Power and legitimacy in pillage games

by

J.S. Jordan

The Pennsylvania State University

July 16, 2006

Abstract

A *pillage game* is a formal model of Hobbesian anarchy as a coalitional game. The technology of pillage is specified by a power function that determines the power of each coalition as a function of its members and their wealth. A coalition can despoil any other coalition less powerful than itself. The present paper studies the extent to which the exercise of power can be constrained by a shared concept of legitimacy. The basic pillage game is augmented by a set of extrinsic variables that can convey information about past behavior. Depending on the power function, the illegitimate use of power can be inhibited by legitimizing the subsequent use of power against the transgressors. Legitimacy is modeled in a static sense, called *quasi-legitimacy*, using the stable set (von Neumann-Morgenstern solution) of the augmented pillage game, and in an explicitly dynamic sense, called simply legitimacy, using a concept of farsighted core. Quasi-legitimacy is shown to be a necessary but not sufficient condition for legitimacy. The sets of quasi-legitimate wealth allocations are characterized, and an iterative process is developed for constructing the largest quasi-legitimate set of allocations for each pillage game. If the power function gives enough weight to coalition size that no individual can be as powerful as the coalition of everyone else, then a natural augmentation of the basic pillage game can legitimize the set of all allocations. However, if the power of each coalition is determined by its total wealth alone, then even the weaker concept of quasi-legitimacy cannot stabilize anything other than the stable set of the basic pillage game. The legitimate sets of wealth allocations are characterized in general, and a weaker condition is shown to characterize the legitimacy of the largest quasi-legitimate set.

1. Introduction

A pillage game is a natural setting for the study of power. There is only one commodity, wealth, which is allocated among a finite number of players. Since wealth is desired by all, reallocation can only be effected by force. The technology of force is specified by a power function that determines the power of each coalition as a function of its members and their wealth. A coalition can take the wealth of any coalition less powerful than itself. Pillage is costless and certain, but due to the absence of a commitment technology, it can also be treacherous. Previous work on pillage games uses the concept of stable set to identify the set of allocations that can be stabilized as a balance of power (Jordan (2005) and Jordan and Obadia (2005)). Stable sets are also characterized as having a dynamic representation as a farsighted core under rational expectations (Jordan (2005, Section 6)). The balance of power in a pillage game is delicate, as evidenced by the fact that stable sets can contain at most finitely many allocations. This raises the question whether the exercise of power can be endogenously constrained to stabilize a larger set of allocations.

The more tightly power is constrained, the larger will be the set of stable allocations. However, constraints on the use of power can only be enforced by the use of power itself. An illegitimate use of power can only be discouraged by legitimizing the subsequent use of power against the transgressors. Loosely speaking, we seek the narrowest concept of legitimacy in the exercise of power that is still broad enough to be self-enforcing.

It is first necessary to extend the basic pillage game to add variables that can convey information about past actions. This information is extrinsic to the environment of the game, in the sense that it is irrelevant to the technology of power and does not enter the players' preferences, which are simply increasing in wealth. In Section 3, below, self-enforcement is represented by the static concept of the stable set of the extended pillage game. A set of allocations is called *quasi-legitimate* if it can be stabilized in the extended game. Theorem 3.7 gives a general characterization of quasi-legitimate sets. Proposition 3.6 shows that the simplest possible extension, adding only a single Boolean variable, suffices to stabilize every quasi-legitimate set. Proposition 3.9 develops an iterative process for constructing the largest quasi-legitimate set. The process is based on an induction argument first used by Roth (1976) to establish the existence of the "supercore" for general abstract games. In the present setting, the process requires only finitely many steps. Section 3 also develops a more interesting extension, called the *citizenship game*. The extrinsic information identifies each player as either a citizen or an outlaw. Any act of pillage that victimizes a citizen causes all who benefit from the pillage to become outlaws. However, any pillage that victimizes only outlaws enables all players to become citizens. The citizenship game thus provides for both the punishment and redemption of outlaws. This concept of legitimacy can fail to be self-enforcing if a player can, through pillage, become wealthy enough to be too powerful to be punished. Accordingly, Proposition 3.17 shows that the set of all allocations in which only citizens possess wealth is a stable set of the citizenship game if the power function gives at least enough weight to coalitional size that no one player, even if possessed of the total wealth in the game, can be as powerful as the coalition of everyone else. This restriction on the power function, called the *no-tyranny* condition, is necessary as well as sufficient. Under the no-tyranny condition, the entire set of allocations is quasi-legitimate, since they are stable if all players are designated as citizens. The no-tyranny condition is obviously violated by the *wealth-is-power* (WIP) game, in which the power of each coalition is determined by its total wealth alone. Proposition 3.15 shows that for the WIP game, legitimacy is a fruitless concept, since no extension is capable of stabilizing any set of allocations other than the unique stable set of the basic game.

The stable set solution concept has a dynamic interpretation but a static definition. Harsanyi (1974) observes that in general, the dynamic interpretation can fail to be supported if players' expectations are explicitly taken into account. As mentioned above, stable sets of the basic pillage game are not vulnerable to Harsanyi's critique, but the same cannot be said in general for stable sets of extended pillage games. Accordingly, Section 4 strengthens the concept of quasi-legitimacy by using the more demanding concept of a farsighted core under rational expectations. A set of allocations that can be stabilized in this fashion is called *legitimate*. Proposition 4.3 shows that quasi-legitimacy is a necessary condition for legitimacy. Theorem 4.5 shows that the citizenship game can be strengthened in such a way that the set of all allocations in which only citizens possess wealth can be stabilized as a farsighted core, assuming the no-tyranny condition. Unfortunately, quasi-legitimacy is not a sufficient condition for legitimacy. Proposition 4.6 shows that no legitimate set exists for the *Cobb-Douglas* pillage game, in which the power of each coalition is the product of its total wealth and the number of its members, despite the fact that this game admits a large quasi-legitimate set. Theorem 4.9 gives a general characterization of legitimate sets.

There is a large and growing theoretical literature on allocation by force, nearly all of which uses noncooperative games as models (e.g., Garfinkel and Skaperdas (1996)). In particular, Dixit, Grossman and Gul (2000) and Jack and Lagunoff (2006) develop repeated game Markov equilibrium models in which a player's current use of force is limited by the prospect of punishment by other players in the future. Articles that bear more directly on the present analysis are mentioned below as they become relevant.

2. Pillage games

This section defines pillage games and records some results from Jordan (2005) that will be used below. Three examples of pillage games are described at the end of this section.

2.1 Definitions: The set of players is the finite set $I = \{1, \ldots, n\}$, where $n \geq 2$. Subsets of I will be called *coalitions*. Let $\mathcal{C}$ denote the set of coalitions. The set of *allocations* is the set $A = \{w \in I\!R^I : w_i \geq 0 \text{ for each } i, \text{and } \sum_i w_i = 1\}$. A *power function* is a function $\pi : \mathcal{C} \times A \rightarrow I\!R$ satisfying

p.1) if $C \subset C'$ then $\pi(C', w) \geq \pi(C, w)$ for all w;

p.2) if $w'_i \geq w_i$ for all $i \in C$ then $\pi(C, w') \geq \pi(C, w)$; and

p.3) if $C \neq \emptyset$ and $w'_i > w_i$ for all $i \in C$ then $\pi(C, w') > \pi(C, w)$.

An allocation w' *dominates* an allocation w if

D) $$\pi(W, w) > \pi(L, w),$$

where $W = \{i : w'_i > w_i\}$ and $L = \{i : w'_i < w_i\}$. Domination is a binary relation on A and will be denoted $\succ$. Domination is asymmetric but typically not transitive or even acyclic.

The property (p.3) implies that power cannot be completely independent of wealth, so the power of each coalition is endogenous. This prevents a pillage game from having a characteristic function. However, the pair $(A, \succ)$ is a special case of an *abstract game*, so the concepts of core and stable set can be defined as in Lucas (1992).

2.2 Definitions: For any set of allocations E, let $U(E)$ denote the set of allocations undominated by E, that is, $U(E) = \{w \in A : \text{no } w' \in E \text{ dominates } w\}$. The *core* of a pillage game is the set of undominated allocations, that is, the core is $U(A)$. A set of allocations E is *internally stable* if no allocation in E is dominated by an allocation in E, that is,

IS) $E \subset U(E)$.

A set of allocations E is *externally stable* if every allocation not in E is dominated by some allocation in E, that is,

ES) $U(E) \subset E$.

A set of allocations E is *stable* if it is both internally and externally stable, that is,

S) $E = U(E)$.

The core always exists, but can be empty. Since $U(\emptyset) = A$, external stability (ES) implies that stable sets cannot be empty, but they can fail to exist. External stability also implies that a stable set must contain the core.

The core is the set of allocations that can be defended by force. That is, each player who holds wealth must be at least as powerful as the coalition of everyone else. The allocations most likely to satisfy this demanding requirement are those that give everything to one player.

2.3 Definition: For each i, let e^i denote the allocation that gives everything to player i, that is, $e^i_i = 1$. The e^i's are called *tyrannical allocations*.

The following result establishes that the core is nonempty only if it contains one or more of the tyrannical allocations.

2.4 Proposition: The core is the set

$$\Big\{w : \{i : w_i > 0\} = \{i : \pi(\{i\}, w) \geq \pi(I\backslash\{i\}, w)\}\Big\}.$$

For each i, the core contains e^i if and only if $\pi(\{i\}, e^i) \geq \pi(I\backslash\{i\}, e^i)$. In particular, the core is empty if and only if

NT) $\pi(\{i\}, e^i) < \pi(I\backslash\{i\}, e^i)$ for all i.

Condition (NT), which is equivalent to the emptiness of the core, will be called the *no tyranny* condition. The following proposition records the analytically useful fact that internally stable sets are finite. This contrasts with characteristic function games, whose stable sets typically contain a continuum of allocations (Lucas (1992)).

2.5 Proposition: An internally stable set can contain at most finitely many allocations.

We close this section with three examples of pillage games, which will be further discussed in subsequent sections. First, suppose the dependence of power on wealth is complete, that is, the power of each coalition is simply its total wealth. It is immediate that the core of this pillage game consists of the tyrannical allocations, together with the allocations that give half the wealth to each of two players. The unique stable set consists of all allocations in which each player with positive wealth has $w_i = (\frac{1}{2})^{k_i}$ for some nonnegative integer k_i. In particular, the stable set for the four-player game consists of all permutations of the allocations $(1, 0, 0, 0)$, $(\frac{1}{2}, \frac{1}{2}, 0, 0)$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$, and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$.

2.6 The wealth-is-power pillage game (Jordan (2005)): The *wealth-is-power* (WIP) game is specified by the power function

WIP) $\pi(C, w) = \Sigma_{i \in C} w_i$.

A number $0 \leq x \leq 1$ is *dyadic* if $x = 0$ or $x = 2^{-k}$ for some nonnegative integer k. An allocation w is *dyadic* if each w_i is dyadic. Let D denote the set of dyadic allocations. For each positive integer k, let $D_k = \{w : w \text{ is dyadic and for each } i, \text{ if } w_i > 0 \text{ then } w_i \geq 2^{-k}\}$. Then $D_k \subset D_{k+1}$ for each k, and $D = \cup_k D_k$. For the WIP game, D_1 is the core and D is the unique stable set.

At the opposite extreme from the WIP game is the traditional majority game, in which the power of each coalition is equal to the number of its members. Since power is independent of wealth in this case, the traditional majority game is not a pillage game. However, there is a closely related pillage game in which the dependence of power on wealth is lexicographically secondary to coalition size, that is, relative wealth only determines the relative power of two coalitions if they have the same size. This *majority pillage* game differs from the traditional majority game in a second important respect, namely that a coalition can commit pillage without having an absolute majority, as long as it is larger than the set of its victims.

2.7 The majority pillage game (Jordan and Obadia (2005)): The *majority pillage* game is specified by any power function of the form

M) $\pi(C, w) = v\#C + \Sigma_{i \in C} w_i$, where $v > 1$.

The core of the majority pillage game is empty. If n is odd, let S_n be the set of allocations consisting of all player-permutations of the allocation $(\frac{2}{n+1}, \ldots, \frac{2}{n+1}, 0, \ldots, 0)$. If $n = 3$, then S_3 is the unique stable set. More generally, if n is odd, then S_n is the unique symmetric stable set. If n is even and $n \geq 4$, then no symmetric stable set exists.

The final example is the game in which the power of each coalition is the product of its size and wealth. The principal interest of this game is that it has no stable set. This is proved in Jordan (2005), but it also follows from the nonexistence of a legitimate set for this game, which is proved in Section 4.

2.8 The Cobb-Douglas pillage game (Jordan (2005)): The *Cobb-Douglas* game is specified by the power function

CD) $\pi(C, w) = \#C \Sigma_{i \in C} w_i$.

For the Cobb-Douglas game with $n > 2$, the core is the set of tyrannical allocations, $\{e^i : i \in I\}$, but no stable set exists.

Jackson and Morelli (2005) develop a model of the willingness of countries to go to war for material gain. The decision whether or not to fight is made by a single pivotal decision-maker in each country. Jackson and Morelli are specifically interested in the influence on the likelihood of war of "political bias," which they model as an asymmetry in the decision-maker's shares of the gain in victory and the loss in defeat. Although the decision-makers are analogous to players in a pillage game, Jackson and Morelli model war as costly and uncertain, so countries can negotiate transfers to remove the incentive for war. Most of the analysis is confined to the two-country case, but they also consider the case in which multiple countries are divided into two alliances, and develop conditions that ensure stability in the sense that neither alliance would go to war against the other in the absence of transfers, and no individual country would prefer to leave its alliance to stand alone or join the other alliance (Proposition 7). Thus, by making pillage costly and uncertain and restricting the possibilities of forming pillaging coalitions, the domination relation can be narrowed to allow a larger core without the need for dynamic considerations.

It may be useful at this point to compare pillage games with a recent model of allocation by force due to Piccione and Rubinstein (2006). In their model, force is also costless and certain, but power is specified by an exogenous linear pecking-order among the players. In contrast to pillage games, there are multiple commodities, and players may have differing preferences, which can be satiated, in the sense that players have compact consumption sets. Allocation proceeds as follows: player 1, the most powerful player, chooses a most-preferred bundle from the attainable set, after which player 2 chooses from the remaining attainable set, and so on. The authors show that under appropriate strict convexity and smoothness conditions, the resulting allocation is unique and Pareto efficient. In pillage games, the question of Pareto efficiency does not arise, since all allocations are trivially Pareto efficient. However, the interest in stabilizing the largest possible set of allocations is motivated by the prospect of extending pillage games to settings which include the possibilities of production and exchange, in order to study the trade-offs between power and efficiency. Loosely speaking, a larger stable set may give greater scope to realize gains from voluntary production and trade.

3. Quasi-legitimacy

This section extends the basic pillage game by adding a set H of extrinsic social information that can distinguish between legitimate and illegitimate uses of power. An allocation w together with social information h constitutes a social state. An act of pillage at a social state (w, h) results in a dominating allocation w', together with new social information h'. The social information plays no role in preferences or the technology of force, so the chance from w to w' must conform to the basic domination relation $w' \succ w$. The set H and the transition from h to h' are interpreted as a social norm that characterizes a particular extension. A given pillage game can have many different extensions, including the trivial extension, which adds no information, or simply ignores the set H.

3.1 Social extensions: A *social extension* of a pillage game π consists of a set H of *social information*, and a *recording function* $\sigma : A \times H \times A \to H$. Define *social domination*, $\succ_s$, by

S) $(w', h') \succ_s (w, h)$ if $w' \succ w$ and $h' = \sigma(w, h; w')$.

A social extension will be denoted by the pair $(A \times H, \succ_s)$, and elements (w, h) of $A \times H$ will be called *states*.

The definition of a social extension embodies an implicit assumption that social information is public and the recording function σ is known to all players. The recording function is interpreted as a way of recording some information about the fact that, at the state (w, h), an act of pillage changed the allocation to w'.

A social extension $(A \times H, \succ_s)$ gives rise to stable sets of social states, that is, sets $S \subset A \times H$ that are internally and externally stable under the relation $\succ_s$. Moreover, different extensions may stabilize different sets of allocations.

3.2 Definition: Given a pillage game π, a set $X \subset A$ is *quasi-legitimate* if there is a social extension $(A \times H, \succ_s)$ of π with a stable set $S \subset A \times H$ with $X = \{w : (w, h) \in S \text{ for some } h \in H\}$.

The definition of a social extension places no limits on the size of H or the amount of information that can be recorded. However, the smallest nontrivial set, $H = \{0, 1\}$, suffices to stabilize every quasi-legitimate set of allocations. This is a consequence of both the mathematical elegance and conceptual weakness of the stable set concept.

3.3 The Boolean extension: Given a pillage game π and a set $E \subset A$, let $H = \{0, 1\}$ and define $\sigma : A \times H \times A \to H$ as follows:

$$\sigma(w, h; w') = \begin{cases} 0 & \text{if } w' \succ w,\ w' \in E,\ \text{and } w \notin E; \\ \neg h & \text{if } w' \succ w \text{ and either } w \in E \text{ or } w' \notin E; \\ h & \text{if } w' \not\succ w, \end{cases}$$

where $\neg h$ denotes the element of H not equal to h.

The Boolean extension is defined for general sets of allocations E, but is only of interest when E can be stabilized. This requires a condition called *self-protection*, which is a generalization of stability. Suppose an element $w \in E$ is dominated by an allocation w'. If an allocation $w'' \in E$ dominates w', then w'' can be very loosely interpreted as protecting w.

3.4 Definition: A set $E \subset A$ is *self-protected* if for each $w \in E$ and each $w' \in A$ such that $w' \succ w$, there exists some $w'' \in E$ such that $w'' \succ w'$. Equivalently, E is self-protected is $E \subset U^2(E)$.

Self-protection does not imply either internal or external stability, but is very useful when combined with either of them. The following proposition, which is an immediate consequence of the definitions, records the two properties that result.

3.5 Proposition: A set $E \subset A$ is externally stable and self-protected if and only if

ESSP) $$U(E) \subset E \subset U^2(E).$$

A set $E \subset A$ is internally stable and self-protected if and only if

ISSP) $$E \subset U^2(E) \subset U(E)$$

We can now show that quasi-legitimacy is completely characterized by external stability and self-protection. We first demonstrate the sufficiency of (ESSP) using the Boolean extension.

3.6 Proposition: Given a pillage game π, let $E \subset A$ and let $(A \times H, \succ_s)$ denote the Boolean extension. Define $S = \big(U(E) \times \{0, 1\}\big) \cup \big((E \backslash U(E) \times \{0\}\big)$. If E is externally stable and self-protected, then S is stable, and thus E is quasi-legitimate.

Proof: We first show that S is internally stable. Let $(w, h), (w', h') \in S$. Suppose by way of contradiction that $(w', h') \succ_s (w, h)$. Then $w' \succ w$ and $w, w' \in E$. Since E is self-protected, there is some $w'' \in E$ such that $w'' \succ w'$. Therefore $w' \notin U(E)$. Hence $w, w' \in E \backslash U(E)$, so $h = h' = 0$. But this contradicts the definition of $\succ_s$, which requires $h' = \neg h$.

We now show that S is externally stable. Let $(w,h) \notin S$. Then $w \notin U(E)$. Therefore there is some $w' \in E$ with $w' \succ w$. If $w \notin E$, then $(w',0) \succ_s (w,h)$. If $w \in E$, then $w \in E \backslash U(E)$, so $h=1$, which also implies that $(w',0) \succ_s (w,h)$. Since $(w',0) \in S$, this proves that S is externally stable.

We now complete the characterization of quasi-legitimacy by showing the necessity of (ESSP).

3.7 Theorem: A set $E \subset A$ is quasi-legitimate if and only if E is externally stable and self-protected.

Proof: Sufficiency is given by Proposition 3.6, so it remains to show necessity. Let $(A \times H, \succ_s)$ be a social extension with a stable set S satisfying $E = \{w : (w,h) \in S \text{ for some } h \in H\}$. We first show that E is externally stable. Let $w \notin E$ and let $h \in H$. Then $(w,h) \notin S$, so there is some $(w',h') \in S$ with $(w',h') \succ_s (w,h)$. Then $w' \in E$ and $w' \succ w$, so E is externally stable. To show that E is self-protected, let $w \in E$, $w' \in A$ with $w' \succ w$. Since $w \in E$, there is some $h \in H$ with $(w,h) \in S$. Since $w' \succ w$, there is some h' with $(w',h') \succ_s (w,h)$. Since S is internally stable, $(w',h') \notin S$. Since S is externally stable, there is some $(w'',h'') \in S$ satisfying $(w'',h'') \succ_s (w',h')$. Then $w'' \in E$ and $w'' \succ w'$, so E is self-protected.

Since stable sets necessarily satisfy (ESSP), the following corollary is immediate.

3.8 Corollary: Every stable set is quasi-legitimate.

A pillage game may have many quasi-legitimate sets, but there is always a unique largest quasi-legitimate set. Moreover, an inductive argument originally used by Roth (1976) for general abstract games (see also Asilis and Kahn (1992) for further discussion and applications) gives a procedure for constructing it. For general abstract games, the process can involve transfinite induction, but the fact that internally stable sets in pillage games are always finite (Proposition 2.5) implies that the process described below stops in a finite number of steps.

3.9 Proposition: Given a pillage game π, let $G_0 = \emptyset$ and let $E_0 = A$. For each integer $t > 0$, let $G_t = U^2(G_{t-1})$ and $E_t = U^2(E_{t-1})$. Then for each $t \geq 0$, $G_t \subset G_{t+1} \subset E_{t+1} \subset E_t$. Also, for each t, $E_t = U(G_t)$ and $G_{t+1} = U(E_t)$. Moreover, there exists $T > 0$ such that $G_t = G_T$ and $E_t = E_T$ for all $t > T$. In particular,

$$G_t \subset G_T \subset E_T \subset E_t \text{ for all } t.$$

Moreover, G_T is internally stable and self-protected, and E_T is externally stable and self-protected.

Proof: Note that for any sets $X, X' \subset A$, if $X \subset X'$ then $U(X') \subset U(X)$, so $U^2(X) \subset U^2(X')$. Since $G_0 = \emptyset$, $G_0 \subset U^2(G_0) = G_1$. Hence by iteration, $G_t \subset G_{t+1}$ for all t.

Since $G_0 = \emptyset$, G_0 is internally stable and self-protected. Since the (ISSP) inclusions are preserved by $U^2(\cdot)$, it follows that G_t is internally stable and self-protected for all t. Let $G^* = \cup_t G_t$. Since each G_t is internally stable and the sequence of sets is increasing, it follows that G^* is internally stable. By Proposition 2.5, all internally stable sets are finite, so G^* is finite, which implies that $G^* = G_T$ for some integer T.

Since $E_0 = A$, $E_1 = U^2(E_0) \subset E_0$. Since $U^2(\cdot)$ preserves inclusion, the $E_{t+1} \subset E_t$ for all t. Also, since $E_0 = A$, E_0 is externally stable, that is, $U(E_0) \subset E_0$. Since $U^2(\cdot)$ preserves this inclusion, E_t is externally stable for all t. Since $\emptyset = G_0 \subset E_0 = A$, applying the $U^2(\cdot)$ operator successively to this inclusion shows that $G_t \subset E_t$ for all t. Note that $E_0 = U(G_0)$. Hence $U(E_0) = U^2(G_0) = G_1$ and $E_1 = U^2(E_0) = U(G_1)$. Hence $E_t = U(G_t)$ for all t, so the required E_T is $E_T = U(G_T)$. Finally, since $E_T = E_{T+1} = U^2(E_T)$, E_T is self-protected.

3.10 Theorem: Given a pillage game π, let $E \subset A$ be quasi-legitimate. Then $G_T \subset E \subset E_T$. In particular, E_T is the largest quasi-legitimate set.

Proof: Since E is quasi-legitimate, Theorem 3.7 implies that E satisfies (ESSP). We will prove the result by induction on t. Suppose that for some t, $G_t \subset E \subset E_t$. Since E is self-protected, $E \subset U^2(E)$. Since $E \subset E_t$, $U^2(E) \subset U^2(E_t) = E_{t+1}$, so $E \subset E_{t+1}$. Since E is externally stable and $E \subset E_t$, $G_{t+1} = U(E_t) \subset E$. Thus, if $G_t \subset E \subset E_t$, then $G_{t+1} \subset E \subset E_{t+1}$. Since $G_0 = \emptyset$ and $E_0 = A$, $G_0 \subset E \subset E_0$, it follows by induction that $G_T \subset E \subset E_T$. The final assertion follows from the fact that E_T satisfies (ESSP).

The following useful corollary is immediate.

3.11 Corollary: If $G_T = E_T$, then E_T is the unique quasi-legitimate set and also the unique stable set.

3.12 Corollary: If π satisfies the no-tyranny condition (NT), then A is quasi-legitimate.

Proof: By Proposition 2.4, (NT) implies that the core, $U(A)$, is empty, so the iterative process terminates at the first step, with $E_T = E_0 = A$.

3.13 Definitions: For any set $E \subset A$, define $D(E) = \{w : w' \succ w \text{ for some } w' \in E\}$. Note that $D(E) = A \backslash U(E)$. Let $F_T = E_T \backslash G_T$, and let $B_T = A \backslash E_T$.

3.14 Proposition:

i) $B_T = D(G_T)$, and

ii) $F_T \subset D(F_T)$.

Proof: By Proposition 3.9, $E_T = U(G_T)$, so $B_T = A \backslash U(G_T)$, which proves (i). By Proposition 3.9, $G_T = U(E_T)$, so $F_T \subset D(E_T)$. Since $F_T \cap D(G_T) = \emptyset$ by (i), it follows that $F_T \subset D(F_T)$.

The decomposition $A = G_T \cup B_T \cup F_T$ is due to Asilis and Kahn (1992) for abstract games. They call G_T the "good set," B_T the "bad set," and F_T the "ugly set." The only originality in the present Propositions 3.9 and 3.14 is the application of Proposition 2.5 to conclude that the iterative construction is finite.

The iterative construction of E_T can be applied to the three specific pillage games defined in the preceding section. The majority pillage game with at least three players satisfies the no tyranny condition, so the entire set of allocations is quasi-legitimate. For the Cobb-Douglas pillage game, the iteration terminates after the first step, yielding a large quasi-legitimate set, despite the absence of a stable set for this game. The WIP game, in contrast, allows no room for quasi-legitimacy beyond the unique stable set.

3.15 Proposition:

a) For the WIP game $G_T = E_T = D$.

b) For the majority game with $n > 2$, $E_T = A$.

c) For the CD game,

$$E_T = \{e^i : i \in I\} \cup \{w : w_i \leq \frac{p(w) - 1}{p(w)} \text{ for all } i\},$$

where $p(w) = \#\{i : w_i > 0\}$.

Proof: The proof given in (Jordan (2005)) that D is the unique stable set for the WIP game uses an iterative construction that can be slightly modified to show (a). The majority pillage game with $n > 2$ is a special case of Corollary 3.12. For the CD game, if $n = 2$, it is immediate that $G_1 = E_1 = \{(1,0),(0,1),(\frac{1}{2},\frac{1}{2})\}$. If $n \geq 3$, the core is $G_1 = U(A) = \{e^i : i \in I\}$. An allocation w is undominated by e^i if and only if $w_i \leq \frac{p(w)-1}{p(w)}$, so $E_1 = U(G_1) = G_1 \cup \{w : w_i \leq \frac{p(w)-1}{p(w)}$ for all $i\}$. Now suppose that $n = 3$. Then $G_2 = U(E_1) = G_1 \cup \{w : w_i = \frac{2}{3}$ and $w_j = w_k = \frac{1}{6}$ for some $i, j, k\}$, and $E_2 = U(G_2) = E_1$, so $E_T = E_1$. If $n > 3$, then $G_2 = U(E_1) = G_1$, so again $E_T = E_1$.

The abstract simplicity of the Boolean extension is helpful in characterizing quasi-legitimacy, but yields little insight about how power can be constrained. The following more explicit, albeit less general, extension is much easier to interpret. Suppose that each player can be designated as either a *citizen* or an *outlaw*. A social state then consists of an allocation together with a given designation for each player. The player designations change with pillage in the following way. If any citizen is pillaged, all players who benefit from the pillage become outlaws. However, if only outlaws are pillaged, then all players become citizens. This social norm protects the property of citizens, provided that the no-tyranny condition is satisfied. Any player who participates in the pillage of a citizen becomes an outlaw, and is thus left open to pillage by anyone. Of course, even under the no tyranny condition, it may not be possible for the remaining citizens to pillage all of the outlaws. For example, in the three-player majority pillage game, suppose that player 1 is a citizen and is pillaged by the coalition of the other two players. Then player 1 will need to enlist the aid of one of the two outlaws, say player 2, to get some of his property back. In this case, player 2 benefits a second time by betraying his fellow outlaw, but this prospect discourages player 3 from participating in the original pillage.

The citizenship extension, like the Boolean extension, makes the entire set of allocations quasi-legitimate under the no-tyranny condition. However, unlike the Boolean extension, the citizenship extension is easily modified to make the entire set of allocations legitimate (Theorem 4.5).

3.16 The citizenship extension: Let $H = \mathcal{C}$ and define $\sigma : A \times H \times A \to H$ as follows:

$$\sigma(w, C; w') = \begin{cases} I \backslash W & \text{if } L \cap C \neq \emptyset; \\ I & \text{if } L \neq \emptyset \text{ and } L \cap C = \emptyset; \\ C & \text{if } L = \emptyset, \end{cases}$$

where $W = \{i : w'_i > w_i\}$ and $L = \{i : w'_i < w_i\}$.

3.17 Proposition: Let $(A \times \mathcal{C}, \succ_s)$ be the citizenship extension of a pillage game π. Let $S = \big\{(w, C) : \{i : w_i > 0\} \subset C\big\}$. The set S is a stable set if and only if π satisfies the no tyranny condition (NT).

Proof: First assume that π has an empty core. Let $(w, C) \in S$ and let (w', C') be a social state that dominates (w, C). Let $W = \{i : w'_i > w_i\}$ and $L = \{i : w'_i < w_i\}$. Since $(w, C) \in S$, $L \subset C$ so $C' \cap W = \emptyset$. Then for each $i \in W$, $w'_i > 0$ and $i \notin C'$, so $(w', C') \notin S$. Thus S satisfies internal stability. Let $(w, C) \notin S$. Then $w_i > 0$ for some $i \notin C$. Let $C' = I \backslash \{i\}$ and let w' satisfy $w'_i = 0$ and $w'_j > w_j$ for each $j \neq i$. Then $(w', C') \in S$. By (NT), $\pi(C', e^i) > \pi(\{i\}, e^i)$, so $w' \succ w$. Hence $(w', C') \succ_s (w, C)$, so S satisfies external stability.

Now suppose that π does not satisfy (NT), so $\pi(I \backslash \{i\}, e^i) \leq \pi(\{i\}, e^i)$ for some i. Then e^i is undominated, so the social state $(e^i, I \backslash \{i\})$ is undominated. Since $(e^i, I \backslash \{i\}) \notin S$, it follows that S does not satisfy external stability.

4. Legitimacy

Quasi-legitimacy is analytically convenient but conceptually inadequate. Quasi-legitimate sets have a simple and complete characterization, and there is a finite iterative procedure that generates the largest quasi-legitimate set. Unfortunately a stable set of a social extension, unlike a stable set of the basic pillage game, is vulnerable to Harsanyi's criticique of the stable set concept (Harsanyi (1974)). To put Harsanyi's criticique in the present context, let $(A \times H, \succ_s)$ be a social extension, and let $S \subset A \times H$ be a stable set. The set S is interpreted as self-enforcing based in part on the following argument. If $(w, h) \in S$ is dominated by some (w', h'), then since S is internally stable, $(w', h') \notin S$. Since S is externally stable, there is some $(w'', h'') \in S$ that dominates (w', h'). Since S is internally stable (w'', h'') does not dominate (w, h), so the prospect of moving to (w'', h'') discourages the move to (w', h'). Harsanyi observed that if the players who benefit from the move to (w', h') benefit still further, or at least don't lose anything in the subsequent move to (w'', h''), they will happily force the move to (w', h') in order to achieve in two moves what internal stability prevents them from achieving in one. Of course, the natural way to resolve this confusion is simply to make the expectation of the subsequent move from (w', h') to (w'', h'') explicit for all players.

Harsanyi's critique of the stable set concept stimulated the development of solution concepts based on indirect dominance, including notable contributions by Greenberg (1990), Chwe (1994) and Xue (1998). The work mostly closely related to the present analysis is that of Konishi and Ray (2003). They model coalition formation as an explicit dynamic process, in which coalition members form expectations over discounted payoff streams. As in pillage games, a coalition can form to make a change in the current state, but cannot make commitments beyond that transition. Whether or not a coalition forms and what action they take are thus explicitly dependent on the expectations of coalition members about what subsequent coalitions will form and what actions they will take. However, the expectations of players outside the coalition are irrelevant, since the set of actions that are feasible for a given coalition is independent of the expectations of outsiders. Konishi and Ray formulate their model for abstract games, in which this assumption is conventional. In pillage games, in contrast, the expectations of outside players are crucial.

For example, consider the three-person wealth-is-power game, at the stable allocation $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. Player 1 has sufficient power to pillage player 2 to obtain the non stable allocation $(\frac{3}{4}, 0, \frac{1}{4})$, as long as player 3 remains neutral, which is assumed in the definition of the domination relation $\succ$. However, if players expect that the initial pillage will be followed by the pillage of player 3, resulting in the allocation $(1, 0, 0)$, then player 3 will join with player 2 to successfully oppose the initial pillage, stabilizing the allocation $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$. More generally, when a pillage game is extended to incorporate players' expectations of future pillages, the neutrality assumption must be recast in terms of the expected final allocation.

For the basic pillage game, in which social information is absent, Jordan (2005, Section 6) shows that stable sets not only survive the Harsanyi critique but can be characterized as the only sets that do. Unfortunately, the stable sets of social extensions fare much worse. The vulnerability of the Boolean extension to the Harsanyi critique is pervasive and in some cases irreparable. The largest quasi-legitimate set for the Cobb-Douglas pillage game is very large, as described by Proposition 3.15, but Proposition 4.6 below shows that the Harsanyi critique prevents the existence of any legitimate set.

The citizenship game, as described in the preceding section, has an interesting deficiency that is subject to the Harsanyi critique as well. The recording function σ uses the allocations w and w' to identify the pillaging coalition and its victims. If actions are motivated by the anticipation of subsequent actions, the allocation w' is only a step on the way to w''. The definition of σ allows farsighted players to force an action w' with the object of diverting punishment from themselves and enabling themselves to benefit from the punishment of players they have caused to be wrongly identified as transgressors. For example, consider the three-player majority pillage game at the stable social state $(w, C) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}; \{1, 2, 3\})$. A pillage resulting in the allocation $(\frac{5}{12}, \frac{1}{6}, \frac{5}{12})$ will identify players 1 and 3 as outlaws for having pillaged player 2, resulting in the social state $(\frac{5}{12}, \frac{1}{6}, \frac{5}{12}; \{2\})$. The ensuing punishment will be effected by player 2 colluding with one of the outlaws to betray the other outlaw. For example, suppose the expected punishment involves players 1 and 2 pillaging player 3 to produce the social state $(\frac{1}{2}, \frac{1}{2}, 0; \{1, 2, 3\})$. This expectation could be manipulated by players 1 and 2. At the initial allocation $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, players 1 and 2 could force player 3 to accept the allocation $(\frac{5}{12}, \frac{1}{6}, \frac{5}{12})$, leading to the eventual allocation $(\frac{1}{2}, \frac{1}{2}, 0)$. This sort of manipulation could be excluded by assuming that no player can be forced to accept more wealth, even temporarily, but that would involve an additional implicit assumption about the technology of power. Instead, we will rely on the assumption of complete information to allow the recording function to correctly identify the coalition that forced an action and the players who unsuccessfully resisted it.

An *expectation*, which is assumed to be the same for all players, is represented as a function $f : A \times H \to A \times H$. We require that $f^2 = f$, so that the step $(w, h) \mapsto f(w, h)$ is final. At a state (w, h), suppose that a coalition is considering forcing a move to an allocation w'. The recording function will generate a new history h', resulting in the new social state (w', h'). Under the expectation f, this may result in a further move to $(w'', h'') = f(w', h')$. Thus all players will evaluate the move from (w, h) to (w', h') by comparing w to w'' rather than w'.

4.1 Definitions: A *dynamic extension* of a pillage game π consists of a set H and a recording function $\delta : A \times H \times \mathcal{C}^2 \times A \to H$. An *expectation* is a function $f : A \times H \to A \times H$ satisfying $f^2 = f$. Since $f^2 = f$, the set $f(A \times H)$ is the set of *stationary states*. Given an expectation f and states (w, h) and (w', h'), let $(w'', h'') = f(w', h')$. The state (w', h') *dominates* (w, h) *in expectation* if there exist coalitions F (the *forcing coalition*) and R (the *resisting coalition*) satisfying

i) $\delta(w, h, F, R, w') = h'$;

ii) $F \cap R = \emptyset$;

iii) $\pi(F, w) > \pi(R, w)$; and

iv) $F \subset \{i : w''_i > w_i\}$ and $\{i : w''_i < w_i\} \subset R$.

Note that if 4.1(ii-iv) are satisfied by F and R, they will remain satisfied if F is replaced by $\{i : w''_i > w_i\}$ and R is replaced by $\{i : w''_i < w_i\}$.

The concept of domination in expectation is only of interest if the expectations are fulfilled. The following definition of *consistent expectations* provides the requisite rational expectation condition.

4.2 Definitions: An expectation f is *consistent* if for each (w, h), either

i) $f(w, h)$ dominates (w, h) in expectation; or

ii) (w, h) is undominated in expectation and $f(w, h) = (w, \delta(w, h, \emptyset, \emptyset, w))$.

Given a consistent expectation f, the *farsighted core* is the set of states $K_f \subset A \times H$ that are undominated in expectation. If K_f is the farsighted core for a consistent expectation f, the set $\{w : (w, h) \in K_f$ for some $h \in H\}$ is said to be *legitimate*. Note that since $f^2 = f$, (i-ii) imply that $f(A \times H) \subset K_f$.

The following result shows that quasi-legitimacy is a necessary condition for legitimacy.

4.3 Proposition: Every legitimate set is also quasi-legitimate.

Proof: Let $E \subset A$ be legitimate, and let K_f be the farsighted core for a consistent expectation f with $E = \{w : (w,h) \in K_f \text{ for some } h \in H\}$. By Theorem 3.7, it suffices to show that E is externally stable and self-protected. Let $w \notin E$ and let $h \in H$. Then $(w,h) \notin K_f$. Then by 4.2(i), $f(w,h) = (w',h')$ for some (w',h') that dominates (w,h) in expectation. Since $f^2 = f$, $(w',h') \in K_f$, so $w' \in E$ and w' dominates w, which proves that E is externally stable. Now let $w \in E$ and let w' dominate w. If $w' \notin E$ then since E is externally stable, there is some $w'' \in E$ that dominates w'. Suppose $w' \in E$. Let $F = \{i : w'_i > w_i\}$, $R = \{i : w'_i < w_i\}$, and $h \in H$ with $(w,h) \in K_f$. Let $h' = \delta(w,h,F,R,w')$. Then $(w',h') \notin K_f$, otherwise (w',h') would dominate (w,h) in expectation. Therefore, as above, there is some $w'' \in E$ that dominates w'.

The following definition adapts the recording function of the citizenship game as required for a dynamic extension. The resulting version of the citizenship game is then shown to stabilize the full set of allocations as a farsighted core, provided, as before, that the no tyranny condition is satisfied.

4.4 Definition: The *dynamic citizenship extension* is defined as follows. Let $H = \mathcal{C}$ and define $\delta : A \times H \times \mathcal{C}^2 \times A \to H$ by

$$\delta(w,C,F,R,w') = \begin{cases} I \backslash F & \text{if } R \cap C \neq \emptyset; \\ I & \text{if } R \neq \emptyset \text{ and } R \cap C = \emptyset; \\ C & \text{if } R = \emptyset. \end{cases}$$

As in Proposition 3.17, let $S = \big\{(w,C) : \{i : w_i > 0\} \subset C\big\}$.

4.5 Theorem: Let π be a pillage game that satisfies the no tyranny condition (NT). For the dynamic citizenship extension, define the expectation $f : A \times H \to A \times H$ by

$$f(w,C) = \begin{cases} (w,C) & \text{if } (w,C) \in S; \\ (w',I) & \text{for some } w' \text{ that dominates } w \text{ and satisfies} \\ & \{i : w'_i < w_i\} \subset I \backslash C \text{ and } w'_i = 0 \text{ for some } i \in I \backslash C, \\ & \text{if } (w,C) \notin S. \end{cases}$$

Then f is consistent and S is the farsighted core. In particular, the entire set A is legitimate.

Proof: We first show that f is a well-defined expectation. By the definition of S, $(w,I) \in S$ for all w, so $f^2 = f$. Second, let $(w,C) \notin S$. We need to show that the allocation w' required in the definition of f exists. Since $(w,C) \notin S$, there is some i^o with $w_{i^o} > 0$ and $i^o \notin C$. Let $w' = (w_1 + \frac{w_{i^o}}{n-1}, \ldots, w_{i^o-1} + \frac{w_{i^o}}{n-1}, 0, w_{i^o+1} + \frac{w_{i^o}}{n-1}, \ldots, w_n + \frac{w_{i^o}}{n-1})$. Since π satisfies (NT), w' dominates w, and w' clearly satisfies the other required properties as well. Therefore f is a well-defined expectation.

We now show that states in S are undominated in expectation. Let $(w, C) \in S$ and $(w', C') \in A \times \mathcal{C}$. First suppose that $(w', C') \in S$. Then (w', C') cannot dominate (w, C) in expectation, since (w', C') is stationary, and since $(w, C) \in S$, if w' dominates w then $\{i : w'_i < w_i\} \subset C$, so by the definition of δ, $(w', C') \notin S$. Now suppose $(w', C') \notin S$ and suppose by way of contradiction that (w', C') dominates (w, C) in expectation. Since $(w', C') \notin S$, $f(w', C')$ dominates (w', C') in expectation and $f(w', C') \in S$. Let $(w'', C'') = f(w', C')$. Since (w', C') dominates (w, C) in expectation, $C' = \delta(w, C, F, R, w')$,where $F \subset \{i : w''_i > w_i\}$ and $\{i : w''_i < w_i\} \subset R$. Since $(w, C) \in S$, $R \cap C \neq \emptyset$, so by the definition of δ, $I \backslash F = C'$. Since $\pi(F, w) > \pi(R, w)$, $F \neq \emptyset$. Then , by the definition of f, there is some $i \in F$ with $w''_i = 0$, which contradicts the requirement that $F \subset \{i : w''_i > w_i\}$, and proves that all states in S are undominated in expectation.

Let $(w, C) \notin S$. We will complete the proof by showing that $f(w, C)$ dominates (w, C) in expectation. Let $(w', I) = f(w, C)$. Then (w', I) is stationary and w' dominates w. Let $F = \{i : w'_i > w_i\}$ and $R = \{i : w'_i < w_i\}$. Then $I = \delta(w, C, F, R, w')$, so (w', I) dominates (w, C) in expectation.

If a pillage game π fails to satisfy the no tyranny condition, then the expectation f in Theorem 4.5 is not well-defined. In particular, if $\pi(\{i\}, e^i) \geq \pi(I \backslash \{i\}, e^i)$, then the required allocation w' in the definition of $f(e^i, \emptyset)$ does not exist. Moreover, for any consistent expectation, the social state $(e^i, \emptyset)$ is undominated in expectation, so S cannot be a farsighted core.

The following result shows that the Cobb-Douglas pillage game with more than two players has no legitimate sets. This result also implies that quasi-legitimacy is not sufficient for legitimacy.

4.6 Proposition: For the CD game with $n > 2$, no legitimate set exists.

Proof: We will show there does not exist a consistent expectation for the Cobb-Douglas game with more than two players. Let $w \in E_T$ satisfy

$$(*) \qquad w_1 \geq w_2 \geq \cdots \geq w_n > 0,\ w_1 = \frac{n-1}{n}, \text{and } \frac{1}{n} > w_2 > \frac{n-2}{n(n-1)}.$$

Then $\pi(\{1\}, w) = \pi(\{2, \ldots, n\}, w)$ and $\pi(\{2\}, w) > \pi(\{3, \ldots, n\}, w)$. Let $w' \in E_T$ also satisfy $(*)$ with $w'_2 > w_2$ and $\Sigma_{i>2} w'_i < w_n$. Let $F = \{2\}$ and $R = \{3, \ldots, n\}$. Suppose by way of contradiction that f is a consistent expectation for some dynamic extension. Let $h \in H$, and let $h' = \delta(w, h, F, R, w')$. We will first show that (w', h') dominates (w, h) in expectation. If $f(w', h') = (w', h')$, this is immediate, since $w' \succ w$ and $h' = \delta(w, h, F, R, w')$. Suppose instead that $f(w', h') = (w'', h'') \neq (w', h')$. Then since $f^2 = f$, $f(w'', h'') = (w'', h'')$, so (w'', h'') is an element of the farsighted core and $w'' \succ w'$. By Proposition 4.3 and Theorem 3.10, $w'' \in E_T$. Then Proposition 3.15, together with the fact that $w'' \succ w'$, implies that w'' also satisfies $(*)$, with $w''_2 \geq w'_2 > w_2$ and $w''_i < \Sigma_{j>2} w'_j < w_i$ for all $i > 2$. Since $h' = \delta(w, h, F, R, w')$, this again shows that (w', h')

dominates (w, h) in expectation. If $(w', h') = f(w', h')$ then since f is consistent, (w', h') is undominated in expectation. However, since w' satisfies $(*)$, the above argument can be repeated to construct a state (w''', h'') that dominates (w', h') in expectation. Therefore $(w'', h'') = f(w', h') \neq (w', h')$ and (w'', h'') is in the farsighted core. However, w'' also satisfies $(*)$, so (w'', h'') is also dominated in expectation, which proves that f is not a consistent expectation.

A stable set of a pillage game is shown to be a farsighted core in Jordan (2005, Section 6). The trivial dynamic extension, $H = \{0\}$, extends this result to show that every stable set of a pillage game is legitimate. For this reason, the nonexistence of a legitimate set in the Cobb-Douglas game also implies the nonexistence of a stable set.

The following definition strengthens the ESSP condition to the concept of *dynamic protection*, which will be shown to characterize legitimate sets.

4.7 Definition: A set $E \subset A$ is *internally dynamically protected* (IDP) if for each $w, w' \in E$ and any coalitions F, R satisfying

i) $F \cap R = \emptyset$;

ii) $\pi(F, w) > \pi(R, w)$; and

iii) $F \subset \{i : w'_i > w_i\}$ and $\{i : w'_i < w_i\} \subset R$,

there exists $w'' \in E$ satisfying

iv) $w'' \succ w'$; and

v) either $F \not\subset \{i : w''_i > w_i\}$ or $\{i : w''_i < w_i\} \not\subset R$.

Note that (i-iii) imply that $w' \succ w$. A set $E \subset A$ is *externally dynamically protected* (EDP) if E is externally stable and for each $w \in E$, $w' \notin E$ and any coalitions F, R satisfying (i) and (ii), there exists $w'' \in E$ satisfying (iv) and (v).

The reason for splitting dynamic protection into IDP and EDP is that the latter property is automatically satisfied by the largest quasi-legitimate set, as is shown below.

4.8 Proposition: For any pillage game π, the set E_T is externally dynamically protected.

Proof: Given a pillage game π, let $w \in E_T$ and let $w' \notin E_T$, that is, $w' \in B_T$. By Proposition 3.14(i), there is some $w'' \in G_T \subset E_T$ satisfying $w'' \succ w'$. In particular, E_T is externally stable. By Proposition 3.9, $E_T = U(G_T)$, so $w'' \not\succ w$. Therefore all coalitions F and R satisfying 4.7(i) and 4.7(ii) must also satisfy 4.7(v).

Dynamic protection supports legitimacy by enabling the punishment of illegitimate pillages. Internal dynamic protection enables the construction of punishments for illegitimate pillages within the set E. Let $w \in E$ and let (w, h) be a "legitimate" social state. Suppose that a coalition is contemplating a pillage to an allocation $w' \in E$. The recording function will generate a new history h' that identifies the pillage as illegitimate. The new history h' will also record the original allocation w, the coalition F that forced the move to w' and the coalition R that unsuccessfully resisted it. IDP allows an appropriate punishment to be constructed as follows. First suppose $F \not\subset \{i : w'_i > w_i\}$. This means that some players in F were looking ahead to a subsequent allocation w''. These players can be punished, and the pillage discouraged, by stabilizing the allocation w' as legitimate. In this case $f(w', h') = (w', h'')$ where h'' marks the social state (w', h'') as legitimate and thus stationary. On the other hand, if $\{i : w'_i < w_i\} \not\subset R$, then some players chose not to resist the move to w' because they expected a further move to an allocation w''. Again, such players can be punished, and the pillage discouraged, by legitimizing the allocation w'. Finally, suppose that $F \subset \{i : w'_i > w_i\}$ and $\{i : w'_i < w_i\} \subset R$. Then the new history h' will mark the allocation w' as illegitimate, leading to a subsequent move to an allocation $w'' \in E$ with $w'' \succ w'$ and either $F \not\subset \{i : w''_i > w_i\}$ or $\{i : w''_i < w_i\} \not\subset R$, the existence of which is ensured by IDP. If the allocation $w' \notin E$, then punishment by legitimizing w' is not an option, so a punishing move to an allocation $w'' \in E$ is required. In this case, the existence of the requisite w'' is ensured by EDP.

4.9 Theorem: For any pillage game π, a set of allocations is legitimate if and only if it is dynamically protected.

Proof: We first prove sufficiency. Given a pillage game π, suppose that $E \subset A$ is dynamically protected. Define a dynamic extension of π as follows. Let $H = \{(w^o, F^o, R^o, \ell) \in A \times \mathcal{C}^2 \times \{0, 1\} : \text{either } F^o = R^o = \emptyset \text{ and } \ell = 0, \text{ or } F^o \cap R^o = \emptyset \text{ and } \pi(F^o, w^o) > \pi(R^o, w^o)\}$. In defining $\delta : A \times H \times \mathcal{C}^2 \times A \to H$, it suffices to consider only points (w, h, F, R, w') satisfying either $F = R = \emptyset$ and $w' = w$, or $F \cap R = \emptyset$, $\pi(F, w) > \pi(R, w)$ and $w' \neq w$, since these are the only two cases that will arise. Let

$$\delta(w; w^o, F^o, R^o, \ell; F, R, w') = \begin{cases} (w, F, R, 1) & \text{if } w' \neq w,\ w \in E,\ \text{and } \ell = 0; \\ (w, F, R, 1) & \text{if } w' \neq w,\ w, w' \in E,\ \ell = 1,\ \text{and either} \\ & F^o \not\subset \{i : w_i > w^o_i\} \ \text{ or } \ \{i : w_i < w^o_i\} \not\subset R^o; \\ (w', \emptyset, \emptyset, 0) & \text{otherwise.} \end{cases}$$

Now define an expectation f as follows. Let $(w, h) = (w; w^o, F^o, R^o, \ell) \in A \times H$. First suppose that $w \notin E$. If $w^o \in E$, $F^o \cap R^o = \emptyset$, and $\pi(F^o, w^o) > \pi(R^o, w^o)$, then since E is EDP, there exists $w' \in E$ satisfying $w' \succ w$ and either $F^o \not\subset \{i : w'_i > w^o_i\}$ or $\{i : w'_i < w^o_i\} \not\subset R^o$. Let $f(w, h) = (w'; w', \emptyset, \emptyset, 0)$. Otherwise, let $f(w, h) = (w'; w', \emptyset, \emptyset, 0)$ for any $w' \in E$, with $w' \succ w$. Such a w' exists since $w \notin E$ and E is EDP, and therefore externally stable.

Now suppose that $w \in E$. If $w = w^o$ or $\ell = 0$, let $f(w, h) = f(w; w, \emptyset, \emptyset, 0)$. Suppose that $w \neq w^o$, $\ell = 1$, $F^o \cap R^o = \emptyset$ and $\pi(F^o, w^o) > \pi(R^o, w^o)$. If either $F^o \not\subset \{i : w_i > w^o_i\}$ or $\{i : w_i < w^o_i\} \not\subset R^o$, then let $f(w, h) = f(w; w, \emptyset, \emptyset, 0)$. Otherwise, there exists $w' \in E$ satisfying $w' \succ w$ and either $F^o \not\subset \{i : w'_i > w^o_i\}$ or $\{i : w'_i < w^o_i\} \not\subset R^o$. The existence of w' follows from IDP if $w^o \in E$ and EDP otherwise. In this case, let $f(w, h) = (w'; w', \emptyset, \emptyset, 0)$.

To see that $f^2 = f$, observe that for every (w, h), $f^2(w, h) = f(w'; w', \emptyset, \emptyset, 0) = (w'; w', \emptyset, \emptyset, 0)$ for some allocation w'. Therefore f is a well-defined expectation.

To show that f is consistent, for any (w, h) let $(w', h') = f(w, h)$. In each case that $w' \neq w$, the definition of f ensures that $w' \succ w$ and $h' = \delta(w, h, F, R, w')$, where

$F = \{i : w'_i > w_i\}$ and $R = \{i : w'_i < w_i\}$. Since $f(w', h') = f^2(w, h) = (h', w')$, it follows that (w', h') dominates (w, h) in expectation. Now suppose $f(w, h) = (w; w, \emptyset, \emptyset, 0)$. We need to show that (w, h) is undominated in expectation. Suppose by way of contradiction that there is some (w', h') that dominates (w, h) in expectation. Let $(w'', h'') = f(w', h')$ and let F and R be the forcing and resisting coalitions, respectively, given by Definition 4.1. Since $(w, \delta(w, h, \emptyset, \emptyset, w)) = f(w, h)$, it follows that $w \in E$. If $w' \notin E$ then by the definition of f, $w'' \in E$, and either $F \not\subset \{i : w''_i > w\}$ or $\{i : w''_i < w_i\} \not\subset R$ in contradiction to 4.1(iv). Hence $w' \in E$. If $w' = w$, then $h' = \delta(w, h, F, R, w) = (w, \emptyset, \emptyset, 0)$, so $f(w', h') = (w, h') = (w'', h'')$, so $w'' \not\succ w$, contradicting 4.1(iii,iv). Hence $w' \neq w$, and $w, w' \in E$. First suppose that $\ell = 0$. Then $h' = \delta(w, h, F, R, w') = (w, F, R, 1)$. If either $F \not\subset \{i : w'_i > w_i\}$ or $\{i : w'_i < w_i\} \not\subset$, then $(w'', h'') = f(w', h') = (w', (w', \emptyset, \emptyset, 0))$, which contradicts 4.1(iv). On the other hand, if $F \subset \{i : w'_i > w_i\}$ and $\{i : w'_i < w_i\} \subset R$, then, by the above definition of f, either $F \not\subset \{i : w''_i > w_i\}$ or $\{i : w''_i < w_i\} \not\subset R$, which again contradicts 4.1(iv). Hence $\ell = 1$. Since $f(w, h) = (w; w, \emptyset, \emptyset, 0)$, it follows from the definition of f that either $F^o \not\subset \{i : w_i > w^o_i\}$ or $\{i : w_i < w^o_i\} \not\subset R^o$ (this includes the case $w = w^o$, since $\ell = 1$ implies that $F \neq \emptyset$). Since $w' \neq w$, the definition of δ implies that $h' = (w, F, R, 1)$. Then repeating the argument for the case $\ell = 0$ again contradicts 4.1. This proves that f is a consistent expectation.

It remains to show that $E = \{w : (w, h) \in K_f \text{ for some } h\}$. Since $K_f = f(A \times H) = \{(w, h) : w \in E \text{ and } h = (w, \emptyset, \emptyset, 0)\}$, the result follows. This proves sufficiency.

To prove necessity, let $E \subset A$ and let π be a pillage game for which E is legitimate. Let (H, δ) be a dynamic extension and let f be a consistent expectation satisfying $E = \{w : (w, h) \in K_f\}$. First let $w \notin E$ and let $h \in H$. Then $(w, h) \notin K_f$. Let $(w', h') = f(w, h)$. Then $(w', h') \in K_f$, so $w' \in E$ and (w', h') dominates (w, h) in expectations, so $w' \succ w$, which implies that E is externally stable. Let $w \in E$ and let $h \in H$ with $(w, h) \in K_f$. Let $w' \in A$, let coalitions F and R satisfy 4.7(i,ii), and let $h' = \delta(w, h, F, R, w')$. Let $(w'', h'') = f(w', h')$. Since $(w, h) \in K_f$, (w', h') does not dominate (w, h) in expectation. First suppose that $w' \in E$ and that F and R satisfy 4.7(iii). Then $w' \succ w$. Since (w', h') does not dominate (w, h) in expectation, $w'' \neq w'$ and 4.7(v) is satisfied. Since f is consistent and $w'' \neq w'$, $w'' \succ w'$, so 4.7(iv) is satisfied. This proves E is IDP. Finally, suppose that $w' \notin E$. Since $(w'', h'') \in f(A \times H) = K_f$, it follows that $w'' \in E$ and $w'' \neq w'$. Since f is consistent $w'' \succ w'$, so 4.7(iv) is satisfied, and since (w', h') does not dominate (w, h) in expectation, 4.7(v) is satisfied. This proves E is EDP.

The following corollary is an immediate consequence of the Theorem and Proposition 4.8.

4.10 Corollary: For any pillage game π, the set E_T is legitimate if and only if it is internally dynamically protected.

References

Asilis, C. and C. Kahn (1992), "Semi-stability in game theory: A survey of ugliness," in *Game theory and economic applications*, B. Dutta et al., (eds), Springer-Verlag, Berlin.

Chwe, M. (1994), "Farsighted coalitional stability," *Journal of economic theory* **63**, 299-325.

Dixit, A., G. Grossman and F. Gul (2000), "The dynamics of political compromise," *Journal of political economy* **108**, 531-568.

Garfinkel,M. and S. Skaperdas (Eds.) (1996), *The political economy of conflict and appropriation*, Cambridge University Press, Cambridge, 1996.

Greenberg, J. (1990), *The theory of social situations*, Cambridge University Press, Cambridge, 1990.

Harsanyi, J. (1974), "An equilibrium-point interpretation of stable sets and an alternative definition," *Management science* **20**, 1472-1495.

Jack, W. and R. Lagunoff (2006), "Dynamic Enfranchisement," *Journal of public economics* **90**, 551-572.

Jackson, M., and M. Morelli (2005), "War, transfers and political bias," mimeo.

Jordan, J. (2005), Pillage and property, *Journal of economic theory*, forthcoming.

Jordan, J. and D. Obadia (2005), "Stable sets in majority pillage games,"mimeo.

Konishi, H. and D. Ray (2003), "Coalition formation as a dynamic process," *Journal of Economic Theory* **110**, 1-41.

Lucas, W. (1992), "Von Neumann-Morgenstern stable sets," in *Handbook of game theory*, Aumann, R. and S. Hart (eds), Elsevier Science Publishers B.V., 543-590.

Piccione, M. and A. Rubinstein (2006), "Equilibrium in the jungle," mimeo.

Roth, A. (1976), "Subsolutions and the supercore of cooperative games," *Mathematics of operations research*, **1**, 43-49.

Xue, L. (1998),"Coalitional stability under perfect foresight," *Economic theory*, **11**, 603-627.