions.

Lemma 3.6 *Suppose that in a game G, a strategy profile $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is (ε, t)-immune and (ε', t)-tolerant. Then $\mathbf{A}$ is an $(\varepsilon + \varepsilon', t + 1)$-coalitional Nash equilibrium.*

Proof: Fix a player i, and consider some subset $S \subset N \setminus \{i\}$ of size $|S| \leq t$. By immunity, we know that for every $b_S \in A^{|S|}$,

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-S} : b_S) - \varepsilon.$$

By tolerance we know that for every strategy $\mathbf{A}'_i$ of player i,

$$u_i(\mathbf{A}_{-S} : b_S) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : \mathbf{A}'_i, b_S) - \varepsilon'.$$

Putting the two together implies that

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \mathbf{A}'_i) - \varepsilon - \varepsilon'.$$

Since this holds for every $b_S \in A^{|S|}$, it also holds that

$$u_i(\mathbf{A}) \geq u_i(\mathbf{A}_{-(S\cup\{i\})} : \mathbf{A}'_{(S\cup\{i\})}) - \varepsilon - \varepsilon'$$

for every strategy $\mathbf{A}'_{(S\cup\{i\})}$ of the players in $S \cup \{i\}$. ■

3.3 Asynchrony

If players do not play simultaneously, then perhaps they are exposed to the chosen actions of other players. They may be able to use this information to their advantage. Note that the difference between our notion of stability under asynchrony and that of tolerance is that in the former the actions of the faulting set S are sampled from the distribution $\mathbf{A}_S$ (and so the probability is over this sample as well), whereas in the latter the actions are chosen in a worst-case manner.

Definition 3.7 ((ε, p, t)-ex post Nash equilibrium) *In a game G, a strategy profile $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is an (ε, p, t)-*ex post Nash equilibrium *if for every set $S \subset N$ of size $|S| \leq t$, and every player $i \notin S$,*

$$\Pr_{a_S \sim \mathbf{A}_S}\left[\exists \mathbf{A}'_i \text{ s.t. } \left|u_i(\mathbf{A}_{-S} : a_S) - u_i(\mathbf{A}_{-(S\cup\{i\})} : a_S, \mathbf{A}'_i)\right| > \varepsilon\right] \leq p.$$

If the equilibrium of a game is an (ε, p, t)-ex post Nash equilibrium, then the original strategy of player i is nearly optimal (up to the additive ε) with probability $1 - p$ even after he sees the chosen actions of players in any set S of size at most t.

Note that if an equilibrium is (ε, t)-tolerant then it is also an $(\varepsilon, 0, t)$-ex post Nash equilibrium. However, by considering a positive p it is possible to obtain this kind of robustness for a larger number of faults: In particular, it is possible to obtain equilibria that are (ε, p, t)-ex post Nash equilibrium that are not $(\varepsilon, 0, t)$-ex post Nash equilibria.

4 λ-Continuous Games

It turns out that λ-continuous games are strongly fault-tolerant, as is shown in the following theorem. Note that we make no assumption on $|A|$ (except that it is discrete).

Theorem 4.1 *Let G be an n-player λ-continuous game and fix a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ of G. Then for any $t \in [n]$ the equilibrium $\mathbf{A}$ is $(\lambda t/n, t)$-immune, $(2\lambda t/n, t)$-tolerant, and $(3\lambda t/n, t+1)$-coalitional.*

Proof: We will show that that $\mathbf{A}$ is $(\lambda t/n, t)$-stable: Namely, that for every $\overline{a} \in A$,

$$\left|u_i(\mathbf{A}_{-i}, \overline{a}) - u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \overline{a})\right| \leq \frac{\lambda t}{n}.$$

Fix some player i and subset $S \subset N \setminus \{i\}$ of size $|S| \leq t$. Also, fix some $b_S \in A^{|S|}$. For every $\overline{a} \in A$,

$$\begin{aligned} u_i^{\overline{a}}(\mathbf{A}_{-i}) &= \sum_{a \in A^{i-1}\times\{\overline{a}\}\times A^{n-i}} u_i(a) \Pr\left[\mathbf{A}_{-i} = a_{-i}\right] \\ &\geq \sum_a \left(u_i(a_{-S} : b_S) \Pr\left[\mathbf{A}_{-i} = a_{-i}\right]\right) - \frac{\lambda t}{n} \\ &= u_i^{\overline{a}}(\mathbf{A}_{-(S\cup\{i\})} : b_S) - \frac{\lambda t}{n}, \end{aligned}$$

where the inequality follows from λ-continuity. A symmetric argument shows that

$$u_i^{\overline{a}}(\mathbf{A}_{-i}) \leq u_i^{\overline{a}}(\mathbf{A}_{-(S\cup\{i\})} : b_S) + \frac{\lambda t}{n}.$$

Thus, we have that for every $\overline{a} \in A$,

$$\left|u_i(\mathbf{A}_{-i}, \overline{a}) - u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \overline{a})\right| \leq \frac{\lambda t}{n},$$

and so $\mathbf{A}$ is $(\lambda t/n, t)$-stable. Immunity and tolerance follow using Lemma 3.4, and Lemma 3.6 implies the resilience against coalitions. ■

The resilience of equilibria in λ-continuous games to asynchrony is no stronger than that guaranteed by tolerance. However, if we further assume that the game is also anonymous, then a much stronger guarantee can be given. More specifically:

Anonymous λ-continuous games A game G is anonymous and λ-continuous if for every player i and every action $\overline{a} \in A$ the function $u_i^{\overline{a}}$ is anonymous and λ-continuous. Such games have the same immunity, tolerance, and coalitional resilience properties as general λ-continuous games. Their advantage, however, lies in their nearly optimal resilience to asynchrony – no player has much incentive to change his strategy even if he sees the chosen actions of **all** other players. This is shown by the following theorem of Kalai (2004).

Theorem 4.2 (Kalai, 2004) *Let G be an anonymous, λ-continuous n-player game, and let $t \in [n]$. Then every Nash equilibrium of G is*

$$\left(\frac{\lambda t|A|}{n}, 2|A| \cdot e^{-\frac{2t^2}{n}}, n-1\right)\text{-ex post.}$$

5 Anonymous Games

We begin this section with statements of our main results on anonymous games, then provide some intuition, and conclude with a discussion.

Theorem 5.1 *For every natural number m and real $\varepsilon > 0$ there exists a constant $C_{5.1}$ for which the following holds. Let G be an anonymous n-player game with $A = \{0, \ldots, m\}$, and fix $t = C_{5.1}\sqrt{n}$. Then G has an ε-Nash equilibrium that is (ε, t)-immune, $(3\varepsilon, t)$-tolerant, and $(4\varepsilon, t+1)$-coalitional.*

Theorem 5.1 relies on the following, which is in itself an interesting indication of fault tolerance in large games (see discussion in Section 5.2).

Theorem 5.2 *For every natural number m and reals $\varepsilon > 0$ and $\alpha > 0$ there exists a constant $C_{5.2}$ for which the following holds.*[2] *Let G be an anonymous n-player game with $A = \{0, \dots, m\}$, let $\mathbf{A}$ be an ε-uniform δ-Nash equilibrium of G, and fix $t = C_{5.2}\sqrt{n}$. Then $\mathbf{A}$ is (α, t)-immune, $(2\alpha + \delta, t)$-tolerant, and $(3\alpha + \delta, t + 1)$-coalitional.*

To obtain Theorem 5.1 from Theorem 5.2, we show that in every game there always exists at least one ε-uniform δ-Nash equilibrium.

Theorem 5.2 relies on a generalization of the following lemma to a larger action space (the generalization appears as Lemma A.5). The idea is that when considering independent Bernoulli random variables, the probability that their sum is x is close to the probability that their sum is $x + t$ (when x is close to the expectation and t is smaller than the standard deviation).

Lemma 5.3 *There exists a universal constant $C_{5.3}$ for which the following holds. Fix any n and any independent Bernoulli random variables $\mathbf{X}_1, \dots, \mathbf{X}_n$ with (possibly different) parameters $p_i \in (0, 1)$. Let $\mathbf{X}$ be the sum of the $\mathbf{X}_i$'s, and let $\sigma^2 = \mathrm{Var}[\mathbf{X}]$. Then the following holds for any $\delta > 0$ and integer t:*

$$|\Pr[\mathbf{X} = x] - \Pr[\mathbf{X} = x + t]| \leq \delta$$

with probability at least

$$1 - \frac{C_{5.3}\,|t|}{\delta\sigma^2}$$

over the choice of x from $\mathbf{X}$.

Since we are studying anonymous games, all utility functions are anonymous, and so their values are determined by the sum of the inputs. Now, if there are t Byzantine faults in such a game, then the sum is off-set by at most t. Intuitively, however, since the probability of the sum being x and $x + t$ is roughly the same, these faults should not alter the expectation of the function by too much.

5.1 Resilience to Asynchrony

Theorem 5.4 *For every natural number m and reals $p > 0$ and $\varepsilon > \delta > 0$ there exists a constant $C_{5.4}$ for which the following holds. Let G be an anonymous n-player game with $A = \{0, \dots, m\}$, let $k(n)$ be some positive and nondecreasing function of n, and fix $t = C_{5.4} \cdot n/k(n)$. Then G has a δ-Nash equilibrium that is $(\varepsilon, pe^{-k(n)}, t)$-ex post.*

Note that if we set $k(n) \equiv 1$ then we get that the equilibrium is (ε, p, t)-ex post with $\Omega(n)$ faulting players. However, if we reduce the number of faulting players to $\Theta(n/\log n)$ by setting $k(n) = \log n$, then a union bound implies that with probability $1 - p$ **no** player can increase his payoff by more than ε after observing the realized actions of t others.

[2] $C_{5.2}$ depends polynomially on $1/\varepsilon$ and $1/\alpha$. However, we do not attempt to optimize this dependence in the present paper.

5.2 ε-Uniformity

Theorem 4.1 shows that *all* Nash equilibria of λ-continuous games are fault tolerant. The corresponding theorem for anonymous games, namely Theorem 5.1, only guarantees the existence of *some* fault tolerant Nash equilibrium. While Theorem 5.2 provides such a guarantee for all equilibria, it does so only for those equilibria that are mixed enough, specifically the ones that are ε-uniform. In this section we argue that ε-uniformity is necessary for fault tolerance of equilibria in anonymous games. We then show that in some cases it is a reasonable assumption.

We begin with some examples of anonymous games with Nash equilibria that are **not** ε-uniform, and show that no fault tolerance is possible:

Example 5.5 In the agreement game, every player has utility 1 if all players output 1, and 0 otherwise. One Nash equilibrium here is when every player outputs 1 with probability 1.

Note that this equilibrium is not 1-immune: If one player faults and outputs 0 instead of 1, then all players get 0 instead of 1.

Example 5.6 In the minority game, players output 0 or 1, and win if and only if their action matches the minority of other players' actions. If n is odd, then a Nash equilibrium of this game is when one player flips a fair coin, half of the others always output 1, and the remaining players always output 0.

This equilibrium is not 2-tolerant: If two players that always output 0 fault and output 1, then it is beneficial for a player who is supposed to always output 1 to output 0. In this case his Nash strategy is very far from optimal.

Our conclusion from these examples is that we can not expect fault tolerance in anonymous games without some restriction on the Nash equilibria. We now describe some situations in which it is reasonable to impose the assumption of ε-uniformity.

First, consider a situation in which a game is played over a noisy channel. Suppose players transmit their actions, but know that there is some small probability ε that their action will chosen at random instead of being correctly communicated. Here all strategies of players are ε-uniform. It is also not difficult to show that there always exists a Nash equilibrium in this situation: Conditioned on all players playing an ε-uniform strategy, there exists an ε-uniform best-response for all players. Consider a new game in which the set of actions of all players is the set of all ε-uniform strategies of the original game. Since this new action space is compact, there exists a Nash equilibrium. Note that an equilibrium in the new game consists of ε-uniform strategies over the original actions.

Second, adding some random noise can be a very easy way to make a mechanism fault tolerant. Suppose we implement a mechanism in a Nash equilibrium. Then we can add random noise into the mechanism by changing each player's actions with probability ε. In this case, the equilibrium will be ε-uniform and thus fault tolerant. Note that for some ranges of parameters ε can be as small as $O(1/n)$. In this case we are making the mechanism fault tolerant at the cost of only a constant number of random errors (on expectation).

Finally, we note that for anonymous games, the requirement of ε-uniformity can be relaxed. All we really require is that for every action $\overline{a} \in A$, there are enough players who output $\overline{a}$ with probability between ε and $1-\varepsilon$.

6 General Games

In general, it is impossible to show the same kind of robustness to Byzantine faults for general games, even if we assume the equilibrium is ε-uniform. The same holds for resilience to asynchrony (see the examples in Section 6.1). However, if we make some assumptions about how faulty or exposed players are chosen, then there is a sort of innate fault tolerance even in general large games.

We begin with a result on resilience to asynchrony, which is a theorem of Gradwohl and Reingold (2010) on partial exposure.

Definition 6.1 (random t-exposure (ε, p)-ex post) *For each $i \in N$, let $\mathbf{R}_i$ be a uniformly random subset $R_i \subset N \setminus \{i\}$ of size $|R_i| = t$. Then a Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is* (ε, p)-ex post Nash with random t-exposure *if for every i and every strategy $\mathbf{A}'_i$ (which may be a function the exposed values $\mathbf{A}_{\mathbf{R}_i}$):*

$$\Pr_{\substack{r\sim\mathbf{R}_i \\ a_r\sim\mathbf{A}_r}} \left[u_i(\mathbf{A}^{-i} : \mathbf{A}'_i(a_r)|\mathbf{A}_r = a_r) - u_i(\mathbf{A}|\mathbf{A}_r = a_r) > \varepsilon\right] < p.$$

In words, the definition states that when a player sees the actions of a random subset of t players, the probability that he will be able to improve his payoff by more than ε is at most p (over the choice of the set R_i and the realized actions of players in R_i). The following theorem shows that t can be rather large:

Theorem 6.2 (Gradwohl and Reingold, 2010) *In any n-player game G, for any $p, \varepsilon > 0$, and positive integer t, every Nash equilibrium $\mathbf{A} = \mathbf{A}_1 \times \ldots \times \mathbf{A}_n$ is*

$$\left(p + \varepsilon, \frac{4t \cdot |A|}{np\varepsilon^2}\right)\textit{-ex post Nash with random } t\textit{-exposure.}$$

Note that for $\mathbf{A}$ to be (ε, p)-ex post Nash with random t-exposure we can take

$$t = \Omega\left(\frac{p^2\varepsilon^2 \cdot n}{|A|}\right).$$

If $|A|$, p and ε are constants, then t can be as large as a constant fraction of all players.

This property of ex post Nash is a weaker form of resilience against asynchrony than what we have for λ-continuous or anonymous games. In particular, Theorem 6.2 implies that "most" players (on expectation, $(1-p)n$) will not be able to improve their expected payoffs by more than ε even after being exposed to the actions of a random subset of t players. In general games it is impossible to obtain fault tolerance for all players simultaneously – see Example 6.4 below.

We now show that general games also exhibit similarly weakened forms of immunity and tolerance.

Theorem 6.3 *For every natural number m and reals $\varepsilon > 0$ and $\delta > 0$ there exists a constant $C_{6.3}$ for which the following holds. Let G be an n-player game with $A = \{0, \ldots, m\}$, and fix $t = C_{6.3}\sqrt{n}$. Then G has a δ-Nash equilibrium $\mathbf{A}$ that satisfies the following: For any player i and positive k,*

$$\Pr\left[\min_{b_S \in A^t} \{u_i(\mathbf{A}_{-S} : b_S)\} < u_i(\mathbf{A}) - k\varepsilon\right] < \frac{|A|}{k}.$$

Also, for any strategy $\mathbf{A}'_i = \mathbf{A}'_i(S, b_S)$ of player i,

$$\Pr\left[\max_{b_S \in A^t} \left\{u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \mathbf{A}'_i) - u_i(\mathbf{A}_{-S} : b_S)\right\} > 2k\varepsilon + \delta\right] < \frac{2|A|}{k}.$$

In the statements above, the probability is over a uniform choice of S of size $|S| = t$ from $N \setminus \{i\}$.

Theorem 6.3 states that if a set of t faulting players is chosen uniformly at random, then they do not decrease a player's utility by more than $k\varepsilon$ with probability more than $|A|/k$. Additionally, the same holds for tolerance – for any player, his strategy is a near best response (up to the additive $2k\varepsilon+\delta$) with probability at least $1 - |A|/k$ even conditioned on the faults of t randomly chosen players.

6.1 Impossibility of Fault Tolerance with Arbitrary Faulting Players

In this section we give an example that shows that it is impossible to have fault tolerance if the faulting players are chosen arbitrarily, as opposed to randomly. It also shows that it is impossible to obtain 1-tolerance or resilience to asynchrony for all players simultaneously (even if the faulting players are chosen randomly):

Example 6.4 In the copycat game, each player i wins if and only if his action matches the action of player $i \pmod{n} + 1$. There are 3 Nash equilibria in this game: All players play 1, all players play 0, and every player flips a fair coin.

Note that the pure equilibria are neither immune nor tolerant, and the mixed equilibrium is not tolerant. So if a faulting player j is chosen arbitrarily as opposed to randomly, then no equilibrium is 1-tolerant for player $j + n - 2 \pmod{n} + 1$. This also implies that, even if the faulting players are chosen randomly, then regardless of the realization of the faulty set of players *some* player will no longer wish to play his prescribed equilibrium strategy.

Acknowledgements

This research was supported in part by grant 1300/05 from the Israel Science Foundation. The authors would also like to thank Ariel Yadin for helpful discussions throughout the course of this research, and the anonymous referees for their comments.

Appendix

A Proofs for Anonymous Games

A.1 Lemma 5.3

Lemma A.1 (Lemma 5.3 restated) *There exists a universal constant $C_{5.3}$ for which the following holds. Fix any n and any independent Bernoulli random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with (possibly different) parameters $p_i \in (0,1)$. Let $\mathbf{X}$ be the sum of the $\mathbf{X}_i$'s, and let $\sigma^2 = \mathrm{Var}[\mathbf{X}]$. Then the following holds for any $\delta > 0$ and integer t:*

$$|\Pr[\mathbf{X} = x] - \Pr[\mathbf{X} = x + t]| \leq \delta$$

with probability at least

$$1 - \frac{C_{5.3}\,|t|}{\delta\sigma^2}$$

over the choice of x from $\mathbf{X}$.

The proof of Lemma 5.3 utilizes the following theorem of Lévy (1937, page 88).

Theorem A.2 (Lévy, 1937) *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent Bernoulli random variables with (possibly different) parameters p_i, and let $\mathbf{X}$ be their sum. Then for $x \geq \mathrm{E}[\mathbf{X}]$ we have that $\Pr[\mathbf{X} = x] > \Pr[\mathbf{X} = x + 1]$. Similarly, for $x \leq \mathrm{E}[\mathbf{X}]$ we have that $\Pr[\mathbf{X} = x - 1] < \Pr[\mathbf{X} = x]$.*

Additionally, we have the following lemma:

Lemma A.3 *There exists a universal constant $C_{A.3}$ for which the following holds. Fix any n and any independent Bernoulli random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with (possibly different) parameters p_i. Let $\mathbf{X}$ be their sum, and let $\sigma^2 = \mathrm{Var}[\mathbf{X}]$. Then for all $x \in \{0, \ldots, n\}$,*

$$\Pr[\mathbf{X} = x] \leq \frac{C_{A.3}}{\sigma}.$$

The proof of the proposition uses the following theorem of Deheuvels et al. (1989).

Theorem A.4 *There exists a universal constant $C_{A.4}$ for which the following holds. Fix any n and any independent Bernoulli random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with (possibly different) parameters p_i. Let $\mathbf{X}$ be their sum, and let $\sigma^2 = \mathrm{Var}[\mathbf{X}]$. Then*

$$\sup_y (1 + |y|^3)\,|F(y) - \phi(y)| \leq \frac{C_{A.4}}{\sigma},$$

where

$$F(y) = \Pr\left[\frac{\mathbf{X} - \mathrm{E}[\mathbf{X}]}{\sigma} \leq y\right]$$

and $\phi(y)$ is the distribution function of a standard normal $N(0,1)$ random variable.

We now prove Lemma A.3.

Proof: Theorem A.2 implies that the element $x \in \{0, \ldots, n\}$ for which $\Pr[\mathbf{X} = x]$ is maximal must be contained in $(\mathrm{E}[\mathbf{X}] - 1, \mathrm{E}[\mathbf{X}] + 1)$. Observe that

$$\begin{aligned}
\Pr[\mathbf{X} = x] &\leq \Pr[\mathbf{X} \in (\mathrm{E}[\mathbf{X}] - 1, \mathrm{E}[\mathbf{X}] + 1)] \\
&= F(1/\sigma) - F(-1/\sigma) \\
&\leq \phi(1/\sigma) - \phi(-1/\sigma) + 2C_{A.4}/\sigma \\
&\leq \frac{1}{\sqrt{2\pi}} \cdot \frac{2}{\sigma} + 2C_{A.4}/\sigma \\
&= \frac{C_{A.3}}{\sigma}
\end{aligned}$$

for constant $C_{A.3} = 2/\sqrt{2\pi} + 2C_{A.4}$ and where the second inequality follows from Theorem A.4. ■

We are now ready to prove Lemma 5.3.

Proof: Assume without loss of generality that $t \geq 0$ – the case of $t < 0$ is symmetric. Let

$$Y = \{x : \Pr[\mathbf{X} = x] \geq \Pr[\mathbf{X} = x + t] + \delta\},$$

and denote the elements of Y by $y_1, \ldots, y_{|Y|}$ such that $\Pr[\mathbf{X} = y_1] \geq \Pr[\mathbf{X} = y_2] \geq \ldots \geq \Pr\left[\mathbf{X} = y_{|Y|}\right]$. Now, due to the monotonicity of $\Pr[\mathbf{X}]$ guaranteed by Theorem A.2, we have that

$$\begin{aligned}
\Pr[\mathbf{X} = y_1] &\geq \Pr[\mathbf{X} = y_{1+t}] + \delta \\
&\geq \Pr[\mathbf{X} = y_{1+2t}] + 2\delta \\
&\ldots \\
&\geq \Pr[\mathbf{X} = y_{1+\iota t}] + \iota\delta \geq \iota\delta,
\end{aligned}$$

where $\iota = \lfloor \frac{|Y|-1}{t} \rfloor$. Combining this with the bound on $\Pr[\mathbf{X} = x]$ given by Lemma A.3 yields

$$\iota\delta \leq \frac{C_{A.3}}{\sigma},$$

and so

$$|Y| \leq \frac{(C_{A.3} + 1)t}{\delta\sigma}.$$

Additionally, and again due to the bound on the maximal probability, we have that

$$\Pr[\mathbf{X} \in Y] \leq \frac{C_{A.3}\,|Y|}{\sigma} \leq \frac{C_{A.3}(C_{A.3} + 1)t}{\delta\sigma^2}.$$

Similarly, let

$$Z = \{x : \Pr[\mathbf{X} = x] \leq \Pr[\mathbf{X} = x + t] - \delta\}.$$

An identical argument as above shows that

$$\Pr[\mathbf{X} \in Z] \leq \frac{C_{A.3}(C_{A.3} + 1)t}{\delta\sigma^2}.$$

Putting these together and fixing $C_{5.3} = 2C_{A.3}(C_{A.3} + 1)$, we get that

$$\Pr_{x \sim \mathbf{X}}[|\Pr[\mathbf{X} = x] - \Pr[\mathbf{X} = x + t]| > \delta] \leq \frac{C_{5.3}t}{\delta\sigma^2}$$

as claimed. ■

A.2 Generalizing Lemma 5.3 to the Multinomial Distribution

In this section we prove the following lemma, which is a generalization of Lemma 5.3 to multinomial random variables. Note that this is not strictly a generalization, since we will additionally require that the parameters p_i be bounded away from 0 and 1. However, this will be sufficient for our purposes.

Lemma A.5 *Fix any n, any natural number m, and any $\varepsilon > 0$. Also fix any independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ taking values in $A = \{0, 1, \ldots, m\}$ such that for each $i \in \{1, \ldots, n\}$ and each $j \in A$ it holds that $\Pr[\mathbf{X}_i = j] \geq \varepsilon$. Let $\mathbf{X}^j$ be the number of $\mathbf{X}_i$'s taking value j. Then the following holds for any $\delta > 0$ and vector $(t_1, \ldots, t_m)$ of integers:*

$$\left|\Pr\left[(\mathbf{X}^1 = x_1) \cap \ldots \cap (\mathbf{X}^m = x_m)\right] - \Pr\left[(\mathbf{X}^1 = x_1 + t_1) \cap \ldots \cap (\mathbf{X}^m = x_m + t_m)\right]\right| \leq \delta$$

with probability at least

$$1 - m\left(\frac{C_{A.3}}{\sqrt{\varepsilon^2(1-\varepsilon)n}} + e^{\frac{-\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}\right)^{m-1}\left(\frac{C_{5.3} \cdot mt}{\delta\varepsilon^2(1-\varepsilon)n}\right) - me^{-\frac{\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}$$

over the choice of $(x_1, \ldots, x_m)$ from $(\mathbf{X}^1, \ldots, \mathbf{X}^m)$, where $t = \max_i |t_i|$ and $C_{A.3}$ and $C_{5.3}$ are the universal constants from Lemmas A.3 and 5.3.

Note that the possible set of actions A includes the action 0, which does not appear in the statement above. The number of players who play 0 can be calculated from the number of players who played actions $1, \ldots, m$.

Proof: To shorten notation, we will write $\Pr[x_j]$ instead of $\Pr\left[\mathbf{X}^j = x_j\right]$ when the $\mathbf{X}^j$ is clear from context. We wish to bound the probability, over the choice of $(x_1, \ldots, x_m)$ from $(\mathbf{X}^1, \ldots, \mathbf{X}^m)$, that

$$|\Pr[x_1 \cap \ldots \cap x_m] - \Pr[(x_1 + t_1) \cap \ldots \cap (x_m + t_m)]| > \delta.$$

The first step is to condition on the event that $(x_1, \ldots, x_m)$ satisfies

$$x_j \leq \mathrm{E}[\mathbf{X}^j] + \frac{\varepsilon n}{m} - |t_j|$$

for all j. By a Chernoff bound this holds with probability at least $1 - m\exp\left(-(\varepsilon^2 n/2m^2) + (t^2/2n)\right)$. We will ignore this error for now, and will add it back in at the end.

Now,

$$
\begin{aligned}
&|\Pr[x_1 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap \ldots \cap (x_m+t_m)]| \\
&= |\Pr[x_1 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m] \\
&\quad + \Pr[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap (x_2+t_2) \cap \ldots \cap x_m] \\
&\quad + \ldots \\
&\quad + \Pr[(x_1+t_1) \cap (x_2+t_2) \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap \ldots \cap (x_m+t_m)]| \\
&\leq |\Pr[x_1 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m]| \\
&\quad + |\Pr[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap (x_2+t_2) \cap \ldots \cap x_m]| \\
&\quad + \ldots \\
&\quad + |\Pr[(x_1+t_1) \cap (x_2+t_2) \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap \ldots \cap (x_m+t_m)]| \,.
\end{aligned}
$$

Thus, we will bound the probability that each one of these summands exceeds a difference of δ/m in probability, and then take a union bound. Formally, we now wish to bound the probability that

$$
|\Pr[x_1 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m]| > \frac{\delta}{m}.
$$

The bounds on the other summands will be identical.

Let $\overline{x} \in (\{2,\ldots,m\} \cup \perp)^n$ denote a vector of realizations of the $\mathbf{X}_i$'s, but only in those indices for which $\mathbf{X}_i \notin \{0,1\}$. The indices in which $\mathbf{X}_i \in \{0,1\}$ are denoted by $\perp$. Now,

$$
\begin{aligned}
&\Pr_{(x_1,\ldots,x_m)}\left[|\Pr[x_1 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m]| > \frac{\delta}{m}\right] \qquad (4) \\
&= \sum_{\overline{x}} \Pr_{(x_1,\ldots,x_m)}\left[\left(|\Pr[x_1 \cap \ldots \cap x_m] - \Pr[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m]| > \frac{\delta}{m}\right) \;\middle|\; \overline{x}\right] \cdot \Pr[\overline{x}] \,.
\end{aligned}
$$

For a fixed $\overline{x}$, observe that if for some $j > 1$ it holds that the number of j's in $\overline{x}$ is not equal to x_j, then

$$
\Pr\left[x_1 \cap \ldots \cap x_m \;\middle|\; \overline{x}\right] = \Pr\left[(x_1+t_1) \cap x_2 \cap \ldots \cap x_m \;\middle|\; \overline{x}\right] = 0,
$$

and so the difference in probabilities is also 0 with probability 1. The probability that for all $j > 1$, the number of j's in $\overline{x}$ is equal to x_j is at most

$$
\left(\frac{C_{A.3}}{\sqrt{\varepsilon^2(1-\varepsilon)n}} + e^{\frac{-\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}\right)^{m-1}. \qquad (5)
$$

This holds because we conditioned on

$$
x_j \leq \mathrm{E}[\mathbf{X}^j] + \frac{\varepsilon n}{m} - |t_j|
$$

for all j, which implies that the variance of $\mathbf{X}^j$ conditional on $\mathbf{X}^k = x_k$ for all $1 < k < j$ is at least $\varepsilon^2(1-\varepsilon)n$. Lemma A.3, together with the fact that we conditioned on $x_j \leq \mathrm{E}[\mathbf{X}^j] + (\varepsilon n)/m - |t_j|$, imply (5).

Now consider some $\overline{x}$ for which the number of j's (for all $j > 1$) is equal to x_j. Let $M = \{i \in \{1, \ldots, n\} : \overline{x}_i = \perp\}$, fix $\mathbf{M} = \sum_{i \in M} \mathbf{X}_m$ and $\sigma_M^2 = \mathrm{Var}[\mathbf{M}]$. The random variable $\mathbf{M}$ counts the number of 1's that occur, conditional on $\overline{x}$ denoting the indices in which some other $j \in \{2, \ldots, m\}$ was realized.

We can now apply Lemma 5.3 to the random variables $\mathbf{X}_i$ for $i \in M$, and obtain that for any t_1 and number δ/m,

$$\left|\Pr\left[x_1 \cap \ldots \cap x_m \;\middle|\; \overline{x}\right] - \Pr\left[(x_1 + t_1) \cap x_2 \cap \ldots \cap x_m \;\middle|\; \overline{x}\right]\right| > \frac{\delta}{m} \tag{6}$$

with probability at most

$$C_{5.3}\left(\frac{at_1}{\delta\sigma_M^2}\right)$$

over the choice of x_1 from $\mathbf{M}$.

Observe now that, again because we conditioned on

$$x_j \leq \mathrm{E}[X^j] + \frac{\varepsilon n}{m} - t_j$$

for all j, it must be the case that $|M| \geq \varepsilon n$, and so $\sigma_M \geq \sqrt{\varepsilon^2(1-\varepsilon)n}$. Thus, inequality (6) holds with probability at most

$$\frac{C_{5.3} \cdot m\,|t_1|}{\delta\varepsilon^2(1-\varepsilon)n}.$$

In fact, this means that (4) is bounded above by

$$\beta \stackrel{\text{def}}{=} \left(\frac{C_{A.3}}{\sqrt{\varepsilon^2(1-\varepsilon)n}} + e^{\frac{-\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}\right)^{m-1}\left(\frac{C_{5.3} \cdot mt}{\delta\varepsilon^2(1-\varepsilon)n}\right),$$

where $t = \max_i |t_i|$. This holds because since it is just a convex combination over all realizations $\overline{x}$ in which the number of j's for $j > 1$ is equal to x_j, and this holds with probability given by (5).

We can now repeat this argument for each $j \in \{2, \ldots, m\}$. A union bound then implies that

$$\begin{aligned}&\Pr\left[\left|\Pr\left[x_1 \cap \ldots \cap x_m\right] - \Pr\left[(x_1 + t_1) \cap \ldots \cap (x_m + t_m)\right]\right| > \delta\right]\\ &\leq m\beta.\end{aligned}$$

Adding in the error of $1 - \exp(-\varepsilon^2 n/2m^2)$ from the Chernoff bound yields a bound of

$$m\beta + me^{-\frac{\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}.$$

Setting $C_{A.5} = \max\{C_{A.3}, C_{5.3}\}$ completes the proof. ∎

A.3 Expectations of Slightly Shifted Summations are Close

In this section we prove Lemmas A.6 and A.7.

Lemma A.6 *Fix any n, any $t < n$, any natural number m, and any $\varepsilon > 0$. Fix any independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_{n-t}$ taking values in $A = \{0, 1, \ldots, m\}$ such that for each $i \in \{1, \ldots, n-t\}$ and each $j \in A$ it holds that $\Pr[\mathbf{X}_i = j] \geq \varepsilon$. Let $\mathbf{X}^j$ be the number of $\mathbf{X}_i$'s taking value j. Finally, fix any anonymous* *function* $f : A^n \mapsto [0,1]$. *Then the following holds for any* $\delta > 0$, $d > 0$, *and any pair of vectors* $\bar{t} = (t_1, \ldots, t_m)$ *and* $\bar{s} = (s_1, \ldots, s_m)$ *of natural numbers satisfying* $\sum_i t_i \le t$ *and* $\sum_i s_i \le t$*:*

$$\begin{aligned}&\left|\mathrm{E}\left[f\left(\mathbf{X}^1 + t_1, \ldots, \mathbf{X}^m + t_m\right)\right] - \mathrm{E}\left[f\left(\mathbf{X}^1 + s_1, \ldots, \mathbf{X}^m + s_m\right)\right]\right| \\ &\le 2\delta\left(2d\sqrt{n} + 2t\right)^m + 2me^{-\frac{d^2}{2}} + 2me^{-\frac{\varepsilon^2(n-t)}{2} + \frac{t^2}{2(n-t)}} \\ &\quad + 2m\left(\frac{C_{A.3}}{\sqrt{\varepsilon^2(1-\varepsilon)n}} + e^{\frac{-\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}\right)^{m-1}\left(\frac{C_{5.3}\cdot mt}{\delta\varepsilon^2(1-\varepsilon)n}\right) + 2me^{-\frac{\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}.\end{aligned}$$

Proof of Lemma A.6: Let $\mathbf{I}_t$ be an indicator for the event that for all $j \in \{1, \ldots, m\}$ it holds that $\mathbf{X}^j > t_j$ and that $\mathrm{E}[\mathbf{X}^j] - d\sqrt{n} - t_j < \mathbf{X}^j < \mathrm{E}[\mathbf{X}^j] + d\sqrt{n} + t_j$. Let $\mathbf{I}_s$ be an indicator the event that for all $j \in \{1, \ldots, m\}$ it holds that $\mathbf{X}^j > s_j$ and that $\mathrm{E}[\mathbf{X}^j] - d\sqrt{n} - s_j < \mathbf{X}^j < \mathrm{E}[\mathbf{X}^j] + d\sqrt{n} + s_j$. Note that both $\Pr[\mathbf{I}_t = 0]$ and $\Pr[\mathbf{I}_s = 0]$ are bounded above by

$$\varepsilon_I = me^{-\frac{d^2}{2}} + me^{-\frac{\varepsilon^2(n-t)}{2} + \frac{t^2}{2(n-t)}}$$

by a pair of Chernoff bounds and the fact that $\mathrm{E}[\mathbf{X}^j] \ge \varepsilon(n-t)$ for all j.

Also, let $\mathbf{J}_t$ be an indicator for the event that

$$\left|\Pr\left[(\mathbf{X}^1 = x_1) \cap \ldots \cap (\mathbf{X}^m = x_m)\right] - \Pr\left[(\mathbf{X}^1 = x_1 - t_1) \cap \ldots \cap (\mathbf{X}^m = x_m - t_m)\right]\right| \le \delta.$$

Let $\mathbf{J}_s$ be an indicator for the event that

$$\left|\Pr\left[(\mathbf{X}^1 = x_1) \cap \ldots \cap (\mathbf{X}^m = x_m)\right] - \Pr\left[(\mathbf{X}^1 = x_1 - s_1) \cap \ldots \cap (\mathbf{X}^m = x_m - s_m)\right]\right| \le \delta.$$

Let

$$\varepsilon_{A.5} = \frac{m^2 t \cdot C_{A.5}^m}{\delta\left(\varepsilon^2(1-\varepsilon)n\right)^{\frac{m+1}{2}}} + 2me^{-\frac{\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}}$$

be the bound on the probabilities of $\mathbf{J}_t = 0$ and $\mathbf{J}_s = 0$ given by Lemma A.5.

Then

$$\begin{aligned}&\left|\mathrm{E}\left[f\left(\mathbf{X}^1 + t_1, \ldots, \mathbf{X}^m + t_m\right)\right] - \mathrm{E}\left[f\left(\mathbf{X}^1 + s_1, \ldots, \mathbf{X}^m + s_m\right)\right]\right| \\ &= \left|\sum_{(x_1,\ldots,x_m)} f(x_1 + t_1, \ldots, x_m + t_m)\Pr\left[x_1 \cap \ldots \cap x_m\right] - \sum_{(x_1,\ldots,x_m)} f(x_1 + s_1, \ldots, x_m + s_m)\Pr\left[x_1 \cap \ldots \cap x_m\right]\right| \\ &\le \left|\sum_{\substack{(x_1,\ldots,x_m)\\ s.t.\ \mathbf{I}_t = \mathbf{J}_t = 1}} f(x_1, \ldots, x_m)\Pr\left[(x_1 - t_1) \cap \ldots \cap (x_m - t_m)\right]\right. \\ &\quad \left. - \sum_{\substack{(x_1,\ldots,x_m)\\ s.t.\ \mathbf{I}_s = \mathbf{J}_s = 1}} f(x_1, \ldots, x_m)\Pr\left[(x_1 - s_1) \cap \ldots \cap (x_m - s_m)\right]\right| + 2\varepsilon_I + 2\varepsilon_{A.5} \\ &\le \sum_{\substack{(x_1,\ldots,x_m)\\ s.t.\ \mathbf{I}_t = \mathbf{J}_t = 1}} f(x_1, \ldots, x_m)\left|\Pr\left[(x_1 - t_1) \cap \ldots \cap (x_m - t_m)\right] - \Pr\left[(x_1 - s_1) \cap \ldots \cap (x_m - s_m)\right]\right| + 2\varepsilon_I + 2\varepsilon_{A.5} \\ &\le 2\delta\left(2d\sqrt{n} + 2t\right)^m + 2\varepsilon_I + 2\varepsilon_{A.5}.\end{aligned}$$

Plugging in the respective values of ε_I and $\varepsilon_{A.5}$ completes the proof of the lemma. ■

We finally have the following lemma, which is derived from Lemma A.6 by setting the parameters.

Lemma A.7 *For any natural number m and reals $\varepsilon > 0$ and $\alpha > 0$ there exists a constant $C_{A.7} = C_{A.7}(m, \varepsilon, \alpha)$ for which the following holds. Fix any natural numbers and n and $t < C_{A.7}\sqrt{n}$, and any independent random variables $\mathbf{X}_1, \ldots, \mathbf{X}_{n-t}$ taking values in $A = \{0, 1, \ldots, m\}$ such that for each $i \in \{1, \ldots, n-t\}$ and each $j \in A$ it holds that $\Pr[\mathbf{X}_i = j] \geq \varepsilon$. Let $\mathbf{X}^j$ be the number of $\mathbf{X}_i$'s taking value j. Finally, fix any anonymous function $f : A^n \mapsto [0, 1]$. Then for any pair of vectors $\overline{t} = (t_1, \ldots, t_m)$ and $\overline{s} = (s_1, \ldots, s_m)$ of natural numbers satisfying $\sum_i t_i \leq t$ and $\sum_i s_i \leq t$ it holds that*

$$\left| \mathrm{E}\left[f\left(\mathbf{X}^1 + t_1, \ldots, \mathbf{X}^m + t_m \right) \right] - \mathrm{E}\left[f\left(\mathbf{X}^1 + s_1, \ldots, \mathbf{X}^m + s_m \right) \right] \right| < \alpha.$$

Proof of Lemma A.7: We will set the parameters so that each of the five summands in the upper bound of Lemma A.6 will be bounded above by $\alpha/5$. We take care of the second summand by fixing $d = \sqrt{2\ln(10m/\alpha)}$. We then take care of the first summand by setting

$$\delta = \frac{\alpha}{10\left(2d\sqrt{n} + 2t\right)^m}.$$

This implies that the fourth summand

$$\begin{aligned} & 2m \left(\frac{C_{A.3}}{\sqrt{\varepsilon^2(1-\varepsilon)n}} + e^{\frac{-\varepsilon^2 n}{2m^2} + \frac{t^2}{2n}} \right)^{m-1} \left(\frac{C_{5.3} \cdot mt}{\delta \varepsilon^2 (1-\varepsilon) n} \right) \\ & \leq \frac{C_4 t d^m}{\sqrt{n}} + \left(e^{\frac{-(m-1)\varepsilon^2 n}{2m^2} + \frac{(m-1)t^2}{2n}} \right) \left(\frac{C_{5.3} \cdot mt}{\delta \varepsilon^2 (1-\varepsilon) n} \right) \\ & \leq \frac{C_4 t d^m}{\sqrt{n}} + \frac{\alpha}{10} \end{aligned}$$

for some constant C_4 and $n > n_4$ for some n_4. This sum is bounded above by $\alpha/5$ if $C_4 t d^m / \sqrt{n} < \alpha/10$, which holds as long as $t < C_4' \sqrt{n}$ for some small enough constant $C_4' > 0$.

Finally, observe that the third and fifth summands are all less than $\alpha/5$ for large enough n, since they are all decreasing (exponentially) with n. Say that the minimal n for the respective summands to be bounded by $\alpha/5$ are n_3 and n_5 respectively.

Thus, we now have that for large enough n, the lemma holds. The last step is to set

$$C_{A.7} < \min\left\{ C_4', \frac{1}{\sqrt{n_3}}, \frac{1}{\sqrt{n_4}}, \frac{1}{\sqrt{n_5}} \right\}.$$

This guarantees that when n is not large enough for the bound to hold we have that $t < 1$, for which the lemma trivially holds. ■

A.4 Proof of Theorem 5.2

Proof: Fix some player i and subset $S \subset N \setminus \{i\}$ with $|S| = t$. Without loss of generality, suppose $S = \{n-t, n-t+1, \ldots, n\}$ and $i = 1$ (this is just a relabeling of the players). Also, fix some $t_S \in A^{|S|}$, and let t_j be the number of j's in t_S. Denote by $\mathbf{X}^j$ the number of j's in $\mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_{n-t-1}$, and a possible value of $\mathbf{X}^j$ by x_j. Finally, denote by $\mathbf{S}^j$ the number of j's in $\mathbf{A}_{n-t}, \ldots, \mathbf{A}_n$, and a possible value of $\mathbf{S}^j$ by s_j. For any $\overline{a} \in A$,

$$
\begin{aligned}
&\left|u_i^{\overline{a}}(\mathbf{A}_{-i}) - u_i^{\overline{a}}(\mathbf{A}_{-(S\cup\{i\})} : t_S)\right| \\
&= \left|\mathrm{E}\left[u_i^{\overline{a}}(\mathbf{X}^1 + \mathbf{S}^1, \ldots, \mathbf{X}^m + \mathbf{S}^m)\right] - \mathrm{E}\left[u_i^{\overline{a}}(\mathbf{X}^1 + t_1, \ldots, \mathbf{X}^m + t_m)\right]\right| \\
&= \left|\sum_{(s_1,\ldots,s_m)} \left(\mathrm{E}\left[u_i^{\overline{a}}(\mathbf{X}^1 + s_1, \ldots, \mathbf{X}^m + s_m)\right] \Pr\left[(s_1, \ldots, s_m)\right]\right) - \mathrm{E}\left[u_i^{\overline{a}}(\mathbf{X}^1 + t_1, \ldots, \mathbf{X}^m + t_m)\right]\right| \\
&= \left|\sum_{(s_1,\ldots,s_m)} \left(\mathrm{E}\left[u_i^{\overline{a}}(\mathbf{X}^1 + s_1, \ldots, \mathbf{X}^m + s_m)\right] - \mathrm{E}\left[u_i^{\overline{a}}(\mathbf{X}^1 + t_1, \ldots, \mathbf{X}^m + t_m)\right]\right) \Pr\left[(s_1, \ldots, s_m)\right]\right| \\
&< \left|\sum_{(s_1,\ldots,s_m)} \alpha \Pr\left[(s_1, \ldots, s_m)\right]\right| = \alpha,
\end{aligned}
$$

where the last inequality follows from Lemma A.7, and setting $C_{5.2} = C_{A.7}$. Because this holds for every $\overline{a} \in A$, $\mathbf{A}$ is (α, t)-stable. Immunity and tolerance follow from Lemma 3.4, and resilience against coalitions from Lemma 3.6. ■

A.5 Proof of Theorem 5.1

Proof: Fix some $0 < \beta < 1/(m+1)$. Consider a game G' that is identical to G, but where the action space is not A but rather $\Delta_\beta(A)$, the set of all β-uniform probability distributions over A. Furthermore, denote payoffs in G' as follows. For $(a_1, \ldots, a_n) \in \Delta_\beta(A)^n$, let

$$u_i'(a_1, \ldots, a_n) = \mathrm{E}[u_i(a_1, \ldots, a_n)].$$

Since the action space of G' is compact and the utility functions are continuous, a theorem of Glicksberg (1952) implies that G' has a Nash equilibrium.

Observe now that a Nash equilibrium $\mathbf{A}'$ of G' is a βm-Nash equilibrium of G. To see this, suppose towards a contradiction that some player i has a deviation $\mathbf{A}''$ that would increase his payoff by more than βm. $\mathbf{A}''$ must put weight at least β on some action j, so consider the following strategy $\mathbf{A}^{(3)}$: play $\mathbf{A}''$ with probability $1 - \beta m$, and actions $k \neq j$ each with probability β. $\mathbf{A}^{(3)}$ is β-uniform, but note that it still improves over $\mathbf{A}_i'$ – the actions $k \neq i$ each yields a minimum payoff of 0, so the loss of $\mathbf{A}^{(3)}$ compared to $\mathbf{A}''$ is at most βm. The strategy $\mathbf{A}^{(3)}$, however, contradicts the fact that $\mathbf{A}'$ is a Nash equilibrium of G'.

Now, since $\mathbf{A}'$ is a β-uniform βm-Nash equilibrium of G, we can use Theorem 5.2 to obtain that, for any α, there exists a constant $C_{5.2}$ for which the following holds: for $t = C_{5.2}\sqrt{n}$, the equilibrium $\mathbf{A}'$ is also $(\alpha + \beta m, t)$-immune, $(2\alpha + \beta m, t)$-tolerant, and $(3\alpha + \beta m, t+1)$-coalitional. Setting

$$\alpha = \beta m = \frac{\varepsilon}{2}$$

and $C_{5.2} = C_{5.1}$ for the appropriate latter constant completes the proof. ■

A.6 Proof of Theorem 5.4

Proof: Assume $\delta \leq 1/2$ (otherwise fix $\delta = 1/2$). Consider a δ-Nash equilibrium $\mathbf{A}$ that is δ/m-uniform, and note that the existence of such an equilibrium is guaranteed for the same reason as in the proof of Theorem 5.1. Fix some player i and some set $S \subseteq N \setminus \{i\}$ of size $|S| = t = C_{5.4} \cdot n/k(n)$, where $C_{5.4}$ will be determined later. Also fix $t' = C_{A.7}\sqrt{n-t}/2$, where $C_{A.7}$ is the constant from Lemma A.7.

Now, for each $j \in \{1, \ldots, m\}$, let

$$\mathbf{X}^j \stackrel{\text{def}}{=} |\{i \in S : \mathbf{X}_i = j\}| .$$

By a Chernoff bound we have that

$$\begin{aligned}
&\Pr\left[\forall j \in \{1, \ldots, m\},\ \left|\mathbf{X}^j - \mathrm{E}[\mathbf{X}^j]\right| < \frac{t'}{m}\right] \\
&> 1 - 2me^{-\frac{C_{A.7}^2(n-t)}{2m^2 t}} \\
&= 1 - 2me^{-\frac{C_{A.7}^2(1-C_{5.4})k(n)\cdot n}{2m^2 C_{5.4} n}} \\
&= 1 - \frac{p}{2e^{k(n)}}
\end{aligned}$$

for a large enough (but, importantly, independent of n) choice of $C_{5.4}$.

We now continue as in the proofs of Theorem 5.2 and Lemma A.6 – the set S, although larger than the one of Theorem 5.2, nevertheless has the property that net shifts in the number of elements of each value are at most within $C_{A.7}\sqrt{n'}$, where $n' = n - t$ is the number of random variables not in S. Thus, its effect is essentially identical to the effect of a Byzantine set of size $C_{A.7}\sqrt{n'}$. ■

B Proofs for General Games

B.1 Proof of Theorem 6.3

The proof utilizes the following lemma, which bounds the number of sets of players who have large influence on the expectation of a function. The proof of the lemma expands on some ideas of Al-Najjar and Smorodinsky (2000), and is a reduction to the case of anonymous functions (about which we already have explicit statements – see Section 5).

Lemma B.1 *Let $f : A^n \mapsto [0,1]$ be any function, let n, δ, and ε be as in Theorem 6.3, fix $t = C_{A.7}(m+2, \delta/2, \varepsilon/2) \cdot \sqrt{n}$, and let $\mathbf{X} = \mathbf{X}_1 \times \ldots \times \mathbf{X}_n$ be a δ-uniform product distribution over A^n. Then*

$$\mathop{\mathrm{E}}_{S}\left[\max_{b_S \in A^t} \left|\mathrm{E}\left[f(\mathbf{X})\right] - \mathrm{E}\left[f(\mathbf{X}_{-S} : b_S)\right]\right|\right] \leq \varepsilon,$$

where S of size $|S| = t$ is chosen uniformly at random from N.

Proof: We first assume that all $\mathbf{X}_i$ are distributed identically over A. We also assume that f is *monotone*: For each i and profile $a_{-i} \in A^{n-1}$ it holds that $f(a_{-i} : 0) \leq f(a_{-i} : 1) \leq \ldots \leq f(a_{-i} : m)$. We will then show how to reduce the problem to this case. Let $\sigma : N \mapsto N$ be a permutation of the players. For a set $S \subset N$, let

$$V_S(f) = \max_{b_S \in A^t} \left| \mathrm{E}\left[f(\mathbf{X})\right] - \mathrm{E}\left[f(\mathbf{X}_{-S} : b_S)\right] \right|.$$

Note that, since f is monotone, the profile of actions b_S at which $V_S(f)$ obtains is always either $(0, \ldots, 0)$ or $(m, \ldots, m)$.

Next, for any positive integer t let

$$V_t(f) = \sum_{\substack{S \subset N \\ |S| = t}} V_S(f).$$

Now let $f^\sigma : A^n \mapsto [0, 1]$ be such that $f^\sigma(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Observe that $V_t(f) = V_t(f^\sigma)$: this holds because in both cases we are summing over all subsets S of size t, and the only difference is that the sum is in a different order. Here we also use the fact that the $\mathbf{X}_i$ are distributed identically.

Now consider the function $g : A^n \mapsto [0, 1]$, where

$$g(x_1, \ldots, x_n) = \frac{1}{n!} \sum_{\sigma} f^\sigma(x_1, \ldots, x_n).$$

First, note that g is an anonymous function, since $g(x_1, \ldots, x_n) = g(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for all permutations σ of the players. Additionally,

$$\begin{aligned}
2V_t(g) &= 2 \sum_{\substack{S \subset N \\ |S| = t}} \frac{1}{n!} \max_{b_S \in A^t} \sum_{\sigma} \left| \mathrm{E}\left[f^\sigma(\mathbf{X})\right] - \mathrm{E}\left[f^\sigma(\mathbf{X}_{-S} : b_S)\right] \right| \\
&\geq \sum_{\substack{S \subset N \\ |S| = t}} \frac{1}{n!} \sum_{\sigma} V_S(f^\sigma) \\
&= \sum_{\substack{S \subset N \\ |S| = t}} \frac{1}{n!} \sum_{\sigma} V_S(f) = V_t(f).
\end{aligned}$$

The factor of 2 comes from the fact that $V_t(g)$ forces each b_S to be the same for every f^σ, whereas this need not be the case. However, due to the monotonization, the b_S achieving the highest $V_t(f^\sigma)$ is always either $(0, \ldots, 0)$ or $(m, \ldots, m)$.

Since g is an anonymous function, every S of size t has the same $V_S(g)$. Thus, on expectation (over the random choice of S), $\mathrm{E}[V_S(f)] \leq 2V_S(g)$. Now, Lemma A.7 shows that for any anonymous function g it holds that $V_S(g) \leq \varepsilon/2$, and this almost completes the proof of the claim.

Recall that in the above, we assumed that all $\mathbf{X}_i$ are distributed identically, and this is not necessarily the case. We now show how to get around this problem by reducing the distribution $\mathbf{X}$ and the function f to a different function and a distribution $\mathbf{Y}$ that is identically distributed, without affecting the influences of sets of players.

For each $i \in N$ and $\overline{a} \in A$, let $p_i^{\overline{a}} = \Pr\left[\mathbf{X}_i = \overline{a}\right]$, and recall that $p_i^{\overline{a}} \in [\delta, 1 - \delta]$. Now, consider the function $f' : [0, 1]^n \mapsto [0, 1]$ defined as follows:

$$f'(r_1, \ldots, r_n) = f(\mathbf{X}_1(r_1), \ldots, \mathbf{X}_n(r_n)),$$

where $\mathbf{X}_i(r_i) = 0$ if $r_i \le p_i^0$, $\mathbf{X}_i(r_i) = 1$ if $p_i^0 < r_i \le p_i^0 + p_i^1$, and so on. Let $\mathbf{U} = \mathbf{U}_1 \times \ldots \times \mathbf{U}_n$ be the uniform distribution over $[0,1]^n$. Then when the inputs to f' are drawn independently from $\mathbf{U}$ we have that $\mathrm{E}[f] = \mathrm{E}[f']$. Also, if we define $V_S(f')$ similarly to $V_S(f)$ (with sup instead of max), then we have $V_S(f') = V_S(f)$.

Now let $\mathbf{Y} = \mathbf{Y}_1 \times \ldots \times \mathbf{Y}_n$ be independent and identically distributed random variables over $\{0,1,\ldots,m,\perp\}$ such that for every $i \in N$ and $\overline{a} \in A$ it holds that $\Pr[\mathbf{Y}_i = \overline{a}] = \delta$, and $\Pr[\mathbf{Y}_i = \perp] = 1-(m+1)\delta$. Consider the function $h : \{0,1,\ldots,m,\perp\}^n \mapsto [0,1]$ such that

$$h(y_1,\ldots,y_n) = \mathrm{E}\left[f'\big(\mathbf{U}|(\mathbf{U}_1 \in y_1),\ldots,(\mathbf{U}_n \in y_n)\big)\right],$$

where

$$(\mathbf{U}_i \in y_i) \overset{\mathrm{def}}{=} \begin{cases} \mathbf{U}_i \in [0,\delta] & \text{if } y_i = 0, \\ \mathbf{U}_i \in [p_i^0, p_i^0+\delta] & \text{if } y_i = 1, \\ \mathbf{U}_i \in [p_i^0+p_i^1, p_i^0+p_i^1+\delta] & \text{if } y_i = 2, \\ \cdots & \\ \mathbf{U}_i \in [p_i^0+\ldots+p_i^{m-1}, p_i^0+\ldots+p_i^{m-1}+\delta] & \text{if } y_i = m, \\ \mathbf{U}_i \in [0,1] \setminus ([0,\delta] \cup [p_i^0, p_i^0+\delta] \cup \ldots) & \text{if } y_i = \perp. \end{cases}$$

We now claim that for every S it holds that $V_S(h) = V_S(f')$. First note that

$$\begin{aligned} \mathrm{E}[h(\mathbf{Y})] &= \sum_{y \in \mathrm{supp}(\mathbf{Y})} \Pr[\mathbf{Y} = y]\, h(y) \\ &= \sum_{y \in \mathrm{supp}(\mathbf{Y})} \Pr[\mathbf{Y} = y]\, \mathrm{E}\left[f'\big(\mathbf{U}|(\mathbf{U}_1 \in y_1),\ldots,(\mathbf{U}_n \in y_n)\big)\right] \\ &= \mathrm{E}[f'(\mathbf{U})]. \end{aligned}$$

Also, for some $S \subset N$ of size $|S| = t$ let $b_S \in \{0,1,\ldots,m,\perp\}^t$ be such that

$$|\mathrm{E}[h(\mathbf{Y})] - \mathrm{E}[h(\mathbf{Y}_{-S} : b_S)]|$$

is maximal. Then note that we can always take $b_S \in A^t$, and

$$\mathrm{E}[h(\mathbf{Y}_{-S} : b_S)] = \mathrm{E}[f'(\mathbf{U}|(\mathbf{U}_i \in b_i\ \forall i \in S))].$$

Hence,

$$\begin{aligned} V_S(h) &= \max_{b_S \in \{0,1,\ldots,m,\perp\}^t} |\mathrm{E}[h(\mathbf{Y})] - \mathrm{E}[h(\mathbf{Y}_{-S} : b_S)]| \\ &= \max_{b_S \in A^t} |\mathrm{E}[h(\mathbf{Y})] - \mathrm{E}[h(\mathbf{Y}_{-S} : b_S)]| \\ &= \max_{b_S \in A^t} \left|\mathrm{E}[f'(\mathbf{U})] - \mathrm{E}\left[f'\big(\mathbf{U}|(\mathbf{U}_i \in b_i\ \forall i \in S)\big)\right]\right| \\ &= V_S(f'). \end{aligned}$$

We now *monotonize* h by iterating the following for each $i \in \{1,\ldots,n\}$: For every profile of actions $a_{-i} \in A^n$, relabel those actions of player i that are in A so that $h(a_{-i} : 0) \le h(a_{-i} : 1) \le \ldots \le h(a_{-i} : m)$.

Note that because each $\mathbf{X}_i$ is uniformly distributed over A this relabeling does not affect the expectation of f. Also, since the vector b_S at which $V_S(h)$ is obtained can always be in A^t, $V_S(h)$ is also unaffected by this monotonization. However, once this has been iterated for all i, it is now the case that the profile of actions b_S at which $V_S(h)$ obtains is always either $(0, \ldots, 0)$ or $(m, \ldots, m)$. (Note that h is not monotone as defined in the beginning of the proof, since the action $\perp$ is not ordered. However, the argument that uses the monotonicity above nonetheless does go through because $b_S \in A^t$.)

So now we have an i.i.d. distribution $\mathbf{Y}$ and a function h such that

$$\mathop{\mathrm{E}}_{\mathbf{Y}}[V_S(h)] = \mathop{\mathrm{E}}_{\mathbf{U}}[V_S(f')] = \mathop{\mathrm{E}}_{\mathbf{X}}[V_S(f)]$$

as desired. Since this distribution is not necessarily δ-uniform (since $\perp$ occurs with probability $1-(a+1)\delta$, which may be smaller than δ), we do all this with $\delta' = \delta/2$. This is reflected in the fact that the parameters to $t = C_{A.7}(a+2, \delta/2, \varepsilon/2)$ in the statement of the theorem include a uniformity parameter of $\delta/2$. Observe also that we have $a+2$ actions, namely the actions in A plus the action $\perp$. ■

We can now prove Theorem 6.3.

Proof: Fix the constant $C_{6.3} = C_{B.1}(\delta/m, \varepsilon)$. Consider a δ-Nash equilibrium $\mathbf{A}$ that is δ/m-uniform, and note that the existence of such an equilibrium is guaranteed for the same reason as in the proof of Theorem 5.1.

Fix some $\overline{a} \in A$ and a player i. Lemma B.1 implies that

$$\mathop{\mathrm{E}}_{S}\left[\max_{b_S \in A^t} \left|u_i^{\overline{a}}(\mathbf{A}_{-i}) - u_i^{\overline{a}}(\mathbf{A}_{-(S\cup\{i\})} : b_S)\right|\right] \leq \varepsilon,$$

where S, $|S| = t$, is chosen uniformly at random from $N \setminus \{i\}$. Now, a Markov bound implies that

$$\Pr\left[\max_{b_S \in A^t} \left|u_i^{\overline{a}}(\mathbf{A}_{-i}) - u_i^{\overline{a}}(\mathbf{A}_{-(S\cup\{i\})} : b_S, \overline{a})\right| > k\varepsilon\right] \leq \frac{1}{k}.$$

The probability that this is true for any of $u_i^0, u_i^1, \ldots, u_i^m$ is at most $(m+1)/k$. Thus, we have that with probability at least $1-(m+1)/k$,

$$\left|u_i(\mathbf{A}_{-i}, \overline{a}) - u_i(\mathbf{A}_{-(S\cup\{i\})} : b_S, \overline{a})\right| \leq k\varepsilon$$

for all $\overline{a} \in A$ and all $b_S \in A^t$, and so $\mathbf{A}$ is $(k\varepsilon, t)$-stable. Immunity and tolerance follow by Lemma 3.4. ■

References

[1] 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*, 2005. Pages 45–58.

[2] 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 the 25th Annual ACM Symposium on Principles of Distributed Computing*, 2006. Pages 53–62.

[3] I. Abraham, D. Dolev, and J. Halpern. Lower bounds on implementing robust and resilient mediators. To appear in *5th IACR Theory of Cryptography Conference*, 2008.

[4] N. I. Al-Najjar and R. Smorodinsky. Pivotal players and the characterization of influence. Journal of Economic Theory 92, 2000. Pages 318–342.

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

[6] E. Azevedo, and E. Budish. Strategyproofness in the large. Working paper, 2012.

[7] Y. Azrieli and E. Shmaya. Lipschitz games. Working paper available on first author's webpage.

[8] M. Babaioff, R. Kleinberg, and C. Papadimitriou. Congestion games with malicious players. *Electronic Commerce*, 2007.

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

[10] J. Chen, S. Micali, and P. Valiant. Robustly leveraging collusion in combinatorial auctions." Manuscript, 2010.

[11] J. Chen and S. Micali. Collusive dominant-strategy truthfulness. Journal of Economic Theory 147.3 (2012): 1300-1312.

[12] J. Deb and E. Kalai. Large games with limited individual impact. Working paper, 2010.

[13] P. Deheuvels, M. Puri, and S. S. Ralescu. Asymptotic expansions for sums of nonidentically distributed Bernoulli random variables. *Journal of Multivariate Analysis* 28, 1989. Pages 282–303.

[14] K. Eliaz. Fault-tolerant implementation. *Review of Economic Studies* 69(3), 2002. Pages 589–610.

[15] I. L. Glicksberg. A further generalization of the Kakutani fixed point theorem, with application to Nash equilibrium points. *Proceedings of the American Mathematical Society* 3, 1952. Pages 170–174.

[16] R. Gradwohl and O. Reingold. Partial exposure in large games. *Games and Economic Behavior* 68(2), 2010. Pages 602–613.

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

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

[19] E. Kalai. Large robust games. *Econometrica*, Vol. 72, No. 6, November 2004. Pages 1631–1665.

[20] I. Kash, E. Friedman, and J. Halpern. Multiagent learning in large anonymous games. Journal of Artificial Intelligence Research 40 (2011): 571-598.

[21] P. Lévy. Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris. 1937.

[22] M. Pease, R. Shostak, and L. Lamport. Reaching agreement in the presence of faults. *J. ACM* 27(2), 1980. Pages 228-234.

[23] T. Moscibroda, S. Schmid, and R. Wattenhofer. When selfish meets evil: Byzantine players in a virus inoculation game. *25th Annual Symposium on Principles of Distributed Computing*, 2006.

[24] D. Monderer and M. Tennenholtz. Distributed games. *Games and Economic Behavior* 28, 1999. Pages 55–72.

[25] D. Monderer and M. Tennenholtz. Distributed games: from mechanisms to protocols. *AAAI*, 1999. Pages 32–37.

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

[27] J. B. Nielsen (editor): Summary Report on Rational Cryptographic Protocols. ECRYPT Report D.PROVI.7, March 2007.

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

[29] A. Roth. The price of malice in linear congestion games. In WINE 2008, *LNCS* 5385, 2008. Pages 118–125.