Hausken, K. (2007) Stubbornness, power, and equlibrium selection in repeated games with multiple equilibra. ***Theory and Decision,*** **62(2), pp. 135-160**

Link to official URL: http://link.springer.com/article/10.1007/s11238-006-9020-4# (Access to content may be restricted)

KJELL HAUSKEN

STUBBORNNESS, POWER, AND EQUILIBRIUM SELECTION IN REPEATED GAMES WITH MULTIPLE EQUILIBRIA

ABSTRACT. Axelord's [(1970), *Conflict of Interest*, Markham Publishers, Chicago] index of conflict in 2 × 2 games with two pure strategy equilibria has the property that a reduction in the cost of holding out corresponds to an increase in conflict. This article takes the opposite view, arguing that if losing becomes less costly, a player is less likely to gamble to win, which means that conflict will be less frequent. This approach leads to a new power index and a new measure of stubbornness, both anchored in strategic reasoning. The win probability defined as power constitutes an equilibrium refinement which differs from Harsanyi and Selten's [(1988), *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge] refinement. In contrast, Axelrod's approach focuses on preferences regarding divergences from imaginary outmost rewards that cannot be obtained jointly. The player who is less powerful in an asymmetric one-shot game becomes more powerful in the repeated game, provided he or she values the future sufficiently more than the opponent. This contrasts with the view that repetition induces cooperation, but conforms with the expectation that a more patient player receives a larger share of the pie.

KEY WORDS: conflict, discounting, equilibrium refinement, equilibrium selection, power index, repeated game, stubbornness incentive

1. INTRODUCTION

The nature of conflict is made up of and can be reduced to two complementary dimensions, the interest dimension and the strategic dimension. Each by itself is insufficient, but together they provide an exhaustive account of conflict. The interest dimension has been worked out by Axelrod (1970) in an insightful way. This article takes the view that Axelrod understood only half of the idea, which means that his picture was incomplete. Accordingly, this article develops and extends Axelrod's work. To supplement Axelrod's focus on the interest dimension, this article conceptualizes conflict as a struggle for preferred equilibria in 2 ×2 conflict games, and analyzes the role of the strategic dimension. Although a moderate percentage of games without a mutually best outcome have two Pareto-superior Nash equilibria, the significance of such games is considerable. Struggle for preferred equilibria occurs when several players within an industry attempt to agree on standards, or agree on procedures for interaction. Knight (1992) argues that struggle for preferred equilibria is a crucial characteristic, for example, at the start-up of most social institutions.

The relative weight assigned to the interest dimension and strategic dimension depends on what is to be explained and the nature of social interaction. The difference between the two dimensions is illustrated by two questions. The strength of conflict in terms of the interest dimension is determined by answering the first question, which refers to the divergence of preferences and thus to the conflict of interest in a game: (1) How large is the payoff one player gets if the other player gets her best payoff, how large is the other's payoff if the first player receives his best payoff, and what are the feasible points which may serve as a compromise? The second question proceeds beyond the first question of how preferences diverge, and refers also to strategic considerations regarding how a player can improve her own payoff by exhibiting aggressive, recalcitrant, stubborn, hardheaded, strategically conflictful behavior:(2) What may a player win if she successfully challenges her opponent and what is her chance of winning the challenge? The difference between the two questions is that the first deals with latent and often hidden conflicts of interest, while the second deals with overt conflictful behavior.

One drawback of limiting attention only to the interest dimension of conflict is that the link between conflict and behavior becomes unclear. Consider, for example, Axelrod's (1970, p. 80), statement, which we refer to as his proposition concerning conflictful behavior, that "other things being equal, the more conflict of interest there is, the more probable it is that conflictful behavior will result." Crucial here is how "conflictful behavior" is defined. Axelrod frequently refers to a prisoner's dilemma in which conflictful behavior is well defined. However, he proceeds to argue that his proposition concerning conflictful behavior holds not only in a prisoner's dilemma, but also in all cases. Unfortunately, Axelrod does not explicitly define conflictful behavior in general social situations. Rather, Axelrod (1970, p. 80) states: "The specific meaning of conflictful behavior will differ from one political process to another, of course, but usually there are types of behavior which are clearly conflictful and hence some hypotheses are easy to specify." We illustrate these points over the next sections, questioning the validity of Axelrod's hypotheses for conflictful behavior by showing that they do not necessarily follow from his conflict measure the way he suggests.

Axelrod's (1970) conflict measure satisfies two properties. One of these is plausible and acceptable, but the other is not, as this article demonstrates. The unacceptable property is replaced by an exact opposite acceptable property. The article argues for the appropriateness of these two properties, and shows how these imply two new conflict measures. One is the stubbornness conflict measure. It accounts for both a player's preference for one equilibrium rather than another,[1] and the other player's inclination to go along with the first player's preferred equilibrium. The second is the power index conflict measure. A player's power is interpreted as his probability of "winning" his preferred equilibrium. The power index conflict measure constitutes an equilibrium refinement technique in normal form games possessing more than one equilibrium.

Another problem with Axelrod's conflict measure is its reliance on the so-called outmost point, which is imaginary since it cannot be jointly obtained. It represents both players' aspiration levels. Axelrod's focus on the outmost point, consistent with his focus on the interest dimension of conflict, implies a de-emphasis of the importance of the Nash equilibria, and a de-emphasis of the role of strategic interaction. Higher incompatibility of preferences between the players does not necessarily imply that they are more inclined to behave challengingly.

We also develop a power index conflict measure for repeated games, summing up the power for each player in a finitely or infinitely repeated game. The implications are illustrated when the players emphasize the future differently. Section 2 presents the three measures of conflict, i.e., Axelrod's measure, the stubbornness measure, and the power index measure. Section 3 compares the conflict measures. Section 4 concludes. The Appendix develops a power index conflict measure in repeated games.

2. THREE MEASURES OF CONFLICT

Consider the game in Table I where $a_1 \geqslant b_1 \geqslant t_1, b_2 \geqslant a_2 \geqslant t_2, a_1 \geqslant d_1, b_2 > d_2$ or $a_1 > d_1, b_2 \geqslant d_2$.[2] Table I encompasses 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 "Let George do it" (game 69). The six games are listed in Table III.[3] The two pure strategy equilibria are (a_1, a_2) and (b_1, b_2). Player 1 (row player) prefers (a_1, a_2) and player 2 (column player) prefers (b_1, b_2). Each side's preferred equilibrium involves its own second strategy and the opponent's first.

The simple bargaining problem when there are two equilibria has a long history. Some of this history is related to the principle of risk dominance, which was originated by Zeuthen (1930) as a dominance relation based on comparing

TABLE I

Two-person two-strategy game with two equilibria

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

the various players' risk limits. Subsequently Ellsberg (1961) discussed the principle related to his paradoxes. Thereafter Harsanyi (1977) analyzed Zeuthen's principle, and Harsanyi and Selten (1988, p. 90) applied the notion of risk dominance as a criterion for equilibrium selection. They state that (a_1, a_2) risk dominates (b_1, b_2) if $(a_1 - d_1)(a_2 - t_2) > (b_1 - t_1)(b_2 - d_2)$, which is a comparison of Nash products, where (d_1, d_2) plays a role. (If the inequality is reversed, (b_1, b_2) risk dominates (a_1, a_2), and if the Nash products are equal, there is no risk dominance between (a_1, a_2) and (b_1, b_2).)

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 characterized by a set of simple axioms." The main problem with Harsanyi and Selten's (1988) conceptualization is that it amounts to assign equal weight to four payoff differences. Choosing equal weight seems as arbitrary as any other choice of weights unless backed with a good argument.

Let us illustrate. Consider first equilibrium (a_1, a_2) in Table I. 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 $b_2 - d_2$. Assigning equal weight to these four payoff differences amounts to assign equal weight to the payoff combinations (t_1, t_2) and (d_1, d_2). That is, the two non-equilibrium outcomes in Table I are placed on an equal footing. One main argument in this article is that (t_1, t_2) and (d_1, d_2) are not on an equal footing since they are positioned differently relative to the equilibrium outcomes (a_1, a_2) and (b_1, b_2).

To see this difference, consider equilibrium (a_1, a_2). Player 1 prefers under no circumstances to switch from II to I since

$a_1 \geqslant d_1$. The payoff d_1 plays no role in player 1's reasoning process. In contrast, player 2 prefers the other equilibrium and may prefer to switch from I to II. Although $a_2 \geqslant t_2$, player 2 observes that since $b_2 \geqslant a_2$, the only way to reach (b_1, b_2) is to switch from I to II. This may cause the payoff t_2, but it may also cause the payoff b_2 if player 1 is thereby induced to switch from II to I as a consequence of player 2's challenge. Hence t_2 indeed plays a role in player 2's reasoning process. Analogously for equilibrium (b_1, b_2), the payoff d_2 plays no role in player 2's reasoning process, while t_1 plays a role in player 1's reasoning process if he challenges player 2.

Although outcome (d_1, d_2) plays no role in the players' reasoning processes when they focus their attention on either of the equilibria, an alternative consideration is whether (d_1, d_2) plays or should play a role when the players consider the game as such. The answer is yes. All outcomes play a role and are possible when players play a game, otherwise it is not a game. The problem is that game theory provides no clear-cut recommendation for how to play a game with at least two equilibria. Without pre-play communication, the players can reason multifariously. All the four outcomes in Table I are possible. This does not mean that all the four payoff differences considered by Harsanyi and Selten (1988) should be assigned equal weight when reasoning about risk dominance, e.g., as a criterion for equilibrium selection or conflict measures. Choosing equal weight seems as arbitrary as any other choice of weights unless backed with a good argument. And, indeed, the argument above suggests that as long as d_1, d_2, t_1, t_2 satisfy the inequalities specified above, (t_1, t_2) does play a role in the players' reasoning process, while (d_1, d_2) does not. The outcome (d_1, d_2) is possible, but the players don't reason about it when attempting to determine risk dominance.

Assume that the players engage in pre-play communication prior to playing the static game in Table I, i.e., prior to making independent, effectively simultaneous, choices. Assume that they during their discussion reason such that each can hold out or give in. The four possible outcomes are such that they either both hold out, one holds out and the other gives in, one gives in and the other holds out, or both give in. It is quite dysfunctional for both of them to give in, so assume that the pre-play communication runs over a time interval such that if one gives in, then the other does not.

The argument above implies that given that d_1, d_2, t_1, t_2 satisfy the specified inequalities, the players can be expected not to assign equal weight to the four payoff differences. Instead, they can be expected to let (t_1, t_2) play a role when determining risk dominance, and let (d_1, d_2) play no role.

Axelrod (1970, p. 20) argues that "there is a single predetermined outcome, which I'll call the no agreement point that occurs if no agreement is reached. In other words, either player can veto anything other than the no agreement point." He defines the "no agreement point" as equivalent to the "null point." From a game-theoretic point of view we most plausibly define Axelrod's null point as the minmax point, hereafter referred to as the threat point (t_1, t_2). Rasmusen (2001, p. 114) and Fudenberg and Tirole (1991, p. 150) define a player's minmax value as his "security value" and "reservation utility," respectively. The question is how incompatible the players' demands are beyond or above the threat point (t_1, t_2).[4]

To fully characterize Axelrod's conception we need to expand the two-strategy game in Table I to the three-strategy game in Table II which additionally assumes $A_1 \geqslant a_1$ and $B_2 \geqslant b_2$, and also $t_i \geqslant x_i, t_i \geqslant y_i, t_i \geqslant z_i, i = 1, 2$.

Axelrod's (1970, p. 5) conflict approach focuses on "the state of incompatibility of the goals of two or more actors,"

TABLE II

Two-person three-strategy game with two equilibria

	I	II	III
I	d_1, d_2	b_1, b_2	y_1, y_2
II	a_1, a_2	t_1, t_2	A_1, t_2
III	x_1, x_2	t_1, B_2	z_1, z_2

a state in which one player can get her best payoff only at the expense of the other player. That is, he refers to "the proportion of the joint demand area which is infeasible" (Axelrod, 1970, p. 57). The area of joint demand above the threat point is the rectangle 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 favorable circumstances. (A_1, B_2) can not be jointly obtained. It represent both players' aspiration levels, and is necessary for Axelrod as an imaginary point in order to determine the conflict of interest between the players. The area of infeasible joint demand in pure strategies is defined as the polygon spanned out by the best possible payoff each player can obtain (A_1, B_2), the two points in which one of the players gets his most preferred payoff (A_1, t_2) and (t_1, B_2), and (t_1, b_2), (b_1, b_2), (b_1, a_2), (a_1, a_2), (a_1, t_2) which closes the poly-gon shown as the sum of the black and dark grey areas in Figure 1. For pure strategies Axelrod define the degree c of conflict as the ratio of the two areas, which is analytically expressed as

$$c = \frac{(A_1 - b_1)(B_2 - a_2) + (A_1 - a_1)(a_2 - t_2) + (b_1 - t_1)(B_2 - b_2)}{(A_1 - t_1)(B_2 - t_2)},$$
$$0 \leqslant c \leqslant 1, \qquad (1)$$

which equals $c = 2/3$ when $(a_1, a_2) = (4, 3)$, $(b_1, b_2) = (3, 4)$, $(t_1, t_2) = (1, 2)$, $(A_1, B_2) = (6, 5)$. Conflict c increases when the area of infeasible joint demand increases and the whole area of joint demand decreases. The first problem with (1) is that the outmost point (A_1, B_2) can be rendered irrelevant by strategic reasoning. (A_1, t_2) and (t_1, B_2), neither of which are equilibria, are irrelevant as points from which either player may challenge the other in order to enforce her most preferred solution. No player may reasonably hope for, nor has the power to induce the other to go along with, his highest possible payoff. Hence the dark grey area in Figure 1 is irrelevant.

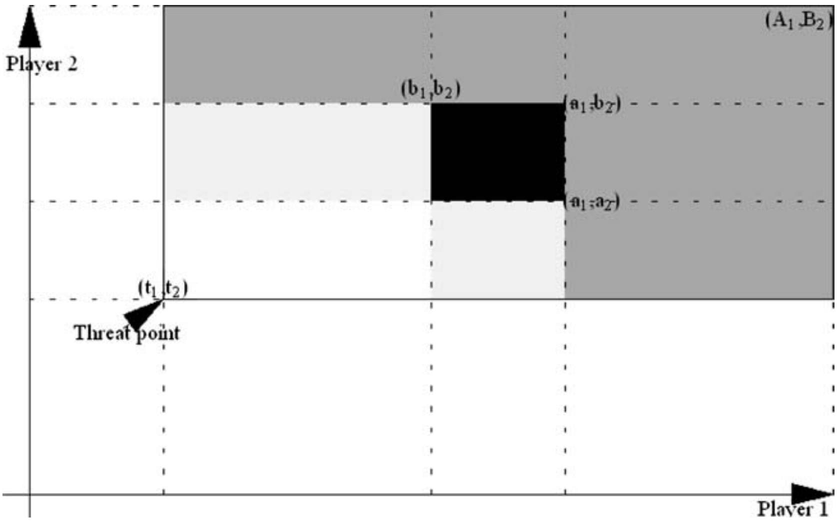

Figure 1. Diagram of the game in Table II, assuming pure strategies.

Inserting $A_1 = a_1$ and $B_2 = b_2$ into (1) gives

$$c = \frac{(a_1 - b_1)(b_2 - a_2)}{(a_1 - t_1)(b_2 - t_2)} \quad \text{when } A_1 = a_1 \text{ and } \quad B_2 = b_2, \tag{2}$$

where the numerator is the black square of infeasible joint demand in Figure 1, and the denominator is the joint demand above the threat point (t_1, t_2) and below the new outmost point (a_1, b_2). Inserting the given values gives conflict $c = 1/6$, which is lower than $c = 2/3$ since the additional strategies (III,III) in Table II sharpens the incompatibility of interest.[5]

The second problem which applies for both (1) and (2) is that the outmost points (A_1, B_2) and (a_1, b_2) are irrelevant, according to the argument in this article, since these cannot be jointly obtained. Furthermore, for the case of pure strategies, the entire black square is irrelevant since the players have no possibility to jointly obtain payoffs within this square.[6] If the players accept this, there is no reason to let the black square play a role in the conflict measure. An analogy may here be drawn to Nash's (1953, p. 137) "independence of irrelevant alternatives" axiom. Whereas Nash rules out alternatives irrelevant for a unique bargaining equilibrium, we remove the black square including the outmost point since strategic reasoning renders these irrelevant. Axelrod presents a procedure for his conflict measure which satisfies the following two properties:

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

The justification is that there is no conflict between the players, which in Table I means $a_1 = b_1$ and $b_2 = a_2$. This article accepts Property 1, and a conflict measure should contain the terms $a_1 - b_1$ and $b_2 - a_2$.

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.*

Property 2 illustrates the fundamental disagreement between the approach in this article and Axelrod's approach. It concerns whether there is more or less "conflict" as the equilibria get closer and closer to the degenerate case of weak dominance, which in Table I means $b_1 = t_1$ for player 1 and $a_2 = t_2$ for player 2. Axelrod's idea is that increases in the level of conflict correspond to smaller costs for holding out. The cost of holding out in (t_1, t_2) is zero for player 1 when $b_1 = t_1$. Analogously, the cost of holding out in (t_1, t_2) is zero for player 2 when $a_2 = t_2$. This means that deviation from the non-preferred equilibrium is costless. That this causes an increase in conflict, as Axelrod suggests, may in one sense seem reasonable if each player is thought of as balancing a motivation to hold out against a motivation to give in. That is, the motivation to give in rises as the potential cost of holding out when the opponent insists falls. Only if the level of "conflict" is very high will the motivation to hold out overcome the motivation to give in. Thus in Axelrod's approach "conflict" refers to a condition under which the bargaining takes place that reflects how much the participants want to win. It some sense conflict then becomes an environmental variable.

This article does not accept Property 2. The contrasting view taken in this article is that as the cost of holding out falls, a player is less likely to do so. Given that the gain from winning is fixed, if losing becomes less costly, a player is less likely to gamble to win. So we observe less "conflict," in the sense of observations of fixed decision makers implementing their threats. Instead of a condition that enhances some motivations relative to others, "conflict" is an observed frequency of disagreement.[7]

To illustrate this further, consider Table I with the change that t_2 is increased to $t_2 = a_2$. It is costless for player 2 to switch from I to II since payoff a_2 is obtained in both cases. Player 1 should accept this fact and realize that player 2 inevitably will switch to II since then a payoff of at least a_2 is guaranteed. This should give no rise to conflict. Why should player 2 be criticized for following her self interest and choose a strategy that guarantees payoff $t_2 = a_2$? Accepting that player 2 inevitably will switch to II, player 1's best response is I which gives (b_1, b_2) and no conflict. Hence we propose the exact opposite of Property 2.

Property 3. *A conflict measure should be lowest when the Nash equilibria in question is weak (weakly dominant), and should increase as the product of the payoff difference each player experiences between his non-preferred equilibrium and the threat point increases.*

An implication of Properties 1 and 3 is that the conflict measure should be highest when the product of four factors is highest. The first factor is player 1's payoff difference $a_1 - b_1$ between his preferred and non-preferred equilibrium. The second factor is player 2's payoff difference $a_2 - t_2$ between his non-preferred equilibrium and the threat point. The third factor is player 2's payoff difference $b_2 - a_2$ between his preferred and non-preferred equilibrium. The fourth factor is player 1's payoff difference $b_1 - t_1$ between his non-preferred equilibrium and the threat point.

The first and third factors are consistent with Property 1. The second and fourth factors are consistent with Property 3, and inconsistent with Property 2. Consider again Table I with the starting point that $t_2 = a_2$. Assume that t_2 decreases which increases $a_2 - t_2$. This is conflict inducing since player 2 becomes less inclined to switch from I to II. Player 1 realizes that player 2 is more inclined to go along with his preferred (a_1, a_2), and can be expected to take advantage of this by insisting more fiercely on his preferred strategy II, which raises the conflict proportional to $a_2 - t_2$. The payoff difference $a_2 - t_2$ is not present in Axelrod's conflict measure, but $b_2 - t_2$ is present in the denominator in (2) as conflict reducing. That is, reducing the threat point payoffs by moving (t_1, t_2) downwards and leftwards in Figure 1 reduces conflict according to Axelrod. In contrast, this article argues that reducing (t_1, t_2) increases the abyss and thus conflict between the two players. A threat point with low payoffs is detrimental for both players and should thus increase conflict, not reduce it. Properties 1 and 2 imply that Axelrod's conflict measure does not depend on the Nash equilibria as equilibria, but on the black square containing the outmost point and the degree to which the payoffs in the preferred equilibria differ for the players. In contrast, Properties 1 and 3 applied in this article imply that only the two equilibria and the threat point are relevant for strategic reasoning.

Player 1's incentive to be stubborn (recalcitrant, hardheaded, tough, aggressive) and induce conflict 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).[8] 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. This expression is shown as a light grey area in Figure 1. The analogous expression $(b_2 - a_2)(b_1 - t_1)$ expresses player 2's incentive to be stubborn, also shown as a light grey area in Figure 1. We divide with the sum of the two expressions to scale the sum of the two players' stubbornness incentives to be equal to one, that is,

$$s_1 = \frac{(a_1 - b_1)(a_2 - t_2)}{(a_1 - b_1)(a_2 - t_2) + (b_2 - a_2)(b_1 - t_1)},$$
$$s_2 = \frac{(b_2 - a_2)(b_1 - t_1)}{(a_1 - b_1)(a_2 - t_2) + (b_2 - a_2)(b_1 - t_1)}, \tag{3}$$

where $s_1 + s_2 = 1$. Graphically in Figure 1, s_1 is expressed as the lower right light grey area divided by the sum of the two light grey areas, and analogously for s_2. The black square plays no role in (3), consistent with our discussion above. Multiplying (3) with 4 since $s_1(1 - s_1) \leqslant 1/4$ when $0 \leqslant s_1 \leqslant 1$, the product c_s of the two stubbornness incentives gives the conflict measure

$$c_s = 4s_1s_2 = \frac{4(a_1 - b_1)(a_2 - t_2)(b_2 - a_2)(b_1 - t_1)}{[(a_1 - b_1)(a_2 - t_2) + (b_2 - a_2)(b_1 - t_1)]^2}, \quad 0 \leqslant c_s \leqslant 1. \tag{4}$$

This conflict measure gives zero when any of the four payoff differences in the numerator equals zero, in contrast to Axelrod's conflict measure which equals zero when any of the two payoff differences $(a_1 - b_1)$ or $(b_2 - a_2)$ in (2) equals zero. More specifically, both c_s in (4) and c in (2) equal zero when $(a_1 - b_1)(b_2 - a_2) = 0$, in accordance with Property 1. However, c_s in (4) also equals zero when $(a_2 - t_2)(b_1 - t_1) = 0$, in accordance with Property 3, and thus not in accordance with Property 2. Axelrod thus predicts conflict in cases when there is no conflict.

Within political economy the concept of power has received the definition of win probability dependent on the players' efforts exerted, resource commitments of players, and conflict technology (Grossman, 1991; Hirshleifer, 1988, 1991, 1995; Skaperdas, 1992, p. 724). I.e., in competition for an object the player with the largest probability of winning has highest power. For non-cooperative games with two equilibria the literature offers to the author's knowledge no power index.[9] Define player 1's stubbornness incentive s_1 as the probability that he is stubborn and plays II. Analogously s_2 is the probability that player 2 plays II. The probability p_1 that player 1 wins his preferred (a_1, a_2) is thus $s_1(1 - s_2)$, and the probability p_2 that player 2 wins her preferred (b_1, b_2) is $(1 - s_1)s_2$.

Defining the two win probabilities as power gives

$$p_1 = s_1(1 - s_2) = s_1^2, \qquad p_2 = (1 - s_1)s_2 = s_2^2, \tag{5}$$

which constitutes an equilibrium refinement. The term equilibrium refinement refers to a definition that picks out a subset of the equilibria by some special property. The subset here is one equilibrium chosen among two pure strategy equilibria. The special property is the stubbornness probability, and thus the stubbornness and power index conflict measures. The equilibrium refinement applies for six especially prominent games listed as games 64–69 in Rapoport and Guyer's (1966, p. 213) taxonomy.

Let us formulate the following equilibrium refinement property.

Property 4. *Equilibrium* (a_1, a_2) *is selected if* $p_1 > p_2$, *equilibrium* (b_1, b_2) *is selected if* $p_1 < p_2$, *and there is no equilibrium selection if* $p_1 = p_2$, *where* p_1 *and* p_2 *are defined in* (5) *and* (3) *based on the game in Table I.*

The product c_p of the two power indices,

$$c_p = 16p_1 p_2 = 16s_1^2 s_2^2 = c_s^2, \quad 0 \leqslant c_p \leqslant 1, \tag{6}$$

is our third measure of conflict. $0 \leqslant 1 - p_1 - p_2 = 2s_1(1 - s_1) \leqslant 1$ is the probability that no player wins a preferred equilibrium, interpreted as general powerlessness among the players. Consistently with Property 1, $\lim_{a_1 \to \infty} s_1 = 1$ and $\lim_{a_1 \to b_1} s_1 = 0$, which means that both the stubbornness incentive and power of player 1 decrease from 1 to 0 as the value of his preferred relative to his non-preferred equilibrium decreases from ∞ to 1. Consistently with Properties 1 and 3, and in contrast to Property 2, $\lim_{t_2 \to -\infty} s_1 = 1$ and $\lim_{t_2 \to a_2} s_1 = 0$. Clearly, $t_2 = -\infty$ means that player 2 is infinitely bad off in her threat point, which makes her maximally vulnerable when threatened, which gives player 1 maximum power over player 2. Conversely, $t_2 = a_2$ means that player 2 is equally well off in her threat point and her non-preferred equilibrium, which gives player 1 no power over player 2.

Rapoport and Guyer (1966) have identified 78 2×2 games with different ordinal ranking of the payoffs. A total of 6 of 12 two-equilibria games numbered 58–63 do not satisfy $a_1 \geqslant b_1 \geqslant t_1, b_2 \geqslant a_2 \geqslant t_2$. These are no-conflict games where both players prefer one equilibrium to the other, which renders a conflict measure inappropriate. The six remaining two-equilibria games satisfying $a_1 \geqslant b_1 \geqslant t_1, b_2 \geqslant a_2 \geqslant t_2$ are conflict games shown in Table III together with Axelrod's measures c and c_m in pure and mixed strategies, and the stubbornness and power index conflict measures c_s and c_p.

Comparing game 68 with a detrimental threat point (1,1) with game 64 with a less detrimental threat point (1,2), the stubbornness and power index conflict measures correctly give higher conflict for the former than the latter, while the reverse is the case for Axelrod's measures. Games 64 and 65 give the same conflict measures $c_s = 8/9$ and $c_p = 64/81$ for the stubbornness and power index conflict measures since $(a_1 - b_1)(b_1 - t_1) = (4 - 3)(3 - 1) = (4 - 2)(2 - 1) = 2$ is equal for the two games (Properties 1 and 3). In contrast, Axelrod's measures give twice as much conflict for game 65 where the area of infeasible joint demand is twice that of game 64, while the area of joint demand above the threat point is equal for games 64 and 65 (Properties 1 and 2). The stubbornness conflict measure $c_s = 16/25$ is relatively low for game 67 due to player 1's low incentive $s_1 = 1/5$ to be stubborn, rendering player 2 likely to get her preferred equilibrium.

For the symmetric games 66, 68, 69, $c_s = c_p = 1$ since $s_1 = s_2 = 1/2$ and $p_1 = p_2 = 1/4$.[10] The philosophical interpretation is that two players with equal incentives to be stubborn, and equal power, disagreeing over one penny or disagreeing over one trillion dollars are in a full-scale conflict in both situations. Assuming players with pure self-interest, it is as difficult to settle on one equilibrium in one symmetric case as another. What they would like to risk, and whether they would like to act challengingly or conciliatory, are irrelevant for stubbornness and power determination.

TABLE III

Six two-equilibria conflict games

Game	64		65		66		67		68		69	
	2,1	**3,4**	3,1	2,4	3,3	2,4	2,3	**3,4**	2,2	3,4	1,1	3,4
	4,3	1,2	**4,3**	1,2	4,2	1,1	4,2	1,1	4,3	1,1	4,3	2,2
c	1/6		1/3		4/9		2/9		1/9		1/4	
c_{m}	1/12		1/6		2/9		1/9		1/18		1/8	
c_{s}	8/9		8/9		1		16/25		1		1	
c_{p}	64/81		64/81		1		256/625		1		1	
p_1	1/9		4/9		1/4		1/25		1/4		1/4	
p_2	4/9		1/9		1/4		16/25		1/4		1/4	
$(a_1-d_1)(a_2-t_2)$	2		1		1		2		4		3	
$(b_1-t_1)(b_2-d_2)$	6		3		1		2		4		3	

The fourth and third last rows in Table III are the win probabilities p_1 and p_2 of players 1 and 2, defined as power. Property 4 dictates the equilibrium (3,4) for games 64 and 67, and the equilibrium (4,3) for game 65, shown in bold. No equilibrium selection is made for games 65,66,69 since $p_1 = p_2$. The reason (3,4) is chosen for game 64 is that 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 willing to insist more fiercely on his preferred equilibrium since the threat point is more acceptable to him than player 1. For game 67 both players receive only 1 in the threat point, but player 2 is more willing to accept it 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 insist most fiercely on his preferred equilibrium.

The last two lines in Table II are the two sides of Harsanyi and Selten's (1988, p. 90) inequality $(a_1 - d_1)(a_2 - t_2) > (b_1 - t_1)(b_2 - d_2)$ for (a_1, a_2) to risk dominate (b_1, b_2). They provide the same equilibrium selection for game 64, the opposite selection for game 65, and no selection for the games 66–69. The reason for the opposite selection for game 65 is the low $d_2 = 1$, which causes a large Nash product $(b_1 - t_1)(b_2 - d_2)$ on the RHS in the inequality.

Table IV shows 12 cases allowing various cardinal rankings of the payoffs, applying L'Hopital's rule to settle indeterminate "0/0" and "∞/∞." The properties of the power index conflict measure are not included in Table IV since these follow straightforwardly from the stubbornness conflict measure, i.e., $p_i = s_i^2, c_\mathrm{p} = c_\mathrm{s}^2$.

(1) Can be interpreted as a conflict game (as Axelrod suggests) and a no-conflict game (as the power index measure does) dependent on the eyes of the beholder.[11] (2) A conflict game where the threat point (t_1, t_2) determines power differentials. (3) Player 1 perceives no equilibrium payoff difference,

TABLE IV

Properties of Axelrod's and the stubbornness conflict measures

		Axelrod's measure	Stubbornness conflict measure
1	$\infty = a_1 > b_1 > t_1, b_2 > a_2 > t_2$	$c > 0$	$s_1 = 1,\ s_2 = 0 = c_s$
2	$\infty = a_1 > b_1 > t_1,\ \infty = b_2 > a_2 > t_2$	$c = 1$	$s_1 = (a_2 - t_2)/((a_2 - t_2) + (b_1 - t_1)),\ s_2 = 1 - s_1,\ c_s > 0$
3	$a_1 = b_1 > t_1,\ b_2 > a_2 > t_2$	$c = 0$	$s_1 = 0 = c_s,\ s_2 = 1$
4	$a_1 = b_1 > t_1, b_2 = a_2 > t_2$	$c = 0$	$s_1 = (a_2 - t_2)/((a_2 - t_2) + (b_1 - t_1)), s_2 = 1 - s_1, c_s > 0$
5	$a_1 > b_1 = t_1, b_2 > a_2 > t_2$	$c > 0$	$s_1 = 1, s_2 = 0 = c_s$
6	$a_1 > b_1 = t_1, b_2 > a_2 = t_2$	$c = 1$	$s_1 = (a_1 - b_1)/((a_1 - b_1) + (b_2 - a_2)), s_2 = 1 - s_1,\ c_s > 0$
7	$a_1 > b_1 > t_1 = -\infty, b_2 > a_2 > t_2$	$c = 0$	$s_1 = 0 = c_s, s_2 = 1$
8	$a_1 > b_1 > t_1 = -\infty, b_2 > a_2 > t_2 = -\infty$	$c = 0$	$s_1 = (a_1 - b_1)/((a_1 - b_1) + (b_2 - a_2)), s_2 = 1 - s_1, c_s > 0$
9	$a_1 = b_1 = t_1, b_2 > a_2 > t_2$	$c > 0$	$s_1 = (a_2 - t_2)/((a_2 - t_2) + (b_2 - a_2)), s_2 = 1 - s_1, c_s > 0$
10	$a_1 = b_1 = t_1, b_2 = a_2 > t_2$	$c = 0$	$s_1 = 1, s_2 = 0 = c_s$
11	$a_1 = b_1 = t_1, b_2 > a_2 = t_2$	$c = 1$	$s_1 = 0 = c_s, s_2 = 1$
12	$a_1 = b_1 = t_1, b_2 = a_2 = t_2$	$c = 1$	$s_1 = s_2 = 1/2, c_s = 1$

and has no power, implying no conflict. (4) No players perceive equilibrium payoff difference, but (t_1, t_2) causes power difference between two equivalent equilibria. (5) The same implication as case 1. (6) (t_1, t_2) is a third equilibrium causing maximum conflict for Axelrod (Property 2), but is irrelevant for power determination in a conflict game. (7) The same implication as case 3 (8) The same implication as case 4 except different power formula. (9) (t_1, t_2) is an equilibrium. Although both measures predict conflict, the players can be expected to settle on (b_1, b_2) assuming player 1 does not prefer to harm player 2. (10) A no-conflict game with two equivalent equilibria where player 1 is omnipotent. (11) The two equivalent equilibria $(a_1, a_2) = (t_1, t_2)$ cause maximum conflict for Axelrod (Property 2), The omnipotent player 2 perceives no power conflict, she chooses strategy II, and the players can be expected to settle on (b_1, b_2). (12) A full-scale conflict over three equivalent equilibria.

4. CONCLUSION

Conflict is conceptualized as a struggle for preferred equilibria. We first illustrate Axelrod's (1970) conflict measure. He assumes that reduced cost of holding out in the no agreement point causes increased conflict. This article takes the opposite view. If losing becomes less costly, a player is less likely to gamble to win, which causes less conflict. Axelrod predicts conflict in cases when there is no conflict. The new approach implies two new conflict measures. The stubbornness conflict measure accounts for a player's incentive to play his preferred equilibrium strategy. The power index conflict measure interprets a player's power as his probability of "winning" his preferred equilibrium. This provides an equilibrium refinement technique in normal form games with more than one equilibrium, where the equilibrium with the largest win probability is selected. The reasons why the equilibrium refinement differs from Harsanyi and Selten's (1988, p. 90) refinement are discussed. The two new conflict measures are anchored in strategic reasoning and the two equilibria. In contrast, Axelrod's measure does not depend on the Nash equilibria as equilibria, but on the divergence of preferences in the outmost point which are the players' aspiration levels which cannot be jointly obtained. Strategic reasoning renders the outmost point irrelevant. The conflict measures are compared for the six possible two-equilibria conflict games assuming ordinal payoff rankings, and subsequently for cardinal rankings of the payoffs for all possible parameter combinations. In a repeated game, one player's power increases if his emphasis on the future increases. For a symmetric game the players are equally powerful when their discount factors are equal. For an asymmetric game, one player, that is, least powerful in a one-shot game, becomes most powerful in the repeated game if he values the future sufficiently more than the other player. This is contrary to the commonly accepted notion where high valuations of the future are conducive to cooperation in long-term relationships, but is consistent with the notion that the more patient player receives a larger share of the pie, and is thus more powerful.

APPENDIX A: POWER INDEX CONFLICT MEASURE IN REPEATED GAMES

Let us consider the following repeated game.[12] At period 0 player 1 wins with probability $p_1 = s_1^2$, and player 2 wins with probability $p_2 = s_2^2$. The remaining probability is $1 - p_1 - p_2 = 2s_1s_2$. Assume that the game ends in period 0 with probability s_1s_2, and that it proceeds to period 1 with the remaining probability s_1s_2 where the same game is played again. We count 1 for a win and 0 for any other outcome. Player 1 discounts at δ_1 and player 2 at δ_2. Player is accumulated absolute power over N periods is

$$\begin{aligned} p_{iNA} &= p_i + \delta_i s_1 s_2 p_i + (\delta_i s_1 s_2)^2 p_i + \cdots + (\delta_i s_1 s_2)^{N-1} p_i \\ &= \frac{1-(\delta_i s_1 s_2)^N}{1-\delta_i s_1 s_2} p_i, \end{aligned} \tag{A1}$$

which is a geometric series. As δ_i approaches infinity, player is absolute power also approaches infinity. We are interested in the relative power balance between players 1 and 2, determined by

$$p_{iN} = \frac{p_{iNA}}{p_{iNA} + p_{jNA}} = \frac{p_i(1-(\delta_i s_1 s_2)^N)(1-\delta_j s_1 s_2)}{p_i(1-(\delta_i s_1 s_2)^N)(1-\delta_j s_1 s_2) + p_j(1-(\delta_j s_1 s_2)^N)(1-\delta_i s_1 s_2)},$$
$$i, j = 1, 2, \quad i \neq j, \quad 0 \leqslant c_{pN} = 4 p_{1N} p_{2N} \leqslant 1. \quad \text{(A2)}$$

Differentiating c_{pN} with respect to δ_i when $N = \infty$ implies that maximum conflict $c_{pN} = 1$ arises when

$$\delta_i = \frac{p_j - p_i(1-\delta_j s_1 s_2)}{p_j s_1 s_2} \quad \Leftrightarrow \quad \frac{p_i}{p_j} = \frac{1-\delta_i s_1 s_2}{1-\delta_j s_1 s_2},$$
$$i, j = 1, 2, \quad i \neq j, \quad \text{(A3)}$$

(A2) is shown in Figures 2 and 3 when $(t_1, t_2) = (2, 2)$ and $(t_1, t_2) = (1.99, 2)$, respectively, as functions of δ_1 when $\delta_2 = 0$ and $\delta_2 = 0.9$, $N = \infty$, $(a_1, a_2) = (4, 3)$, $(b_1, b_2) = (3, 4)$.

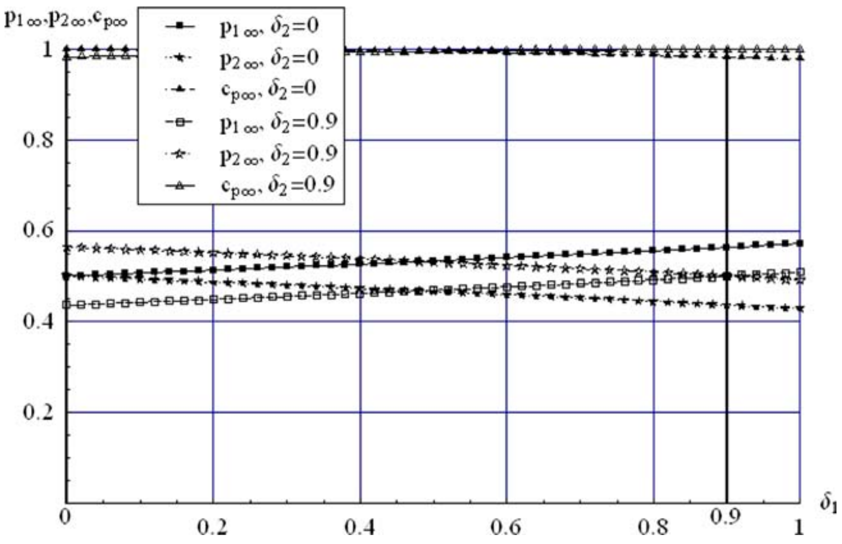

Figure 2. $p_{1\infty}, p_{2\infty}, c_{p\infty}$ as functions of $\delta_1, \delta_2 = 0$ and $\delta_2 = 0.9, N = \infty, (t_1, t_2) = (2, 2)$.

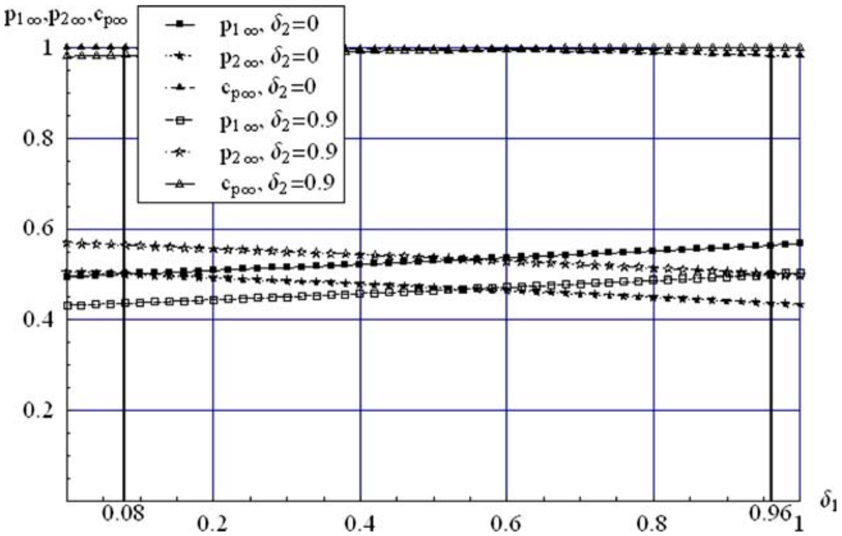

Figure 3. $p_{1\infty}, p_{2\infty}, c_{p\infty}$ as functions of $\delta_1, \delta_2 = 0$ and $\delta_2 = 0.9, N = \infty, (t_1, t_2) = (1.99, 2)$.

Player 1's (2's) power increases (decreases) in δ_1 and maximum conflict $c_{p\infty} = 1$ arises when the players are equally powerful $p_{1\infty} = p_{2\infty} = 0.5$ illustrated with the two thick vertical lines in each figure, for $\delta_2 = 0$ and $\delta_2 = 0.9$, respectively. For the symmetric game the players are equally powerful when their discount factors $0 \leqslant \delta_1 = \delta_2 \leqslant 1$ are equal (Figure 2). For the asymmetric game the least powerful player 1 in the 1-period game ($t_1 = 1.99 < t_2 = 2$) needs a sufficiently larger δ_1 ($\delta_1 = 0.08$ when $\delta_2 = 0, \delta_1 = 0.96$ when $\delta_2 = 0.9$) to be equally powerful $p_{1\infty} = p_{2\infty} = 0.5$ in the infinitely repeated game (Figure 3). Thus note the rightward shift of the two vertical lines in Figure 3. This result contrasts with the common view, often supported by Folk Theorem arguments (Fudenberg and Maskin, 1986), that high valuations of the future are conducive to cooperation in long-term relationships. However, the result bears resemblance to Rubinstein's (1982) result in a complete information alternating offers bargaining model where the more patient player receives a larger share of the pie, and is thus more powerful.[13]

ACKNOWLEDGEMENTS

I thank the editor and two anonymous referees of this journal for useful comments. I thank Dr. Matthias Mohr for useful discussions. The discussion of the "outmost point" in Section 2 and the Axelrod (1970) conflict measure are further developments of Mohr and Hausken (1996).

NOTES

1. See Brams' (1994) account of a player's "threat power."
2. The four diagonal payoffs are sometimes considered to be equivalent, i.e., $d_1 = d_2 = t_1 = t_2$. We distinguish between these because our objective is to model situations where the players have two kinds of preferences: (1) Each has a strong preference for a certain state, independent of the choice of the other player. (2) Each prefers a coordinated to an uncoordinated outcome. If the diagonal payoffs were equivalent, the importance of the second preference would be emphasized rather than the first. The conflict measures developed in this article of course also holds when $d_1 = d_2 = t_1 = t_2$.
3. Game 69 is sometimes called "Apology," "Hero," "Sacrificed leader." The games 64, 65, and 67 are hybrids and mixtures which likely don't have names because of the asymmetries.
4. (t_1, t_2) constitutes a benchmark relevant for the players' anchoring and adjustment (Kahneman and Tversky, 1979) of perceptions.
5. Axelrod considers his calculation of conflict c to be valid for pure and mixed strategies. For mixed strategies all points on the line connecting the equilibrium points (a_1, a_2) and (b_1, b_2) are feasible. The area of infeasible joint demand shrinks to the triangle spanned by $(a_1, a_2), (a_1, b_2), (b_1, b_2)$ which when substituted instead of the black square in Figure 1 gives conflict $c_m = 1/12$. It is unclear how Axelrod would define conflict in mixed strategies for the case in (1) where (A_1, t_2) and (t_1, B_2) are not equilibria.
6. For mixed strategies, the triangle spanned by $(b_1, b_2), (a_1, b_2), (a_1, a_2)$ is irrelevant.
7. I thank an anonymous referee of this journal for the formulations in the previous two paragraphs.
8. As an alternative formulation, player 1's incentive to induce conflict increases if $a_1 - t_1$ increases and he is then more likely to hold on in the threat point. Conversely, player 1's incentive to induce conflict decreases if $b_1 - t_1$ increases, and he is then more likely to acquiesce to player 2's preference. Adding these two effects gives $a_1 - b_1$.

9. Within non-cooperative game theory assuming one unique equilibrium Steunenberg et al. (1999) have suggested a power index based on the distance between a player's ideal point and the equilibrium outcome of a game. Within cooperative game theory power indices are more common (Banzhaf, 1965; Johnston, 1978; Holler, 1984; Shapley, 1953; Shapley and Shubik, 1954). The concept of power has been analyzed from a resource-dependence perspective (Pfeffer, 1981), as power bases (Machiavelli, 1532; Hobbes, 1651; French and Raven, 1968; Hersey and Blanchard, 1982), in exchange networks (Bonacich, 1987; Skvoretz and Willer, 1993; Bienenstock and Bonacich, 1993), and within mathematical sociology Emerson (1962, p. 32) defines "the power of actor A over actor B as the amount of resistance on the part of B which can be potentially overcome by A."
10. Game 64 is not symmetric in our account since (d_1, d_2) does not enter the players' reasoning processes.
11. In the former sense players compete for the two equilibria which cause different payoffs. In the latter sense player 1 is omnipotent and plays his dominant strategy II regardless of player 2's strategy. Maximizing von-Neumann and Morgenstern utility, player 1 perceives strategy I as irrelevant, observes no dilemma, and experiences no conflict. Player 2 may perceive a conflict, but it takes two to cause conflict.
12. I thank the editor for suggesting this game which is an improvement of the repeated game first analyzed.
13. Skaperdas and Syropoulos (1996) and Hausken (2005) also show that increased importance of the future may harm cooperation.

REFERENCES

Axelrod, R. (1970), *Conflict of Interest*, Chicago: Markham Publishers.

Banzhaf, J.F. (1965), Weighted voting doesn't work: a mathematical analysis, *Rutgers Law Review* 19, 317–343.

Bonacich, P. (1987), Power and centrality: a family of measures, *American Journal of Sociology* 92, 1170–1182.

Bienenstock, E.J. and Bonacich, P. (1993), Game-theory models for exchange networks: experimental results, *Sociological Perspectives* 36(2), 117–135.

Brams, S.J. (1994), *Theory of Moves*, Cambridge: Cambridge University Press.

Ellsberg, D. (1961), Risk, ambiguity and the savage axioms, *Quarterly Journal of Economics* 75, 643–649.

Emerson, R.M. (1962), Power-dependence relations, *American Sociological Review* 27, 31–41.

French, J.R.P. and Raven, B. (1968), The bases of social power, in Cartwright, D. and Zander, A. (eds.), *Group Dynamics* (3. ed.), New York: Harper and Row.
Fudenberg, D.M. and Maskin, E. (1986), The folk theorem in repeated games with discounting or with incomplete information, *Econometrica* 54, 533–554.
Fudenberg, D.M. and Tirole, J. (1991), *Game Theory*, Cambridge: MIT Press.
Grossman, H.I. (1991), A general equilibrium model of insurrections, *American Economic Review* 81, 912–921.
Harsanyi, J.C. (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.
Harsanyi, J.C. and Selten, R. (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge: MIT Press.
Hausken, K. (2005), The battle of the sexes when the future is important, *Economics Letters* 87(1), 89–93.
Hersey, P. and Blanchard, K.H. (1982), *Management of Organizational Behavior*, New Jersey: Prentice-Hall Inc.
Hirshleifer, J. (1988), The analytics of continuing conflict, *Synthese* 76, 201–233.
Hirshleifer, J. (1991), The paradox of power, *Economics and Politics* 3, 177–200.
Hirshleifer, J. (1995), Anarchy and its breakdown, *Journal of Political Economy* 103(1), 26–52.
Hobbes, T. (1651), *Leviathan*, London: J.M. Dent and Sons Ltd. (1973).
Holler, M.J. (1984), A public good power index, in Holler, M.J. (ed.), *Coalitions and Collective Action*, Wuerzburg: Physica Verlag, pp. 51–59.
Johnston, R.J. (1978), On the measurement of power: some reactions to laver, *Environment and Planning* 10, 907–914.
Kahneman, D. and Tversky, A. (1979), Prospect theory: an analysis of decision under risk, *Econometrica* 47, 263–291.
Knight, J. (1992), *Institutions and Social Conflict*, Cambridge: Cambridge University Press.
Machiavelli, N. (1532), *The Prince*, New York: Penguin Books (1981).
Mohr, M. and Hausken, K. (1996), Conflict, Interest and Strategy: A Risk-Limit Approach to Conflict, Discussion Paper 96/7, Max Planck Institute for the Study of Societies.
Nash, J. (1953), Two-person cooperative games, *Econometrica* 21, 128–140.
Pfeffer, J. (1981), *Power in Organizations*, Boston: Pitman.
Rapoport, A. and Guyer, M. (1966), A taxonomy of 2×2 games, *General Systems* 11, 203–214.
Rasmusen, E. (2001), *Games and Information*, Cambridge: Basil Blackwell Inc.

Rubinstein, A. (1982), Perfect equilibrium in a bargaining model, *Econometrica* 50(1), 97–109.

Shapley, L. (1953), A value for n-person games, in Kuhn, H.W. and Tucker, A.W. (eds.), *Contributions to the Theory of Games II*, Princeton: Princeton University Press.

Shapley, L. and Shubik, M. (1954), A method for evaluating the distribution of power in a committee system, *American Political Science Review* 48, 787–792.

Skaperdas, S. (1992), Cooperation, conflict, and power in the absence of property rights, *American Economic Review* 82, 720–739.

Skaperdas, S. and Syropoulos, C. (1996), Can the shadow of the future harm cooperation?, *Journal of Economic Behavior and Organization* 29, 355–372.

Skvoretz, J. and Willer, D. (1993), Exclusion and power: a test of four theories of power in exchange networks, *American Sociological Review* 58, 801–818.

Steunenberg, B., Schmidtchen, D., and Koboldt, C. (1999), Strategic power in the european union, *Journal of Theoretical Politics* 11 (3), 339–366.

Zeuthen, F. (1930), *Problems of Monopoly and Economics*, London: Routledge.