A Risk-Dominant Allocation: Maximizing Coalition Stability

MATTHEW McGINTY
University of Wisconsin-Milwaukee

Abstract

This paper presents a rule to allocate a coalition's worth for superadditive games with positive externalities. The allocation rule awards each member their outside payoff, plus an equal share of the surplus. The resulting allocation maximizes coalition stability. Stable coalitions are Strong Nash equilibria since no subset of members has an incentive to leave. Similarly, no subset of non-members has an incentive to join a stable coalition if the game is concave in this region. The allocation is risk-dominant. All stable coalitions are robust to the maximum probability of 50% that players' deviate from their individual best-responses. The paper compares the allocation to the Shapley value and the Nash bargaining solution, and illustrates why these traditional rules result in small coalitions when applied to issues such as international environmental agreements.

1. Introduction

Coalition games are convex (concave) when the marginal contribution of a given player increases (decreases) in the size of the coalition they join. For convex characteristic function form games the core is non-empty, the Shapley value lies in the core, and the grand coalition is a unique equilibrium. However, for concave games with positive externalities the grand coalition may not be an equilibrium as some, or all, players would earn a higher payoff by leaving. Achieving large coalitions becomes more difficult when positive externalities generate free-rider incentives.

Matthew McGinty, University of Wisconsin-Milwaukee (mmcginty@umw.edu).

This paper has greatly benefited from the comments of an anonymous Associate Editor and referees, Dan Friedman, John Heywood, Frans de Vries, Scott Drewianka and a line of encouragement from Herve Moulin. All remaining errors are my own.

Received October 8, 2008; Accepted March 9, 2010.

Journal of Public Economic Theory, 13 (2), 2011, pp. 311–325.

Many real world situations correspond to superadditive, but concave, coalition games with positive externalities. Superadditivity implies aggregate payoff increases as coalitions increase in size. With positive externalities, the payoff to those outside the coalition increases in coalition size. For example, international environmental agreements (IEAs), such as the Kyoto Protocol, specify levels of global public good provision. IEAs typically suffer from the free-rider problem, since the payoff to those outside the coalition increases in membership (Barrett 1994). The grand coalition (full participation) is socially optimal for superadditive, positive externality games (Hafalir 2008), but generally is not stable (Yi 1997, Maskin 2003). Thus, any allocation of the coalition's worth must recognize the payoff outside the coalition, and must be appropriate in the event that the grand coalition is not an equilibrium.

This paper considers a single coalition game and proposes an allocation that maximizes stability. Of primary importance is what players would earn if they were to leave the coalition, not what they add to a coalition when they join. The approach seeks to minimize the incentive to deviate from a stable coalition. The set of Nash equilibria is determined using the non-cooperative approach from the cartel stability analysis of d'Aspremont et al. (1983). The set of equilibrium coalitions is generally a non-singleton, and the allocation has the same stability and robustness properties for all players and for all elements of this set. The allocation results in Strong Nash equilibria, since no subset of members has an incentive to leave a stable coalition. Similarly, no subset of non-members has an incentive to collectively join a stable coalition if the game is concave in this region.

The allocation presented below is robust to the maximum probability of 50% that players deviate from their individual best-response (intended action). Under the allocation rule, each coalition member receives their payoff outside the coalition plus an equal share of the coalition's surplus. Thus, what a player earns outside the coalition increases their payoff, in contrast to traditional allocations which are increasing in what a player contributes to the coalition. The Shapley value (1953) is a weighted average of the marginal contributions across all paths leading to the grand coalition that include that player. Unlike the rule presented in this paper, the Nash bargaining (1953) and Shapley (1953) allocations may divide the worth in the wrong direction, potentially making a stable coalition unstable.

The Myerson (1977) generalization of the Shapley value allows for externalities and multiple coalitions, but explicitly assumes that the grand coalition will form. More recently, Maskin (2003) has extended the Shapley value to games with externalities and shows that the grand coalition may not form. This approach derives an allocation from a randomization across all possible coalition formation paths. In Maskin's sequential bidding process each player's allocation is determined by their marginal contribution, thus the allocation reduces to the Shapley value in the absence of externalities. Similarly, Macho-Stadler et al. (2007) consider marginal contributions and derive an allocation which is an average of the Shapley value, whereas de Clippel and Serrano (2008) determine an allocation derived from both the marginal contributions and the externalities. Their allocation is unique when the externalities are symmetric, but results in a set with asymmetry. Again, all of these papers assume that the grand coalition will form.

Bloch (1996) and Yi (1997) consider coalition stability under fixed allocation rules and find that the grand coalition may not form. Yi considers games among (*ex ante*) symmetric players with an equal division of the coalition's worth. Bloch (1996) specifies an ordering where players propose a coalition, and the coalition forms if unanimously agreed upon. However, players are bound by their commitments to remain in the coalition, and the sharing rule is exogenously fixed.

Traditional allocation rules work well for convex characteristic function games without externalities (Winter 2002). In such games collective rationality requires that the grand coalition form since the core is non-empty. The core consists of all imputations (allocations of the grand coalition's worth) that remain after eliminating those blocked by all possible subcoalitions. However, the Shapley value and the Nash (1953) bargaining solution fare poorly when applied to single coalition games with positive externalities (Barrett 1997, Botteon and Carraro 2001). Large coalitions are typically not stable due to the free-rider problem. The resulting coalition is smaller than could be obtained with a rule that recognizes this issue.

Recent work (McGinty 2007, Weikard 2009) shows that there exists a set of allocations that satisfy the internal and external stability requirements of d'Aspremont et al. (1983). McGinty (2007) chooses a unique allocation from this set, with an arbitrary allocation based on a benefit–cost ratio for an IEA model. The allocation involves an abatement requirement under a system of tradable pollution permits. This new class of rules results in greater IEA participation than found in Barrett (1997) and Botteon and Carraro (2001), which implement the Shapley and Nash bargaining allocations. However, the allocation in McGinty (2007) does not maximize the stability of a stable coalition, nor is it the most robust when the game is subjected to uncertainty.

The rest of the paper is structured as follows. Section 2 defines coalition surplus and stability. Section 3 presents the new allocation rule and a simple example that compares the allocation with the Shapley value and the Nash bargaining solution. Section 4 shows the rule risk-dominates any other allocation. Section 5 provides an example with multiple equilibria and shows how the allocation can be applied to an IEA. The final section concludes.

2. Defining Coalition Stability

Consider a finite set N of players, with cardinality $n = |N|$. Coalitions are subsets of N, denoted as S, and contain $s = |S|$ elements. In a single coalition game there are 2^n possible coalitions comprising the powerset of N. A single coalition game, $\Gamma(N, S, v)$ specifies the worth of the coalition, $v(S)$, and the payoff to the $n - s$ players outside the coalition $v_i(S)$, $i \in N \backslash S$. With positive externalities, the payoff to those outside the coalition, $v_i(S)$, strictly increases in s. Let S_{-i} denote the resulting coalition when member i leaves $(S\backslash\{i\})$, let S_{+i} denote the resulting coalition when non-member i joins $(S \cup \{i\})$, and let N_{-S} denote the set of players outside the coalition $(N\backslash S)$. An efficient allocation rule, $x(S)$, distributes the entire worth among coalition members. The set of all efficient allocations is X:

$$X = \left\{ x \in \mathbb{R}^s : \sum_{i \in S} x_i(S) = v(S) \right\}. \tag{1}$$

Nash equilibria of the open-membership game are called stable coalitions. In an open-membership game existing coalition members may not block the accession of new members (Yi 1997). Nash equilibria satisfy the internal and external stability requirements from the cartel literature (d'Aspremont et al. 1983). Each member earns at least as much as they could be leaving: $x_i(S) \geq v_i(S_{-i}) \forall i \in S$ and each outside player would earn less by joining: $x_i(S_{+i}) < v_i(S) \forall i \notin S$. A coalition is called *essential* if there is some non-negative *surplus*, $\sigma(S)$, to distribute. The surplus is defined as the worth of the coalition minus the sum of payoffs from individually leaving the coalition:

$$\sigma(S) \equiv v(S) - \sum_{i \in S} v_i(S_{-i}). \tag{2}$$

If a coalition is essential, then the worth is sufficient such that there exists some efficient allocation rule $x \in X$ that satisfies internal stability for all coalition members. Similarly, if an enlarged coalition S_{+i} has a negative surplus for all $i \notin S$, then coalition S is externally stable.

For superadditive games with positive externalities, Weikard (2009) shows that there must be some stable (non-singleton) coalition. With superadditivity, at least a two-member coalition must be stable.

LEMMA 1: *For positive externality games, the set of stable coalitions is non-empty if the game $\Gamma(N, S, v)$ is superadditive.*

Proof: See Weikard (2009).

3. The New Allocation Rule

The new allocation rule awards each member their outside payoff, plus an equal share of the surplus. A simple example then shows that the Shapley value and Nash bargaining solution may divide the worth in the wrong direction, making a potentially stable coalition unstable. The rule is then compared to the nucleolus, and it is shown that the allocation results in strong Nash equilibria. Finally, if the game is concave in s beyond an equilibrium, then stable coalitions are robust to multiple non-members joining.

The new allocation rule, $x^*(S)$, is:

$$x_i^*(S) = v_i(S_{-i}) + \frac{1}{s}\left[v(S) - \sum_{i \in S} v_i(S_{-i})\right] \qquad \forall i \in S. \tag{3}$$

The stability of a coalition is determined by the member with the smallest payoff advantage to membership. Formally, a coalition S under some allocation rule r is internally stable up to the payoff difference $\pi_i^r(S) \equiv r_i(S) - v_i(S_{-i})$, where:

$$Stability(S) = \min_{i \in S} \pi_i^r(S). \tag{4}$$

PROPOSITION 1: *Coalition S is stable if and only if it is stable with respect to allocation rule $x^*(S)$.*

Proof: Under x^* the payoff advantage of coalition membership to player i is: $\pi_i^{x^*}(S) \equiv x_i^*(S) - v_i(S_{-i}) = \frac{1}{s}[v(S) - \sum_{i \in S} v_i(S_{-i})]$ for all $i \in S$. By definition, the coalition's surplus is: $\sigma(S) \equiv v(S) - \sum_{i \in S} v_i(S_{-i})$, so the payoff advantage is: $\pi_i^{x^*}(S) = \frac{\sigma(S)}{s} \forall i \in S$. If the surplus is non-negative, $\sigma(S) \geq 0$, then the coalition is internally stable. External stability ensures no non-member has an incentive to join $S: x_i^*(S_{+i}) < v_i(S) \forall i \in N_{-S}$, that is, $\sigma_{S+i} < 0 \forall i \in N_{-S}$. ■

The allocation $x^*(S)$ awards each member an equal share of the surplus. The new allocation rule (3) is unique, always exists and maximizes the stability of an internally stable coalition. Any other efficient allocation is less stable since the member with a smaller than equal share of the surplus has a smaller incentive to remain a member. Furthermore, $x^*(S)$ is unique and exists even if the core is empty or the coalition is non-essential.[1]

A simple three-player example illustrates differences with traditional allocations.

Superadditivity is seen by adding the appropriate elements in the rows of Table 1 to create an enlarged coalition. In all cases the worth of the enlarged coalition exceeds the sum of the payoffs for the members in that row. For example, $v\{1\} + v_2\{1\} < v\{1, 2\}$, since $10 < 28$. Positive externalities imply that outside payoff is strictly increasing in coalition size. Table 1 shows the grand coalition is the unique Nash equilibrium under x^*. The surplus $\sigma_{\{1,2,3\}} \equiv v(N) - \sum_{i \in N} v_i(N_{-i}) = 15$, thus in Table 2 each player receives five more than their outside payoff under x^*.

[1] I thank an anonymous Associate Editor for suggesting the current form of Proposition 1, and an anonymous referee for pointing out that $x_i^*(S)$ could be negative, even in a game where all worths and outside payoffs are strictly non-negative. This could occur if a coalition is unstable, $v_i(S)$ is small, and the surplus σ_S is sufficiently negative such that: $v_i(S) < \frac{\sigma_S}{s}$.

Table 1: Example 1

Coalition worth $v(S)$	**Outside payoff $v_i(S)$, $i \notin S$**
$v(\{\emptyset\}) = 0$	$v_i = 0\ \forall i \in N$
$v(\{1\}) = 4$	$v_2 = 6$, $v_3 = 8$
$v(\{2\}) = 8$	$v_1 = 10$, $v_3 = 12$
$v(\{3\}) = 12$	$v_1 = 14$, $v_2 = 16$
$v(\{1, 2\}) = 28$	$v_3 = 20$
$v(\{1, 3\}) = 44$	$v_2 = 38$
$v(\{2, 3\}) = 64$	$v_1 = 27$
$v(\{1, 2, 3\}) = 100$	

Table 2: Allocation rules for Example 1

		Shapley value	**Nash bargaining**
S	x_i^*	sv_i	nb_i
$\{1\}$	$x_1 = 4$	$sv_1 = 4$	$nb_1 = 4$
$\{2\}$	$x_2 = 8$	$sv_2 = 8$	$nb_2 = 8$
$\{3\}$	$x_3 = 12$	$sv_3 = 12$	$nb_3 = 12$
$\{1, 2\}$	$x_1 = 16$, $x_2 = 12$	$\mathbf{sv_1 = 12}$, $\mathbf{sv_2 = 16}$	$\mathbf{nb_1 = 12}$, $\mathbf{nb_2 = 16}$
$\{1, 3\}$	$x_1 = 25$, $x_3 = 19$	$\mathbf{sv_1 = 18}$, $\mathbf{sv_3 = 26}$	$\mathbf{nb_1 = 18}$, $\mathbf{nb_3 = 26}$
$\{2, 3\}$	$x_2 = 34$, $x_3 = 30$	$\mathbf{sv_2 = 30}$, $\mathbf{sv_3 = 34}$	$\mathbf{nb_2 = 30}$, $\mathbf{nb_3 = 34}$
$\{1, 2, 3\}$	$\mathbf{x_1 = 32}$, $\mathbf{x_2 = 43}$, $\mathbf{x_3 = 25}$	$sv_1 = 22$, $sv_2 = 34$, $sv_3 = 44$	$nb_1 = 29.3$, $nb_2 = 33.3$, $nb_3 = 37.3$

Note. Payoffs for the Nash equilibrium coalition structures are in bold.

The grand coalition has a positive surplus, but is not stable under the Shapley value or the Nash bargaining solution. Under both the Shapley value and Nash bargaining solution, the set of Nash equilibria consists of all the two-member coalitions in Table 2. For Shapley, both players 1 and 2 have an incentive to leave the grand coalition, while under Nash bargaining player 2 has an incentive to leave. Player 3 has the highest marginal contributions in Table 1, but the lowest payoff outside the grand coalition. Thus, the Shapley value awards the most to player 3 and that player has no incentive to leave the grand coalition.

The allocation x^* bears resemblance to the nucleolus (Schmeidler 1969). The nucleolus is an efficient division of the grand coalitions worth (imputation) which minimizes the largest deficit $d(x)$ across all coalitions. The deficits are the difference between the allocation of the grand coalition's worth and the worths of all possible coalitions, i.e., for any coalition S and imputation x, $d(x) = v(S) - \sum_{i \in S} x_i(N)$. The nucleolus seeks to minimize the dissatisfaction of any subset of players by choosing the imputation that minimizes the largest deficit. By contrast, the allocation x^* minimizes the dissatisfaction relative to a player's outside payoff $v_i(S)$, not the worth of coalitions containing player i. Furthermore, x^* is coalition specific, i.e.,

x^* for player i changes in S, while the nucleolus is determined by the grand coalition. When the grand coalition is a Nash equilibrium, x^* only considers the worth of the grand coalition, $v(N)$, and the outside payoffs, $v_i(N_{-i})$, not the worth of any other coalition, $v(S)$.

A further refinement is that of a strong Nash equilibrium (SNE). An SNE is robust to deviations by multiple members.

PROPOSITION 2: *Stable coalitions under allocation x^* are strong Nash equilibria in a superadditive single coalition game with positive externalities.*

Proof: Internally stable coalitions S satisfy: $x_i^*(S) \geq v_i(S_{-i}) \forall i \in S$. Suppose a proper subset of coalition members, $R \subset S$, with cardinality $r = |R| \geq 2$, leaves the coalition and earns payoff $v_i(S_{-R}) \forall i \in R$. With positive externalities, $v_i(S)$ is strictly increasing in s. Thus, $v_i(S_{-R}) < v_i(S_{-i}) < x_i^*(S)$ and no subset of S has an incentive to collectively leave the coalition. Therefore, under $x^{*,}$ all stable coalitions are strong Nash equilibria. ■

A similar argument shows that multiple non-members have no incentive to join a stable coalition S if the game is concave for coalitions larger than S.

PROPOSITION 3: *Suppose the game is concave for coalitions larger than S. Then no subset of noncoalition members $R \subseteq N_{-S}$ has an incentive to join S.*

Proof: Positive externalities imply outside payoff strictly increases in coalition size, $v_i(S_{+j-i}) > v_i(S_{-i}) \forall i \in S,\ j \in N_{-S}$. That is, all coalition members have a greater outside payoff once non-member j joins. Concavity over this range implies that an outside player contributes less to the worth, the larger the coalition: $v(S_{+i+j}) - v(S_{+i}) < v(S_{+j}) - v(S)$. The proof of Proposition 1 shows coalition S is internally stable when $\sigma_S > 0$ and externally stable when $x_i^*(S_{+i}) < v_i(S) \forall i \in N_{-S}$, that is, $\sigma_{S+i} < 0 \forall i \in N_{-S}$. Therefore, when S is stable, $\sigma_{S+i} - \sigma_S < 0$. Thus, $\sigma_{S+i} - \sigma_S < 0$ implies $\sigma_{S+i+j} - \sigma_{S_{+i}} < \sigma_{S+i} - \sigma_S < 0$, since $\sigma_{S+i+j} - \sigma_{S_{+i}} = v(S_{+i+j}) - \sum_{k \in S_{+j+i}} v_k(S_{+j+i-k}) - v(S_{+i}) + \sum_{j \in S_{+i}} v_j(S_{+i-j})$. By concavity, $v(S_{+i+j}) - v(S_{+i}) < v(S_{+i}) - v(S)$, and by positive externalities $\sum_{k \in S_{+j+i}} v_k(S_{+j+i-k}) > \sum_{j \in S_{+i}} v_j(S_{+i-j})$, thus $\sigma_{S+i+j} - \sigma_{S_{+i}} < \sigma_{S+i} - \sigma_S$. Similarly, $\sigma_{S_{+R}} - \sigma_{S_{+i}} < \sigma_{S+i} - \sigma_S$. Thus, if coalition S is stable, then S_{+R} is unstable for all $R \subseteq N_{-S}$. ■

However, as example 2 in Section 5 will illustrate, when there are multiple Nash equilibria, stable coalitions can be obtained when a member leaves and a non-member joins. That is, multiple deviations from one Nash may result in coalitions that are also Nash equilibria. Section 4 shows that the allocation risk-dominates any other efficient allocation, since it is robust to the largest possible deviation probability. It considers all possible deviations, and the robustness of stable coalitions under x^*.

4. Risk Dominance

Proposition 1 showed that the allocation x^* maximizes stability, compared to all other efficient allocation rules. Next, the risk properties of the allocation rule are considered. Given uncertainty about the actions of other players, it is shown that the allocation rule risk dominates all other efficient allocations (Harsanyi and Selton 1988).[2] An allocation risk dominates if membership remains a best-response for the greatest degree of uncertainty regarding other players membership decisions. From the game in the previous sections $\Gamma(N, S, v)$ we add the (independent) probability $\epsilon \in (0, 1)$ that each player chooses a different action, given a stable coalition S. Thus, there is an ϵ probability that each member leaves and an ϵ probability that each non-member joins.[3] Note that these decisions are individually irrational since, by definition, coalition S is stable. However, multiple deviations may lead to stable coalitions, that is, both S and S_{-i+j} may be stable.

Define the set of players that deviate as T, where $T \subseteq N$ and $|T| = t$. The probability of a given number of deviations t is determined by a binomial distribution. The resulting coalition depends on which players tremble, not just the number. Previously, the payoff advantage to coalition membership under x^* was $\pi_i(S) \equiv x_i^*(S) - v_i(S_{-i})$. With deviations, the game is then $G(N, S, v, \epsilon, T)$ and the payoff advantage becomes an expected value: $\pi_i(S, T, \epsilon)$. It is assumed that the deviations are independent across players (Mas-Colell et al. 1995) and that players intend to play their best-response from the action set $A = \{\text{in}, \text{out}\}$. A stable coalition S, under allocation rule r, is robust to a deviation rate $\epsilon^*(S)$ where:

$$\epsilon^*(S) = \min_{\epsilon_i \in S} \max_{\epsilon_i \in (0,1)} \pi_i^r(S, T, \epsilon) \geq 0. \tag{5}$$

Proposition 1 showed that if a coalition is stable for one member under the allocation rule $x^*(S)$, then it is stable for all members.

With deviation rate ϵ, all 2^n coalitions are obtained with probability given by a binomial distribution. Under $x^*(S)$ the expected payoff advantage for member i is

$$\pi_i^{x^*}(S, T, \epsilon) = \sum_{t=0}^{n} \frac{n!}{t!(n-t)!}\epsilon^t\,(1-\epsilon)^{n-t}\left[x_i^*(S, T) - v_i(S_{-i}, T)\right].$$

[2] Harsanyi and Selton (1988) define risk-dominance as a relationship between two Nash equilibria. In this paper, risk-dominance is used to compare allocations at a given Nash equilibrium. It is shown that $x^*(S)$ risk-dominates all other allocations, at all Nash equilibria.

[3] This notion of stability is similar to trembling-hand perfection. Robustness of a coalition is the minimum tremble rate such that there is a payoff advantage to remaining in a stable coalition. In this context, there is probability $\epsilon \in (0, 1)$ that each player trembles and chooses a different action.

PROPOSITION 4: *Under allocation rule $x^*(S)$, the payoff advantage to coalition membership $\pi_i^{x^*}(S, T, \epsilon)$ in the game with uncertainty is an expected value of the surplus shares containing player i.*

Proof: Using the definitions of σ_S and x^* in Equations (2) and (3), and expanding the binomial distribution, each surplus containing player i is realized twice. Each surplus, σ_S, has two terms, with the following relationship between coefficients: (i) $\epsilon^j(1-\epsilon)^{n-t}$ and (ii) $\epsilon^k(1-\epsilon)^{n-k}$, where $|j-k| = 1$. ■

For the example in Tables 1 and 2, the expected payoff to player 1 from remaining a member of the stable grand coalition is

$$\begin{aligned}
&\sum_{t=0}^{n} \frac{n!}{t!(n-t)!}\epsilon^t(1-\epsilon)^{n-t}\left[x_1^*(\{1,2,3\}, T)\right] \\
&\quad = (1-\epsilon)^3\left[x_1^*(\{1,2,3\})\right] \\
&\qquad + \epsilon(1-\epsilon)^2\left[x_1^*(\{1,2\}) + x_1^*(\{1,3\}) + v_1(\{2,3\})\right] \\
&\qquad + \epsilon^2(1-\epsilon)\left[x_1^*(\{1\}) + v_1(\{2\}) + v_1(\{3\})\right] + \epsilon^3\left[v_1(\emptyset)\right].
\end{aligned} \tag{6}$$

Note that all eight coalitions are reached with a positive probability, with the stable grand coalition most likely, and the empty coalition least likely. The expected payoff to player 1 from leaving the grand coalition is

$$\begin{aligned}
&\sum_{t=0}^{n} \frac{n!}{t!(n-t)!}\epsilon^t(1-\epsilon)^{n-t}\left[v_1(\{2,3\}, T)\right] \\
&\quad = (1-\epsilon)^3 v_1(\{2,3\}) + \epsilon(1-\epsilon)^2\left[x_1^*(\{1,2,3\}) + v_1(\{2\}) + v_1(\{3\})\right] \\
&\qquad + \epsilon^2(1-\epsilon)\left[x_1^*(\{1,2\}) + x_1^*(\{1,3\}) + v_1(\emptyset)\right] + \epsilon^3\left[x_1^*(\{1\})\right].
\end{aligned} \tag{7}$$

Thus, the expected payoff advantage to coalition membership, the difference between (6) and (7), is

$$\begin{aligned}
\pi_1^{x^*}(S, T, \epsilon) = {} & (1-\epsilon)^3\left[x_1^*(\{1,2,3\}) - v_1(\{2,3\})\right] \\
& + \epsilon(1-\epsilon)^2\left[x_1^*(\{1,2\}) - v_1(\{2\}) + x_1^*(\{1,3\}) - v_1(\{3\})\right. \\
& \qquad\qquad \left. + v_1(\{2,3\}) - x_1^*(\{1,2,3\})\right] \\
& + \epsilon^2(1-\epsilon)\left[x_1^*(\{1\}) - v_1(\emptyset)\right. \\
& \qquad\qquad \left. + v_1(\{2\}) - x_1^*(\{1,2\}) + v_1(\{3\}) - x_1^*(\{1,3\})\right] \\
& + \epsilon^3\left[v_1(\emptyset) - x_1^*(\{1\})\right].
\end{aligned} \tag{8}$$

Using the definitions of the allocation x^* and the surplus σ_S in Equations (3) and (2), the expected payoff advantage reduces to

$$\pi_1^{x^*}(S, T, \epsilon) = (1-\epsilon)^3\left[\frac{\sigma_{\{1,2,3\}}}{3}\right] + \epsilon(1-\epsilon)^2\left[\frac{\sigma_{\{1,2\}}}{2} + \frac{\sigma_{\{1,3\}}}{2} - \frac{\sigma_{\{1,2,3\}}}{3}\right]$$
$$+ \epsilon^2(1-\epsilon)\left[\sigma_{\{1\}} - \frac{\sigma_{\{1,2\}}}{2} - \frac{\sigma_{\{1,3\}}}{2}\right] + \epsilon^3\left[-\sigma_{\{1\}}\right]. \quad (9)$$

Note that under x^* each surplus containing player 1 appears twice, once with a positive and once with a negative coefficient. Furthermore, each surplus differs by one power of ϵ. Analogous expressions can be obtained for players 2 and 3, containing all four surpluses that each player contributes to. This leads directly to the robustness of allocation rule x^*.

PROPOSITION 5: *The allocation rule x^* is risk-dominant. It is robust to a deviation rate of 0.5, larger than any other allocation.*

Proof: The expected payoff difference $\pi_i^{x^*}(S, T, \epsilon)$ is: $\sum_{t=0}^{n}\frac{n!}{t!(n-t)!}\epsilon^t(1-\epsilon)^{n-t}[x_i^*(S, T) - v_i(S_{-i}, T)]$. Thus, $\pi_i^{x^*}(S, T, \epsilon) = 0$ for $\epsilon = 0.5$. This is true for all $i \in S$, and for all stable coalitions S. Consequently, the coalition is stable up to the maximum deviation rate 0.5. Consider the set $X(S)$ of all efficient allocations that are not $x^*(S)$. $X(S) = \{\sum_{i\in S} x_i(S) = v(S) : x_i(S) \neq x_i^*(S)$ for some $i \in S\}$. By efficiency, $\sum_{i\in S} x_i(S) = \sum_{i\in S} x_i^*(S) = v(S)$ and $x_i(S) \neq x_i^*(S)$ for some $i \in S$. Then there exists some $i \in S$, such that $x_i(S) < x_i^*(S)$. Thus, for all efficient rules $x(S)$ that are not $x^*(S)$, $\pi_i^{x}(S, T, \epsilon) = 0$ for $\epsilon < 0.5$. ■

5. An Illustration and Comparison with Alternative Allocation Rules

Example 2 in Table 3 considers a four-player ($i = 1, 2, 3, 4$) superadditive game with positive externalities, generating multiple stable coalitions under x^*. The grand coalition is not stable under any allocation since $\sigma_{\{1,2,3,4\}} < 0$.

To show risk-dominance, consider the stable coalition $\{1, 2, 3\}$ and probability ϵ that each player deviates. The payoff advantage for player 1 remaining in stable coalition $\{1, 2, 3\}$ is

$$\pi_1^{x^*}(\{1, 2, 3\}, T, \epsilon)$$
$$= (1-\epsilon)^4\left[\frac{\sigma_{\{1,2,3\}}}{3}\right] + \epsilon(1-\epsilon)^3\left[\frac{\sigma_{\{1,2,3,4\}}}{4} + \frac{\sigma_{\{1,2\}}}{2} + \frac{\sigma_{\{1,3\}}}{2} - \frac{\sigma_{\{1,2,3\}}}{3}\right]$$
$$+ \epsilon^2(1-\epsilon)^2\left[\sigma_1 + \frac{\sigma_{\{1,2,4\}}}{3} + \frac{\sigma_{\{1,3,4\}}}{3} - \frac{\sigma_{\{1,2,3,4\}}}{4} - \frac{\sigma_{\{1,2\}}}{2} - \frac{\sigma_{\{1,3\}}}{2}\right]$$
$$+ \epsilon^3(1-\epsilon)\left[\frac{\sigma_{\{1,4\}}}{2} - \frac{\sigma_{\{1,2,4\}}}{3} - \frac{\sigma_{\{1,3,4\}}}{3} - \sigma_{\{1\}}\right] + \epsilon^4\left[-\frac{\sigma_{\{1,4\}}}{2}\right]. \quad (10)$$

Table 3: Example 2

Coalition and worth	Outside payoff	Surplus	Allocation
$v(S)$	$v_i(S),\ i \notin S$	$\sigma(S)$	$x_i^*(S)$
$v(\{\emptyset\}) = 0$	$v_i = 0 \forall i \in N$		
$v(\{1\}) = 30$	$v_i = 10 \forall i \neq 1$	$\sigma_{\{1\}} = 30$	$x_1 = 30$
$v(\{2\}) = 40$	$v_i = 20 \forall i \neq 2$	$\sigma_{\{2\}} = 40$	$x_2 = 40$
$v(\{3\}) = 50$	$v_i = 30 \forall i \neq 3$	$\sigma_{\{3\}} = 50$	$x_3 = 50$
$v(\{4\}) = 60$	$v_i = 40 \forall i \neq 4$	$\sigma_{\{4\}} = 60$	$x_4 = 60$
$v(\{1, 2\}) = 110$	$v_3 = 40,\ v_4 = 50$	$\sigma_{\{1,2\}} = 80$	$x_1 = 60,\ x_2 = 50$
$v(\{1, 3\}) = 130$	$v_2 = 60,\ v_4 = 80$	$\sigma_{\{1,3\}} = 90$	$x_1 = 75,\ x_3 = 55$
$v(\{1, 4\}) = 150$	$v_2 = 80,\ v_3 = 90$	$\sigma_{\{1,4\}} = 100$	$x_1 = 90,\ x_4 = 60$
$v(\{2, 3\}) = 180$	$v_1 = 50,\ v_4 = 60$	$\sigma_{\{2,3\}} = 130$	$x_2 = 95,\ x_3 = 85$
$v(\{2, 4\}) = 210$	$v_1 = 60,\ v_3 = 90$	$\sigma_{\{2,4\}} = 150$	$x_2 = 115,\ x_4 = 95$
$v(\{3, 4\}) = 240$	$v_1 = 70,\ v_2 = 80$	$\sigma_{\{3,4\}} = 170$	$x_3 = 125,\ x_4 = 115$
$v(\{\mathbf{1, 2, 3}\}) = 270$	$v_4 = 160$	$\sigma_{\{1,2,3\}} = 120$	$\mathbf{x_1 = 90, x_2 = 100, x_3 = 80}$
$v(\{\mathbf{1, 2, 4}\}) = 310$	$v_3 = 150$	$\sigma_{\{1,2,4\}} = 120$	$\mathbf{x_1 = 100, x_2 = 120, x_4 = 90}$
$v(\{\mathbf{1, 3, 4}\}) = 330$	$v_2 = 130$	$\sigma_{\{1,3,4\}} = 90$	$\mathbf{x_1 = 100, x_3 = 120, x_4 = 110}$
$v(\{\mathbf{2, 3, 4}\}) = 350$	$v_1 = 140$	$\sigma_{\{2,3,4\}} = 120$	$\mathbf{x_2 = 120, x_3 = 130, x_4 = 100}$
$v(\{1, 2, 3, 4\}) = 500$		$\sigma_{\{1,2,3,4\}} = -80$	$x_1 = 120,\ x_2 = 110,\ x_3 = 130,\ x_4 = 140$

Note. Stable coalitions are in bold.

Each σ that contains player 1 appears twice, with one positive and one negative coefficient. Hence, it is straightforward to show that the payoff difference is zero when $\epsilon = 0.5$, for all possible values of the σs. Using the example in Table 3, (10) simplifies to

$$\pi_1^{x^*}(\{1, 2, 3\}, T, \epsilon = 40(1-\epsilon))^4 + 25\epsilon(1-\epsilon)^3 + 35\epsilon^2(1-\epsilon)^2 - 50\epsilon^3(1-\epsilon) - 50\epsilon^4, \qquad (11)$$

where $\pi_1^{x^*}(\{1, 2, 3\}, T, \epsilon) = 0$ for $\epsilon = 0.5$, and is positive for all $\epsilon \in (0, 0.5)$. The partial derivative of $\pi_1^{x^*}(\{1, 2, 3\}, T, \epsilon)$ with respect to any σ results in an expression of the form $\frac{\partial \pi_1^{x^*}(\{1,2,3\},T,\epsilon)}{\partial \sigma_{\{1,2,3\}}} = \frac{(1-\epsilon)^4}{3} - \frac{\epsilon(1-\epsilon)^3}{3}$, which will equal zero only for $\epsilon = 0.5$. Similarly, for all other players $\pi_i^{x^*}$ yields a payoff difference in ϵ and all σs that contain player i. Again, the critical value is $\epsilon = 0.5$ for all $i \in N$, as shown formally in the previous section. Using the definition of dominance in Equation (5) the coalition is robust up to the maximum value of $\epsilon^* = 0.5$.

5.1. Comparison with Alternative Allocation Rules

Table 4 compares the allocations under x^*, the Shapley value and the Nash bargaining solution for Example 2.

Table 4: Example 2 allocations

		x∗(S)				Nash bargaining solution				Shapley value			
Coalition	**Worth**	x_1	x_2	x_3	x_4	nb_1	nb_2	nb_3	nb_4	sv_1	sv_2	sv_3	sv_4
∅	0												
{1}	30	30				30				30			
{2}	40		40				40				40		
{3}	50			50				50				50	
{4}	60				60				60				60
{12}	110	60	50			50	60			50	60		
{13}	130	75		55		55		75		55		75	
{14}	150	90			60	60			90	60			90
{23}	180		95	85			85	95			85	95	
{24}	210		115		95		95		115		95		115
{34}	240			125	115			115	125			115	125
{123}	270	90	100	80		80	90	100		65	95	110	
{124}	310	100	120		90	90	100		120	70	105		135
{134}	330	100		120	100	93.3		113.3	123.3	68.3		123.3	138.3
{234}	350		120	130	100		106.7	116.7	126.7		96.7	116.7	136.7
{1234}	500	120	110	130	140	110	120	130	140	88.3	116.7	135	160

The ordinal ranking of payoffs differs across rules. For example, coalition $\{1, 2, 3\}$ shows that under x^* the highest payoff, of 100, is for player 2. By contrast, player 3 has the highest payoff under the Shapley and Nash bargaining allocations. Player 3 receives the highest payoff under Shapley and Nash, but the lowest payoff under x^*. Except for the singletons, Example 2 shows this result is obtained systematically. This opposite ordinal ranking is due to the different threatpoints. For the Nash bargaining rule the relevant threatpoints are the singletons. The Shapley value is composed of the marginal contribution across all possible subcoalitions. For instance, this means that player 1 gets a greater share of the worth of $\{1, 2\}$ under x^* than player 2, even though $v\{1\} = 30$ and $v\{2\} = 40$. Unlike the Shapley value, one's contribution to subcoalitions is not what increases one's share of the allocation. This result occurs because the game is open-membership. While player 2 would prefer to join a coalition with 4 instead, it cannot stop 1 from joining. Open membership describes most international agreements where the decision is from {in, out} and one cannot dictate *in* with one player rather than another. This also highlights why the Shapley value has fared so poorly in IEA applications (e.g., Barrett 1997, Botteon and Carraro 2001). Note that the relevant question is not "what would you bring if you were to join," but rather "what would I get if I were to leave." A final example shows how the allocation can be implemented to generate abatement requirements in an environmental agreement.

5.2. International Environmental Agreements

The allocation $x^*(S)$ has many practical applications, including the design of IEAs such as the Kyoto Protocol. Greenhouse gas abatement is a global public good, generating positive externalities to nations outside the agreement (free-riders). The Kyoto Protocol is an open-membership single-coalition game where existing coalition members may not prevent a nation from joining. Signatories are coalition members who choose an aggregate abatement level and then allocate this among members via an abatement requirement. Pollution permit trading among members allows the aggregate abatement level to be reached at the lowest cost, thus maximizing the coalition's worth. The individual abatement requirements then define an allocation of the coalition's worth given actual abatement and permit revenue.

The seminal IEA paper is Barrett's (1994) model with declining marginal benefit and increasing marginal cost. McGinty (2007) extends the model to allow for asymmetric benefit shares and marginal abatement cost slopes. Global benefit is a concave function of aggregate abatement: $B(Q) = b(aQ - \frac{Q^2}{2})$, where $\sum_{i \in N} q_i = Q$, $a > 0$, and $b > 0$. Nation i receives benefit share α_i, thus $B_i(Q) = b\alpha_i(aQ - \frac{Q^2}{2})$, where $\sum_{i \in N} \alpha_i = 1$. Abatement costs are asymmetric $C_i(q_i) = \frac{c_i q_i^2}{2}$, and marginal abatement costs are rays from the origin with slope c_i. Net benefit to nation i is $B_i(Q) - C_i(q_i)$. The

IEA is a coalition of signatories that choose a coalition level of abatement $Q(S) = \sum_{i \in S} q_i$ to maximize the worth: $v(S) = \sum_{i \in S}[b\alpha_i(aQ - \frac{Q^2}{2}) - \frac{c_i q_i^2}{2}]$. All non-signatories choose q_i to maximize individual payoff: $B_i(Q) - C_i(q_i)$. Coalition worth is maximized by choosing the least cost allocation of abatement level $Q(S)$. This occurs where the marginal abatement cost of the last unit is equated across members: $c_i q_i = p(S) \forall i \in S$.

The allocation rule $x^*(S)$ can be implemented by a system of tradable pollution permits with price $p(S)$. The coalition specifies an abatement requirement q_i^r for each member, such that the sum of requirements is optimal for that coalition, $Q(S) = \sum_{i \in S} q_i^r(S)$. Therefore, nation i earns permit revenue equal to: $p(S)[q_i(S) - q_i^r(S)]$. Note that the permit revenue is a zero-sum system of transfers. In this context, allocation rules differ by their abatement requirements. The payoff from leaving the coalition is: $v_i(S_{-i}) = B_i(Q) - C_i(q_i)$ and the surplus of any coalition is: $\sigma(S) = v(S) - \sum_{i \in S} v_i(S_{-i})$. The outside payoff is $B_i(Q) - C_i(q_i)$, where abatement is obtained for coalition structure S_{-i} and $\sum_{i \in S} x_i^*(S) = v(S)$. The allocation $x^*(S)$ then results in abatement requirements that solve: $x_i^*(S) = v_i(S_{-i}) + \frac{\sigma(S)}{s}$.

6. Conclusion

This paper presents a new allocation rule that maximizes coalition stability in the presence of positive externalities. The rule awards each player their payoff if they were to individually leave the coalition, plus an equal share of the coalition's surplus. Contrary to classical rules like the Shapley value, the relevant consideration is not what a player adds to a coalition, but rather what a player would get if they were to leave the coalition. The rule is unique, always exists, and is robust up to the maximum degree of uncertainty regarding players' decisions. Under the allocation, coalitions are robust to multiple defections and, when the game is concave, to multiple non-members joining.

The rule has practical application such as, but not limited to, the provision of global public goods. Provision generates positive externalities to non-members and hence the free-rider problem implies the socially optimal grand coalition is not an equilibrium. Under the actual Kyoto treaty the individual abatement requirements are *ad hoc*. They do not address coalition stability, nor do they consider robustness if a nation chooses an action that is not a best-response. Clearly, the proper abatement requirements can improve both participation and stability of the agreement. The more robust the allocation, the more participation remains a best-response given uncertainty about others' actions.

References

d'ASPREMONT, C., A. JACQUEMIN, J. J. GABSZEWICZ, and J. WEYMARK (1983) On the stability of collusive price leadership, *Canadian Journal of Economics* **16**, 17–25.

BARRETT, S. (1994) Self-enforcing international environmental agreements, *Oxford Economic Papers* **46**, 878–894.

BARRETT, S. (1997) Heterogeneous international environmental agreements, in *International Environmental Negotiations: Strategic Policy Issues*, C. Carraro, ed. Cheltenham, UK: Edward Elgar.

BLOCH, F. (1996) Sequential formation of coalitions in games with externalities and fixed payoff division, *Games and Economic Behavior* **14**, 90–123.

BOTTEON, M., and C. CARRARO (2001) Environmental coalitions with heterogenous countries: Burden sharing and carbon leakage, in *Environmental Policy, International Agreements, and International Trade*, A. Ulph, ed. New York: Oxford University Press, 38–65.

de CLIPPEL, G., and R. SERRANO (2008) Marginal contributions and externalities in the value, *Econometrica* **76**(6), 1413–1436.

MAS-COLELL, A., M. WHINSTON, and J. GREEN (1995) *Microeconomic Theory*. New York: Oxford University Press.

HAFALIR, I. (2007) Efficiency in coalition games with externalities, *Games and Economic Behavior* **61**(2), 242–258.

HARSANYI, J., and R. SELTON (1988) *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

MACHO-STADLER, I., D. PÉREZ-CASTRILLO, and D. WETTSTEIN (2007) Sharing the surplus: An extension of the Shapley value for environments with externalities, *Journal of Economic Theory* **135**, 339–356.

MASKIN, E. (2003) Bargaining, coalitions, and externalities, Working paper and Presidential Address to the Econometric Society, 2003.

McGINTY, M. (2007) International environmental agreements among asymmetric nations, *Oxford Economic Papers* **59**(1), 45–62.

MYERSON, R. B. (1977) Values of games in partition function form, *International Journal of Game Theory* **6**(1), 23–31.

NASH, J. (1953) Two-person cooperative games, *Econometrica* **21**(1), 128–140.

SCHMEIDLER, D. (1969) The nucleolus of a characteristic function game, *SIAM Journal on Applied Mathematics* **17**(6), 1163–1170.

SHAPLEY, L. (1953) A value for n-person games, in *Contributions to the Theory of Games II*, H. W. Kuhn and A. W. Tucker, eds. (Annals of Mathematics Studies 28), Princeton University Press, Princeton, 307–317.

WEIKARD, H-P. (2009) Cartel stability under an optimal sharing rule, *The Manchester School* **77**(5), 575–593.

WINTER, E. (2002) The Shapley value, in *The Handbook of Game Theory*, R. J. Aumann and S. Hart, eds. North-Holland.

YI, S-S. (1997) Stable coalition structures with externalities, *Games and Economic Behavior* **20**, 201–237.