SOBRAPO

# MIXED EQUILIBRIUM IN 2 × 2 NORMAL FORM GAMES: WHEN BURNING MONEY IS RATIONAL

Filipe Costa de Souza[1]* and Leandro Chaves Rêgo[2]

**ABSTRACT.** We discuss the rationality of burning money behavior from a new perspective: the mixed Nash equilibrium in 2 × 2 normal form games. We support our argument analyzing the first-order derivatives of the players' mixed equilibrium expected utility payoffs with respect to their own utility payoffs. We establish necessary and sufficient conditions that guarantee the existence of negative derivatives. In particular, games with negative derivatives are the ones that create incentives for burning money behavior since such behavior in these games improves the player's mixed equilibrium expected utility payoff. We show that a negative derivative of some player $i$'s mixed equilibrium expected utility payoff occurs if, and only if, he has a strict preference for one of the strategies of the other player. Moreover, negative derivatives always occur when they are taken with respect to player $i$'s highest and lowest game utility payoffs. We also present sufficient conditions that ensure that such derivatives are always non-negative in finite normal form games.

**Keywords**: mixed Nash equilibrium, payoff reduction, collaborative dominance, security dilemma.

## 1 INTRODUCTION

Based on the concept of forward induction proposed by Kohlberg & Mertens (1986) and, especially, on the idea of iterative elimination of weakly dominated strategies, Van Damme (1989) and Ben-Porath & Dekel (1992) studied the effects of burning money as a way to signal future actions allowing (in some games) the player who burned money to achieve his or her preferred equilibrium point in the game. Ben-Porath & Dekel (1992) proved that in games in which a player has a strict preference for an equilibrium point, and if this player can self-sacrifice, then, based on the forward induction rationality and iterative elimination of weakly dominated strategies, such player will achieve his most preferred outcome.

*Corresponding author.

[1]Universidade Federal de Pernambuco, Departamento de Ciências Contábeis e Atuariais, 50670-901 Recife, PE, Brasil. E-mail: filipe.costas@ufpe.br

[2]Universidade Federal do Ceará, Departamento de Estatística e Matemática Aplicada, Bloco 910, Campus do Pici, 60455-760 Fortaleza, CE, Brasil. E-mail: leandro@dema.ufc.br.

To test the theoretical prediction that the opportunity of burning money brought an advantage for the player that can do it, Huck & Müller (2005) developed an experimental study based on the Battle of the Sexes game. Initially, the authors considered a sequential burning money game where, first, player 1 had to choose between two utility payoff matrices, one in which the payoffs (an aside: in this paper we use the terms utility and payoffs as synonyms since we never refer to material payoffs) were identical to the original Battle of the Sexes (i.e. without burning money) and another in which all payoffs of player 1 were reduced (i.e. a game with burning money), then, player 2 observes the initial choice made by player 1 and, thereafter, both choose their strategies simultaneously. With this experiment, the authors were able to detect the first mover advantage (even though the observed frequency was lower than the one expected by the theory, which completely solves the coordination problem). On the other hand, when the game was exposed in its reduced normal form representation, the result was very different and no first move advantage was detected. As an explanation for this divergent result, the authors suggested that the participants were unable to perform the iterative elimination of weakly dominated strategies. Other discussions on forward induction and elimination of dominated strategies can be found in Brandts & Holt (1995).

In a critical analysis of Ben-Porath & Dekel's ideas, Myerson (1991, p. 194-195) argues that in the context of sequential equilibrium, player 1's act of burning money can be interpreted by player 2 as irrational or as an error and, for this reason, should not be considered in the prediction of player 1's future behavior.

In addition, Van Damme (1989) and Ben-Porath & Dekel (1992) recognize that if all players can signal their intentions by burning money, then the final outcome of the game may be inefficient. The authors also emphasize that the order in which players can burn money define their power on the game, since the last one always has the opportunity to make a counter-signal that makes the early signal invalid. For this reason, the last player to signal has the greater advantage.

Burning money behavior can also occur just for some specific strategy profiles, as discussed in Fudenberg & Tirole (1991, p. 9). The authors propose the game presented in Figure 1. In this game, there is a unique (and inefficient) pure equilibrium $(U, L)$.

|   | $L$ | $R$ |
|---|-----|-----|
| $U$ | (1, 3) | (4, 1) |
| $D$ | (0, 2) | (3, 4) |

**Figure 1**

But suppose that player 1 can show player 2 (before player 2 chooses her action) that the strategy $U$ is not strongly dominant for him. For example, suppose that player 1 changes the game and reduce his payoffs if $U$ occurs by 2 utils. So in this new game, there is also a unique pure equilibrium point $(D, R)$, which is now efficient. For more details, see the original argumentation in Fudenberg & Tirole (1991).

It is important to highlight that, as in Fudenberg & Tirole (1991), we focus on burning money behavior before the beginning of the game, and this is common knowledge among players. One can think of it as a pre-contract which imposes him a predefined penalty if some specified strategy

profile is played. Moreover, the existence of such contract is common knowledge among players. We emphasize that in this paper we do not study burning money behavior in the context of signaling games, but we see it as a form of self-sacrificing reducing the player's utility payoff.

Based on the exposed arguments, a burning money behavior may be an important mechanism of cooperation, and also allow players to achieve efficient outcomes. Moreover, once we assume that players are capable to self-sacrifice (it is easier to suppose that players can reduce their own payoff than that they can increase it) it is natural to assume that if the same penalty is imposed by an external and impartial agent, the same result will emerge.

In this paper, we discuss the rationality of burning money behavior from a new perspective: the mixed Nash equilibrium in $2 \times 2$ normal form games. We establish necessary and sufficient conditions for the existence of negative and non-positive derivatives of some player $i$'s mixed equilibrium expected utility with respect to his own payoffs. In particular, games in which negative derivatives occur are the ones that create incentives for burning money behavior since it would improve player $i$'s mixed equilibrium expected utility.

Supported by the concept of collaborative dominance, proposed by Souza & Rêgo (2013), we show that a negative derivative of some player $i$'s mixed equilibrium expected utility occurs if, and only if, player $j$ has a strictly collaboratively dominant strategy for player $i$ (i.e., if player $i$ has a strict preference for one of the strategies of player $j$). Moreover, negative derivatives always occur with respect to player $i$'s highest and lowest game utility payoffs. We also evaluate how player $j$ reacts to the reduction in a payoff made by player $i$, i.e., how $j$'s mixed equilibrium strategy varies given a change in player $i$'s utility payoffs. We show that if the derivative of player $i$'s mixed equilibrium expected utility taken with respect to some of his utility payoff, say $a$, is negative, then if player $i$ reduces $a$, he is inducing player $j$ to choose more often the strategy that is strongly collaboratively dominant for him, player $i$. This fact allows player $i$ to achieve a more desired result. Therefore, player $i$ should reduce utility with respect to the payoff that provides the greater increase in the probability of player $j$ choosing the strategy that is strongly collaboratively dominant for him, player $i$.

We also point out some difficulties to extend the proposed analysis for more general games, especially regarding how players will react to burning money behavior of the other players. On the other hand, we present sufficient conditions for the existence of non-negative derivatives of some player's equilibrium expected utility taken with respect to one of him payoffs in finite normal form games; and finally, to illustrate our approach, we revised some of Jervis (1978) conclusions about what makes cooperation more likely in the security dilemma.

For this purpose, the remaining of the paper is structured as follows: in Section 2, we analyze the first-order derivative of some player $i$'s mixed equilibrium expected utility in a $2 \times 2$ normal form game with respect to his own utility payoffs; in Section 3, we discuss the necessary and sufficient conditions that guarantee the existence of negative (or at least non-positive) derivatives of mixed equilibrium expected utility which would justify the burning money behavior; in Section 4, we study the problem of finding the best utility reduction strategy. In Section 5, we discuss the difficulties that prevent the extension of our conclusions to a more general class of games; and in

Section 6, to illustrate some of the applications of our results, we analyze the security dilemma in light of our conclusions about burning money behavior in $2 \times 2$ games. Finally, the conclusions are presented in Section 7.

## 2  THE ANALYSIS OF FIRST-ORDER DERIVATIVES

We start presenting some general definitions about normal form games. Let $G$ be a finite normal form game, where $G = (N, (S_i)_{i \in N}, (U_i)_{i \in N})$ and $N = \{1, \dots, n\}$ is a finite set of players, $S_i$ is a finite set of (pure) strategies available for player $i \in N$, and $U_i : \times_{i \in N} S_i \to \mathbb{R}$ is the utility function of player $i \in N$. A strategy profile $s = \times_{i \in N} s_i$ is a collection of pure strategies, one for each player in $G$, and $S = \times_{i \in N} S_i$ is the set of all pure strategy profiles. Additionally, let $\Delta(S_k)$ be the set of all mixed strategies of player $k$, where $\sigma_k(s_k)$ indicates the probability that $k$ gives to his pure strategy $s_k$ when implementing the mixed strategy $\sigma_k \in \Delta(S_k)$. The support of a mixed strategy $\sigma_k$ is the subset of the pure strategies, $S_k$, that receive positive probability according to $\sigma_k$. Therefore, player $i$'s expected utility when the mixed strategy profile $\sigma$ is implemented is determined by $U_i(\sigma) = \sum_{s \in S} (\prod_{j \in N} \sigma_j(s_j)) U_i(s)$. Finally, let $\sigma_{-k}$ be the set of all mixed strategies for players different from $k$ in $\sigma$.

Consider now the particular structure of a $2 \times 2$ normal form game, as shown in Figure 2.

|   | $L$ | $R$ |
|---|---|---|
| $U$ | $(a, e)$ | $(b, f)$ |
| $D$ | $(c, g)$ | $(d, h)$ |

**Figure 2**

Let $p$ be the probability of player 1 choosing pure strategy $U$ and $1 - p$ the probability of choosing pure strategy $D$. Similarly, let $q$ be the probability of player 2 choosing pure strategy $L$ and $1 - q$ the probability of choosing pure strategy $R$. We want to restrict our attention to the case where there is only one mixed equilibrium in the non-degenerated sense (no restriction is made on the number of pure equilibrium). In this case, it is well-known (see, e.g., Rasmusen, 1996) that the mixed equilibrium strategies are given by:

$$p = \frac{h - g}{e - f - g + h} \quad \text{and} \tag{2.1}$$

$$q = \frac{d - b}{a - b - c + d}. \tag{2.2}$$

Thus, we can write the expected utility of the players in the mixed Nash equilibrium as a function of the utilities payoffs of each player, as follows:

$$EU_1 = \frac{ad - bc}{a - b - c + d} \quad \text{and} \tag{2.3}$$

$$EU_2 = \frac{eh - fg}{e - f - g + h}. \tag{2.4}$$

Once the mixed equilibrium expected utilities are written only in terms of each player own utility payoffs, we can study the variation of the mixed equilibrium expected utility with respect to changes in a given utility payoff through a first-order derivative analysis.

We restrict our attention to marginal changes in a single utility payoff of the game, i.e., changes that do not alter more than one payoff simultaneously and do not alter the general order of player's preferences. By general order of the payoffs we mean: if $u$ and $v$ are two utility payoffs of some player in the game and $u > v$, then, after a marginal change in payoffs, it should not happen that $v > u$. There are two main ideas behind marginal changes: the first one is that the nature of the equilibrium set remains the same, i.e., none mixed or pure equilibrium is added or removed from the original equilibrium set, and changes only affect the probabilities of mixed strategies without altering its support (this condition is enough to guarantee this fact in our restricted domain of $2 \times 2$ games with a unique mixed equilibrium in the non-degenerate sense); the second one is that we are assuming that, there is no functional relationship among two or more utilities payoffs. These assumptions are needed because we use the payoff order to characterize the games that provide incentives to burning money behavior. Furthermore, for those readers interested in the discussion about payoff changes that alter the general order of players' preferences, we recommend Englemann (1994). Thus, for player 1, we have:

$$\frac{\partial EU_1}{\partial a} = \frac{(c-d)(b-d)}{(a-b-c+d)^2}; \tag{2.5}$$

$$\frac{\partial EU_1}{\partial b} = \frac{(c-d)(c-a)}{(a-b-c+d)^2}; \tag{2.6}$$

$$\frac{\partial EU_1}{\partial c} = \frac{(b-a)(b-d)}{(a-b-c+d)^2}; \tag{2.7}$$

$$\frac{\partial EU_1}{\partial d} = \frac{(b-a)(c-a)}{(a-b-c+d)^2}; \quad \text{and} \tag{2.8}$$

$$\frac{\partial EU_1}{\partial e} = \frac{\partial EU_1}{\partial f} = \frac{\partial EU_1}{\partial g} = \frac{\partial EU_1}{\partial h} = 0.$$

On the other hand, for player 2, we have:

$$\frac{\partial EU_2}{\partial e} = \frac{(g-h)(f-h)}{(e-f-g+h)^2}; \tag{2.9}$$

$$\frac{\partial EU_2}{\partial f} = \frac{(g-h)(g-e)}{(e-f-g+h)^2}; \tag{2.10}$$

$$\frac{\partial EU_2}{\partial g} = \frac{(f-e)(f-h)}{(e-f-g+h)^2}; \tag{2.11}$$

$$\frac{\partial EU_2}{\partial h} = \frac{(f-e)(g-e)}{(e-f-g+h)^2}; \quad \text{and} \tag{2.12}$$

$$\frac{\partial EU_2}{\partial a} = \frac{\partial EU_2}{\partial b} = \frac{\partial EU_2}{\partial c} = \frac{\partial EU_2}{\partial d} = 0.$$

Through these general expressions, we can evaluate how players' mixed equilibrium expected utilities vary, when their respective utility payoffs change, analyzing the sign of the derivatives in Equations (2.5)-(2.8) and (2.9)-(2.12).

## 3    ANALYZING THE SIGN OF THE DERIVATIVES

In this section, we discuss necessary and sufficient conditions that guarantee that the derivative of some player expected utility with respect to one of his utility payoff is negative (or at least non-positive). Bur first, in Lemma 1, we summarize the conditions that guarantee that a 2 × 2 game has a unique mixed equilibrium in the non-degenerated sense. Before stating this Lemma, let us present a definition.

**Definition 1.** *Imagine a 2 × 2 game in strategic form as stated in Figure 2. We say that player 1 (resp. 2) is always indifferent between strategies U and D (resp. L and R) if $U_1(U, s_2) = U_1(D, s_2)$ for all $s_2$ in $\{L, R\}$ (resp. if $U_2(s_1, L) = U_2(s_1, R)$ for all $s_1$ in $\{U, D\}$).*

**Lemma 1.** *In any 2 × 2 normal form game, there is a unique mixed Nash equilibrium in the non-degenerated sense if, and only if, no player is always indifferent between his strategies nor has a weakly (or strongly) dominant strategy.*

**Proof.**    In a 2×2 normal form game where no player is always indifferent between his strategies nor has a weakly (or strongly) dominant strategy, we have two possible scenarios for player 1 ($a > c$ and $d > b$; or $c > a$ and $b > d$) and two for player 2 ($e > f$ and $h > g$, or $f > e$ and $g > h$). In both scenarios for player 1, player 2 has a unique mixed strategy of the form $(q, 1 - q)$, where $q = \frac{d-b}{a-b-c+d}$, that makes player 1 indifferent between $U$ and $D$. Since $a > c$ and $d > b$, then $q$ is in the interval $(0, 1)$, which proves that the mixed equilibrium strategy for player 2 is unique and non-degenerated. Similarly, in both scenarios for player 2, one can show that there is a unique non-degenerated strategy for player 1 that makes player 2 indifferent between $L$ and $R$. Thus, there is a unique mixed Nash equilibrium in the non-degenerated sense.

Suppose now that a given 2×2 game has a unique mixed Nash equilibrium in the non-degenerated sense. Then player 2 has a unique mixed equilibrium strategy $(q, 1 - q)$, with $q$ in the $(0, 1)$ interval, that makes player 1 indifferent between strategies $U$ and $D$. An analogous reasoning is applied for player 1 and therefore will be omitted. The only way to do that is to play a mixed strategy where

$$q = \frac{d - b}{a - b - c + d} = \frac{1}{1 + \frac{a-c}{d-b}}$$

with $a - b - c + d \neq 0$. Thus, in order for $q$ belong to the interval $(0, 1)$, we must have that $\frac{a-c}{d-b} > 0$. Therefore, either $a > c$ and $d > b$, or $c > a$ and $b > d$. In both cases, neither player 1 is indifferent between his strategies nor he has a weakly (or strongly) dominant strategy.    □

Now analyzing the conditions that guarantee non-positive derivatives and strictly negative derivatives, it will be shown that the negative derivatives occurred only in games in which players have a preference that the other uses a particular strategy, regardless of his own choice. To make this idea more formal, consider the concept of collaborative dominance proposed by Souza & Rêgo (2013, p. 470). For more details see the original paper.

*Weak Collaborative Dominance (or non-strict):* For a $2 \times 2$ normal form game, we say that the strategy $s_j$ from player $j$ is weakly collaborative dominant for player $i$ (with $j \neq i$) if there $U_i(s_i, s_j) = U_i(s_i, s'_j)$ for all $s_i$ in player $i$'s strategy set, with at least one strict inequality. In this case we say that $s_j$ weakly collaboratively dominates $s'_j$ for player $i$.

For a strong version of the collaborative dominance definition, we should use a strict inequality signal ($>$) in the definition.

Theorems 1 and 2 show that, in games with a unique mixed equilibrium in the non-degenereted sense, a non-positive (respectively, negative) derivative of player $i$'s mixed equilibrium expected utility with respect to his own utility payoffs occurs if, and only if, player $j$ has a strategy that is weakly (respectively, strongly) collaboratively dominant for him, player $i$. Moreover, non-positive (respectively, negative) derivatives always occur when are taken with respect to player $i$'s utility payoffs associated with the strategy that is the best response to the weakly (resp. strongly) collaboratively dominant strategy of player $j$ (and those are the highest and lowest payoffs of player $i$ in the game).

**Theorem 1.** *Suppose a $2 \times 2$ normal form game and that player i does not have a strongly or a weakly dominant strategy and is not always indifferent between his strategies. Thus there are two derivatives of player i's mixed equilibrium expected utility taken with respect to one of his utility payoffs that are non-positive and two that are positive if, and only if, player j has a weakly collaboratively dominant strategy for player i. Moreover, the non-positive derivatives are always with respect to player i's utility payoffs associated with the strategy that is the best response to the weakly collaboratively dominant strategy of player j (and those are the highest and lowest utility payoffs of player i in the game).*

**Proof.**    Without loss of generality, assume that $i = 1$. So, given the assumptions of Theorem 1, there are two possibilities for partially ordering the payoffs of player 1, as follows: (A) $a > c$ and $b < d$ or (B) $a < c$ and $b > d$. Let us consider case (A). It follows that $\frac{\partial EU_1}{\partial a} \leq 0 \leftrightarrow c \geq d$, $\frac{\partial EU_1}{\partial b} \leq 0 \leftrightarrow c \geq d$, $\frac{\partial EU_1}{\partial c} \leq 0 \leftrightarrow b \geq a$ and $\frac{\partial EU_1}{\partial d} \leq 0 \leftrightarrow b \geq a$, so we should consider the following three sub-cases:

(A1)  $c \geq d$: In this case, $a > c \geq d > b$ and strategy $L$ is weakly collaboratively dominant for player 1. Furthermore, strategy $U$ of player 1 is the best response to $L$ and $\frac{\partial EU_1}{\partial a}$ and $\frac{\partial EU_1}{\partial b}$ are non-positive ($a$ is the highest payoff and $b$ is the lowest), while the other derivatives are positive.

(A2)  $b \geq a$: In this case, $d > b \geq a > c$, and strategy $R$ is weakly collaboratively dominant for player 1. Furthermore, strategy $D$ of player 1 is the best response to $R$ and $\frac{\partial EU_1}{\partial c}$ and $\frac{\partial EU_1}{\partial d}$

are non-positive ($d$ is the highest payoff and $c$ is the lowest), while the other derivatives are positive.

(A3) $d > c$ and $a > b$: In this case, there are no weakly collaboratively dominant strategies and all derivatives are positive.

The proof of condition (B) is analogous and is left to the reader.    □

**Theorem 2.** *Suppose a* $2 \times 2$ *normal form game and that player i does not have a strongly or a weakly dominant strategy and is not always indifferent between his strategies. Thus there are two derivatives of player i's mixed equilibrium expected utility taken with respect to one of his utility payoffs that are negative and two that are positive if, and only if, player j has a strongly collaboratively dominant strategy for player i. Moreover, the negative derivatives are always with respect to player i's utility payoffs associated with the strategy that is the best response to the strongly collaboratively dominant strategy of player j (and those are the highest and lowest utility payoffs of player i in the game).*

**Proof.**    Following the same idea of the proof of Theorem 1, we have: (A) $a > c$ and $b < d$ or (B) $a < c$ and $b > d$. Suppose that we are in case (A). It follows that $\frac{\partial EU_1}{\partial a} < 0 \leftrightarrow c > d$, $\frac{\partial EU_1}{\partial b} < 0 \leftrightarrow c > d$, $\frac{\partial EU_1}{\partial c} < 0 \leftrightarrow b > a$ and $\frac{\partial EU_1}{\partial d} < 0 \leftrightarrow b > a$. Therefore, consider the following three sub-cases:

(A1) $c > d$: In this case, $a > c > d > b$ and strategy $L$ is strongly collaboratively dominant for player 1. Furthermore, strategy $U$ of player 1 is the best response to $L$ and $\frac{\partial EU_1}{\partial a}$ and $\frac{\partial EU_1}{\partial b}$ are negative ($a$ is the highest payoff and $b$ is the lowest), while the other derivatives are positive.

(A2) $b > a$: In this case, $d > b > a > c$, and strategy $R$ is strongly collaboratively dominant for player 1. Furthermore, strategy $D$ of player 1 is the best response to $R$ and $\frac{\partial EU_1}{\partial c}$ and $\frac{\partial EU_1}{\partial d}$ are negative ($d$ is the highest payoff and $c$ is the lowest), while the other derivatives are positive.

(A3) $d \geq c$ and $a \geq b$. In this case, there are no strongly collaboratively dominant strategies and all derivatives are non-negative.

Again, the analysis of case (B) is analogous and is left to the reader.    □

## 4    BEST UTILITY REDUCING STRATEGY

In the previous section, we showed that whenever negative derivatives of some player mixed equilibrium expected utility happen, they are with respect to the highest and lowest payoffs of such player. In this section, we answer the following question: assuming that players will play according to the mixed equilibrium, and there are two negative derivatives for a given player, if

this player has the opportunity to reduce $x$ units from a given strategy profile, then what is the best utility reducing strategy that the player can adopt? We prove that he should reduce utility in the case that he uses a strategy that is a best response to the strategy of the other player that is strongly collaboratively dominant for him. However, as we show next, for some cases the player should only reduce utility if the other player *indeed* chooses the strongly collaboratively dominant strategy for him (this situation corresponds to reduce utility in his highest utility payoff in the game), while in other cases the opposite should happen (this situation corresponds to reduce utility in his lowest utility payoff in the game).

In order to show those results, we should look initially at how the mixed equilibrium strategy of a given player reacts to changes in the payoffs of the other player. Therefore, for player 2, using Equation (2.2), we have:

$$\frac{\partial q}{\partial a} = \frac{\partial (1-q)}{\partial c} = \frac{b-d}{(a-b-c+d)^2}; \tag{4.1}$$

$$\frac{\partial q}{\partial b} = \frac{\partial (1-q)}{\partial d} = \frac{c-a}{(a-b-c+d)^2}; \tag{4.2}$$

$$\frac{\partial q}{\partial c} = \frac{\partial (1-q)}{\partial a} = \frac{d-b}{(a-b-c+d)^2}; \quad \text{and} \tag{4.3}$$

$$\frac{\partial q}{\partial d} = \frac{\partial (1-q)}{\partial b} = \frac{a-c}{(a-b-c+d)^2}. \tag{4.4}$$

Now, we can rewrite Equations (2.5), (2.6), (2.7) and (2.8), as shown in Equations (4.5), (4.6), (4.7), (4.8), respectively. From these latter equations, it can be seen that the derivative of the expected utility of player 1 is a function of the derivative of player 2's mixed equilibrium strategy.

$$\frac{\partial EU_1}{\partial a} = (c-d)\frac{\partial q}{\partial a}; \tag{4.5}$$

$$\frac{\partial EU_1}{\partial b} = (c-d)\frac{\partial q}{\partial b}; \tag{4.6}$$

$$\frac{\partial EU_1}{\partial c} = (a-b)\frac{\partial q}{\partial c}; \quad \text{and} \tag{4.7}$$

$$\frac{\partial EU_1}{\partial d} = (a-b)\frac{\partial q}{\partial d}. \tag{4.8}$$

Theorem 2 states that negative derivatives of player 1's mixed equilibrium expected utility taken with respect to his payoffs occur if, and only if, one of these four orderings of payoffs happens: (1) $d > b > a > c$; (2) $a > c > d > b$; (3) $b > d > c > a$; (4) $c > a > b > d$. Let us consider Case (1).

**Case 1:** $d > b > a > c$**.**   In this case, strategy $R$ of player 2 is strongly collaboratively dominant for player 1. Thus, it follows that the derivative of the expected utility of player 1 is negative

with respect to payoffs $d$ and $c$, and $\frac{\partial q}{\partial c}$ and $\frac{\partial q}{\partial d}$ are positive, implying that a reduction in one of these payoffs also reduces the chance of player 2 choosing strategy L and therefore increases the chance of player 2 choosing strategy $R$, which, in this case, is strongly collaboratively dominant for player 1. The analyses of the remaining cases are analogous.

Therefore, player 1 should reduce utility with respect to the payoff that provide the greater increase in the probability of player 2 choosing the strategy that is strongly collaboratively dominant for him, player 1.

To emphasize this conclusion, let us analyze the same problem from another perspective. Now imagine that the player 1 has $x > 0$ units of utility to reduce in any payoff. Then, assuming that the general ordering of payoffs in the game is maintained, with respect to what payoff should he reduce these $x$ units of utility?

Suppose, for example, that we are in Case 1, where $d > b > a > c$, also suppose that player 1 decided to reduce $\alpha x$ units of utility in $c$ and $(1 - \alpha)x$ units in $d$, with $\alpha \in [0, 1]$. To maintain the order of the payoffs we must ensure that $(1 - \alpha)x \leq d - b$. Thus, we have that player 1's expected utility is given by:

$$EU_1 = \frac{a(d - x(1 - \alpha)) - b(c - x\alpha)}{a - b - (c - x\alpha) + (d - x(1 - \alpha))} . \tag{4.9}$$

We want to find the value of $\alpha$ that maximizes $EU_1$. Differentiating $EU_1$ with respect to $\alpha$, we have:

$$\frac{EU_1}{\partial \alpha} = \frac{x(a - b)(x + a + b - c - d)}{(a - b - c + d - x + 2x\alpha)^2} . \tag{4.10}$$

Based on Equation (4.10), it can be seen that the derivative is positive if $0 < x < (d-b)+(c-a)$, and the player should reduce the $x$ units in the lowest payoff, $c$. On the other hand, if $d-b \geq x > (d-b)+(c-a)$, then he should reduce the $x$ units in the highest payoff, $d$. If $x = (d-b)+(c-a)$, then the derivative is equal to zero and, consequently, it does not make difference in what payoff to reduce utility. Note also that for a small value of $x$, as expected, the conclusions are the same that we obtained with the analysis of the derivatives made above, that is, player 1 should reduce utility with respect to $c$ if $\frac{\partial q}{\partial c} > \frac{\partial q}{\partial d}$, which is equivalent to $d - b > a - c$, or reduce utility with respect to $d$ in the other case. By a similar analysis, we can describe what should be player 1's behavior in each of the four cases where he has incentive to reduce utility. Thus, suppose that player 1 can reduce $x$ units of utility:

**Case 1:** $d > b > a > c$. If $x < (d - b) + (c - a)$, then he should reduce the $x$ units of utility with respect to the payoff $c$, while if $d - b \geq x > (d - b) + (c - a)$, then he should reduce it in $d$.

**Case 2:** $a > c > d > b$. If $x < (a - c) + (b - d)$, then he should reduce the $x$ units of utility with respect to the payoff $b$, while if $a - c \geq x > (a - c) + (b - d)$, then he should reduce it in $a$.

**Case 3:** $b > d > c > a$**.**    If $x < (b - d) + (a - c)$, then he should reduce the x units of utility with respect to the payoff a, while if $b - d \geq x > (b - d) + (a - c)$, then he should reduce it in $b$.

**Case 4:** $c > a > b > d$**.**    If $x < (c - a) + (d - b)$, then he should reduce the $x$ units of utility with respect to the payoff $d$, while if $c - a \geq x > (c - a) + (d - b)$, then he should reduce it in $c$.

Thus, if a player has small power and cannot reduce a great amount of utility, then he should invest all his efforts to reduce his lowest utility payoff in the game. On the other hand, if the player has a greater power, he should invest all his efforts to reduce highest utility payoff in the game.

Assume that the conditions of Theorem 2 are satisfied. *It is interesting to point out that in games with no pure equilibria and where both players have a strongly collaboratively dominant strategy, if we measure the value of participating in the game by the mixed equilibrium expected utility, then we showed that the value of participating in the game decreases as the highest and lowest utility payoffs of a player increases.* Additionally, once a player knows that a reduction in some of his payoffs increases his mixed equilibrium expected utility, he may be tempted to lie about his true utility and that can cause a serious problem for utility elicitation in strategic settings and/or game modeling. However, it is beyond the scope of this paper to go deeper into this discussion.

## 5    DISCUSSIONS

Until this section, we restricted our analysis of mixed equilibrium (and the problem of payoff reduction) only to $2 \times 2$ games with a single mixed equilibrium. This type of analysis is relevant in game theory, and an important example is the work of Engelmann & Steiner (2007), that studied the effects of risk preferences in mixed-strategy equilibria of $2 \times 2$ games. Now, we present numerical examples that help us understand the fundamental limitations that prevent us from extending the results already exposed to more general games. We finish this section presenting sufficient conditions that guarantee that derivatives of some player's equilibrium expected utility taken with respect to one of his payoffs are always non-negative.

### 5.1    Games with Mixed Equilibrium with $2 \times 2$ Support

We begin the discussion by analyzing the game shown in Figure 3, for which the conclusions of Section 3 are still valid (with the appropriate adjustments). The game in Figure 3 has only one mixed equilibrium $((1/3, 2/3), (1/3, 0, 2/3))$ and its support is $\{U, D\} \times \{L, R\}$. Moreover, the expected utility of the players are $(13/3, 17/3)$. Also note that in this game, the strategy $L$ is strongly collaboratively dominant with respect to strategy $R$ for player 1 and, removing $M$, since it is outside the equilibrium support, the strategy $D$ is strongly collaboratively dominant with respect to $U$, for player 2. So for this game, we can use the results of Theorem 2 which indicates, for example, that a reduction in utility $U_1(U, L)$ in two utils would increase the expected utility

of player 1 to 5 and a reduction of utility $U_2(D, L)$ in one util would increase the expected utility of player 2 to 6, i.e., both players would like to reduce utility if they could.

|   | L | M | R |
|---|---|---|---|
| U | (7, 3) | (4, 7) | (3, 5) |
| D | (5, 7) | (6, 2) | (4, 6) |

**Figure 3**

However, in this particular case, the game has a unique mixed equilibrium, whose support is composed of two pure strategies of each player, making it similar to a $2 \times 2$ game.

### 5.2    Games with Mixed Strategy Equilibrium with More than 2 Pure Strategies in the Support

Now, we analyze a game in which all three pure strategies of player 2 are in the equilibrium support, as shown in Figure 4.

|   | L | M | R |
|---|---|---|---|
| U | (8, 0) | (3, 1) | (2, 0.5) |
| D | (6, 1) | (4, 0) | (5, 0.5) |

**Figure 4**

Before calculating the mixed equilibrium of this game, let us define some notation. Let $\sigma(U)$ be the probability of player 1 choosing $U$ (hence $\sigma(D) = 1 - \sigma(U)$ is the probability that he chooses $D$) and let $\sigma(L)$ be the probability of player 2 choosing $L$ and $\sigma(M)$ be the probability of choosing $M$ (indeed, $\sigma(R) = 1 - \sigma(L) - \sigma(M)$). Thus we can characterize the mixed equilibrium of this game as follows:

$$\left( (1/2, 1/2), \left( \sigma(L), \frac{3 - 5\sigma(L)}{2}, \frac{3\sigma(L) - 1}{2} \right) \right),$$

where $\sigma(L) \in [1/3, 3/5]$. Moreover, the expected utility of the mixed equilibrium for player 1 is given by $EU_1 = \frac{7\sigma(L)+7}{2}$, and depending on the value of $\sigma(L)$, it can vary in the range $EU_1 = [14/3, 28/5]$.

Now, consider the mixed equilibrium $((1/2, 1/2), (1/2, 1/4, 1/4))$. In this case, the expected utility of player 1 is 5.25. Note that in this game, strategy $L$ of player 2 is strongly collaboratively dominant (with respect to all others of player 2) for player 1. Thus, we may be tempted to apply our previous results and think that player 1 could reduce, for example, his highest payoff to induce player 2 to choose $L$ more frequently. Suppose that player 1 reduces $U_1(U, L)$ from 8 to 7. Then, we can characterize the mixed equilibrium of the new game as follows:

$$\left( (1/2, 1/2), \left( \sigma(L), \frac{3 - 4\sigma(L)}{2}, \frac{2\sigma(L) - 1}{2} \right) \right),$$

where $\sigma(L) \in [1/2, 3/4]$. Indeed, player 1's mixed equilibrium expected utility is $EU_1 = \frac{6\sigma(L)+7}{2}$, and depending on the value of $\sigma(L)$, it can vary in the range $EU_1 \in [5, 23/4]$. In this case, it is easy to see that player 2 may, for example, keep $\sigma(L)$ fixed (just changing the values of $\sigma(M)$ and $\sigma(R)$). In such situation, player 1's expected utility reduces by $\sigma(L)/2$. Since player 2 has a range of values ($\sigma(L) \in [1/2, 3/4]$) for which he can manipulate $\sigma(L)$, in general, it is impossible to say how he will react to any change in payoffs made by player 1.

Formally, this problem always happens every time that we are unable to establish a well defined functional form of how the mixed strategy chosen by player $j$ depends on player $i$'s payoffs. In fact this could also happen in $2 \times 2$ games, for example: Based on Figure 2, suppose that player 1 has a strongly dominant strategy, say strategy $U$. Moreover, for player 2, assume that $e = f$. If this condition occurs, then the game has two pure equilibria, $(U, L)$ and $(U, R)$, and infinitely many mixed equilibria on the form $(M, N)$, where $M = (1, 0)$ and $N = (q^*, 1 - q^*)$ for $q^* \in [0, 1]$. Thus, the expected utility of player 1 is $EU_1 = aq^* + b(1 - q^*)$. Now suppose that player 2 chooses $q^*$ depending on $a$. Thus the derivative of player 1's mixed equilibrium expected utility with respect to payoff a varies accordingly to the functional form that $q^*$ depends on $a$ and may assume either negative or positive values.

### 5.3    Games with 3 Players and 2 Pure Strategies Each

Now consider a game with three players each one with two strategies, as shown in Figure 5. Admit that the payoff a from the strategy profile $(U, L, L)$ is a value between 6 and 9, $a \in [6, 9)$. Thus, this game has two pure equilibria, $(U, L, L)$ and $(D, R, R)$ and one mixed equilibrium.

|   | L | R | L |   |   | L | R | R |
|---|---|---|---|---|---|---|---|---|
| U | $(a, 8, 8)$ | $(5, 7, 5)$ |   |   | U | $(0, 3, 7)$ | $(1, 4, 6)$ |   |
| D | $(3, 5, 3)$ | $(6, 6, 1)$ |   |   | D | $(4, 1, 4)$ | $(2, 2, 2)$ |   |

**Figure 5**

By making the payoff a vary between 6 and 9, we can analyze how the mixed equilibrium expected utility of player 1 reacts. In particular, we are interested if the expected utility is a increasing or a decreasing function of $a$. Player 1's expected utility of is given by:

$$EU_1 = (a - 4)\left(\frac{3 + \sqrt{13 + 4a}}{2a + 2}\right)^2 + 3\left(\frac{3 + \sqrt{13 + 4a}}{2a + 2}\right) + 1. \qquad (5.1)$$

Thus, for any value of $a$ higher than 61/9 and lower then 9, a reduction in $a$ will lead to an increase in the expected utility of player 1. On the other hand, for any value of $a$ lower than 61/9 and higher than 6, a reduction in $a$ will also lead to a reduction in the mixed equilibrium expected utility of player 1. Furthermore, if we assume that, initially, $a$ is equal 6, any reduction in any payoff of player 1 will also lead to a reduction in his expected utility. But if, for example, we assume an initial value of 8, a *small reduction* in any payoff of player 1 related to the pure

strategy $U$, will lead to an increase of player's 1 mixed equilibrium expected utility, even though none of the players 2 or 3 has a strategy that is collaboratively dominant for player 1.

Consequently, in more general class of games, the existence of negative derivatives does not depend on the existence of collaboratively dominant strategies. Moreover, since $a$ is always the highest payoff of player 1, the existence of negative derivatives does not depend only on the order of the payoffs.

Finally, based on Figure 5, assume that the payoff of player 2 from the strategy profile $(D, R, L)$ was reduced from 6 to 4, and $a$ is equal to 8, as shown in Figure 6.

| | $L$ | $R$ | $L$ | | | $L$ | $R$ | $R$ |
|---|---|---|---|---|---|---|---|---|
| $U$ | (8, 8, 8) | (5, 7, 5) | | | $U$ | (0, 3, 7) | (1, 4, 6) | |
| $D$ | (3, 5, 3) | (6, 4, 1) | | | $D$ | (4, 1, 4) | (2, 2, 2) | |

**Figure 6**

This new game also has two pure equilibria, $(U, L, L)$ and $(D, R, R)$ and one mixed equilibrium, $((3/4; 1/4), (2/3; 1/3); (1/2, 1/2))$. However, in this new game, when we make a small reduction in any payoff of player 1, then the expected utility from the mixed equilibrium also reduces. This example shows us that in a more general class of games, the existence of negative derivatives of the expected utility of a given player with respect to one of his payoff may depend on the payoffs of other players, as opposed to the result of the $2 \times 2$ games. Thus these facts prevent us to make extensions of Theorems 1 and 2 to a more general class of games.

## 5.4   Sufficient conditions for the existence of non-negatives derivatives

On the above subsections, we presented numerical examples indicating difficulties to extend our earlier results from Theorem 1 and 2 about the conditions that guarantee the existence of negative derivatives of some player equilibrium expected utility with respect to one of his payoffs. On the other hand, we now present sufficient conditions that guarantee that such derivatives are always non-negative in finite normal form games.

**Proposition 1.** *In a finite normal form game G, if every player j in $N \setminus \{i\}$ uses an equilibrium strategy that does not depend on player i's payoffs, then none of the derivatives of player i's equilibrium expected utility taken with respect to one of his utility payoffs is negative.*

**Proof.**   Let $\sigma$ be a Nash equilibrium that satisfies that hypothesis of Proposition 1. It is well known that if every player $j$ in $N \setminus \{i\}$ follows his mixed equilibrium strategy $\sigma_j(s_j)$, then for any pure strategy $t_i$ in the support of $\sigma_i$, it follows that

$$U_i(\sigma) = U_i(t_i, \sigma_{-i}) = \sum_{s_{-i} \in S_{-i}} \left( \prod_{j \in N \setminus \{i\}} \sigma_j(s_j) \right) U_i(t_i, s_{-i}).$$

Taking the derivatives of player $i$'s mixed equilibrium expected utility with respect to one of his utility payoffs, say $U_i(s')$, with $s'$ in $S$, then we have:

$$\frac{\partial U_i(\sigma)}{\partial U_i(s')} = \frac{\partial(\sum_{s_{-i} \in S_{-i}} (\prod_{j \in N \setminus \{i\}} \sigma_j(s_j)) U_i(t_i, s_{-i}))}{U_i(s')}.$$

Since none of the $\sigma_j(s_j)$ for every player $j$ in $N \setminus \{i\}$ depends on $U_i(s')$ and there is no functional relationship between the payoffs of a given player, then for every non-zero term on the sum, the multiplicand term are non-negative constants, so the sign of the derivative will depend only on the sign of the sum of $\frac{\partial U_i(t_i, s_{-i})}{\partial U_i(s')}$, for all $s_{-i}$ in $S_{-i}$. Since we are assuming only marginal changes when we take derivatives, we have that $\frac{\partial U_i(t_i, s_{-i})}{\partial U_i(s')}$ is equal to one if $(t_i, s_{-i}) = s'$ and equal to zero otherwise. Thus $\frac{\partial U_i(\sigma)}{\partial U_i(s')}$ must be always non-negative.    □

Finally, it is easy to see that any pure Nash equilibrium satisfies the hypothesis of Proposition 1, then an increase in some of the payoffs of any player will never reduce his expected utility in that pure equilibrium.

## 6    SOME THOUGHTS ON JERVIS (1978)

Aumann (1990) proposed a discussion on when a Nash equilibrium can be considered self-enforcing based on a verbal agreement among players, i.e. how we can ensure that players will choose a given Nash equilibrium since they announced that they will. To develop his argument, Aumann uses as his main example the Stag-Hunt game.

For Aumann, there are two ways to encourage a player to perform a given choice. The first one is related to a change in the information available to the player and the second one is related to a change in payoffs. Aumann decided to dedicate his analysis to the first case. Thus, based on the Stag-Hunt game, he concludes that even if the players claim that they will play $(U, L)$ it does not increase the incentive of them actually choosing this strategy profile. For example, when player 1 declares that he will play $U$, it does not add any information to player 2, because, since $L$ is a strongly collaboratively dominant strategy for player 1, player 2 knows that player 1 prefers that he (player 2) plays $L$. Thus, player 2 knows that player 1 would state that consents to any agreement in which player 2 plays $L$, but this fact does not guarantee that player 1 will really fulfill the agreement and play $U$. For example, player 1 may prefer to play $D$, since this is a safer option. Similar reasoning also applies to player 2.

Now, we discuss an application of our results by exploiting the gap left by Aumann (1990), i.e., we evaluate how to encourage players to make a given choice based on changes in the payoffs. For this, we also illustrate our argumentation with the Stag-Hunt game which is used in international politics literature to discuss the security dilemma as done by Jervis (1978).

It is important to point out that Jervis's work is one o the most influential paper in international politics and an essential reading for those interested in the security dilemma (see, e.g., Glaser, 1997 and Tang, 2009) and his canonical model is still used for theoretical game theory modeling about the security dilemma (Acharya & Ramsay, 2013). But our purpose is not to discuss exhaustively Jervis's (1978) work. In fact, we aim to critically analyze some passages from

Jervis (1978) regarding to *what makes cooperation more likely* (which is the title of section 2 of his work) in the security dilemma, revisiting the author's conclusions with a game theoretic perspective.

To summarize the main idea of the security dilemma, imagine two nations that go through a period of international tension. They have two strategic options, namely: do not make investment in weapons (cooperate, C) or perform military investment (non-cooperate, D – defecting). The order of preferences for the possible strategies profiles is equivalent to that of the stag-hunt game, as stated before. However, Jervis (1978) states that nations will only cooperate if they believe that the other will too and points out some possible explanations for the players to sacrifice the most desired option (CC), namely: the fear of being attacked and not being able to defend itself, political uncertainty in the neighboring nations and even coercion opportunities and participation in international affairs because of the military power (reputation).

Jervis starts studying what could make mutual cooperation more likely by listing a set of conditions. For the author, the chance of achieving cooperation would increase by:

> "(1) anything that increases incentives to cooperate by increasing the gains of mutual cooperation (CC) and/or decreasing the cost the actor will pay if he cooperates and the other does not (CD); (2) anything that decreases the incentives for defecting by decreasing the gains of taking advantage of the other (DC) and/or increasing the cost of mutual noncooperation (DD); (3) anything that increases each side's expectation that other will cooperate." (Jervis, 1978, p. 171).

We will now evaluate the effects of these affirmations, especially regarding conditions (1) and (2). The idea of *what makes cooperation more likely* can raise various interpretations, e.g., we may think about the concept of equilibrium selection or focal point, but to apply these concepts, it is not necessary to make any changes in payoffs, i.e., if the players were determined to apply any equilibrium selection criterion (or identify a focal point), then a marginal change in some payoff should not alter the original decision, except if the change in payoffs is such that it modifies the original equilibrium set of the game (see Englemann, 1994). Therefore, we must analyze this idea from the perspective of the mixed equilibrium.

Based on Figure 2, the order of the payoffs for the Stag-hunt game (security dilemma) is $a > c > d > b$ (for player 1) and $e > f > h > g$ (for player 2). Thus, by condition (1) Jervis suggests that cooperation would be more likely if the players were able to increase the payoffs $a$ and $e$ or if they were able to increase the payoffs $b$ and $g$. However, by Case 2 in Section 4, we saw that $\partial q / \partial a$ and $\partial q / \partial b$ are negative (the same holds for $\partial p / \partial e$ and $\partial p / \partial g$) and, thereby, any increase in these payoffs, in fact, would make cooperation less likely. In turn, condition (2) states that cooperation would be more likely to occur if the players would reduce the payoffs $c$ and $f$ or reduce the payoffs $d$ and $h$: but, since $\partial q / \partial c$ and $\partial q / \partial d$ are positive, the effect is reversed and cooperation, again, would be less likely. In particular, by condition (2), the cooperation would only become more likely if, for example, the reduction in payoffs $d$ and $h$ were of such intensity that turn them in the lowest payoff of the game and, consequently, the new game will have a unique Nash equilibrium (CC).

Later in his study, Jervis discusses what a player (nation) should do to increase the likelihood that the other player will cooperate, stating:

> "The variables discussed so far influence the payoff for each of the four possible outcomes. To decide what to do, the state has to go further and calculate the expected value of cooperating or defecting. Because such calculations involve estimating the probability that the other will cooperate, the state will have to judge how the variables discussed so far act on the other. To encourage the other to cooperate, a state may try to manipulate these variables. It can lower the other's incentives to defect by decreasing what it could gain by exploiting the state (DC) [...]" (Jervis, 1978, p. 179).

The author follows his argument by pointing another example:

> "The state can also try to increase the gains that will accrue to the other from mutual cooperation (CC). Although the state will of course gain if it receives a share of any new benefits, even an increment that accrues entirely to the other will aid the state by increasing the likelihood that the other will cooperate." (Jervis, 1978, p. 180).

Again, we must focus on players' mixed strategies. As it was shown in Section 2, the mixed equilibrium strategy of a given player depends only on the utility payoffs of the other player. Thus, increasing the utility payoff from mutual cooperation of a given player does not change the mixed equilibrium strategy of such player. In fact, what happens is a change in the mixed equilibrium strategy of the other player, which will now choose to cooperate less likely, as opposed to what was expected by Jervis.

We recognize that the problems of international cooperation are far more complex than as exposed above, because they involve aspects of reputation, long-term relationship and uncertainty, for example. However, we hope that our approach can contribute to the better understanding of some aspects of the problem.

## 7 FINAL REMARKS

In this paper, we propose a new approach to analyze payoff reduction behavior through the analysis of the mixed Nash equilibrium in $2 \times 2$ normal form games. We provide a necessary and sufficient condition for the existence of negative derivatives of the expected utility that justify a (pre-commitment) payoff reduction behavior. In particular, games with negative derivatives are the ones that create incentives for payoff reduction behavior since such behavior improves player's mixed equilibrium expected utility. We show that a negative derivative for the mixed equilibrium expected utility of a given player $i$ occurs if, and only if, he has a strict preference for one of the strategies of the other player. Moreover, negative derivatives always occur when they are taken with respect to player $i$'s highest and lowest game utility payoffs. Additionally, we pointed out the difficulty to extend our results for a more general class of games. On the other hand, we present sufficient conditions for the existence of non-negative derivatives of some

player's equilibrium expected utility taken with respect to one of his payoffs in finite normal form games. Nevertheless, we showed the relevance of the present study using our insights to analyze the security dilemma, revisiting some conclusions made by Jervis (1978).

In another related study, Souza & Rêgo (2014) experimentally studied if those theoretical statements were true when played by real subjects. As one of the results, authors concluded that the payoff reduction opportunity was overlooked by the players because they do not played according to the mixed equilibrium and, therefore, to reduce utility could not bring benefits. Other interesting experimental results about related subjects can be found in Goeree & Holt (2001). In this paper the authors pointed that in some strategic situations, subjects are unable to achieve the theoretical result because these results are sometimes unintuitive, as the payoff reduction behavior (in mixed equilibrium) seems to be.

## REFERENCES

[1]   ACHARYA A & RAMSAY HW. 2013. The Calculus of the Security Dilemma. *Quarterly Journal of Political Science*, **8**: 183–203.

[2]   AUMANN RJ. 1990. Nash Equilibria are not Self-Enforcing. In Gabszewicz JJ, Richard JF & Wolsey L. (ed) Economic Decision Making, Econometrics, and Optimization: Essays in Honor of Jacques Dreze. Elsevier Science Publishers, Amsterdam, pp. 201–206.

[3]   BEN-PORATH E & DEKEL E. 1992. Signaling future actions and the potential for sacrifice. *Journal of Economic Theory*, **57**: 36–51.

[4]   BRANDTS J & HOLT A. 1995. Limitations of dominance and forward induction: Experimental evidence. *Economics Letters*, **49**: 391–395.

[5]   ENGLEMANN W. 1994. Conditions for Disarmament: A Game Theoretical Model. *Group Decision and Negotiation*, **3**: 321–332.

[6]   ENGELMANN D & STEINER J. 2007. The effects of risk preferences in mixed-strategy equilibria of $2 \times 2$ games. *Games and Economic Behavior*, **60**: 381–388.

[7]   FUNDENBERG D & TIROLE J. 1991. Game Theory. MIT Press, Cambridge.

[8]   GLASER CL. 1997. The Security Dilemma Revisited. *World Politics*, **50**: 171–201.

[9]   GOEREE JK & HOLT CA. 2001. Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. *The American Economic Review*, **91**: 1402–1422.

[10]  HUCK S & MÜLLER W. 2005. Burning money and (pseudo) first-mover advantages: an experimental study on forward induction. *Games and Economic Behavior*, **51**: 109–127.

[11]  JERVIS R. 1978. Cooperation under the Security Dilemma. *World Politics*, **30**: 167–214.

[12]  KOHLBERG E & MERTENS JF. 1986. On the Strategic Stability of Equilibria. *Econometrica*, **54**: 1003–1037.

[13]  MYERSON RB. 1991. Game theory: analysis of conflict. Harvard University Press, London.

[14]  RASMUSEN E. 1996. Game and Information: an introduction to game theory. Blackwell, Cambridge.

[15]  SOUZA FC & RÊGO LC. 2013. Collaborative dominance: when doing unto others as you would have them do unto you is reasonable. *Pesquisa Operacional*, **33**: 467–476.

[16]  Souza FC & Rêgo LC. 2014. Mixed Equilibrium, Collaborative Dominance and Burning Money: An Experimental Study. *Group Decision and Negotiation*, **23**: 377–400.

[17]  Van Damme E. 1989. Stable equilibria and forward induction. *Journal of Economic Theory*, **48**: 476–496.

[18]  Tang S. 2009. The Security Dilemma: A Conceptual Analysis. *Security Studies*, **18**: 587–623.