Discussion Paper #1605

"CONTAGIOUS STRATEGIES IMPLEMENT CONFESSION IN PRISONER'S DILEMMA GAMES"

Ehud Kalai
Northwestern University

November 1, 2023

J.L. Kellogg School of Management at Northwestern University
2211 Campus Drive, Evanston, Illinois 60208 | 847.491.3300 | kellogg.northwestern.edu

CONTAGIOUS STRATEGIES IMPLEMENT CONFESSION IN PRISONER'S DILEMMA GAMES

EHUD KALAI

ABSTRACT. A contagious strategy determines the most resilient equilibrium of the game in which it is played. As in dominant strategies, strong incentives motivate players to play a contagious strategy. This is observed in markets, auctions, and political interactions.

In generalized prisoner's dilemma games confession is a contagious strategy. We illustrate how this can be exploited to implement confession, even when confession is not a dominant strategy. Our illustration is based on an actual recent prosecution of a RICO trial in the US.

1. Introduction

Prisoner's dilemma (PD) games are designed by strategically-minded prosecutors, to induce confessions prior to the start of multi-defendent trials. RICO trials are excellent examples.[1] In the classical illustration of PD games, guilty defendants confess because the confession strategy is dominant. However, a more careful examination reveals that confession strategies are *contagious* and satisfy a condition weaker than dominance, they are *the most resilient*. In a sense similar to dominant strategy equilibrium, the most resilient equilibria also induces confessions. Moreover, the most resilient equilibria exist in a broader class of PD games, including ones in which dominant strategies equilibria do not. Prosecutors of RICO trials successfully design games in which confession strategies are the most resilient and induce confession, even when they cannot design games in which confession strategies are dominant.

Section 2 of this paper formalizes the notion of a contagious strategy and studies properties of the associated contagious equilibria. The main property of a contagious strategy is that its play by one player creates strong incentives to play it for the others. Moreover, a theorem in this section illustrates that contagious equilibria are distinctly the most resilient equilibria in their games. We point to the applicability of contagious equilibria in politics, markets, and auctions, see Kalai and Kalai (2002) for additional examples. However, the main application studied in

Date: Novermber 1, 2023.

2000 *Mathematics Subject Classification.* Primary 05C38, 15A15; Secondary 05A15, 15A18.

Key words and phrases. prisoner's dilemma, mechanism design, Nash equilibrium, equilibrium reselience, critical mass.

The author thanks Fern Moss Kalai, who inspired the writing of this paper; Adam Tauman Kalai, who proposed the orignial Conspiracy of Silence game, see Kalai and Kalai (2022): and Julien Manili and Professor Timothy Feddersen for important comments on earlier versions.
See full text at https://www.researchgate.net/publication/377148182.

[1] The Racketeer Influenced and Corrupt Organizations Act (RICO Act for short) was enacted in 1970 to combat organized crime in the United States. Testimony and evidence provided in a RICO trial against one defendant can be used against all defendants.

Sections 3-5 is to confession strategies that are contagious and are the most resilient in PD games.

Section 3 reviews the classical presentation of PD games and the role of dominant strategies in their analysis. Section 4 presents a generalized PD game as one with n-defendants who have a contagious confession strategy. The main example presented in this section is a RICO trial with ten defendants. In the pre-trial game, the confession strategy is contagious and thus the most resilient. This example serves two purposes. First, it applies the theorem about resilience of contagious strategies of Section 2 to argue that defendants would confess. Conversely, as discussed in Section 5, the PD application illustrates and elaborates on the various notions of stability that may support or contradict the rationale for confession. It also discusses the fragility of the competing equilibrium, in which all the defendants deny the charges against them. Section 6 discusses types of interactions in which individual incentives lead to other fragile equilibria, ones with undesirable social outcomes.

Section 7 points to earlier papers that dealt with the issue of resilience. In particular, Eliaz (2002) and Abraham et al (2006) who show that high resilience enables robust implementation of problems in economics and computer science. In addition, this section suggests needed further research on of the notion of resilience and its possible applications.

2. Contagious strategies and equilibria

Definition 1. *An individual strategy c in an n-person game Γ ($n \geq 2$) is called* contagious *if two conditions hold: (1) c is a* common strategy*,i.e., it is in the set of individual strategies of each of the n players, (2) c is* pairwise contagious*, i.e., for any two different players i and j, c is the strict best response of player i to any (strategy) profile θ in which $\theta_j = c$.*

A profile of strategies π is called contagious if for some contagious strategy c, $\pi_i = c$ for every player i, i.e., $\pi = \widehat{c} \equiv$ the profile in which every player plays the strategy c.

It is easy to see that any contagious profile is a Nash equilibrium. In the rest of this section we discuss the applicability, stability, uniqueness and focality of contagious equilibria.

Despite their highly restrictive definition, contagious strategies and equilibria are of interest in a variety of different areas. In price competition for example, the lowest profitable price p charged for a good sold by n symmetric sellers, is a contagious strategy. In auction theory, the highest bid b that each of n symmetric bidders is willing to pay for a common-valued item being auctioned, is a contagious strategy. Our lead example in this paper is a prisoner's dilemma game, studied in greater detail in Section 4. Such a game is constructed by a prosecutor to incentivize defendants who are guilty of a joint crime, to adopt and play a confession strategy, $\boldsymbol{y}$, that is contagious.

Whether the players will play a contagious strategy c depends on its stability, resilience and focality properties that are discussed in the theorem below. But before we state the Theorem and its proof, it is useful to review the *resilience hierarchy* of strategy profiles presented in Kalai and Kalai (2022).

As summarized in the table below, any profile π in an n-person game Γ may be classified and described according to its *resilience level*, $\rho(\pi)$; or equivalently according to its *dual* level of *critical mass* $\kappa(\pi)$, $\kappa(\pi) = n - \rho(\pi)$.

$\rho(\pi) = -1$	$\rho(\pi) = 0$	$\rho(\pi) = 1, 2, ..., n-3$	$\rho(\pi) = n-2$	$\rho(\pi) = n-1$
$\kappa(\pi) = n+1$	$\kappa(\pi) = n$	$\kappa(\pi) = n-1, ..., 4, 3$	$\kappa(\pi) = 2$	$\kappa(\pi) = 1$
π is not Nash eqm	π is a fragile Nash eqm	Nash eq'a π arranged in increasing resilience	π is nearly dom't eqm	π is dom't stgy eqm

The number $\rho(\pi)$ specifies the maximal number of π-defectors, i.e., players i who choose a strategy different from π_i, who "cannot disrupt the best response property of π" in two equivalent formal senses: (R1) $\rho(\pi)$ is the largest integer $d = 0, 1, ...n-1$, such that in any profile θ with d or fewer π-defectors π_i is a best response of every π-*loyalist*, i.e., player i with $\theta_i = \pi_i$. Equivalently, (R2) $\rho(\pi)$ is the largest integer d such that at any profile θ with d or fewer π-defectors π_i is a best response of *every player i*.

The number $\kappa(\pi)$ specifies the minimal number of π-loyalists l needed to justify the play of π as equilibrium in two equivalent senses: (K1) $\kappa(\pi)$ is the smallest integer l such that at any profile θ with l or more π-loyalists, π_i is a best response of *every π-loyalist*. Equivalently, (K2) $\kappa(\pi)$ is the smallest integer l such that at any profile θ with l or more π-loyalists, π_i is a best response of *every player i*.

The resilience of π may also be interpreted as the strict lower bound on a number of players x, i.e., $\rho(\pi) < x$, that a malicious agent needs to control ex-ante in order to create a profile θ at which π_i is not optimal for some player i. In other words, to assure the ability to induce a defection from π, a malicious agent needs to control the strategy choices of $x \geq \rho(\pi) + 1$ players.

It is easy to see that $\rho(\pi) \geq 0$ (or $\kappa(\pi) \leq n$) iff π is a Nash equilibrium. The reference to πs with $\rho(\pi) = 0$ as fragile equilibria, captures the property that a *single defector* from such a π can incentivize the defection of other(s). It is also easy to see that if $\rho(\pi) = n-1$ then π_i of every player i is a dominant strategy. The term nearly dominant when $\rho(\pi) = n-2$ captures the property that "the play π_j by any single player j is enough to makes π_i a dominant strategy for every player $i \neq j$," that is, at every profile θ with $\theta_j = \pi_j$, π_i is a best response of player i. Trivially, a profiles π is not a Nash equilibrium iff $\rho(\pi) = -1$, or equivalently $\kappa(\pi) = n+1$.

When the conditions of resilience are applied to contagious equilibrium $\widehat{c}$, $\widehat{c}$ becomes the most prominent resilient equilibrium in the sense described in the following theorem.

Theorem 1. *The* ***Resilience Focality of Contagious Equilibrium.***[2] *Consider any n-person game with a contagious strategy c: (1) the contagious profile $\widehat{c}$ is of resilience $\rho(\widehat{c}) \geq n-2$; moreover (2) any profile $\pi \neq \widehat{c}$ has resilience $\rho(\pi) \leq 0$.*

Proof. (1) follows from the definitions of κ and the contagiousness condition, first $\kappa(\widehat{c}) \leq 2$, and thus $\rho(\widehat{c}) \geq n-2$.

To prove (2), we assume that π is a profile of resilience $\rho(\pi) \geq 1$ and show that $\pi = \widehat{c}$. It suffices to show that for every player i, $\pi_i = c$.

[2] The author thanks Julien Manili for suggesting the generalized version of the theorem and its proof.

Construct any profile π' that coincides with π for all players, except for some player $j \neq i$ (recall that $n \geq 2$) define $\pi'_j \equiv c$. Since π' is obtained from π by at most 1 defection, π's best response properties are sustained at π'.

(1) π is a Nash equilibrium, because $\rho(\pi) \geq 1$,

(2) $\pi'_i = \pi_i$ is a best response of player i at π; and

(3) π'_i is a best response to π', because $\rho(\pi) \geq 1$, and

(4) $\pi'_i = c$, by the strict best response property of the contagious strategy c. Thus,

$\pi_i = \pi'_i = c$. □

Therefore, we can summarize the possible resilience levels of all pure or mixed strategy profiles in a game with a contagious strategy as follows:

Resilience of profiles $\pi \neq \widehat{c}$ in an n-person game with a contagious strategy c:

$\rho(\pi) = -1$ or 0	$\rho(\pi) = 1, 2..., n-3$	$\rho(\pi) = n-2$ or $n-1$
$\kappa(\pi) = n+1$ or n	$\kappa(\pi) = n-1, ..., 4, 3$	$\kappa(\pi) = 2$ or 1
$\pi \neq \widehat{c}$ **fragile or not Nash**	$\varnothing$ no πs	$\widehat{c}$ **dom't or nearly dom't**

Remark 1. *The theorem leads to the following conclusions.*

A. From the first part of the theorem, a contagious equilibrium $\widehat{c}$ of a game is the most resilient profile: $\rho(\widehat{c})$ must be either $n-1$ or $n-2$. In other words $\widehat{c}$ must be the unique dominant strategy equilibrium if one exists; or a nearly-dominant strategy equilibrium, if dominant strategy equilibrium does not exist.

B. Every profile $\pi \neq \widehat{c}$ must be a fragile Nash equilibrium or not a Nash equilibrium.

C. For $n > 2$, $\widehat{c}$ must be the unique most resilient profile *of the game in two strong senses: it is the unique dominant strategy equilibrium if one exists, and it is the unique nearly-dominant-strategy equilibrium otherwise. Moreover, c is the only contagious strategy in the game.*

D. If $n = 2$, the game may have a multiplicity of most-resilient equilibria.

3. PRISONER'S DILEMMA GAMES

			II	
		x		***y***
	x	0,0		-6,0
I				
	y	0,-6		-5,-5

A traditional PD game is described by the payoff table above that specifies possible jail sentences to two suspects in a joint crime, I and II. While each of the two suspects has no incentive to confess, a clever prosecutor can construct a prisoner's dilemma game in which rational selfish play by two guilty criminals leads to confessions.

Example 1. ***Classical PD***. *Simultaneously, prior to their trial, the prosecutor offers each accused criminal two choices:* ***y****, confess and present convincing evidence that the joint crime was committed; or* ***x****, deny the occurrence of the crime; It is assumed that if they committed the crime, each one can play* ***x*** *or* ***y****; but if they did not commit the crime, each can only play* ***x****. The judge agrees to impose the jail sentences specified in the table. Thus if both partners deny the crime, they* *will obtain the payoffs* $u(\boldsymbol{x},\boldsymbol{x}) = [u_1(\boldsymbol{x},\boldsymbol{x}), u_2(\boldsymbol{x},\boldsymbol{x})] = [0,0]$*. Similarly* $u(\boldsymbol{y},\boldsymbol{y}) = [-5,-5]$*, and* $u(\boldsymbol{x},\boldsymbol{y}) = [-6,0]$*.*

The classical PD argument is that if they did not commit the crime, they can only play $(\boldsymbol{x},\boldsymbol{x})$ and be released. However, if they did commit the crime they will play the $(\boldsymbol{y},\boldsymbol{y})$ strategies and each would end up with a 5 year jail sentence. Why would the guilty criminals choose $(\boldsymbol{y},\boldsymbol{y})$?

The game has two pure strategy equilibria: both players deny, $\widehat{\boldsymbol{x}} = (\boldsymbol{x},\boldsymbol{x})$, and both confess, $\widehat{\boldsymbol{y}} = (\boldsymbol{y},\boldsymbol{y})$. It is easy to see that the strategy $\boldsymbol{y}$ (weakly) dominates the strategy $\boldsymbol{x}$. Equivalently, expressed in terms of resilience, $\widehat{\boldsymbol{y}}$ is more resilient than $\widehat{\boldsymbol{x}}$, $\rho(\widehat{\boldsymbol{y}}) = 1 > \rho(\widehat{\boldsymbol{x}}) = 0$. Thus, even though $\widehat{\boldsymbol{x}}$ is Pareto superior in payoffs, it is the less viable play of the game.

4. A Dilemma of more than two prisoners

The next example shows that resilience superiority goes further than just ruling out weakly dominated strategies. It enables the imposition of jail sentences on guilty parties in RICO trials with more than two conspirators, even when confession is not a dominant strategy. In the example, M is a mastermind who might have recruited ten conspirators, $i = 1, 2, ..., 10$, to commit a crime. The prosecutor's objective is to impose a 5 year jail sentence on M if the crime was committed, but to let M go free otherwise. It is assumed that the judge will not impose jail sentence on any person without evidence of the committed crime.

Example 2. ***Ten defendant RICO Trial****: Prior to the start of the trial, simultaneously and privately the prosecutor offers each of the ten accused conspirators the options* $\boldsymbol{y}$ *or* $\boldsymbol{x}$ *as above. The judge agrees to follow the following punishment scheme:*

Case 1: complete denial: If all accused conspirators deny the crime, $\widehat{\boldsymbol{x}}$*, then no jail sentence will be imposed on anyone including* M*. Moreover the accused conspirators will be declared innocent, with payoffs* $u_i(\widehat{\boldsymbol{x}}) = 1$ *for* $i = 1, 2, ..., 10$*. (* $u_i(\widehat{\boldsymbol{x}}) > 0$ *indicates that for these accused conspirators being declared innocent has added value beyond the* 0 *jail time.)*

Case 2: some confessions: If some of the conspirators $i = 1, 2, ..., 10$ *confess, then: every confessor among* $i = 1, ..., 10$ *is let go with no jail time, i.e.,* $u_i = 0$*; and every denier among* $i = 1, ..., 10$ *is sentenced to* 5 *years, i.e.,* $u_i = -5$*. The mastermind* M *is also sentenced to* 5 *years.*

Remark 2. *In the current example it is assumed that* $u_i(\widehat{\boldsymbol{x}}) = 1$*, as opposed to* $u_i(\widehat{\boldsymbol{x}}) = 0$ *that was assumed in the classical two-defendant example. This serves to illustrate the point that under the more nuanced resilience analysis, player will play the strategy* $\boldsymbol{y}$*, even when it does not dominate* $\boldsymbol{x}$*.*

How would the game above be played by the ten conspirators?

First, if the crime was not committed, they can only play the profile $\widehat{\boldsymbol{x}}$, and everybody will be released. However, if the crime was committed the conspirators game is more interesting.

The reader can verify that the game played by the ten guilty conspirators has two pure strategy Nash equilibria: everybody confesses, $\widehat{\boldsymbol{y}}$, and everybody denies, $\widehat{\boldsymbol{x}}$. As discussed informally below, it is easy to see that resilience explains why $\widehat{\boldsymbol{y}}$ is significantly more robust than $\widehat{\boldsymbol{x}}$:

Any defection from $\widehat{\boldsymbol{x}}$ to $\boldsymbol{y}$, even by a **single** opponent, is a strong motivation for any denier to also defect from $\widehat{\boldsymbol{x}}$ to $\boldsymbol{y}$. Adopting the terminology in the resilience scale of Kalai and Kalai (2022), we refer to $\widehat{\boldsymbol{x}}$ as *fragile equilibrium.*

On the other hand, only a defection $\widehat{\boldsymbol{y}}$ to $\boldsymbol{x}$ by **all nine** opponents can motivate a confessor to also defect from $\widehat{\boldsymbol{y}}$ to $\boldsymbol{x}$. Put differently, "conditional on the choice of $\boldsymbol{y}$ by one conspirator" it is a dominant strategy for any other conspirator to also choose $\boldsymbol{y}$. Adopting the terminology in Kalai and Kalai (2022), we refer to such a $\widehat{\boldsymbol{y}}$ as a *nearly dominant strategy equilibrium.*

As discussed formally in the next section, the near dominance property of $\widehat{\boldsymbol{y}}$ versus the fragility of $\widehat{\boldsymbol{x}}$, serves as a focal point that incentivize conspirators to confess. Indeed, it is not surprising that prosecutors prefer a larger number of conspirators (e.g., greater than ten in the example above), as this increases the significance of both: (1) the fragility of $\widehat{\boldsymbol{x}}$ and (2) the near-dominance property of $\widehat{\boldsymbol{y}}$.

5. Stability and resilience in PD games

The fragility of $\widehat{\boldsymbol{x}}$ discussed above is in stark contrast to strong stability arguments of traditional game theory that favor of $\widehat{\boldsymbol{x}}$: $\widehat{\boldsymbol{x}}$ is a strict Nash equilibrium that strictly Pareto dominates $\widehat{\boldsymbol{y}}$, $\widehat{\boldsymbol{x}}$ is *strong* in the sense of Aumann (1957), it is *coalition proof* in the sense of Bernheim, Whinston, and Peleg (1987), and it is trembling hand *perfect* in the sense of Selten (1975).

It is easy to see the rationale for the refinements that are based on Pareto superiority. First, when playing $\widehat{\boldsymbol{x}}$ every conspirator's payoff is 1. Moreover, any other play of this game yields every conspirator a payoff of 0 or -5. Thus any alternative play, chosen individually or even coordinated by a group, would result in a strict loss to every participant. In this sense $\widehat{\boldsymbol{x}}$ is the *uniquely strict Pareto superior* profile in this game. This is a strong justification for individual and coalitional rationale for the play of $\widehat{\boldsymbol{x}}$.

However, a player who fears defections may ask the additional question: If some opponent defects from $\widehat{\boldsymbol{x}}$ and plays $\boldsymbol{y}$, is it still optimal for me to play $\boldsymbol{x}$? The negative answer to this question may lead to an equilibrium "mutiny," in which the equilibrium play of $\widehat{\boldsymbol{y}}$ becomes more plausible than that of $\widehat{\boldsymbol{x}}$. Thus, it is not clear which of the two equilibria $\widehat{\boldsymbol{y}}$ or $\widehat{\boldsymbol{x}}$ is more viable.

To assess the viability of the two, we compare the resilience of the two. It is easy to check that the confession strategy $\boldsymbol{y}$ is contagious. Thus we may substitute $\boldsymbol{y}$ for c in the resilience scale of contagious strategies from section 2 , to obtain the resilience scale below for the confession equilibrium $\widehat{\boldsymbol{y}}$, and for any profile $\pi \neq \widehat{\boldsymbol{y}}$.

Resilience of profiles π in an n-defendant prisoner's dilemma game with a contagious confession strategy $\boldsymbol{y}$:

$\rho(\pi) = -1$ or 0	$\rho(\pi) = 1, 2..., n-3$	$\rho(\pi) = n-2$ or $n-1$
$\kappa(\pi) = n+1$ or n	$\kappa(\pi) = n-1, ..., 4, 3$	$\kappa(\pi) = 2$ or 1
every $\pi \neq \widehat{\boldsymbol{y}}$. not Nash or fragile Nash	$\varnothing$ no πs	**only $\widehat{\boldsymbol{y}}$.** dom't or nearly dom't

One can easily compare the resilience of the two equilibria in our 10 defendants RICO game to see that $\rho(\widehat{\boldsymbol{y}}) = 8$ defectors and $\rho(\widehat{\boldsymbol{x}}) = 0$ defectors. Equivalently from the dual index of critical mass, it is rational for any $\kappa(\widehat{\boldsymbol{y}}) = 2$ $(= 10 - 8)$ or more players to play $\boldsymbol{y}$, whereas the play of any other profile π is rational only if all

$\kappa(\pi) = 10$ $(= 10 - 0)$ players play it. In other words, the play of $\widehat{\boldsymbol{y}}$ relies on having a minimum of two confessors and it can withstand up to eight defectors, whereas the play of $\widehat{\boldsymbol{x}}$ relies on the participation of all ten deniers and it cannot withstand any defectors. Notice that at $n = 3$, this viability advantage of $\widehat{\boldsymbol{y}}$ over any alternative profiles is strictly positive and increases with n.

The zero resilience of $\widehat{\boldsymbol{x}}$ is also in direct contrast with Selten's view, who considers $\widehat{\boldsymbol{x}}$ as a trembling hand *perfect equilibrium.* The conflict is due to the different notion of robustness that underlies the definitions in Kalai and Kalai (2022) and the ones in Selten (1975).

Selten's view of robustness requires $\boldsymbol{x}$ to be a best response to profiles $\widetilde{\boldsymbol{x}}$ in which *every player has an infinitesimal probability of deviating from the choice of* $\boldsymbol{x}$*.* Kalai and Kalai's view on the other hand, is that $\boldsymbol{x}$ should be a best response to profiles $\widetilde{\boldsymbol{x}}$ in which *all players play* $\boldsymbol{x}$*, except for a random small group of unrestricted deviators.* It is easy to see that $\boldsymbol{x}$ is an optimal strategy when all the opponents are infinitesimal deviators; but $\boldsymbol{x}$ is not optimal, even if only one opponent is an unrestricted deviator.

In support of the reasoning above it is natural to assume that in the recent publicized RICO trial, conspirators who chose to confess were motivated by fear of unrestricted defectors, ones who may defect from $\boldsymbol{x}$ to $\boldsymbol{y}$ with significant probability. The substantial legal experience of the confessors in this trial suggests that their fears were rational.

6. Fear-of-defection in related applications

In general terms, if fear-of-defection to a contagious strategy motivates players to defect, then fear-of-defection can become a self-fulfilling equilibrium. But while such equilibrium plays a positive role on the outcomes of prisoner's dilemma games, fear-of-defection often gives rise to negative social outcomes. This phenomenon was cleverly addressed in the famous "The Only Thing We Have to Fear is Fear Itself" speech, the inaugural presidential address in which F.D. Roosevelt tried to undo an equilibrium of fear. Similarly, in many markets, fear of price drops motivates further selling that leads to additional price drops. Fear of inflation may lead to excessive buying that accelerates the inflation and higher levels of excessive buying. Government insurance of bank deposits is an example of a mechanism to combat bad outcomes that may result from equilibrium of fears.

Furthermore, beyond equilibrium explanations, fear of defection is directly justified in games that are not fully specified. In an example of centralized production game discussed in Kalai and Kalai (2022), a single producer of chips, e.g., Taiwan, chooses the type of chips to produce, and n users of chips, e.g., car manufacturers, choose to produce items that use the type of chips the producer makes. The equilibrium in which all players focus on the same type of chips is fragile, because the chip users fear a switch by the producer to another type of chips (ones that cannot be used in their products). Their fear may be based on concrete physical or political issues that are not modeled in the game. For example, raw material needed for the production of the equilibrium type of chips is depleted, or a foreign entity takes control of the production facilities. A more complete description of the game that takes all physical and political concerns into account is most-often intractable.

7. CONCLUSION, PAST AND FUTURE RESEARCH

Mechanism designers and social planners can learn from prosecutors, and use the more nuanced equilibrium considerations in their designs. In our example of a RICO trial, the prosecutor constructs a punishment scheme that results in two possible equilibrium play of guilty conspirators: a higher resilience equilibrium in which they all confess, $\widehat{\boldsymbol{y}}$, versus lower resilience ones in which they deny the charges. The prosecutor's goals are accomplished under the assumption that the conspirators would choose to play the one of higher resilience.

A small number of theoretical mechanism-design papers applied and studied applications similar to the approach of the prosecutor above. One early application is in the design of distributed computing systems that should work properly despite the possibility of a bounded number of faulty components, see for example Goldreich et al (1998). Eliaz (2002) showed that a mechanism designer who uses a resilient game theoretic equilibrium can perform "fault tolerant implementation," i.e., implement social economic goals even in the presence of a bounded number of irrational or poorly informed economic agents. Abraham et al (2006) were first to coin the term *resilient* to mean the ability to withstand opponent defections. Similar to Eliaz, they showed that resilient equilibria implement computer science tasks, even in the presence of a bounded number of agents who may be faulty. Schelling (1973) studied cardinal computations of resilience in an n-prisoner's dilemma game. Kalai and Kalai (2022) introduced and studied the measure of resilience for the strategic equilibria of general n-person games, together with its dual measure of critical mass.

Prosecutors who prosecute RICO trials deal with practical and theoretical issues that go well beyond our analysis. The colloquial saying that "the first to squeal gets the best deal" captures the fact that confessions in such trials are decided at critical points in continuous time. Moreover, the games are subject to unfolding incomplete information on the part of the prosecutor and all the participants. For example, based on earlier confession/denial decisions and information provided in earlier confessions, the prosecutor must decide at any point in time on: (1) to whom, if at all, she should offer the next deal, and (2) what should be the terms of the deals to follow.

As known from earlier papers on PD games, dynamic play with random moves and incomplete information may have drastic effect on the incentives of conspirators to confess, see for example Kalai (1981) and Nishihara (1997). The resilience index, used in the analysis of this paper was restricted to one-shot simultaneous-move games of complete information. A straightforward extension of the same analysis to general dynamic pre-trial games, would requires a generalization of the Kalai and Kalai (2022) resilience/critical mass indices to dynamic games.

8. References

Abraham, I., D. Dolev, R. Gonen, and J. Halpern (2006), “Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation.” In *Proceedings of the. 25th ACM Symposium on Principles of Distributed Computing*, 53–62.

Aumann, R. (1959), “Acceptable points in general cooperative n-player games,” in Tucker, A. and Luce, R., Eds., *Contributions to the Theory of Games IV*, Princeton Univ. Press, Princeton.

Bernheim, B. D.,B. Peleg, and M. Whinston (1987), “Coalition-Proof Nash Equilibria I. Concepts,” *Journal of Economic Theory*, Volume 42, Issue 1, 1–12.

Eliaz, K. (2002), “Fault-tolerant implementation," *Review of Economic Studies*, 69(3), 589–610.

Goldreich, O., S. Goldwasser, and N. Linial (1998), “Fault-tolerant computation in the full information model,” *SIAM Journal on Computing* 27.2: 506–544.

Kalai, A.T. and E. Kalai (2022), "Beyond Dominance and Nash: Ranking Strategic Equilibrium by Critical Mass," https://www.researchgate.net/publication/371782905, forthcoming in *Games and Economic Behavior*.

Kalai, E. (1981), "Preplay Negotiations and the Prisoner’s Dilemma," *Mathematical Social Sciences* 1, 375-379.

Nishihara, K. (1997), "A Resolution of N-person Prisoners’ Dilemma," *Economic Theory*, 10, 531-540.

Selten, R. (1975), “Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games,” *International Journal of Game Theory* (4), 22–55.

Schelling, T. (1973), “ Hockey Helmets, Concealed Weapons, and Daylight Saving, a Study of Binary Choices With Externalities," *Journal of Conflict Resolution*, 17, No. 3: 381-42.

Northwestern University
E-mail address: kalai@kellogg.northwestern.edu