Governance as a strategy in state-of-nature games

Jason Briggeman

George Mason University

This paper will be available at www.springerlink.com
once it has been published in *Public Choice.*

Abstract

Public choice scholars routinely claim that coercion can be used to solve the social dilemma. However, while social contract theorists have frequently described state-of-nature societies using game theory, they have not used game theory to show how coercive action within such societies can improve outcomes. Here I operationalize the concepts of coercion and governance within a Prisoners' Dilemma (PD) framework; governance is operationalized as coercion to compliance, and to adopt a coercive strategy is to impose a strategy choice upon another player. I show that, under certain conditions, adding governance strategies to a noncoercive one-shot PD game can improve outcomes.

JEL classification: C72, D7, H1, H41

Keywords: Governance; coercion; Prisoners' Dilemma; game theory; Thomas Hobbes; state of nature; anarchy; public choice

1. Introduction

Public choice scholars have long claimed that coercion can be used to solve the social dilemma. In particular, the social contract theory of government holds that governments are instituted by society to force itself out of an anarchic state of nature.[1] However, while social contract theorists have frequently illustrated a state-of-nature society using the tools of game theory, those tools have not been used to show how coercive action taken by individuals within that society could lead to a desirable result; instead, the *deus ex machina* of an "external" government is often called upon to remake the game by directly altering payoffs for the "internal" players.[2] This would be a sensible approach if we believed that real-world governments could alter the utilities of actors in society directly, but even Thomas Hobbes himself was never so optimistic about the capabilities of his "Leviathan"; Hobbes readily acknowledged that any absolute monarch would be but a man, chosen from among the men within the society itself.[3] To model this Hobbesian situation with the tools of game theory, it would seem to be required that society's governor be an internal *player* who chooses *strategies*, not an external model-builder who sets payoffs. To that end, in this paper I operationalize the concepts of coercion and governance within a game-theoretic context, specifically that of the famous Prisoners' Dilemma (PD) game.

Much previous research has considered the potential for sustained cooperation under anarchy, often using a noncoercive repeated PD game (*e.g.*, Greif 1993). However, there is no standard formal benchmark of performance under government against which the performance of these model anarchies can be measured; in lieu of such a benchmark, governed societies are generally presumed to achieve perpetual universal cooperation while incurring some unspecified-but-low

cost of governance.[4] Here I consider the conditions under which adding governance strategies to a noncoercive one-shot PD game can improve social outcomes, developing a game called the "Prisoners' Dilemma with Coercion" in an effort to begin building up the sort of formal benchmark described above.

In Section 2 I discuss scholarly usage of the PD game to model anarchic societies. Section 3 makes a logical argument that the structure of the state-of-nature PD game restricts its applicability to permanently anarchic societies, setting the table for the "PD with Coercion" variant that I introduce formally in Section 4. Section 5 illustrates the applicability of the model, and Section 6 concludes.

2. State-of-nature games in the public choice literature

Many surveys of rational choice-based subfields in political economy introduce a PD game as a formal representation of a society without government. Mueller (2003: 9-10) imagines members of a "Hobbesian state of nature" as choosing whether to refrain from stealing each other's corn, with mutual defection prevailing in an "anarchic environment." Dixit (2004: 17-20), in introducing the PD game, cites Gambetta (1993: 15) in suggesting that a two-player PD game can represent a story of a cattle breeder and a butcher; both players would be better off if the breeder can sell an animal to the butcher, but the breeder would like to get full price while giving up a low-quality animal, while the butcher would like to get an animal without following through on payment. Uninvolved in this two-way interaction is at least one other relevant element of

society: the Sicilian Mafia, which may wish to get involved in the deal, quite possibly to the benefit of all three parties. The Mafia is not modelled as being party to this PD game, and therefore its governance services can be introduced from the outside in order to improve outcomes.

Political philosophers have made similar use of the PD game in studying the work of Hobbes.[5] Hampton (1986) and Taylor (1987) both used the PD game to model two-way interactions within a Hobbesian state-of-nature. In the game, players are supposed to choose from two options; Taylor (1987: 141-142), using the words of Hobbes, defines the two options as "laying aside one's natural right to all things" or not so doing. Each of these scholars, after concluding that his or her model is a close characterization of the Hobbesian state of nature, points out Hobbes's suggestion that the players in fact have a third option: unanimously[6] to choose from among themselves an absolute sovereign who will, at an insignificant cost, impose upon them the Pareto-efficient cooperative solution.

Curiously, despite the popularity of the formal PD game representation of anarchy, I have been unable to locate any significant efforts in either the public choice or political philosophy literature where the PD game is formally extended to illustrate the advent of governance.[7] The state, the Mafia and the Leviathan are discussed as means of transcending the PD game, but they routinely fail to appear as players within any formal game. What might such a formalization look like? In Figure 1 I sketch a normal-form game of the Hobbesian state of nature along the lines suggested by Hampton and Taylor:

Figure 1.

		Player 2		
		Lay aside natural right	Retain natural right	Install a sovereign
Player 1	Lay aside natural right	both receive Pareto-superior payoff	P. 1: min. payoff P. 2: max. payoff	
	Retain natural right	P. 1: max. payoff P. 2: min. payoff	both receive Pareto-inferior payoff	
	Install a sovereign			both receive Pareto-superior payoff

Players realize that playing the PD game will result in the Pareto-inefficient payoff; they opt not to play it, thus receiving the cooperative payoff. This extended formalization makes it apparent that to add the Leviathan option is to assume away the PD game, releasing the societal Prisoners for an endless day at the beach under the benevolent gaze of an unpaid yet tireless lifeguard.

Despite the dodge of the state-of-nature problem that the Leviathan option represents, much of the public choice literature – at least that share of it which flows out of social contract theory – is built on the conflicting notions that the state-of-nature problem is serious and that the Leviathan solution to that problem has been adopted. In the public choice version of the Hobbesian story, the lifeguard is not always benevolent, is well compensated for his efforts, and often grows weary – but it is still true that society built his tower to facilitate adult swim. Mueller (2003, p. 10) describes the construction process as it has been put forward in the public choice literature:

> "[I]ndividuals become better off by tacitly or formally agreeing not to steal, provided that the enforcement of such an agreement costs less than they jointly gain from it. The movement from cell 3 [mutual defection] to cell 1 [mutual cooperation] is a Pareto move

> that lifts the individuals out of a Hobbesian state of nature (Bush, 1972; Bush and Mayer, 1974; Buchanan, [1975]; Schotter, 1981)."

Figure 1 applies almost perfectly to the public choice story, with only the names of the strategies being different: "Lay aside natural right" becomes "Do not steal", "Retain natural right" becomes "Steal", and "Install a sovereign" becomes "Tacitly or formally agree not to steal and enforce the agreement." Having made the no-brainer "Pareto move", society then confronts the very different problems of choosing and constraining the all-too-human enforcers, the study and modeling of such problems constituting the bulk of public choice scholarship.

If the Leviathan option or Pareto move is so trivial, so obviously the right move, then the state of nature could never have existed save in a land of incompetents. For a political scientist to draw a Prisoners' Dilemma matrix on the blackboard is a pointless exercise if at the end of the lecture he is to erase it and say, "Thankfully, people are not so stupid as to actually *play* that game!" If the Dilemma had never cost anyone a wink of sleep, one would think it would have long since been dropped from the pedagogical narrative – but considering the consensus that the Dilemma is real and important, one also would think it should not be assumed away.

3. The state-of-nature Prisoners' Dilemma is a model of a permanently anarchic society

Dixit (2004: 14), giving the standard interpretation of the PD game *qua* model of an anarchic society, writes:

> "The need for governance arises because, in its absence, individuals pursuing their own interests would generate an inferior equilibrium outcome. Game theory studies many instances of this, most notably the prisoner's dilemma."

Contra this view, the state-of-nature PD game does not demonstrate a "need for governance." The PD game is intended to represent society in the absence of governance, and thus its mutual-defection equilibrium is demonstrably "inferior" only to the outcome associated with mutual compliance *in the absence of governance.* Without any model purporting to represent a society with governance, the superiority of such a model society cannot be established.

The state-of-nature PD game, however, is a two-choice world. Players can choose only to cooperate or defect; there is no third choice as in Figure 1, and thus *the possibility of governance is in fact precluded by the structure of the state-of-nature PD game itself.* An unmodified PD game, then, cannot be an appropriate representation of a pre-governmental society; it could be an appropriate representation of the state of nature only if governance is in fact impossible.[8] Even so, fundamental intuitions about the state of nature remain unchanged: there is of course excellent reason to believe that the PD game captures the essence of a real social dilemma, where

mutual cooperation is Pareto superior to mutual defection but unilateral defection is a dominant strategy. This suggests that, in seeking to formalize the possibility of governance, one should not reinvent the state-of-nature game-theoretic wheel, but rather seek only to alter the basic PD game to allow players to coerce one another.

Within the context of the state-of-nature PD game (and similar games), I propose that governance be operationalized as *coercion to compliance*, where to adopt a coercive strategy is *to impose a strategy choice upon another player*. Each player's strategy profile is thus expanded from two to four choices: in addition to choosing to cooperate or defect, he or she must also choose whether to coerce compliance from the other player.

4. The Prisoners' Dilemma with Coercion

For the basic structure of a PD game, including payoffs, I follow Molander (1992: 759):

> In the *n*−person prisoner's dilemma, each player faces a choice between two alternatives: to cooperate (C) or to defect (D). The payoffs are functions of the number i of other players cooperating, and are denoted by c_i and d_i, respectively (i = 0, 1, …, $n - 1$). The payoff functions satisfy the following conditions:
>
> 1. monotonicity
>
> $$c_i > c_{i-1}$$
>
> $$d_i > d_{i-1} \qquad i = 1, 2, \ldots, n - 1.$$
>
> 2. dominance of the D alternative
>
> $$d_i > c_i \qquad i = 0, 1, \ldots, n - 1.$$
>
> 3. efficiency of cooperation

$$(i + 1)c_i + (n - i - 1)d_{i+1} > ic_{i-1} + (n - i)d_i \quad i = 1, 2, \ldots, n - 1.$$

$$c_{n-1} > d_0.$$

The model developed in this section is a one-shot game involving two players. In addition to the standard pure strategies C and D, a player may opt to coerce the other player. To avoid confusion, I refer to the coercive option as "Impose" or "I". Thus there are four possible pure strategies:

1. Cooperate (and do not coerce other player)
2. Cooperate/Impose (i.e., cooperate while coercing other player to cooperate)
3. Defect/Impose (i.e., defect while coercing other player to cooperate)
4. Defect (and do not coerce other player)

A player must bear a cost $X \geq 0$ to impose cooperation on the other player, who experiences a disutility $Y \geq 0$ from being coerced.[9] It is assumed that paying the cost X is sufficient to assure success in imposition.

By these changes, the standard PD game is transformed into the PD with Coercion. While the solution concept for the standard PD game is straightforward, the solution concepts that apply to the PD with Coercion change with the values of X and Y.

Figure 2 is the symmetric two-player PD with Coercion in normal form.

Figure 2.

		Player 2			
		Cooperate	Cooperate/Impose	Defect/Impose	Defect
Player 1	Cooperate	c_1, c_1	c_1-Y, c_1-X	c_0-Y, d_1-X	c_0, d_1
	Cooperate/ Impose	c_1-X, c_1-Y	c_1-X-Y, c_1-X-Y	c_1-X-Y, c_1-X-Y	c_1-X, c_1-Y
	Defect/ Impose	d_1-X, c_0-Y	c_1-X-Y, c_1-X-Y	c_1-X-Y, c_1-X-Y	d_1-X, c_0-Y
	Defect	d_1, c_0	c_1-Y, c_1-X	c_0-Y, d_1-X	d_0, d_0

Cooperate is weakly dominated by Defect, and Cooperate/Impose is weakly dominated by Defect/Impose, for all values of X and Y.[10] If the players adopt iterated elimination of weakly dominated strategies as a solution concept, they face the reduced game in Figure 3.

Figure 3.

		Player 2	
		Defect/Impose	Defect
Player 1	Defect/ Impose	c_1-X-Y, c_1-X-Y	d_1-X, c_0-Y
	Defect	c_0-Y, d_1-X	d_0, d_0

The nature of this reduced game depends on the payoff values, as detailed in Figure 4. Generally, when the cost to impose cooperation X falls below the value d_1-d_0, standard solution concepts predict a positive probability that coercion will occur; if X also falls below c_1-c_0, mutual coercion is assured. (I will refer to the case in which $X < d_1-d_0$ and $X < c_1-c_0$ as "cheap coercion".) The disutility from being coerced, Y, is never relevant to the game outcome, but it does have implications for efficiency and welfare. The Pareto efficiency of the outcome under cheap coercion, for example, is dependent on the relationship between X and the value c_1-d_0-Y; in the case where $X < c_1-d_0-Y$ ("very cheap coercion"?), the mutually coercive outcome results

in a Pareto improvement over normal PD payoffs. By contrast, when coercion is still "cheap" but $X > c_1 - d_0 - Y$, mutual coercion results in a Pareto loss relative to normal payoffs.

Regardless of efficiency considerations, however, we do under certain conditions see unilateral adoption of governance strategies. Furthermore, under one subset of values the symmetric solution concept is a mixed-strategies Nash equilibrium: should it be the case that $c_1 - c_0 < X < d_1 - d_0$, each player would be expected to randomize over Defect and Defect/Impose.[11] Here there is some nonzero probability that one player will impose cooperation while the other effectively chooses to be coerced; in some sense, this could be viewed as the emergence of an individual governor.[12]

Figure 4.

Value restrictions	Reduced game is…	Solution concept	Unique solution	Relative to (D,D), solution is…
$X < c_1 - d_0 - Y$	trivial	dominance	(D/I, D/I)	Pareto improvement
$c_1 - d_0 - Y < X$; $X < c_1 - c_0$; $X < d_1 - d_0$	PD	dominance	(D/I, D/I)	Pareto loss
$c_1 - c_0 < X < d_1 - d_0$	Chicken	MSNE	n/a	Pareto loss*
$d_1 - d_0 < X < c_1 - c_0$	Coordination	focal PSNE	(D, D)	No change
$c_1 - c_0 < X$; $d_1 - d_0 < X$	trivial	dominance	(D, D)	No change

dominance = elimination of strictly dominated strategy
MSNE = mixed-strategy Nash equilibrium
PSNE = pure-strategy Nash equilibrium
D/I = Defect/Impose
D = Defect

* See appendix for algebraic demonstration.

5. Discussion of the Prisoners' Dilemma with Coercion

In his discussion of public choice ideas in the context of central banking, Wagner (1986: 522) relates this story:

> Suppose someone were to say that there would be a social saving from replacing our present system of personal security, in which resources are tied up in locks, guns, dogs, police, and the like, with a system of trust and love. This might be thought of as buffoonery, and it might provide material for political speeches or churchly sermons, but it would not be thought of as being grounded in reality. Merely pointing out the potential social saving that would result if people did not feel a need for investing in various forms of personal security does not imply that there is any way of realizing that saving.

The unmodified state-of-nature PD game effectively points out a potential social saving without showing any coercive method by which players can realize that saving; by contrast, the state-of-nature PD with Coercion game incorporates players' unilateral investments in locks, guns, and dogs. If a householder is concerned that his neighbor may commit theft against his property, he may seek to acquire a watchdog; if the neighbor entertains a reciprocal concern, he too may obtain a watchdog. If both neighbors' concerns are warranted and the dogs effective, the situation neatly corresponds to a PD with Coercion where both players move Defect/Impose; the effort to commit theft is tantamount to defection (*a la* Mueller 2003) while the dogs are a means of coercion to compliance. The resulting scenario – both pay the dog-cost X and endure neighbor-dog-disutility Y, but both retain their property and thus receive the cooperative payoff

c_1 – corresponds to the standard solution under cheap coercion and is a plausible outcome in a world of highly effective and inexpensive guard dogs. If one of the thief-neighbors fails to hire the dog – *i.e.*, he plays Defect while his neighbor plays Defect/Impose – he saves the cost X but still bears the disutility Y from having to put up with his neighbor's dog, and he receives the sucker payoff c_0 when he is denied success in his efforts at thievery and suffers the theft of his own property. Such is the typical lot of those who neglect their personal security.

While the model presented in Section 4 is symmetric, the PD with Coercion does invite us to consider the realities of asymmetric power. One's imagination is immediately drawn to the possibility that one player will face a lower cost to impose cooperation on the other player, thus emerging as relatively dominant. This thought prompts consideration of the nature of the costs involved: What does it take to coerce another person? What would it mean to face a "lower cost" in that regard? These are difficult questions and are largely beyond the scope of the present paper, but I believe that as a first approximation the symmetric model can be at least minimally justified. Consider that the social contract approach has effectively dodged the question of asymmetric power, escaping to the egalitarian world of imaginary constitutional conventions and assigning to public choice the task of institutional analysis for one-man-one-vote democracies; by comparison, a symmetric PD with Coercion is at least not competitively inferior. But there also may be good theoretical support for a prediction of symmetry in the realm of physical coercion within a state of nature, as argued by Hobbes himself. He laid great emphasis on the fragility of the human body, claiming that the ability of even the very weak to kill the strong (possibly while asleep) has significant implications for equality.[13] While equating imposition with murder might alter the way one conceptualizes a state-of-nature game, Hobbes's

point does seem to limit the extent to which coercion that stops short of murder can be an effective means of dominating a weaker person.

6. Conclusion

The PD with Coercion can be used to represent an anarchic society in which individuals need not wait for hypothetical unanimous consent or even real plurality consent before embarking on an entrepreneurial venture in forcing their compatriots to behave themselves. Where in the public choice account, a collective governance institution is presumed to arise when (in Mueller's terms) "the enforcement of…an agreement costs less than they jointly gain from it", in the PD with Coercion an individual player is simply presumed to adopt a coercive strategy when he or she expects it to result in a payoff superior to that yielded by noncoercive strategies. The presence or absence of governance in the society thus becomes a function of the costs and benefits of governance to individual actors as perceived by individual actors, not a function of an economist's calculations of social cost or benefit. In this way, the state-of-nature PD with Coercion provides a formal illustration of the advent of governance that is consonant with conventional rational-choice reasoning.

Acknowledgments

Thanks for helpful or encouraging comments by Charles Rowley, Ayse Evrensel, participants in the Back Row Seminar at George Mason (especially Adam Martin and Dan D'Amico), and three anonymous referees. Errors are my own.

Appendix

Following is a proof by contradiction that the mixed-strategies Nash equilibrium payoff in the Chicken reduced game (that is, the reduced game when $c_1-c_0 < X < d_1-d_0$) constitutes a Pareto loss when compared to the payoff from both players playing Defect as a pure strategy.

To find the MSNE, define P_I as the probability that Player 2 moves Defect/Impose. The following equation can be solved for the value of P_I for which Player 1 will be indifferent between the two pure strategies:

$$(P_I)(c_1-X-Y) + (1-P_I)(d_1-X) = (P_I)(c_0-Y) + (1-P_I)(d_0)$$

This can be simplified algebraically to:

$$P_I = (d_1-d_0-X) / [d_1-d_0-(c_1-c_0)]$$

By symmetry, the MSNE holds when both players move Defect/Impose with probability P_I. In this situation, the expected payoff to each player is given by:

$$(P_I)^2(c_1-X-Y) + (P_I)(1-P_I)(d_1-X) + (P_I)(1-P_I)(c_0-Y) + (1-P_I)^2(d_0)$$

Now assume that this expected payoff is greater or equal to the payoff from (Defect, Defect):

$$(P_I)^2(c_1-X-Y) + (P_I)(1-P_I)(d_1-X) + (P_I)(1-P_I)(c_0-Y) + (1-P_I)^2(d_0) \geq d_0$$

Subtracting from both sides and dividing by P_I:

$$(1-P_I)(d_1-X) \geq (2-P_I)(d_0) - (P_I)(c_1-X-Y) - (1-P_I)(c_0-Y)$$

$$d_1-X - (P_I)(d_1-X) \geq 2d_0 - (c_0-Y) - (P_I)[d_0 + (c_1-X-Y) - (c_0-Y)]$$

$$d_1-X \geq 2d_0 - (c_0-Y) - (P_I)[d_0 + (c_1-X-Y) - (c_0-Y) - (d_1-X)]$$

$$d_1-X \geq 2d_0 - (c_0-Y) + (d_1-d_0-X)$$

Simplifying leaves us with $c_0 \geq d_0 + Y$, a false statement.

Notes

1. As noted by Leeson (2006), the main rival to the social contract theory of government is predatory theory, which is associated primarily with the work of Olson (*e.g.*, 2000). A sort of hybrid theory has been offered by North, Wallis, and Weingast (2006), who describe the state as an uneasy coalition of "violence specialists".

2. For sample expositions in which external governments are presumed directly to alter game payoffs, see discussions by Munger (2006: 134-135) or Rutten (1997: 1152-1154). On the general unrealism of an external government, see Buchanan (1975: 130).

3. Slomp (2000: 148-152) does argue that Hobbes's sovereign is in fact an elite person, an educated "geometer" who transcends rational choice to unilaterally impose peace rather than dominion; cf. Taylor (1987: 145), who sees Hobbes as considering the selection of a particular sovereign to be an insignificant matter that can safely be left to either of two collective-choice mechanisms, and Hampton (1986: 206), who ultimately claims that Hobbes's social contract argument is untenable and the installation of a sovereign impossible.

4. The famous complaint of Demsetz (1969) comes to mind. Coyne (2006) makes a similar point in his discussion of foreign interventions to reconstruct "failed states" such as Somalia.

5. Scholars have used games other than the PD to model interpersonal interactions in the absence of a property-rights assumption, including Chicken and Assurance games, but by far the most common representation is the PD game (Slomp 2000: 121, 126).

6. As Taylor (1987: 145) puts it: "Hobbes gives two accounts of how a particular man or assembly of men is to be made Sovereign. In the first, the Sovereign is specified in the covenants of each man with every other man and is thus unanimously agreed on (*Lev*[*iathan*]

132). In the second, there is in effect a unanimous agreement, in the form of the covenants between every pair of men, to abide by a majority choice of a particular Sovereign (*Lev* 133)."

7. Schotter (1981: 45-51), citing Nozick (1975), puts forward a PD-like "state-of-nature theft game" from which he claims that government emerges endogenously. In Schotter's model, players can costlessly organize a mutual defense against "attacks", which is sensible given his assumption that returns to collaboration are increasing; group benefits obviously exceed group costs in the Schotter model, but Olson (1971) argues that this can be insufficient for a group to in fact form. It is hard to understand in what sense a government can be said to have emerged from a model in which, at equilibrium, no one does anything differently from anyone else; Schotter does make textual reference to "the organizers of the protective association, who have been granted judicial and enforcement powers as part of the agreement to form the agency," *i.e.*, the government and the things the government does, but these are not modelled. See also Jiborn (1999: 138-146) for a model where a governance agency is external to the societal PD game and does not itself face any game or any choice between alternate strategies. While Jiborn says, "The idea of this section is to take some of Tocqueville's Sing-Sing prison guards and place them in a population of 'prisoners', captured in an ongoing n-player Prisoner's Dilemma, with the single task of detecting and punishing defections", the section does not in fact follow through on this admirable plan. The guards are never mentioned again; rather, the payoffs of PD-game defectors are directly reduced (and the costs of coercive action are directly and uniformly imposed on all PD-game players) by a formless "sanction system". The constituent actors making up this system, their possible courses of action, and the incentives facing them are not modelled. Cf. Hirshleifer and Rasmusen (1989) who develop a "Repeated Prisoners' Dilemma with Ostracism" that is similar in spirit to my PD with Coercion; while they provide excellent discussion of

"repeated social dilemmas" and even mention imprisonment as possibly falling under the rubric of ostracism, Hirshleifer and Rasmusen stop short of relating their game to the creation or existence of governments (in fact, the words *govern* and *government* do not appear in the paper, while the noun *state* appears only in an etymological discussion of the word *ostracism*). Cf. also Stephens (1992), who recasts the prisoners' original dilemma as a three-player game that includes a superordinate "district attorney" whose dominant strategy is to renege on his initial promises of lenient treatment.

8. This would hold particularly in the extreme case where the PD is used to represent a desperate war of all against all, in which to choose the "defect" move is to explore the *anything* of "anything goes". Adding a government to such a model cannot change outcomes; the government must be unable to compel cooperation because the model already has specified that no possible action is effective in altering the move of another player. In other words: if a player is able to defect *and* finds it best to defect in the face of even the other player's most warlike behavior, there are no actions left with which one might coerce that player into cooperation. Hampton (1986: 62) uses a PD game to represent such a situation, which she refers to as Hobbes's "rationality account of conflict", which is distinguished from a separate Hobbesian "passions account of conflict".

9. Primarily for simplicity of presentation, I have chosen to model a player who incurs the disutility from being coerced regardless of whether his planned move is Cooperate or Defect. In versions of the model where the disutility is only incurred when a player plans to move Defect, Cooperate and Cooperate/Impose are not weakly dominated for all possible payoff values; the solutions thus often involve more complex mixed strategies than solutions to the game presented herein. While those more complex models would be even more favorable for a Pareto-

improving emergence of governance and thus did hold appeal for me, the simpler model proved sufficiently favorable to illustrate the possibility of Pareto-efficient emergent governance.

10. For high values of X, Cooperate/Impose and Defect/Impose are strictly dominated by Defect. Specifically, Cooperate/Impose is strictly dominated when $X > c_1 - c_0$ and $X > c_1 - d_0$, and Defect/Impose is strictly dominated when $X > c_1 - c_0$ and $X > d_1 - d_0$.

11. This section presents a symmetric game, so I emphasize the symmetric solution concept; however, the reduced game under these values of X is Chicken, for which there are of course two asymmetric pure-strategies Nash equilibria.

12. Primarily for simplicity of presentation, I have restricted this paper to outlining only a 2-person example of the PD with Coercion. With $n > 2$ players, further complicating assumptions are necessary, *e.g.*, with regard to whether one player can pay X to coerce all other players or, say, only to coerce one other player; perhaps X could be modelled as a function of n. Such models may well make plainer the possible emergence of a government as comprising one player or subset of players, but the current model does minimally suffice in that regard.

13. An effective condensed exposition of Hobbes's claim is offered by Slomp (2000: 25-26).

References

Buchanan, J. M. (1975). *The limits of liberty: between anarchy and leviathan*. (Chicago: University of Chicago Press)

Bush, W. C. (1972). Individual welfare in anarchy. (In G. Tullock (Ed.), *Explorations in the theory of anarchy* (pp. 5-17). Blacksburg, Va.: Center for the Study of Public Choice)

Bush, W. C. and Mayer, L. S. (1974). Some implications of anarchy for the distribution of property. *Journal of Economic Theory, 8,* 401-412

Coyne, C. (2006). Reconstructing weak and failed states: foreign intervention and the nirvana fallacy. *Foreign Policy Analysis, 2,* 343-360

Demsetz, H. (1969). Information and efficiency: another viewpoint. *Journal of Law and Economics, 12,* 1-22

Dixit, A. (2004). *Lawlessness and economics: alternative modes of governance.* (Princeton, NJ: Princeton University Press)

Gambetta, D. (1993). *The Sicilian Mafia: the business of protection*. (Cambridge, Mass.: Harvard University Press)

Greif, A. (1993). Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition. *American Economic Review, 83,* 525-548

Hampton, J. (1986). *Hobbes and the social contract tradition.* (New York: Cambridge University Press)

Hirshleifer, D. and Rasmusen, E. (1989). Cooperation in a repeated prisoners' dilemma with ostracism. *Journal of Economic Behavior and Organization, 12,* 87-106

Jiborn, M. (1999). Voluntary coercion: collective action and the social contract. Dissertation, Lund University

Leeson, P. (2006). Efficient anarchy. *Public Choice, 130,* 41-53

Molander, P. (1992). The prevalence of free riding. *The Journal of Conflict Resolution, 36*, 756-771

Mueller, D. (2003). *Public choice iii.* (New York: Cambridge University Press)

Munger, M. (2006). Preference modification vs. incentive manipulation as tools of terrorist recruitment: the role of culture. *Public Choice, 128,* 131-146

North, D.; Wallis, J.; and Weingast, B. (2006). A conceptual framework for interpreting recorded human history. (Cambridge, Mass.: National Bureau for Economic Research, working paper 12795.)

Nozick, R. (1975). *Anarchy, state and utopia.* (New York: Basic Books)

Olson, M. Jr. (1971). *The logic of collective action.* Revised edition. (New York: Schocken Books)

Olson, M. Jr. (2000). *Power and prosperity: outgrowing communist and capitalist dictatorships.* (New York: Basic Books)

Rutten, A. (1997). Anarchy, order, and the law: a post-Hobbesian view. *Cornell Law Review, 82,* 1150-1164

Schotter, A. (1981). *The economic theory of social institutions.* (New York: Cambridge University Press)

Slomp, G. (2000). *Thomas Hobbes and the political philosophy of glory*. (Basingstoke, UK: Macmillan Press Ltd.)

Stephens, G. H. III (1992). Sell out, no sell out: the politics of artificial social traps. Dissertation, University of California, Los Angeles

Taylor, M. (1987). *The possibility of cooperation.* (Cambridge: Cambridge University Press)

Wagner, R. E. (1986). Central banking and the Fed: a public choice perspective. *Cato Journal, 6,* 519-543

List of Figures