Mixed Equilibrium: When Burning Money is Rational

Souza, Filipe and Rêgo, Leandro

Federal University of Pernambuco

10 February 2012

Mixed Equilibrium: When Burning Money is Rational

Filipe Costa de Souza
Federal University of Pernambuco, Accounting and Actuarial Science Department.
filipe.costas@ufpe.br

Leandro Chaves Rêgo
Federal University of Pernambuco, Statistics Department.
leandro@de.ufpe.br
Fax: +55-81-2126-7438.

Abstract. We discuss the rationality of burning money behavior from a new perspective: the mixed Nash equilibrium. We support our argument analyzing the first-order derivatives of the mixed equilibrium expected utility of the players with respect to their own utility payoffs in a 2x2 normal form game. 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. 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.

Keywords: Mixed Nash Equilibrium, Burning Money, Collaborative Dominance, Security Dilemma.

JEL Classification: C72.

1. Introduction

Based on the concept of forward induction proposed by Kohlberg and Mertens (1986) and, especially, the idea of iterative elimination of weakly dominated strategies, Van Damme (1989) and Ben-Porath and Dekel (1992) studied the effects of burning utility as a way to signal future actions. First, the authors analyzed the Battle of the Sexes game in which player 1 had the opportunity, before the beginning of the game, to signal player 2 his ability to burn utility. The Battle of the Sexes game is shown in Figure 1.

		Player 2	
		W	*Z*
Player 1	*X*	*(3, 1)*	*(0, 0)*
	Y	*(0, 0)*	*(1, 3)*

Figure 1

The game in Figure 1 has three equilibria, two in pure strategies, (*X, W*) and (*Y, Z*), and one in mixed strategy $E=(M, N)$, where $M = (3/4, 1/4)$[1] and $N = (1/4, 3/4)$. Now, consider that player 1 can burn one utility unit before the Battle of the Sexes game starts. The normal form representation of the new game is shown in Figure 2. In this game, *B* indicates that player 1 burned utility and *NB* indicates that he did not burn. Moreover, the second letter, *X* or *Y*, indicates the strategy chosen by him after deciding whether to burn utility or not. In turn, for player 2, the first letter indicates the chosen strategy if player 1 burns utility and the second letter indicates the chosen strategy if player 1 does not burn utility.

		Player 2			
		WW	*WZ*	*ZW*	*ZZ*
Player 1	*BX*	*(2, 1)*	*(2, 1)*	*(-1, 0)*	*(-1, 0)*
	BY	*(-1, 0)*	*(-1, 0)*	*(0, 3)*	*(0, 3)*
	NBX	*(3, 1)*	*(0, 0)*	*(3, 1)*	*(0, 0)*
	NBY	*(0, 0)*	*(1, 3)*	*(0, 0)*	*(1, 3)*

Figure 2

Based on the new game, and supported by the principle of iterative elimination of weakly dominated strategies, it is easy to see that the only remaining equilibrium will be (*NBX, WW*), i.e., player 1 will not burn utility and will choose strategy *X*, while player 2 will choose strategy *W* no matter what player 1 does. Thus, the opportunity to burn utility allows player 1 to achieve his preferred equilibrium point in the game. Ben-Porath and Dekel (1992) generalized this result reaching the follow conclusion: in games in which a player has a strict preference for a equilibrium point, and if this player can self-sacrifice (burning utility), then, based on the forward induction rationality and iterative elimination of weakly dominated strategies, such player will achieve his (or

[1] We denote the mixed strategy for player 1 that chooses *X* with probability *p* and *Y* with probability *1-p*, by (*p, 1-p*).

her) most preferred outcome.[2] This conclusion was supported by Huck and Müller (2005) in an experimental study[3].

However, Myerson (1991, p.194-195) argues that in the context of sequential equilibrium, player 1's act of burning utility 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. Another argument against the conclusions of Ben-Porath and Dekel is inspired on Luce and Raiffa (1989). These authors argue that any effort of communication between the players before the beginning of the game may change the utility payoff matrix, i.e., can lead individuals to play a different game in the future. In the game context, the option to burn money for player 1 can be seen as a threat by player 2 and, consequently, could change player 2's mood. Moreover, this change of mood could also change the utility payoffs of the game and, possibly, its equilibrium points.

In addition, Van Damme (1989) and Ben-Porath and Dekel (1992) recognize that if all players can signal their intentions by burning utility then the final outcome of the game may be inefficient. To confirm this idea, observe the games in Figures 3 and 4 that were proposed by Ben-Porath and Dekel (1992). On the first, we have a stag-hunt game.

	Player 2		
Player 1		*W*	*Z*
	X	*(9, 9)*	*(0, 7)*
	Y	*(7, 0)*	*(6, 6)*

Figure 3

Now suppose that both players can signal their future intentions by burning 1.5 units of utility and after that, they started to play the stag-hunt game. The result of this new game is presented in Figure 4.

	Player 2								
		BWW	*BWZ*	*BZW*	*BZZ*	*NBWW*	*NBWZ*	*NBZW*	*NBZZ*
Player 1	*BXX*	*(7.5, 7.5)*	*(7.5, 7.5)*	*(-1.5, 5.5)*	*(-1.5, 5.5)*	*(7.5, 9)*	*(7.5, 9)*	*(-1.5, 7)*	*(-1.5, 7)*
	BXY	*(7.5, 7.5)*	*(7.5, 7.5)*	*(-1.5, 5.5)*	*(-1.5, 5.5)*	*(5.5, 0)*	*(5.5, 0)*	*(4.5, 6)*	*(4.5, 6)*
	BYX	*(5.5, -1.5)*	*(5.5, -1.5)*	*(4.5, 4.5)*	*(4.5, 4.5)*	*(7.5, 9)*	*(7.5, 9)*	*(-1.5, 7)*	*(-1.5, 7)*
	BYY	*(5.5, -1.5)*	*(5.5, -1.5)*	*(4.5, 4.5)*	*(4.5, 4.5)*	*(5.5, 0)*	*(5.5, 0)*	*(4.5, 6)*	*(4.5, 6)*
	NBXX	*(9, 7.5)*	*(0, 5.5)*	*(9, 7.5)*	*(0, 5.5)*	*(9, 9)*	*(0, 7)*	*(9, 9)*	*(0, 7)*
	NBXY	*(9, 7.5)*	*(0, 5.5)*	*(9, 7.5)*	*(0, 5.5)*	*(7, 0)*	*(6, 6)*	*(7, 0)*	*(6, 6)*
	NBYX	*(7, -1.5)*	*(6, 4.5)*	*(7, -1.5)*	*(6, 4.5)*	*(9, 9)*	*(0, 7)*	*(9, 9)*	*(0, 7)*
	NBYY	*(7, -1.5)*	*(6, 4.5)*	*(7, -1.5)*	*(6, 4.5)*	*(7, 0)*	*(6, 6)*	*(7, 0)*	*(6, 6)*

Figure 4

Note that by iterative elimination of weakly dominated strategies only the strategies *BYY* and *BZZ* can be eliminated. Thus, if both players had the opportunity to burn utility at the same time, the iterative elimination of weakly dominated strategies does not lead them to an efficient equilibrium. Furthermore, the authors emphasize that the order in which the players can burn utility also 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.

---

[2] For more information about burning money games and about forward induction rationality, we recommend: Gersbach (2004), Shimoji (2002), Stalnaker (1998) and Hammond (1993).

[3] For other experimental results, we also recommend Brandts and Holt (1995).

Burning money games can also be seen in another approach, which involves burning utility just for some specific strategy profiles, as discussed in Fudenberg and Tirole (1991, p.9). The authors propose the game presented in Figure 5. In this game, there is a unique (and inefficient) pure equilibrium (*X, W*).

	Player 2		
Player 1		*W*	*Z*
	X	*(1, 3)*	*(4, 1)*
	Y	*(0, 2)*	*(3, 4)*

Figure 5

But suppose that player 1 can show for player 2 that the strategy *X* is not strongly dominant for him, i.e., suppose that player 1 sign a contract that will force him to burn two units of utility if he chooses the strategy *X*. So, the new game is shown in Figure 6. In this game, there is also a unique pure equilibrium point (*Y, Z*), but now this equilibrium is efficient.

	Player 2		
Player 1		*W*	*Z*
	X	*(-1, 3)*	*(2, 1)*
	Y	*(0, 2)*	*(3, 4)*

Figure 6

Based on the exposed examples, 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. For example, Laffont and Martimort (2002) affirm that a basic hypothesis of a principal-agent model is the existence of an external and impartial mediator who can monitor and punish any part that violets the contract. Therefore, burning money behavior can be applied in more general economics contexts.

In this paper, we discuss the rationality of burning money games from a new perspective: the mixed Nash equilibrium. We establish necessary and sufficient conditions for the existence of negative and non-positive derivatives of the mixed equilibrium expected utility of a given player *i* with respect to his (or her) own payoffs. In particular, games in which negative derivatives occur are the ones that create incentives for burning utility behavior since such behavior would improve player *i*'s mixed equilibrium expected utility. We show that a negative derivative of 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, in such a case, as defined in Souza and Rêgo (2010), we say that player *j* has a strongly (or strictly) collaboratively dominant strategy for player *i*. 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 act of burning money 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 the mixed equilibrium expected payoff of player *i* taken with respect to a given utility payoff of player *i*, say *a*, is negative, then if player *i* burns utility with respect to *a*, i.e., 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 burn utility with respect to the payoff that makes player *j* converge faster to the strategy that is strongly collaboratively dominant for him. We also point out the difficulties of extending the proposed analysis for more general games, especially regarding of how players will react to burning utility behavior of the other players. Finally, we present an example of how our results can be used to evaluate the cooperation between players, reviewing some conclusions about the security dilemma obtained by Jervis (1978).

For this purpose, the remaining of the paper is structured as follows: in Section 2, we analyze the first-order derivative of the mixed equilibrium expected utility of a given player in a 2x2 normal form game with respect to his (or her) own utility payoffs; in Section 3, we discuss the necessary and sufficient conditions that guarantee the existence of negative derivatives of mixed equilibrium expected utility (or at least non-positive) which would justify the burning utility behavior; in Section 4, we study the problem of finding the best burning utility 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 2x2 games. Finally, the conclusions are presented in Section 7.

2. The analysis of first-order derivatives

We start considering the general structure of a 2x2 game in normal form, as shown in Figure 7.

	Player 2		
Player 1		*W*	*Z*
	X	*(a, e)*	*(b, f)*
	Y	*(c, g)*	*(d, h)*

Figure 7

Let *p* be the probability of player 1 choosing pure strategy *X* and *1-p* the probability of choosing pure strategy *Y*. Similarly, let *q* be the probability of player 2 choosing pure strategy *W* and *1-q* the probability of choosing pure strategy *Z*. 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 that the mixed equilibrium strategies are given by:

$$p = \frac{h-g}{e-f-g+h} \quad (2.1) \quad \text{and} \quad q = \frac{d-b}{a-b-c+d} \quad (2.2)$$

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:

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

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

Once the mixed equilibrium expected utilities are written only in terms of each player own utility payoff, 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. Furthermore, we restrict our attention to changes in payoffs that do not change the general order of the player's preferences to maintain the same strategic situation. By *general order* of the payoffs we mean: if $u>v$, then, after a change in payoffs, it should not happen that $v>u$ and if $u=v$, then this relationship must be maintained after the changes. Thus, for player 1 we have:

$$\frac{\partial EU_1}{\partial a}=\frac{(c-d)(b-d)}{(a-b-c+d)^2} \quad (2.5) \qquad \frac{\partial EU_1}{\partial b}=\frac{(c-d)(c-a)}{(a-b-c+d)^2} \quad (2.6)$$

$$\frac{\partial EU_1}{\partial c}=\frac{(b-a)(b-d)}{(a-b-c+d)^2} \quad (2.7) \qquad \frac{\partial EU_1}{\partial d}=\frac{(b-a)(c-a)}{(a-b-c+d)^2} \quad (2.8)$$

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

On the other hand, for player 2 we have:

$$\frac{\partial EU_2}{\partial e}=\frac{(g-h)(f-h)}{(e-f-g+h)^2} \quad (2.9) \qquad \frac{\partial EU_2}{\partial f}=\frac{(g-h)(g-e)}{(e-f-g+h)^2} \quad (2.10)$$

$$\frac{\partial EU_2}{\partial g}=\frac{(f-e)(f-h)}{(e-f-g+h)^2} \quad (2.11) \qquad \frac{\partial EU_2}{\partial h}=\frac{(f-e)(g-e)}{(e-f-g+h)^2} \quad (2.12)$$

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

Through these general expressions, we can evaluate how the mixed equilibrium expected utility of each player varies, when their respective utility payoffs change, analyzing the sign of the derivative. Before drawing some general conclusions, let us consider what happens in some classic games.

Battle of the sexes game[4]:

Based on Figure 7, the ordering of payoffs for this game is: $a>d>c=b=0$ and $h>e>f=g=0$. In this game there are two pure equilibria (X, W) and (Y, Z) and one mixed equilibrium. So, we have for player 1 (since the analysis is similar to player 2, it will be omitted):

$$\frac{\partial EU_1}{\partial a}=\frac{d^2}{(a+d)^2}>0, \quad \frac{\partial EU_1}{\partial b}=\frac{\partial EU_1}{\partial c}=\frac{2ad}{(a+d)^2}>0, \quad \frac{\partial EU_1}{\partial d}=\frac{a^2}{(a+d)^2}>0.$$

[4] For a better understanding of the history of the Battle of the sexes game, see Luce and Raiffa (1989).

In this game, we conclude that an increase in any of the utility payoffs also leads to an increase in the mixed equilibrium expected utility of its respective player. But does this always happen? That means, an increase in one of the utility payoffs always leads to an increasing of the mixed equilibrium expected utility of a given player? The next example shows that this is not true.

The Stag-hunt game[5]:

Based on Figure 7, the ordering of payoffs is: $a>c>d>b=0$ and $e>f>h>g=0$. In this game there are two pure equilibria (X, W) and (Y, Z) and one mixed equilibrium, but the pair (X, W) is payoff dominant[6], i.e., it is Pareto efficient. Therefore, we have for player 1:

$$\frac{\partial EU_1}{\partial a} = \frac{d(d-c)}{(a-c+d)^2} < 0, \quad \frac{\partial EU_1}{\partial b} = \frac{(c-d)(c-a)}{(a-c+d)^2} < 0,$$

$$\frac{\partial EU_1}{\partial c} = \frac{ad}{(a-c+d)^2} > 0, \quad \frac{\partial EU_1}{\partial d} = \frac{a(a-c)}{(a-c+d)^2} > 0.$$

Now, a strange result emerges. In some cases, when the utility payoff of a given player increases, his mixed equilibrium expected utility decreases. For example[7], when the utility payoff a (resp., e) increases, the mixed equilibrium expected utility of player 1 (resp., 2) decreases, even the pair of strategies (X, W) being a Nash equilibrium. To highlight the problem, imagine the following case: suppose the utility payoffs a and e increase indefinitely, making $a \to \infty$ and $e \to \infty$, while the other payoffs remain constant. So, $\lim_{a\to\infty} q = \lim_{e\to\infty} p = 0$ and, consequently, players will converge to the equilibrium (Y, Z). Thus, when the utilities of the equilibrium (X, W) become extraordinarily higher than the other utility payoffs of the game, the mixed equilibrium strategy profile recommends the players to choose it with extremely low probability. It can be shown that positive variations in the highest and lowest utilities payoffs of each player, would lead to a reduction in their respective mixed equilibrium expected utility. But before discussing the causes of these results, it is appropriate to consider other examples, verifying the similarities between them.

Games without pure equilibrium:

Now, we analyze two games without pure equilibrium. In the first game, the ordering of payoffs is: $d>a>c>b=0$ and $f>g>h>e=0$. Thus, we have for player 1:

$$\frac{\partial EU_1}{\partial a} = \frac{d(d-c)}{(a-c+d)^2} > 0, \quad \frac{\partial EU_1}{\partial b} = \frac{(c-d)(c-a)}{(a-c+d)^2} > 0,$$

[5] See Binmore (1994).
[6] While (Y, Z) is risk dominant. For a complete discussion see Harsanyi and Selten (1988).
[7] The reader is invited to test other cases when the payoffs tend to specific limits and check other strange results.

$$\frac{\partial EU_1}{\partial c} = \frac{ad}{(a-c+d)^2} > 0, \quad \frac{\partial EU_1}{\partial d} = \frac{a(a-c)}{(a-c+d)^2} > 0.$$

In the second game, the ordering of payoffs is: *a>c>d>b=0* and *g>h>f>e=0*. Thus, for player 1 we have:

$$\frac{\partial EU_1}{\partial a} = \frac{d(d-c)}{(a-c+d)^2} < 0, \quad \frac{\partial EU_1}{\partial b} = \frac{(c-d)(c-a)}{(a-c+d)^2} < 0,$$

$$\frac{\partial EU_1}{\partial c} = \frac{ad}{(a-c+d)^2} > 0, \quad \frac{\partial EU_1}{\partial d} = \frac{a(a-c)}{(a-c+d)^2} > 0.$$

Although in the first game all derivatives are positive, in the second game there are negative derivatives which are the ones taken with respect the highest and the lowest utility payoffs of the game.
The chicken game:

In our last example, the ordering of payoffs is: *b>d>c>a=0*, *g>h>f>e=0*. In this game, there are two pure equilibria (*Y, W*) and (*X, Z*) and one mixed equilibrium. Thus, we have for player 1:

$$\frac{\partial EU_1}{\partial a} = \frac{(c-d)(b-d)}{(d-b-c)^2} < 0, \quad \frac{\partial EU_1}{\partial b} = \frac{c(c-d)}{(d-b-c)^2} < 0,$$

$$\frac{\partial EU_1}{\partial c} = \frac{b(b-d)}{(d-b-c)^2} > 0, \quad \frac{\partial EU_1}{\partial d} = \frac{bc)}{(d-b-c)^2} > 0.$$

In this game, we also have some negative derivatives. It is important to observe that, in all examples, whenever a negative derivative exists, it is when it is taken with respect to either the highest or the lowest payoff of the player. In the next section, we prove some necessary and sufficient conditions for the existence of negative derivatives and we show that this result always occurs and it is not a mere coincidence of our examples.

3. Analyzing the sign of the derivatives

In this section, we discuss necessary and sufficient conditions that guarantee that the derivative of expected utility of a given player with respect to a given utility payoff is negative (or at least non-positive). Initially, we analyze the case of expected utility for a given pure equilibrium, as summarized in Lemma 1.

Lemma 1: In any pure Nash equilibrium, the derivatives of the expected utilities of a given player with respect to his (or her) own payoffs are always non-negative.

Proof: The proof of this Lemma is very simple and intuitive. Suppose that the strategy profile (s_i, s_j) is a pure equilibrium of a given game, resulting in a utility $U_i(s_i, s_j) = x$ for player i. Therefore, the derivative of expected utility with respect to player i's payoff is equal to one for any payoff equal to x, and is equal to zero for the other cases.□

This result indicates that, when we analyze a pure equilibrium (when there is at least one pure equilibrium in the game), an increase in some of the payoffs of any player will never reduce his expected utility in that pure equilibrium. Moreover, we can also ensure non-negative derivatives for the following cases, as summarized in Lemma 2. But before, let us make a definition.

Definition 1: Let $\Gamma = (\mathrm{K}, (\mathrm{S_i})_{\mathrm{i}\in\mathrm{K}}, (\mathrm{U_i})_{\mathrm{i}\in\mathrm{K}})$ be a two person game in strategic form and $s_i \neq \hat{s}_i$ be two strategies in S_i, then we say that s_i and $\hat{s}_i$ are *always indifferent* for player $i \in K$ if $\mathrm{U_i}(s_i, \mathrm{s_j}) = \mathrm{U_i}(\hat{s}_i, \mathrm{s_j}), \forall\, \mathrm{s_j} \in \mathrm{S_j}$.

Lemma 2: Based on Figure 7, if player i has a strongly or weakly dominant strategy or is always indifferent between his strategies then the derivatives of the mixed equilibrium expected utility of player i are non-negative.

Proof: Without loss of generality, let us consider that i=1. First, let us analyze the case in which such player has a strongly dominant strategy, say strategy X.[8] Let us consider the possible cases: (a) $e>f$, (b) $f>e$ or (c) $f=e$. If $e>f$ or $f>e$, then the game has a unique pure equilibrium, (X, W) or (X, Z), respectively, and as shown in Lemma 1, in any pure Nash equilibrium, the derivatives of the expected utility are non-negative. Thus, we must check only the case where $e=f$. If this condition occurs, then the game has two pure equilibria and infinitely many mixed equilibria, $E = (M, N)$, where $M = (1, 0)$ and $N = (q^*, 1-q^*)$ for $q^* \in [0, 1]$. Thus, the expected utility of player 1 would be: $EU_1 = aq^* + b(1-q^*)$ and consequently, the derivatives of the expected utility for the payoffs are also non-negative. Secondly, consider the case where X is a weakly dominant strategy (we already know from Lemma 1 that the derivatives of the expected utility of pure equilibrium are at least non-negative, so we will not analyze these cases anymore). Now we have two possible cases: (A) $a=c$ and $b>d$ or (B) $a>c$ and $b=d$. In case (A), the mixed equilibrium expected utility of player 1 is $EU_1 = a = c$; on the other hand, in case (B), the mixed equilibrium expected utility of player 1 is $EU_1 = d = b$. Therefore, it is easy to see that the derivatives are non-negative. Finally, we must examine the case in which the strategies X and Y are always indifferent for player 1, that is, when $a=c$ and $b=d$. In this case, regardless of the mixed strategy chosen by player 2, the mixed equilibrium will result in an expected utility for player 1 equal to $EU_1 = aq - bq + b$. Thus, the derivatives are non-negative.□

Returning to the analysis of the conditions that guarantee non-positive derivatives and strictly negative derivatives, we request the reader to re-examine the games discussed in the beginning of this Section. There, it can be seen that the negative derivatives occurred only in games in which players have a preference that the other

[8] We could also assume Y as the strongly dominant strategy.

uses a particular strategy, regardless of his own choice. To make this idea more formal, consider the concept of collaborative dominance proposed by Souza and Rêgo (2010):

Strong (or Strict) Collaborative Dominance: For game Γ, we say that strategy $s_j^* \in S_j$ is *strongly collaboratively dominant with respect to strategy* $s_j \in S_j$ for player *i* if $U_i(s_i, s_j^*) > U_i(s_i, s_j)$, $\forall\, s_i \in S_i$.

Weak (or Non-Strict) Collaborative Dominance: For game Γ, we say that strategy $s_j^* \in S_j$ is *weakly collaboratively dominant with respect to strategy* $s_j \in S_j$ for player *i* if $U_i(s_i, s_j^*) \geq U_i(s_i, s_j)$, $\forall\, s_i \in S_i$ and, for at least one $\hat{s}_i \in S_i$, $U_i(\hat{s}_i, s_j^*) > U_i(\hat{s}_i, s_j)$.

Theorems 1 and 2 show that a non-positive (respectively, negative) derivative of the mixed equilibrium expected utility of a given player i 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 utility payoffs of player *i*).

Theorem 1: Suppose that player *i* does not have a strongly or a weakly dominant strategy and is not always indifferent between his strategies. So there are two derivatives of the mixed equilibrium expected utility of player *i* taken with respect to player *i*'s 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*).

Proof: Without loss of generality, assume that *i*=1. So, given the assumptions of the Theorem, 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 *W* is weakly collaboratively dominant for player 1. Furthermore, strategy *X* of player 1 is the best response to *W* and $\frac{\partial \mathrm{EU}_1}{\partial \mathrm{a}}$ and $\frac{\partial \mathrm{EU}_1}{\partial \mathrm{b}}$ 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 *Z* is weakly collaboratively dominant for player 1. Furthermore, strategy *Y* of player 1 is the best response to *Z* 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 that player i does not have a strongly or a weakly dominant strategy and is not always indifferent between his strategies. So there are two derivatives of the mixed equilibrium expected utility of player i taken with respect to player i's 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).

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

(A1) $c>d$: In this case, $a>c>d>b$ and strategy W is strongly collaboratively dominant for player 1. Furthermore, strategy X of player 1 is the best response to W 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 Z is strongly collaboratively dominant for player 1. Furthermore, strategy Y of player 1 is the best response to Z 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. Burning money

In the previous section, we showed that whenever negative derivatives happen, they are with respect to the highest and lowest payoffs of a given 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 burn x utility payoff units, then what is the best burning utility strategy that the player can adopt? We now prove that he should burn 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 burn utility if the other player indeed chooses the strongly collaboratively dominant strategy for him (this situation corresponds to burning utility in his highest utility payoff in the game), while in other cases the opposite should happen (this situation corresponds to burning utility in his lowest utility payoff in the game).

In order to show this, 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 we have: [9]

[9] The analysis for player 1 is similar and therefore will be omitted.

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

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

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

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

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.

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

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

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

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

Theorem 2 states that negative derivatives of the 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 Z 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 W and therefore increases the chance of player 2 choosing strategy Z, which, in this case, is strongly collaboratively dominant for player 1. The analysis of the remaining cases is analogous.

So, it is evident that the player 1 should reduce the utility payoff (burning utility) that makes the player 2 converge faster to 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 burn in any payoff. Then, assuming that the general ordering of payoffs in the game is maintained, with respect to what payoff should he burn 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 burn αx units of utility in c and $(1 - \alpha)x$ units in d, with $\alpha \in [0,1]$. The reader should realize that to maintain the order of the payoffs we must ensure that $(1 - \alpha)x \leq d-b$. Thus, we have that the expected utility of player 1:

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

We want to find the value of α that maximizes the expected utility of player 1. Deriving EU_1 with respect to α we have:

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

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 burn the x units in the lowest payoff, c. On the other hand, if $d-b\geq x>(d-b)+(c-a)$, then he should burn 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 burn 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 burn utility with respect to c while $\frac{\partial q}{\partial c} > \frac{\partial q}{\partial d}$, which is equivalent to $d-b>a-c$, or burn 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 burn money. Thus, suppose that player 1 can burn x units of utility:

Case 1: $d>b>a>c$. If $x<(d-b)+(c-a)$, then he should burn the x units of utility with respect to the payoff c, while if $d-b\geq x>(d-b)+(c-a)$, then he should burn it in d.
Case 2: $a>c>d>b$. If $x<(a-c)+(b-d)$, then he should burn the x units of utility with respect to the payoff b, while if $a-c\geq x>(a-c)+(b-d)$, then he should burn it in a.
Case 3: $b>d>c>a$. If $x<(b-d)+(a-c)$, then he should burn the x units of utility with respect to the payoff a, while if $b-d\geq x>(b-d)+(a-c)$, then he should burn it in b.
Case 4: $c>a>b>d$. If $x<(c-a)+(d-b)$, then he should burn the x units of utility with respect to the payoff d, while if $c-a>x>(c-a)+(d-b)$, then he should burn it in c.

Thus, if a player has small power and cannot burn a great amount of utility, then he should invest all his efforts in burning money when he plays the best response (*Y*) to the other player‟s strongly collaboratively dominant strategy (*Z*), but the other player does not play such strategy, which corresponds to his lowest utility payoff in the game. On the other hand, if the player has a greater power, he should invest all his efforts in burning money when he plays the best response (*Y*) to the other player‟s strongly collaboratively dominant strategy and the other player indeed plays such strategy (*Z*), which corresponds to his 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 expected utility of the mixed equilibrium, 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.

In a recent research, Engelmann and Steiner (2007) developed a study that evaluated how the expected *material payoff* of a mixed equilibrium (for a given player) increases or decreases with the degree of risk aversion of this given player. For this purpose, the authors focused on 2x2 games with two pure equilibria and one mixed equilibrium, restricting their analysis to the mixed equilibrium[10]. As their main contribution, the authors identified conditions, with respect to the *material payoffs*, that guarantee that the expected material equilibrium payoff of a given player is an increasing function of his risk aversion degree, as summarized by the following propositions.

Proposition 1 (ENGELMANN and STEINER, 2007, p.383-384): When, $a>c>d>b$ or $a>d>c>d,$ the equilibrium probability q that player 2 chooses strategy W increases in the degree of risk aversion of player 1.

Proposition 2 (ENGELMANN and STEINER, 2007, p.385): In any mixed equilibrium of a 2x2 game, if $a>c>d>b,$ then the expected material payoff of player 1 increases in his degree of risk aversion.

The intuition behind Proposition 2 is as follows. Since $a>c>d>b,$ then, based on Proposition 1, we kwon that the probability q that player 2 chooses strategy W increases in the degree of risk aversion of player 1. Since strategy W of player 2 is strongly collaboratively dominant for player 1, then player 1 will always benefit (in terms of *material payoff*) by any increase in q. For a formal proof, see Engelmann and Steiner (2007). The authors also admit that their approach does not allow them to make any conclusion about the expected utility, and that because a variation in risk preference will lead to a variation in the utility of each (or some) pure strategy profile, and depending on the aggregate change, the new expected utility can increase, decrease or remain unchanged. In our paper, we provide a new contribution in the sense that we do not deal with *material payoffs.* We discussed how variations in utility may increase the mixed equilibrium expected utility of a given player.

5. Discussions

Until this section, we restricted our analysis of mixed equilibrium and the problem of burning utility only to 2x2 games with a single mixed equilibrium. 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 begin the discussion by analyzing the game shown in Figure 8, for which the conclusions of Section 4 are still valid (with the appropriate adjustments).

---

[10] The authors also make some restriction on the payoffs to simplify the analysis: $|a-b|>|c-d|$ and $a-b>0$, with sign($a-c$)= sign($d-b$) and sign($e-f$)= sign($h-g$).

In this game we only have one mixed equilibrium ((1/3, 2/3), (1/3, 0, 2/3)) and its support is $\{\alpha_1, \alpha_2\} \times \{\beta_1, \beta_3\}$. Moreover, the expected utility of the players are (13/3, 17/3). Also note that in this game, the strategy β_1 is strongly collaboratively dominant with respect to strategy β_3 for player 1 and, without consider β_2 since it is outside the equilibrium support, the strategy α_2 is strongly collaboratively dominant with respect to α_1, for player 2. So for this game, we can use the results of Theorem 2 which indicates, for example, that a reduction in utility $U_1(\alpha_1, \beta_1)$ in two units would increase the expected utility of player 1 to 5 and a reduction of utility $U_2(\alpha_2, \beta_1)$ in one unit would increase the expected utility of player 2 to 6, i.e., both players would like to burn utility if they could.

Player 1	*Player 2*			
		β_1	β_2	β_3
	α_1	(7, 3)	(4, 7)	(3, 5)
	α_2	(5, 7)	(6, 2)	(4, 6)

Figure 8

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 2x2 game. Now, we analyze a game in which all three pure strategies of player 2 are in the equilibrium support, as shown in Figure 9.

Player 1	*Player 2*			
		β_1	β_2	β_3
	α_1	*(8, 0)*	*(3, 1)*	*(2, 1)*
	α_2	*(6, 1)*	*(4, 0)*	*(5, 0)*

Figure 9

Before calculating the mixed equilibrium of this game, let us define some notation. Let $\sigma(\alpha_1)$ be the probability of player 1 choosing α_1 (hence $\sigma(\alpha_2) = 1 - \sigma(\alpha_1)$ is the probability that he chooses α_2) and let $\sigma(\beta_1)$ be the probability of player 2 choosing β_1 and $\sigma(\beta_2)$ be the probability of choosing β_2 (indeed, $\sigma(\beta_3) = 1 - \sigma(\beta_1) - \sigma(\beta_2)$). Thus we can characterize the mixed equilibrium of this game as follows: $\left((½, ½), \left(\sigma(\beta_1), \frac{3-5\sigma(\beta_1)}{2}, \frac{3\sigma(\beta_1)-1}{2}\right)\right)$, where $\sigma(\beta_1) \in \left[\frac{1}{3}, \frac{3}{5}\right]$. Moreover, the expected utility of the mixed equilibrium for player 1 is given by $EU_1 = \frac{7\sigma(\beta_1)+7}{2}$, and depending on the value of $\sigma(\beta_1)$, it can vary in the range $EU_1 \in \left[\frac{14}{3}, \frac{28}{5}\right]$.

Now, consider the mixed equilibrium ((½, ½), (1/2, 1/4, 1/4). In this case, the expected utility of player 1 is 5.25. Note that in this game, the strategy β_1 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 β_1 more frequently. Suppose that player 1 reduces $U_1(\alpha_1, \beta_1)$ from 8 to 7. Then, we can characterize the mixed equilibrium of the new game as follows: $\left((½, ½), \left(\sigma(\beta_1), \frac{3-4\sigma(\beta_1)}{2}, \frac{2\sigma(\beta_1)-1}{2}\right)\right)$, where $\sigma(\beta_1) \in \left[\frac{1}{2}, \frac{3}{4}\right]$. Indeed, player 1's mixed equilibrium expected utility is $EU_1 = \frac{6\sigma(\beta_1)+7}{2}$, and depending on the value of $\sigma(\beta_1)$, it can vary in the range $EU_1 \in \left[5, \frac{23}{4}\right]$. In this case, is easy to see that player 2 may, for example, keep $\sigma(\beta_1)$ equal to ½ (just changing the values of $\sigma(\beta_2)$ and $\sigma(\beta_3)$). In such situation, player 1's expected utility reduces to 5. Once player 2 has a range[11] of values for which he can manipulate $\sigma(\beta_1)$, in general, it is impossible to say how he will react to any change in payoffs made by player 1.

Now consider a game with three players each one with two strategies, as shown in Figure 10. Admit that the payoff a from the strategy profile $(\alpha_1, \beta_1, \gamma_1)$ is a value between 6 and 9, $a \in [6, 9]$. Thus, this game has two pure equilibria, $(\alpha_1, \beta_1, \gamma_1)$ and $(\alpha_2, \beta_2, \gamma_2)$ and one mixed equilibrium.

γ_1

	β_1	β_2
α_1	(a, 8, 8)	(5, 7, 5)
α_2	(3, 5, 3)	(6, 6, 1)

γ_2

	β_1	β_2
α_1	(0, 3, 7)	(1, 4, 6)
α_2	(4, 1, 4)	(2, 2, 2)

Figure 10

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. The expected utility of player 1 is given by Equation 5.1; additionally, Figure 11 shows the mixed equilibrium expected utility of player 1 when a varies from 6 to 9.

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

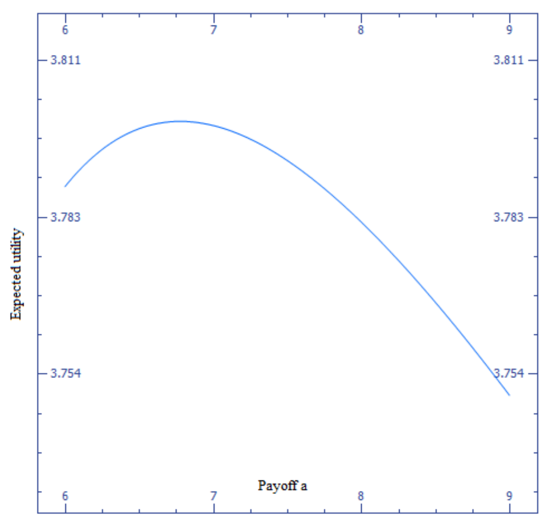

Figure 11

So for any value of a higher than 6.9 (approximately) 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 6.9 (approximately) 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

[11] There are an intersection between the two cases, $\sigma(\beta_1) \in \left[\frac{1}{2}, \frac{3}{5}\right]$.

assume a initial value of 8, a *small reduction* in any payoff of player 1 related to the pure strategy α_1, will lead an increase of player's 1 mixed equilibrium expected utility, even neither of the players (2 nor 3) having a strategy that is collaboratively dominant for player 1. Consequently, in more general class of games, the existence of negative derivatives does not depend of 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.

Additionaly, based on Figure 10, supposes that the payoff of player 2 from the strategy profile $(\alpha_2, \beta_2, \gamma_1)$ was reduced from 6 to 4, and *a* is equal to 8, as shown in Figure 12.

γ_1

	β_1	β_2
α_1	(8, 8, 8)	(5, 7, 5)
α_2	(3, 5, 3)	(6, 4, 1)

γ_2

	β_1	β_2
α_1	(0, 3, 7)	(1, 4, 6)
α_2	(4, 1, 4)	(2, 2, 2)

Figure 12

This new game also has two pure equilibria, $(\alpha_1, \beta_1, \gamma_1)$ and $(\alpha_2, \beta_2, \gamma_2)$ 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 does not depend only of his own payoffs, as was the case in the 2x2 games. So, these facts prevent us to make extensions of Theorems 1 and 2 to a more general class of games.

6. An application: The Security dilemma

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. A numerical example of this game is shown in Figure 3.

For the author, 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 (*X, W*) it does not increase the incentive of them actually choose those respective strategies. For example, when player 1 declares that he will play *X*, it does not add any information to player 2, because, since *W* is a strongly collaboratively dominant strategy for player 1, player 2 knows that player 1 prefers that he (player 2) plays *W*. Thus, player 2 knows that player 1 would state that consents to any agreement in which player 2 plays *W*, but this fact does not guarantee that the player 1 will really fulfill the agreement and play *X*. For example, player 1 may prefer to play *Y*, 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 on the payoffs. For this, we also illustrate our argumentation with the stag-hunt game which is also known as the security dilemma due to the work of Jervis

(1978). Furthermore, we will critically analyze some passages from the work of Jervis, 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)[12]. The order of preferences for the possible strategies profiles are equivalent to 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 be 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 to study 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 change in payoffs should not alter the original decision, except that the change in payoffs is such that modify the original equilibrium set of the game. Therefore, we must analyze 'what makes cooperation more likely' from the perspective of the mixed equilibrium.

In Section 3, Figure 7, we saw that 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 5, we saw that $\frac{\partial q}{\partial a}$ and $\frac{\partial q}{\partial b}$ are negative (the same holds for $\frac{\partial p}{\partial e}$ and $\frac{\partial p}{\partial g}$) 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 $\frac{\partial q}{\partial c}$ and $\frac{\partial q}{\partial d}$ 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).

[12] In our early version of the stag-hunt game, cooperation is represented by the strategies *X* and *W*, and non-cooperation is represented by strategies *Y* and *Z*.

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 and long-term relationship, 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 burning money behavior through the analysis of the mixed Nash equilibrium in normal form games. We provide a necessary and sufficient condition for the existence of negative derivatives of the expected utility that justify burning money behavior. Furthermore, we use our insights to analyze the security dilemma revisiting some conclusions made by Jervis (1978).

References

Aumann, R. J., 1990 Nash Equilibria are not Self-Enforcing. In Gabszewicz J. J., Richard J. F., Wolsey L. (ed) Economic Decision Making, Econometrics, and Optimisation: Essays in Honor of Jacques Dreze. Elsevier Science Publishers, Amsterdam, pp.201-206.

Ben-Porath, E., Dekel, E., 1992. Signaling future actions and the potential for sacrifice, Journal of Economic Theory. 57, 36-51.

Binmore, K., 1994. Game theory and the Social Contract Volume I: Playing Fair, MIT Press, Cambridge.
Brandts, J., Holt, A., 1995. Limitations of dominance and forward induction: Experimental evidence, Economics Letters, 49, 391-395.
Engelmann, D., Steiner, J., 2007. The effects of risk preferences in mixed-strategy equilibria of 2x2 games, Games and Economic Behavior, 60, 381-388.
Fundenberg, D., Tirole, J., 1991. Game Theory. MIT Press, Cambridge.
Gersbach, H., 2004. The money-burning refinement: with an applicationto a political signalling game, International Journal of Game Theory, 33, 67–87.
Hammond. P.J., 1993. Aspects of rational behavior, In Binmore, K., Kirman, A. and Tani, P. (ed) Frontiers of Games Theory, pp.307-320.
Harsanyi, J. C., Selten, R., 1988. A General Theory of Equilibrium Selection in Games, MIT Press, London.
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.
Jervis, R., 1978. Cooperation under the Security Dilemma, World Politics, 30,167-214.
Kohlberg, E., Mertens, J. F., 1986. On the Strategic Stability of Equilibria, Econometrica, 54, 1003-1037.
Laffont, J-J., Martimort, D., 2002. The Theory of Incentive: The principal-agent model, Princeton University Press, Princeton.
Luce, R. D., Raiffa, H., 1989. Games and Decision: Introduction and Critical Survey, Dover, New York.
Myerson, R. B., 1991). Game theory: analysis of conflict, Harvard University Press London.
Shimoji. M., 2002. On forward induction in money-burning games, Economic Theory, 19, 637–648.
Souza, F. C.; Rêgo, L. C., 2010. Collaborative Dominance: When Doing Unto Others as You Would Have Them Do Unto You Is Rational, working paper.
Stalnaker, R., 1998. Belief revision in games: forward and backward induction, Mathematical Social Sciences, 36, 31–56.
Van Damme, E., 1989. Stable equilibria and forward induction, Journal of Economic Theory, 48, 476-496.