

# Risk limits, conflict, and equilibrium selection in games with multiple equilibria

Kjell Hausken

**Abstract:** A risk limit conflict measure is developed as the product of the two players' maximum probabilities of being recalcitrant when pursuing a preferred equilibrium. Although the justification for it is different, the measure is equivalent to Axelrod's (1970) measure, which is the ratio of infeasible joint demand and joint demand above the threat point which he illustrated graphically. Axelrod did not justify his measure beyond informal verbal descriptions. The article furthermore offers an equilibrium selection in favour of the player with the largest risk limit. The equilibrium selection is different from Harsanyi and Selten's (1988, p.90) equilibrium selection, which assigns equal weight to four payoff differences, which the article argues is not realistic. The equilibrium selection is also compared with Hausken (2007).

**Keywords:** conflict; risk limit; equilibrium selection; equilibrium refinement.

## 1 Introduction

Struggle for preferred equilibria is common when players attempt to agree on standards, or agree on procedures for interaction. It is a crucial characteristic, for example, at the start-up of most social institutions (Knight, 1992). The significance of games with two Pareto-superior Nash equilibria is substantial. This article develops a risk limit conflict measure for such games, and presents a method for selecting between the two equilibria.

Axelrod (1970) presented a conflict measure, which is the ratio of the area of infeasible joint demand, and the area of joint demand, for two players. His presentation is gametheoretic, but lacks mathematical rigour, and he provides a verbal reference to five properties. This article considers the risk limits of two players, and uses these to develop a risk limit conflict measure which, remarkably, is shown to be equivalent to Axelrod (1970) measure. The measure is demonstrated to satisfy two properties.

The article proceeds to consider Harsanyi and Selten's (1988) risk dominance criterion, which is a comparison of Nash products, which determines which equilibrium gets selected. Harsanyi and Selten (1988) assign equal weight to four payoff differences. In contrast, this article uses the risk limit approach to argue that these four payoff differences play different roles in the players' reasoning processes. The reasoning implies a new equilibrium selection method where the player with the highest risk limit gets his preferred equilibrium. The article further compares with Hausken's (2007) equilibrium selection based on the players' win probabilities. The conflict measure and three equilibrium selection methods are illustrated for the six games 64–69 in Rapoport and Guyer's (1966, p.213) taxonomy, which are conflict games with two equilibria where neither player has a dominating strategy. See van Damme (2002) and van Huyck (1997) for further discussions of equilibrium selection.

Section 2 develops a risk limit conflict measure. Section 3 presents the two properties of the conflict measure. Section 4 develops a new equilibrium selection proposition. Section 5 compares with Harsanyi and Selten's (1988) risk dominance criterion. Section 6 compares with Hausken's (2007) equilibrium selection. Section 7 illustrates the conflict measure and three equilibrium selection methods. Section 8 concludes.

## 2 A risk limit conflict measure

Consider the game in Table 1 where  $a_1 \geq b_1 \geq t_1$ ,  $b_2 \geq a_2 \geq t_2$ ,  $a_1 \geq d_1$ ,  $b_2 > d_2$  or  $a_1 > d_1$ ,  $b_2 \geq d_2$ .<sup>1</sup> The two pure strategy equilibria are  $(a_1, a_2)$  and  $(b_1, b_2)$ . Row Player 1 prefers  $(a_1, a_2)$ , and column Player 2 prefers  $(b_1, b_2)$ .

**Table 1** Two-person two-strategy game with two equilibria

	I	II
I	$a_1, a_2$	$t_1, t_2$
II	$d_1, d_2$	$b_1, b_2$

Let us first link Table 1 to Harsanyi and Selten's (1988) application of the notion of risk dominance<sup>2</sup> as a criterion for equilibrium selection. They state that  $(a_1, a_2)$  risk dominates  $(b_1, b_2)$  if

$$(a_1 - d_1)(a_2 - t_2) > (b_1 - t_1)(b_2 - d_2) \quad (1)$$

which is a comparison of Nash products.<sup>3</sup> Mathematically, Harsanyi and Selten (1988) assign equal weight to the four payoff differences in (1). This means that only the four payoff differences as such matter, not whether the differences pertain to equilibrium or non-equilibrium outcomes. Consider first equilibrium  $(a_1, a_2)$ . Player 1 prefers  $a_1$  relative to  $d_1$  which gives the difference  $a_1 - d_1$ . Player 2 prefers  $a_2$  relative to  $t_2$  which gives the difference  $a_2 - t_2$ . Then consider equilibrium  $(b_1, b_2)$ . Player 1 prefers  $b_1$  relative to  $t_1$  which gives the difference  $b_1 - t_1$ . Player 2 prefers  $b_2$  relative to  $d_2$  which gives the difference  $b_2 - d_2$ . Equal weight to these four payoff differences means equal weight to  $(t_1, t_2)$  and  $(d_1, d_2)$  which places these two non-equilibrium outcomes on an equal footing.

This article argues that  $(t_1, t_2)$  and  $(d_1, d_2)$  are not on an equal footing. Consider equilibrium  $(a_1, a_2)$ . Player 1 does not prefer to switch from I to II since  $a_1 \geq d_1$ . The payoff  $d_1$  is irrelevant for Player 1's process of reasoning. In contrast, Player 2 prefers the other equilibrium and may prefer to switch from I to II. Although  $a_2 \geq t_2$ , Player 2 observes that since  $b_2 \geq a_2$ , the only way to reach  $(b_1, b_2)$  is to switch from I to II. This may cause  $t_2$ , but it may also cause  $b_2$  if Player 1 is thereby induced to switch from I to II as a consequence of Player 2's challenge. Hence,  $t_2$  indeed plays a role in Player 2's process of reasoning. Analogously for equilibrium  $(b_1, b_2)$ , the payoff  $d_2$  is irrelevant in Player 2's reasoning, while  $t_1$  is relevant in Player 1's reasoning if he challenges Player 2.

Consequently,  $(d_1, d_2)$  is irrelevant for the players' processes of reasoning when they focus their attention on either of the two equilibria. Prior to playing the static game in Table 1, i.e., prior to making independent, effectively simultaneous, choices, we assume that the players engage in pre-play communication. They reason such that each can hold out or acquiesce. They either both hold out, one holds out and the other acquiesces, one acquiesces and the other holds out, or both acquiesce. It is dysfunctional for both players to acquiesce. We assume that the pre-play communication takes place over a time interval such that if one acquiesces, then the other does not. Hence, the players consider  $(t_1, t_2)$  as relevant for their reasoning, while  $(d_1, d_2)$  is not relevant.

Let us determine the highest probability or risk of conflict between  $(a_1, a_2)$  and  $(b_1, b_2)$ . Assume without loss of generality that the players anchor their focus on equilibrium  $(a_1, a_2)$ .  $q_1$  is the probability that Player 1 sticks to his first strategy given that Player 2 challenges, and  $q_2$  is the probability that Player 2 sticks to her challenge given that Player 1 does not acquiesce. Hence,  $q_1q_2$  is the probability or risk of conflict, implying the threat point  $(t_1, t_2)$ . In the threat point, each player may acquiesce, or exhibit aggressive, recalcitrant, stubborn, hard-headed, conflictful behaviour by attempting to induce (persuade, convince, force) the other player to accept one's own preferred equilibrium. Player 2's expected payoff of a challenge is  $q_1t_2 + (1 - q_1)b_2$ , and she gets  $a_2$  otherwise. Player 2 challenges when

$$q_1t_2 + (1 - q_1)b_2 > a_2 \Rightarrow q_1 < \frac{b_2 - a_2}{b_2 - t_2} = q_1^* = r_2. \quad (2)$$

Equation (2) does not mean that the players play conditional strategies. We consider a static game, but pre-play communication allows the players to reason about their payoffs dependent on which strategy the other player chooses. When  $q_1 < q_1^*$ , Player 2 should challenge, otherwise she should not. Increasing  $q_1$  means that Player 2 incurs a higher risk of conflict when challenging. The maximum risk that Player 2 is willing to incur is  $r_2 < q_1^*$ , defined as Player 2's upper acceptable risk limit. The numerator in (2) is zero when  $b_2 = a_2$ . This implies zero risk limit  $r_2 = q_1^* = 0$ . That Player 2 accepts a risk limit of 0, means that she accepts no risk, she never challenges, and she gets  $a_2$ . Conversely,  $r_2 = q_1^* = 1$  when  $a_2 = t_2$ . That Player 2 accepts a risk limit of 1, means that she accepts maximum risk, she always challenges, and she gets  $t_2$ . Given that Player 2 challenges, the expected payoff to Player 1 if he resists is  $q_2t_1 + (1 - q_2)a_1$ , and he gets  $b_1$  otherwise. Player 1 resists if

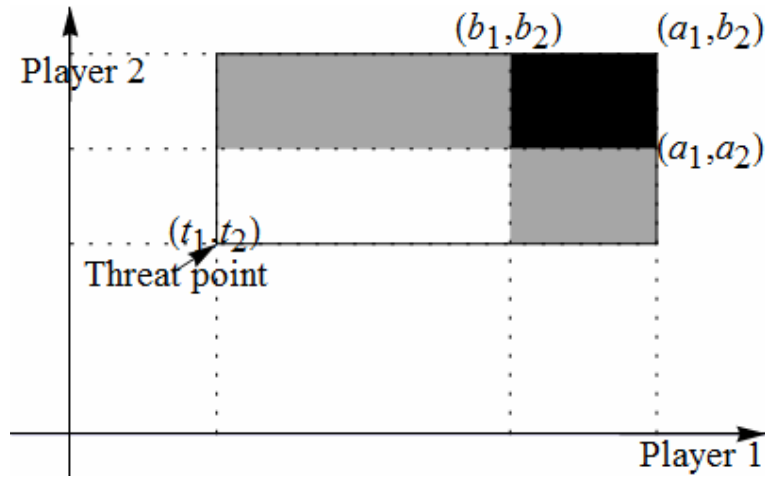
$$q_2t_1 + (1 - q_2)a_1 > b_1 \Rightarrow q_2 < \frac{a_1 - b_1}{a_1 - t_1} = q_2^* = r_1. \quad (3)$$

When  $q_2 < q_2^*$ , Player 1 should resist the challenge, otherwise he should acquiesce. Increasing  $q_2$  means that Player 1 incurs a higher risk of conflict when resisting. The maximum risk that Player 1 is willing to incur is  $r_1 < q_2^*$ , defined as Player 1's upper acceptable risk limit. The numerator in (3) is zero when  $a_1 = b_1$ . This implies zero risk limit  $r_1 = q_2^* = 0$ . Hence, Player 1 accepts no risk, acquiesces, and gets  $b_1$ . Conversely,  $r_1 = q_2^* = 1$  when  $b_1 = t_1$ . Thus, Player 1 accepts maximum risk, resists the challenge, and gets  $t_1$ .

When both  $q_1$  and  $q_2$  are low, the risk  $q_1 q_2$  of conflict is low.  $q_1^*$  and  $q_2^*$  are maximum individual probabilities of individual conflictful behaviour, and  $r_1$  and  $r_2$  are the maximum individually acceptable risks. As  $q_1$  and  $q_2$  increase, conflict gets avoided when  $q_1 > q_1^*$  or  $q_2 > q_2^*$ . The most interesting point is the highest  $q_1$  where Payer 2 challenges combined with the highest  $q_2$  where Payer 1 resists, which gives conflict. The probability of this occurrence is  $q_1^* q_2^* = r_1 r_2$ , which is the joint maximum probability that both players choose their respective conflictful strategies. It is a maximum since each of the constituents  $q_1^*$  and  $q_2^*$  are maximum individual probabilities of conflictful behaviour. The degree of conflict cannot get any higher than this, since if  $q_1$  increases above  $q_1^*$ , Player 2 does not challenge, and if  $q_2$  increases above  $q_2^*$ , Player 1 acquiesces. However, if  $q_1^*$  and  $q_2^*$  increase, the degree of conflict increases. Given this, we propose the risk limit conflict measure

$$c = q_1^* q_2^* = r_1 r_2 = \frac{(a_1 - b_1)(b_2 - a_2)}{(a_1 - t_1)(b_2 - t_2)}. \quad (4)$$

**Figure 1** Illustration of risk limit conflict measure



The product  $c$  of the risk limits is the product of the maximum degrees of recalcitrance the players assign to each other, which is the maximum degree of conflict between the players. Equation (4) has a very remarkable feature. Namely, Axelrod (1970) has suggested the same conflict measure in 1970, though without the risk limit considerations in (2) and (3). Let us briefly recapitulate his analysis which is intuitive, graphic with utility diagrams, and laced with examples of feasible and unfeasible outcomes using apples and oranges. Axelrod (1970, Chapter 2, pp.19–57) develops his measure based on a dialogue between an empiricist and a theoretician. First, they develop a 'no agreement point', which corresponds to  $(t_1, t_2)$  in Figure 1.

‘Either player can veto anything other than the no agreement point’ (p.20). Then they agree that ‘normalising a game has no effect on the amount of conflict of interest’ (p.28). The next step is to agree graphically, referring to Figure 1, that an area of joint demand exists spanned out by the threat point  $(t_1, t_2)$  and the outmost point determined by the best payoff  $(a_1, b_2)$  each player can possibly obtain under his most favourable circumstances. Furthermore, the small black rectangle spanned out by  $(b_1, a_2)$  and  $(a_1, b_2)$  is ‘the proportion of the joint demand area which is infeasible’ (p.57). Axelrod thereafter concludes that a conflict measure should equal the latter small black rectangle divided by the larger rectangle spanned out by  $(b_1, a_2)$  and  $(a_1, b_2)$ . This, amazingly, expresses the same area as that developed mathematically in (4) using the risk limit approach. But, the notion of risk limits is absent in Axelrod’s analysis. Axelrod further argues without proof that his intuitive procedure, in the form of a narrative of what conflict entails, satisfies the five properties symmetry, independence, continuity, boundedness, additivity, and that ‘there is no other way to satisfy’ these five properties (p.32). Axelrod concludes (p.57) that the measure ‘makes sense intuitively’ ‘because the less incompatible are the goals of the players, the more the region of feasible agreements bulges and thus the less conflict of interest there is’.

The contribution of this article is to provide a more rigorous foundation for the conflict measure in (4) in terms of how the two players reason strategically by focusing on their risk limits. That the conflict measure is equivalent to that suggested by Axelrod strengthens it since the robustness of a result benefits from reaching that result multifariously. Let us link the risk limit approach to Axelrod’s analysis. Equation (4) consists of a small rectangle in the numerator, and a larger rectangle in the denominator. In the two-dimensional utility diagram for the two players in Figure 1, the risk limit conflict measure is the relation of the small black rectangle  $(a_1 - b_1)(b_2 - a_2)$  of risky conflictful behaviour to the large rectangle  $(a_1 - t_1)(b_2 - t_2)$  of joint demand.<sup>4</sup> Axelrod (1970, p.57) refers to the small rectangle as ‘the proportion of the joint demand area which is infeasible’,<sup>5</sup> and to the large rectangle as the area of joint demand spanned out by the threat point  $(t_1, t_2)$  and the outmost point determined by the best payoff  $(a_1, b_2)$  each player can possibly obtain under his most favourable circumstances.

The risk limit conflict measure increases when the black rectangle (area of infeasible joint demand) becomes larger or the large rectangle (area of joint demand) becomes smaller by increasing  $(t_1, t_2)$ . Altering  $(b_1, a_2)$  changes only the small rectangle, altering  $(t_1, t_2)$  changes only the large rectangle, while altering  $(a_1, b_2)$  changes both rectangles. At the limit when the two rectangles overlap, which means  $b_1 = t_1$  and  $a_2 = t_2$ , the conflict measure is 1.

Conversely, the risk limit conflict measure decreases when the black rectangle becomes smaller or the large rectangle becomes larger by decreasing  $(t_1, t_2)$ . At the limit when the small rectangle disappears, which means  $a_1 = b_1$  and  $b_2 = a_2$ , or the large rectangle becomes infinitely large while the small rectangle is finite, which means  $t_1 \rightarrow -\infty$  or  $t_2 \rightarrow -\infty$ , the conflict measure is 0.

### 3 Two properties of the conflict measure

Equation (4) satisfies the following two properties.

Property 1 A conflict measure should be lowest when the two Nash equilibria are identical or non-discriminating, and should increase as the product of the payoff difference each player experiences between the two equilibria increases.

Property 2 A conflict measure should be highest when the Nash equilibria in question are weak (weakly dominant), and should decrease as the product of the payoff difference each player experiences between his non-preferred equilibrium and the threat point increases.

Let us link the two properties to the risk limit approach. Property 1 implies no conflict when  $a_1 = b_1$  and  $b_2 = a_2$  (the small black rectangle disappears). Property 2 implies maximum conflict when  $b_1 = t_1$  and  $a_2 = t_2$  (the two rectangles overlap). Player 2's upper acceptable risk limit in (2) equals zero when the numerator is zero, which occurs when  $b_2 = a_2$  (Property 1). Conversely, Player 2's upper acceptable risk limit equals one when  $a_2 = t_2$  (Property 2). Analogously, Player 1's upper acceptable risk limit in (3) is zero when  $a_1 = b_1$  (Property 1), and one when  $b_1 = t_1$  (Property 2).

### 4 A new equilibrium selection proposition

The conflict measure in (4) is such that  $(d_1, d_2)$  plays no role, and  $(d_1, d_2)$  is also absent in (2) and (3). The reason is that Player 2 (and not Player 1) prefers to switch strategy to avoid  $(a_1, a_2)$ , which may give the threat point  $(t_1, t_2)$ , and that Player 1 (and not Player 2) prefers to switch strategy to avoid  $(b_1, b_2)$ , which may also give  $(t_1, t_2)$ . The players' risk limit reasoning processes cause transition between the two equilibria through  $(t_1, t_2)$ , and not through  $(d_1, d_2)$ . Hence,  $(d_1, d_2)$  does not enter the players' thought and evaluation processes when choosing among  $(a_1, a_2)$  and  $(b_1, b_2)$ . Since  $(d_1, d_2)$  is irrelevant, we propose that  $(a_1, a_2)$  risk dominates  $(b_1, b_2)$  if<sup>6</sup>

$$r_1 > r_2 \Leftrightarrow \frac{a_1 - b_1}{a_1 - t_1} > \frac{b_2 - a_2}{b_2 - t_2}. \quad (5)$$

The player with the highest risk limit gets his preferred equilibrium. No risk, no gain. A player with zero tolerance for risk can most reasonably expect his non-preferred equilibrium. Let us formulate (5) as an equilibrium selection proposition.

*Proposition:* Equilibrium  $(a_1, a_2)$  is selected if  $r_1 > r_2$ , equilibrium  $(b_1, b_2)$  is selected if  $r_1 < r_2$ , and there is no equilibrium selection if  $r_1 = r_2$ , where  $r_1$  and  $r_2$  are defined in (2),(3), and (5) based on the game in Table 1.

*Proof:* Follows from (2), (3) and (5).

### 5 Comparison with Harsanyi and Selten's (1988) risk dominance criterion

The absence of  $(d_1, d_2)$  in the risk limit and conflict measures amounts to criticising Harsanyi and Selten's (1988, p.90) application of the notion of risk dominance as a criterion for equilibrium selection, as expressed in (1), where  $(d_1, d_2)$  plays a role. The problem with Harsanyi and Selten's (1988) conceptualisation is that it mathematically assigns equal weight to four payoff differences.

## 6 Comparison with Hausken's (2007) equilibrium selection

Hausken (2007) argues that Player 1's incentive to be stubborn in insisting on his preferred equilibrium depends on two factors. First, it depends on  $a_1 - b_1$ , which expresses the extent to which Player 1 prefers equilibrium  $(a_1, a_2)$  rather than  $(b_1, b_2)$ . Second, it depends on  $a_2 - t_2$ , which is Player 2's lack of inclination to go along with Player 1's preferred equilibrium  $(a_1, a_2)$ . Multiplying these two effects,  $(a_1 - b_1)(a_2 - t_2)$  expresses Player 1's incentive to be stubborn. The analogous expression  $(b_2 - a_2)(b_1 - t_1)$  expresses Player 2's incentive to be stubborn. Dividing with the sum of the two expressions to scale the sum of the two players' stubbornness incentives to be equal to one gives the players' stubbornness

$$s_1 = \frac{(a_1 - b_1)(a_2 - t_2)}{(a_1 - b_1)(a_2 - t_2) + (b_2 - a_2)(b_1 - t_1)}, s_2 = \frac{(b_2 - a_2)(b_1 - t_1)}{(a_1 - b_1)(a_2 - t_2) + (b_2 - a_2)(b_1 - t_1)}. \quad (6)$$

The probability that Player 1 wins his preferred  $(a_1, a_2)$  is  $s_1(1 - s_2)$ , and the probability that Player 2 wins her preferred  $(b_1, b_2)$  is  $(1 - s_1)s_2$ . Hence, Hausken (2007) defines  $s_1^2$  and  $s_2^2$  as the players' win probabilities or power. He further suggests the equilibrium selection

$$s_1 > s_2 \Leftrightarrow (a_1 - b_1)(a_2 - t_2) > (b_2 - a_2)(b_1 - t_1) \quad (7)$$

which differ from both (5) and (1).

## 7 Illustration of the conflict measure and three equilibrium selection methods

Rapoport and Guyer (1966) developed a taxonomy of 78  $2 \times 2$  games, which is the number of non-equivalent games when accounting for ordinal preference orderings.<sup>7</sup> The 78 games are divided into three classes.

Class I Each player has a dominating strategy

- No-conflict games (1–6)
- Games with strongly stable equilibria (7–11)
- Games with strongly stable deficient equilibrium (12, prisoner's dilemma)
- Games with stable equilibria (13–18)
- Games with threat-vulnerable equilibria (19–21)

Class II One player has a dominating strategy

- No conflict games (22–30)
- Games with stable equilibria (31–36)
- Games with threat-vulnerable equilibria (37–39)
- Games with force-vulnerable equilibria (40–48)
- Games with unstable equilibria (49–57)

Class III Neither player has a dominating strategy

- Two-equilibria no conflict games (58–63)<sup>8</sup>
- Two-equilibria games with equilibrium outcome (64–65)
- Two-equilibria games with non-equilibrium outcome (66–69)
- Games without equilibria (70–78)

Rapoport and Guyer (1966, p.209) argue for Game 64, shown in Table 2, that (3, 4) is the natural outcome since Player 1 has no threat at his disposal at (4, 3). If Player 1 switches from I to II, Player 2 immediately switches from I to II causing (3, 4). In contrast, Player 2 has both threat and force at his disposal at (3, 4). It is to Player 1's advantage to switch voluntarily from I to II so that the return to (3, ) will be via an outcome that is better for him [(2, 1) rather than (1, 2)]. Rapoport and Guyer (1966, p.209) argue with one sentence that Game 65 is similar to Game 64, which is more controversial (note that their objective is to classify games, not to provide necessarily convincing arguments for the most controversial games). Below, we show how the risk limits are different for Games 64 and 65.

**Table 2** Six two-equilibria conflict games

<i>Game</i>	<i>64</i>		<i>65</i>		<i>66</i>		<i>67</i>		<i>68</i>		<i>69</i>	
	4, 3	1, 2	4, 3	1, 2	4, 2	1, 1	4, 2	1, 1	4, 3	1, 1	4, 3	2, 2
	2, 1	3, 4	3, 1	2, 4	3, 3	2, 4	2, 3	3, 4	2, 2	3, 4	1, 1	3, 4
$r_1$	1/3		2/3		2/3		1/3		1/3		1/2	
$r_2$	1/2		1/2		2/3		2/3		1/3		1/2	
$c$	1/6		1/3		4/9		2/9		1/9		1/4	
Equilibrium selection	(3, 4)		(4, 3)		Indifferent		(3, 4)		Indifferent		Indifferent	
$s_1$	1/3		2/3		1/2		1/5		1/2		1/2	
$s_2$	2/3		1/3		1/2		4/5		1/2		1/2	
Hausken (2007)	(3, 4)		(4, 3)		Indifferent		(3, 4)		Indifferent		Indifferent	
$(a_1 - d_1)(a_2 - t_2)$	2		1		1		2		4		3	
$(b_1 - t_1)(b_2 - d_2)$	6		3		1		2		4		3	
Harsanyi and Selten (1988)	(3, 4)		(2, 4)		Indifferent		Indifferent		Indifferent		Indifferent	



Rapoport and Guyer (1966, p.209) argue that Games 66–69 are preemption games with no natural outcome. Game 66 is the chicken game (Fudenberg and Tirole, 1991), Rasmusen (2001, pp.71, 85). If one preempts (daredevil) and the other does not (chicken), the first earns a high payoff and the second a low payoff. If none preempt, their payoffs are intermediate. If both preempt, the result is disastrous for both. One example is two teenagers driving against each other in stolen cars. Another example [Taylor, (1987), p.37] is two factories discharging polluted material into a lake. If both pollute, an ecological catastrophe follows. No pollution is beneficial, but costly for each. Each prefers to free ride on the other’s pollution restraint.

Game 67 is similar but preemption by Player 2 gives (3, 4) which improves the payoff for both players compared with the non-Pareto optimal payoff (2, 3). Game 68 is the battle of the sexes (Hausken, 1998; Taylor, 1987), which Rapoport and Guyer (1966, p.209) refer to as a ‘benevolent’ chicken game. Each player, by preempting, increases the payoff of both. This makes coordination more prominent making the game relevant for standard setting. For example, consider two dominant companies within an industry with different software preferences, due to sunk costs in training and earlier software investments. Both companies realise that mutual gains can be reaped within the industry by agreeing on one type of software, but each benefits additionally if its own preference gets implemented. A second example is a couple discussing whether to spend the evening at the movies or the theatre. They prefer to be together, but the wife prefers the cinema and the husband prefers the theatre.

In Game 69, each player prefers to preempt, but it prefers even more that the other player does the preemption. Assume that the natural outcome is a quarrel, which both players seek to settle. Although the player choosing a conciliatory apology earns a higher payoff than if the quarrel persists, he earns a lower payoff than if he waits for the other player to choose the conciliatory apology. The conciliatory apology by the preemptor can be interpreted as ‘losing face’. But, the preemptor can also be interpreted as a hero who seeks higher benefits for others than for himself. Rapoport and Guyer (1966, p.210) observe that the hero in this game is different from a martyr who seeks to benefit others and harm himself. (A player switching from the equilibrium in the prisoner’s dilemma is a martyr. His motivation may be to ‘teach by example’, since the Pareto optimal outcome follows if the other player follows suit.)

In Table 2, we illustrate the proposition in Section 4 by characterising the six games 64–69.

The first two rows below the six matrices are the risk limits  $r_1$  and  $r_2$  of Players 1 and 2. The third line is the conflict measure in (4). The proposition and (5) dictate the equilibrium (3, 4) for Games 64 and 67, and the equilibrium (4, 3) for Game 65, shown in italics. No equilibrium selection is made for Games 65, 66, 69 since  $r_1 = r_2$ . These equilibrium selections are shown in the fourth line. (3, 4) is chosen for Game 64 since Player 2 receives 2 in the threat point ( $t_1, t_2$ ), while Player 1 receives only 1. Hence, Player 2 is in a stronger position and is more willing to risk the threat point than Player 1. For Game 67, both players receive only 1 in the threat point, but Player 2 accepts it more willingly since earning only 2 in the non-preferred equilibrium is only marginally better than the threat point. Game 65 selects (4, 3). Player 1 loses 2 through the equilibrium switch, and loses 1 more in the threat point. Player 2 loses only 1 through the equilibrium switch, and loses 1 more in the threat point. Hence, player 1 is willing to risk the most to secure his preferred equilibrium.

The fifth and sixth lines are the players' stubbornness defined in (6) based on Hausken (2007), which gives the equilibrium selection in the seventh line. The equilibrium selection is the same as in this paper for these examples, but since stubbornness is different from risk limit, the equilibrium selection methods are generally different.

The eighth and ninth lines in Table 2 are the two sides of Harsanyi and Selten's (1988, p.90) inequality in (1). Their equilibrium selections are shown in the last line. They select equivalently with this paper for Game 64, do not select for Games 66–69, and select oppositely to this paper for Game 65 because of the low  $d_2 = 1$ , which causes a large Nash product on the RHS in (1). As argued in this article,  $(d_1, d_2)$  is irrelevant for equilibrium selection since it does not enter the players' reasoning processes when choosing among the equilibria. Observe the symmetry, for both the equilibria and the non-equilibria, in the three games, 66, 68, 69 where none of the methods offer an equilibrium selection. In contrast, observe the equilibrium asymmetry combined with the symmetric threat point  $(t_1, t_2) = (1, 1)$  in Game 67 where the risk limit conflict measure selects the equilibrium (3, 4) over equilibrium (4, 2), while Harsanyi and Selten (1988) make no equilibrium selection since the (irrelevant) asymmetric  $(d_1, d_2) = (2, 3)$  compensates for the equilibrium asymmetry generating indifference.

## 8 Conclusions

This article develops a risk limit conflict measure for games with two Pareto-superior Nash equilibria, and presents a method for selecting between the equilibria. The conflict measure is the product of the two players' maximum probabilities of being recalcitrant when pursuing a preferred equilibrium. The measure is equivalent to Axelrod's (1970) measure, which is the ratio of infeasible joint demand and joint demand above the threat point. Axelrod provided no justification beyond informal verbal descriptions for his conflict measure. Two properties are presented which support the measure. These state that a conflict measure should be lowest when the two Nash equilibria are identical or non-discriminating, and highest when the Nash equilibria in question are weakly dominant. The risk limit approach implies a new equilibrium selection method where the player with the highest risk limit gets his preferred equilibrium.

Aside from the two Nash equilibria and the threat point, the  $2 \times 2$  conflict game has a fourth outcome, which plays no role in the players' reasoning, and thus plays no role in the conflict measure and equilibrium selection. This stands in contrast to Harsanyi and Selten's (1988, p.90) equilibrium selection which assigns equal weight to four payoff differences, which means that all the four outcomes in the game are influential. Harsanyi and Selten (1988) present a risk dominance criterion, which is a comparison of Nash products, which determines which equilibrium gets selected. The article further compares with Hausken's (2007) equilibrium selection based on how stubbornly the players' insist on their preferred equilibria. The conflict measure and three equilibrium selection methods are illustrated and discussed for the six games 64–69 in Rapoport and Guyer's (1966, p.213) taxonomy.

## Acknowledgements

The author would like to thank the two anonymous referees of this journal for useful comments, and Dr. Matthias Mohr for useful discussions. The discussion of risk limits and Axelrod's (1970) conflict measure are further developments of Mohr and Hausken (1996).

## References

- Axelrod, R. (1970) *Conflict of Interest*, Markham Publishers, Chicago.
- Ellsberg, D. (1961) 'Risk, ambiguity and the savage axioms', *Quarterly Journal of Economics*, Vol. 75, pp.643–649.
- Février, P. and Linnemer, L. (2006) 'Equilibrium selection: payoff or risk dominance?: The case of the 'weakest link'', *Journal of Economic Behavior & Organization*, Vol. 60, No. 2, pp.164–181.
- Fudenberg, D.M. and Tirole, J. (1991) *Game Theory*, MIT Press, Cambridge.
- Harsanyi, J.C. (1977) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, MIT Press, Cambridge.
- Harsanyi, J.C. and Selten, R. (1988) *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge.
- Hausken, K. (1998) 'Collective rent seeking and division of labor', *European Journal of Political Economy*, Vol. 14, No. 4, pp.739–768.
- Hausken, K. (2007) 'Stubbornness, power, and equilibrium selection in repeated games with multiple equilibria', *Theory and Decision*, Vol. 62, No. 2, pp.135–160.
- Knight, J. (1992) *Institutions and Social Conflict*, Cambridge University Press, Cambridge.
- Mohr, M. and Hausken, K. (1996) 'Conflict, interest and strategy: a risk-limit approach to conflict', Max Planck Institute for the Study of Societies, Discussion Paper 96/7.
- Rapoport, A. and Guyer, M. (1966) 'A taxonomy of  $2 \times 2$  games', *General Systems*, Vol. 11, pp.203–214.
- Rasmusen, E. (2001) *Games and Information*, Basil Blackwell Inc., Cambridge.
- Taylor, M. (1987) *The Possibility of Cooperation*, Cambridge University Press, Cambridge.
- van Damme, E. (2002) 'Strategic equilibrium', *Handbook of Game Theory*, North Holland, Vol. 3, pp.1521–1596.
- van Huyck, J. (1997) 'Emergent conventions in evolutionary games', in Plott, C.R., Smith, V.L. (Eds.): *Prepared for Handbook of Experimental Economics Results*.
- Zeuthen, F. (1930) *Problems of Monopoly and Economics*, Routledge, London.

## Notes

- 1 Table 1 represents Games 64–69 in Rapoport and Guyer's (1966, p.213) ordinal taxonomy. The most well-known of these are the Battle of the Sexes (Game 68), Chicken (Game 66), and Game 69 with several names such as 'Let George do it', 'Apology', 'Hero', 'Sacrificed leader'. The Games 64, 65, 67 are hybrid asymmetric games.
- 2 Zeuthen (1930) originated the principle of risk dominance as a dominance relation based on comparing the various players' risk limits. Subsequently, Ellsberg (1961) discussed the principle related to his paradoxes, and Harsanyi (1977, pp.280–288) analysed it. Harsanyi and Selten (1988, p.86ff) provide three axioms, which uniquely determine the given risk dominance relationship. These are invariance with respect to isomorphisms, best-reply invariance, and payoff monotonicity. There is no axiom of independence of irrelevant alternatives. As Harsanyi and Selten (1988, p.86ff) point out, 'the nature of the problem of equilibrium point selection in non-cooperative games does not seem to permit a satisfactory solution concept that can be characterised by a set of simple axioms'.

- 3 If the inequality is reversed,  $(b_1, b_2)$  risk dominates  $(a_1, a_2)$ , and if the Nash products in (1) are equal,  $(a_1 - d_1)(a_2 - t_2) = (b_1 - t_1)(b_2 - d_2)$ , there is no risk dominance between  $(a_1, a_2)$  and  $(b_1, b_2)$ .
- 4 The risk limit conflict measure does not depend on the equilibrium from which one starts. The areas would have been the same had we started from the equilibrium  $(b_1, b_2)$  and let Player 1 challenge Player 2.
- 5 It is the jointly infeasible expectation of an additional gain in case of a conflict.
- 6 If the inequality is reversed,  $(b_1, b_2)$  risk dominates  $(a_1, a_2)$ , and when the risk limits in (5) are equal,  $r_1 = r_2$ , there is no risk dominance.
- 7 Each player has  $4! = 24$  preference orderings, which gives  $24 \times 24 = 576$  games. Some of these are equivalent by interchanging rows or columns. Each of 66 non-equivalent games generates 8 equivalent games, and each of 12 non-equivalent games generates 4 equivalent games, i.e.,  $66 \times 8 + 12 \times 4 = 576$ .
- 8 Games 58–63 are labelled by Rapoport and Guyer (1966) as no conflict games since one equilibrium Pareto dominates the other equilibrium. These do not satisfy our requirements for Table 1 and are excluded. One example is the stag hunt game 61. Harsanyi and Selten (1988, p.359) correctly argue that the stag-stag equilibrium should be selected since it dominates the hare-hare equilibrium, although the latter risk dominates the former. See Février and Linnemer (2006) for a recent study.