

# Cooperating Democrats, Defecting Autocrats<sup>a</sup>

Fiona McGillivray

Department of Political Science

Yale University

124 Prospect St

New Haven CT 06520

Alastair Smith

Department of Political Science

Yale University

124 Prospect St

New Haven CT 06520

15th October 1998

---

<sup>a</sup>The ideas in this paper were inspired by a conversation with Helen Milner. Our thanks to her. Our thanks also to David Cameron and Matt Gabel for their comments.

## ABSTRACT

Using the infinitely repeated prisoners' dilemma as a modeling platform, we examine how the domestic political institutions of nations affect their ability to cooperate internationally. We propose a strategy, the Leader Specific Grim Trigger, in which leaders direct punishments for past defections at the leader responsible rather than at the nation she represents. Leaders refuse to cooperate with those leaders that have cheated them in the past. However, by being prepared to cooperate with new leaders, cooperation can be restored. The focus of punishment on specific leaders rather than the nation means that domestic electorates want to remove leaders that defect. Hence, leaders that are held domestically accountable pay audience costs for failing to cooperate. In addition to showing that democratic dyads can cooperate to a greater extent than other pairs of nations, the model produces hypotheses about both the domestic consequences of international cooperation and the relationship between leadership change and the restoration of cooperative agreements.

## INTRODUCTION

We examine how domestic political institutions affect the ability of nations to cooperate at the international level. In common with much of the previous literature, we use the metaphor of the prisoners' dilemma as a model of international cooperation (Axelrod 1984; Axelrod and Keohane 1986; Milner 1992). Although in such games nations have myopic incentives to cheat each other, if nations are sufficiently patient then cooperation can be maintained by making future cooperation contingent on current behavior. Well known examples of these contingent strategies include Grim Trigger and Tit-For-Tat strategies. Our interest is in how domestic political institutions affect the range of conditions under which cooperation can occur. By incorporating a model of the domestic selection of leaders we show that democratic

nations can cooperate with each other under conditions where other pairs of nations find cooperation impossible.

The enhanced ability to cooperate arises when leaders use punishment strategies that focus punishment against specific leaders rather than against the nations that they represent. Should leader A exploit leader B's attempts to cooperate then leader B withdraws all future cooperation with leader A. However, if leader B holds no grudges against future leaders, then once leader A is replaced, cooperation can be restarted. Such a strategy provides voters with an incentive to remove leaders that fail to cooperate. As long as the incumbent remains in office, they realize that cooperation is unlikely to restart. However, the voters can readily restore cooperation by removing the offending leader. This provides an endogenous explanation for domestic audience costs (Eyerman and Hart 1996; Fearon 1994; Leeds 1998; Mansfield et. al. 1998; Martin 1993; Smith 1998a, b). Since voters remove leaders who fail to cooperate, office seeking leaders cooperate for fear of being removed. The greater the ease with which citizens can replace leaders and the greater the rewards of holding office, the greater the range of conditions under which cooperation can be supported. Hence, if punishments are directed at individual leaders rather than at the nations they represent then democratic dyads can cooperate at much higher levels than other pairs of states. Hence our model contributes to the growing literature on the democratic peace (Bremer 1992; Bueno de Mesquita et. al. 1998; Dixon 1994; Levy 1988; Maoz and Abdolai 1989; Maoz and Russett 1993; Ray 1995).

## INTERNATIONAL COOPERATION

The interactions between nations can take numerous forms. However, we focus on interactions where nations want to cooperate to achieve mutual benefits, but where each nation has an incentive to exploit the cooperative behavior of the other. Such circumstances are typically characterized as prisoners' dilemma. We do not assert that

all the interactions between nations have this strategic pattern. Many interactions are either directly conflictual, coordinational or lack the possibility of exploitation (Morrow 1994). Rather we focus on the prisoner's dilemma type interaction because it is within this setting that credible commitment is problematic. In contrast to this general setup, Mansfield et. al. (1998) and Milner and Rosendorff (1997) model the specific economic consequence of cooperation on trade issues.

The prisoner's dilemma game is a common metaphor. Both nation, 1 and 2, independently and simultaneously decide whether to cooperate or to defect. If neither side cooperates, both defect, then neither side gains any benefits. Nation 1's payoff from this outcome is  $P_1$ , often referred to as the punishment payoff for failing to cooperate.  $P_2$  represents nation 2's payoff from this non-cooperative outcome. If both decide to cooperate then they realize the gains of mutual cooperation. The payoffs associated with this outcome are  $R_1$  and  $R_2$ , often called the rewards of cooperation. There are benefits from cooperation, so  $R_1 > P_1$  and  $R_2 > P_2$ .

To illustrate what the gains from cooperation are consider the following two examples. First suppose that nations 1 and 2 fear each other militarily. To address these fears both sides buy large militaries. Since both sides purchase large amounts of weapons, neither side has the advantage over the other. Both sides would be better off if they both simultaneously scaled back their military forces. Under this contingency, both sides still have similar sized forces and so a similar level of security to that when they both bought large militaries. The advantage is that for the same level of security they both spend less. Hence by cooperating, both nations maintain the same level of security but reduce their defense budget. International trade offers another example. If both nations lower trade barriers then both sides gain from increased trade. The incentives for cooperation are clear, both sides are better off. Yet, the potential for exploitation means that cooperation is difficult to implement.

If nation 1 cooperates but nation 2 defects then nation 1 is exploited. Being ex-

exploited is costly, so nation 1 is worse off than if it had not cooperated in the first place. It receives, what is commonly called, the sucker's payoff  $S_1$ . Nation 2, on the other hand gains from exploiting nation 1, receiving the temptation payoff  $T_2$ . If nation 2's cooperation is exploited by nation 1, then nation 2 receives the sucker's payoff ( $S_2$ ), while nation 1 gets the temptation payoff ( $T_1$ ).

Figure 1: The Prisoners' Dilemma Game

		Nation 2	
		cooperate, C	defect, D
Nation 1	cooperate, C	$R_1; R_2$	$S_1; T_2$
	defect, D	$T_1; S_2$	$P_1; P_2$

Where  $T_i > R_i > P_i > S_i$  for  $i = 1; 2$

In spite of the value of mutual cooperation, each nation has a dominant strategy to defect. By this we mean that whatever the other side does, it is always better to play defect than cooperate. To illustrate, suppose nation 2 plays cooperate. If nation 1 also cooperates then both nations receive  $R$ . However, nation 1 can improve its payoff to  $T$  by defecting. Similarly, since  $P$  is greater than  $S$ , if nation 2 is defecting, nation 1 also wants to defect rather than be exploited. Defecting is a dominant strategy. The outcome (D; D) is often referred to as the single shot Nash equilibrium. Given that the game is played only once, both players defect and cooperation never occurs. Both nations would prefer to cooperate, but in the Hobbesian world of international relations, the lack of enforceable contracts makes cooperation impossible.

Yet, the future is not as bleak as the single shot game suggests. Nations can cooperate providing they anticipate repeated interactions. Nation 1 may forego the opportunity to capitalize on nation 2's present cooperation, if the possibility of long term cooperation depends upon not exploiting 2 in the short term. Thus cooperation is possible providing the long term rewards of cooperation outweigh the short term gains of cheating the other nation. In order to understand what this means, we need

to precisely define what it means to care about the future. Game theory allows us to assess the balance between these short and long term gains via the concept of the discount factor{ the relative value of the future.

Viewed from the present, payoffs in the future are not as valuable as payoffs in the present. Stated more directly, it is better to receive a dollar today than a dollar tomorrow. Although this can be justified on several levels, perhaps the two most common justifications are the concepts of an interest or inflation rate and the probability of the game continuing. We will use the mathematical terminology to explain these principles. The discount factor,  $\delta$ , represents the relative value of future payoffs. Thus, if we receive a payoff of one in the next period, viewed from the current period it has only value  $\delta$ , where  $0 < \delta < 1$ . It is better to receive payoffs in the present, rather than the future. The interest rate justification for the discount factor is straightforward. If we had  $\delta$  dollars today then we could invest it and have one dollar in the next period. Alternatively, phrased in terms of inflation, the purchasing power of a dollar today is greater than the purchasing power of a dollar tomorrow. Thus, it is better to have the rewards as soon as possible. The discount factor can also express the probability of a player being around to collect rewards in the future. If the game could possibly stop, then future rewards may not be realized. Hence, players prefer to collect today rather than possibly collect tomorrow. The discount factor reflect the extent to which players discount the future. If  $\delta$  is close to one, then players are patient, with payoffs tomorrow having a similar value to payoffs today. If  $\delta$  is low (close to zero), then players are impatient.

Using the discount factor, we have a metric with which to compare short and long run gains. In particular, we will find out how patient players need to be in order to sustain cooperation. There are many possible ways in which nations can structure their interactions to facilitate cooperation.<sup>1</sup> Yet, the motivation behind

---

<sup>1</sup>This result is often referred to as the folk theorem (Fudenberg and Maskin 1986).

all these strategies is that future cooperation is withdrawn if either nation defects. Providing both nations regard the benefits of future cooperation as more important than the short run gains to be had from defection, cooperation continues. Although cooperation can be supported through a variety of strategies, we focus on a single mechanism, the Grim Trigger strategy. We adopt this focus for three reasons: First, it has a simple behavioral interpretation. Second, it describes the limits of cooperation. Put simply, if nations can not cooperate via the grim trigger strategy then they can not cooperate at all. Third, we will use grim trigger like strategies when we examine how different political systems behave.

### Grim Trigger strategy

In this section we describe the Grim Trigger strategy and calculate the conditions under which it is a subgame perfect equilibrium in the infinitely repeated prisoners' dilemma game. We describe the strategy for nation 1, nation 2 behaves analogously.

1. In the initial period,  $t = 0$ , nation 1 plays Cooperate.
2. In every subsequent period,  $t > 0$ , nation 1 plays Cooperate providing that no nation has ever defected in the past. If a defection has ever occurred, then nation 1 plays Defect.

Under this strategy, nations start cooperating and cooperation is sustained provided neither ever defects. Should defection ever occur, then both nations defect in every subsequent round and cooperation is ended forever. Hence, a single incidence of defection results in the outcome (D; D) in all subsequent rounds. It is this threat of permanently ending cooperation, that prevents nations from exploiting each other in the short run.

A brief analysis on the mathematics illustrates the balance between long run gains from cooperation and the short run incentives for exploitation. If both nations play

the GT strategy then they both start cooperating, and continue to do so. In each period, nation 1 receives a payoff of  $R_1$ , the current value of which is  $R_1 + \delta R_1 + \delta^2 R_1 + \delta^3 R_1 + \dots = \sum_{t=0}^{\infty} \delta^t R_1 = \frac{R_1}{1-\delta}$ .<sup>2</sup> This is the expected value of playing the GT strategy given that the other nation also plays the GT strategy. Yet, in the short term, nation 1 gets a higher payoff by defecting. Unfortunately, the long term consequences of doing so are that cooperation never occurs again. Formally, if 1 defects then it receives  $T_1$  today and then  $P_1$  in every subsequent period, the value of which is  $T_1 + \delta P_1 + \delta^2 P_1 + \dots = T_1 + \frac{\delta P_1}{1-\delta}$ . By comparing these payoffs we calculate the conditions under which nation 1 wants to cooperate:  $\frac{R_1}{1-\delta} \geq T_1 + \frac{\delta P_1}{1-\delta}$ . Providing that  $\delta \geq \frac{T_1 - R_1}{R_1 - P_1}$ , nation 1 wants to cooperate.<sup>3</sup> Put more straightforwardly, if nations care sufficiently about the future then cooperation is possible. Using a standard example, of  $T = 4$ ;  $R = 3$ ,  $P = 2$  and  $S = 1$ , cooperation is only possible if  $\delta \geq \frac{4-3}{3-2} = 1$ . This represents the limit of cooperation; there is no way to maintain cooperation if nations are less patient.

Many scholars suggest that the GT strategy is inappropriate because it requires perfect information and no noise. In the real world it is often difficult to determine exactly what happened and so interpretation mistakes lead to the end of cooperation. Additionally, although the prisoner's dilemma captures the general incentives of nations, the interaction in each period might not be identical. In some period the relative value of cooperation or temptation might be greater than in other periods. To add this feature of the real world to the game, we might suppose that the payoffs

---

<sup>2</sup>A useful mathematical result is that infinite sum  $x + \delta x + \delta^2 x + \dots = \frac{x}{1-\delta}$ , ( $\delta < 1$ ). If we want to calculate an infinite stream of payoffs  $x$ , but have the payoff not start until two periods have elapsed then the current value of the infinite sum is simply discounted twice (i.e. the current value is  $\frac{\delta^2 x}{1-\delta}$ ).

<sup>3</sup>There are other conditions that we also need to check. For example, we need to show that once defection occurs, defection in all subsequent rounds is optimal. There are corresponding conditions for nation 2.

in each cell of the payoff matrix vary randomly over time. Instead of fixed payoffs in each period, the payoffs might vary. On average the temptation payoff might be  $T_1$ , but in each period it might be randomly drawn. This places much greater strain on the condition  $\delta > \frac{T_{1;t} R_1}{T_1 P_1}$ . Suppose that in a particular period  $t$ , the temptation to defect was particularly high,  $T_{1;t} \gg T_1$ . As a numerical example, suppose that nation 1's temptation reward,  $T_{1;t}$ , is 10. Under this condition, nation 1 must really value the future if cooperation is to be maintained:  $\delta > \frac{T_{1;t} R_1}{T_1 P_1} = \frac{10 \cdot 3}{10 \cdot 2} = 1.5$ . In the real world, with its associated noise, variation, and opportunities, cooperation is hard to maintain.

Of course, if we assume that  $\delta$  is very large for both nations then we can support cooperation through a variety of means. Yet, this is not our goal. Rather we look at conditions where cooperation is liable to fail. In this context, we examine how domestic political institutions affect the extent to which nations can cooperate.

## DOMESTIC INSTITUTIONS, AUDIENCE COSTS AND COOPERATION

Polities differ institutionally and this affects the ease with which the citizen can remove leaders. Rather than enter a discussion of the effects of different institutional features, we consider two extreme types, which we label autocrats and democrats. We assume that autocrats are immune from removal and democrats serve only at the discretion of the citizens (Bueno de Mesquita and Siverson, 1995). Obviously, in reality, no system is at either extreme. Yet, the focus on the limiting cases allows us to clearly demonstrate how electoral accountability generates audience costs.

If the citizens are able to punish leaders then leaders that take undesirable actions, such as escalating a crisis and then backing down, must pay the costs for their actions. These costs have been labeled audience costs (Bueno de Mesquita and Lalman 1992; Eytzinger and Hart 1996; Fearon 1994; Leeds 1998; Mansfield et al. 1998; Martin

1993; Smith 1998a, b). Fearon (1994) shows how democratic leaders, because of their heightened vulnerability to domestic audience costs, are advantaged in their ability to signal their intentions during international crises. Leeds (1998) uses a similar model of international cooperation to the one used here. She shows that as agreements become more costly to break, nations are better able to cooperate (see Gaubatz (1996) and Smith (1995) for similar arguments with respect to alliances and see Mansfield et. al. (1998) for similar arguments with respect to trade policy). Using the argument that electoral accountability produces the greatest audience costs, she predicts, and finds support for, the hypothesis that democratic dyads are more likely to cooperate than other pairs of states. In parallel with the other studies, Leeds asserts that democratic leaders pay greater audience costs than autocrats for breaking agreements. While it is clear that democrats are more sensitive to domestic pressures than autocrats, why the citizens should want to punish leaders for breaking agreements is less clear (Smith 1998a, b). The value added in this paper is making this link clear. In doing so we provide an endogenous theory for the production of domestic audience costs and we also generate predictions about the domestic ramifications of the success or failure of international cooperation.

Before moving on, we pause to consider why voters should want to punish their leaders. In common with the literature, we have glossed over the redistributive aspects of international cooperation. Hence, everyone in the nation receives the same payoff from the prisoner's dilemma. A national leader wants to cooperate providing the long run benefits are greater than the short term gains that the leader must forego (i.e.  $\frac{R_1}{1+i} > T_1 + \frac{P_1}{1+i}$ ). The voters have the same preferences and so, if they were to choose for themselves, would make the same decisions as their leader. In this calculation, a leader only defects when the citizens would also defect. Why then should the voters punish such a leader, she is, after all, only carrying out their wishes. We rectify this inconsistency by simultaneously modeling foreign policy choices within

the prisoners' dilemma and the voters' electoral decision.

## A MODEL OF COOPERATION AND ELECTIONS

In this section we propose a simple stylized model of the electoral process embedded within the context of international cooperation, as represented by the prisoners dilemma. We start with a simple stage game. In this game, the leaders of each nation decide whether or not to cooperate. We assume the strategic nature of this interaction is similar to the prisoners' dilemma. Having played the international part of the game, each leader faces domestic reselection. At this point the incumbents might be replaced. The game is then infinitely repeated.

Notationally, we represent the current leader of nation 1 at time  $t$  as  $A$ . The corresponding leader in nation 2 is labelled  $B$ . We assume that these leaders are office seeking. Hence for each period that leader  $A$  remains in power, she receives an office holding benefit  $\alpha$ . Each nation has an electorate,  $V_1$  and  $V_2$ . In the domestic reselection phase these voters decide whether to retain the incumbent or replace her with another leader.

The modified stage game is as follows:

1. The current national leaders,  $A$  and  $B$ , play the prisoners dilemma.
2. The members of each polity receive their payoffs
3. Elections occur. In democratic systems the electorate can remove their leader and replace her with another leader. In autocratic systems the leader can not be replaced.

### Modeling Elections.

The key feature of this paper is the degree to which leaders are accountable to the electorate. In democratic regimes, voters can readily remove leaders. In autocratic

regimes, elections, if they are even held, are largely meaningless. As a first cut we shall stick with this dichotomy and assume that autocrats can not be removed by elections. We assume that there is huge pool of potential leaders. If the voters decide to remove the current leader then the next leader is drawn from this pool (for mathematical simplicity, the pool of candidates is infinite and leaders do not enter the pool of candidates once removed).

We assume that elections occur in every period of the game. However, it is a straightforward generalization of the model to have elections only in certain periods. For example, we could assume that elections occur every fourth period. This could be used to represent elections which occur every four years, while still considering foreign policy as made annually. However, such elaborations add few substantive results and so are ignored here.

### The Leader Specific Grim Trigger strategy (LSGT)

Here we consider a strategy, the Leader Specific Grim Trigger strategy, and show how it leads to cooperative behavior. This strategy specifies the conditions under which leaders cooperate and defect and specifies the voters' electoral choice in the modified stage game.

The Leader Specific Grim Trigger strategy (LSGT) is as follows:

1. The current leader of nation 1, A, cooperates unless she or the current leader in nation 2, B, has previously unilaterally defected. If either current leader has unilaterally defected in the past then A defects. (Definition of unilateral defection: we say that leader A unilaterally defects if she plays D while leader B plays C.)
2. The current leader of nation 1, B, plays an analogous strategy to leader A, defecting only if either current leader has unilaterally defected in the past;

otherwise B cooperates.

3. The electorate in nation 1,  $V_1$ , retain their leader providing that she has never unilaterally defected against nation 2 (independent of who is the leader in 2 at the current time). If A has ever unilaterally defect in any previous period then  $V_1$  replaces her with a new leader.
4. The voters in nation 2,  $V_2$ , remove their current leader, B, if he has ever unilaterally defected against nation 1. Under all other conditions,  $V_2$  retains their current leader.

Providing that  $\pm > \frac{T_i R}{(T + \pm_i R)}$ , the LSGT is a subgame perfect equilibrium. Before exploring the substantive implications of this result, we outline aspects of the mathematical proof to this proposition. In the process of doing so we explain the logic behind the equilibrium.

We start by analyzing  $V_1$ 's decision to remove their leader if she has previously unilaterally defected against nation 2. Given the LSGT, once leader A unilaterally defects, no leader in nation 2 will ever play cooperate against A. As long as leader A remains in power the outcome will be (D;D) in every period. Thus, if leader A unilaterally defects and  $V_1$  retains her indefinitely, then  $V_1$  receives a payoff of  $T + \pm P + \pm^2 P + \dots = T + \frac{\pm P}{1 \pm}$ . Alternatively the voters might decide to give her another chance but remove her at the next election. Their reward for doing this is  $T + \pm P + \frac{\pm^2 R}{1 \pm}$ . Under the LSGT, once leader A is replaced nation 2's leader will restart cooperation. The feature of the LSGT that differentiates it from the conventional GT is that punishment is directed at a defecting leader and not the nation that she represents. Once the individual responsible for the harm is removed, nation 2 no longer has any incentive to forego cooperation; the responsible party has been punished. Hence, the longer the voters retain leader A after she has defected, the longer it takes to restore cooperation. If the voters immediately remove A then

their payoff is  $T + \frac{\pm R}{1_i \pm}$ . This is clearly superior to any payoff in which A is kept. Hence, it is optimal for the voters to remove any leader that unilaterally defects.

Note that in equilibrium, both leaders A and B always cooperate. Hence the voters receive the payoff associated with cooperation in every period, which is worth  $\frac{R}{1_i \pm}$ . Interestingly, the voters receive a higher payoff,  $T + \frac{\pm R}{1_i \pm}$ , if their leader defects. The voters actually want their leaders to cheat. This leads to the perverse consequence that if leaders carry out the voters wishes they will subsequently be punished by the voters. Of course, in equilibrium, office seeking leaders never cheat if the electorate can remove them.

We now examine the incentives of leaders. If leader A unilaterally defects then she is removed from office. Suppose that her foreign rival has not unilaterally defected in the past. According to the LSGT she should cooperate. If she does so then her nation receives R in the first period. In addition, she receives a payoff of  $a$  associated with holding office. Thus, her total first period payoff is  $R + a$ . In equilibrium, she continues to receive this payoff in every period, which is worth  $(R + a) + \pm(R + a) + \pm^2(R + a) + \dots = \frac{R}{1_i \pm} + \frac{a}{1_i \pm}$ . If leader A defects then her immediate reward is larger. However, under LSGT, the voters remove her at the next election. Since under LSGT the next leader implements cooperation, A's payoff from defecting is  $(T + a) + \pm R + \pm^2 R + \dots = (T + a) + \frac{\pm R}{1_i \pm}$ . This represents being in office and getting the temptation payoff in the first period, but then being replaced by a new leader who cooperates in every subsequent period.

Leader A cooperates providing the reward from doing so outweighs the short term benefits of defection:  $\frac{R}{1_i \pm} + \frac{a}{1_i \pm} > (T + a) + \frac{\pm R}{1_i \pm}$ . This condition is satisfied providing  $\pm > \frac{T_i R}{(T + a_i R)}$ . Providing leaders care about office holding cooperation is possible through the LSGT.

We still need to check a few remaining conditions to ensure that this is a subgame perfect equilibrium, but these conditions are straightforward. For example, if leader

B unilaterally defected in a previous period then according to LSGT, leader B will defect in the future. A's optimal response therefore is to also defect. Note she will not be domestically punished for this because she is not responsible for the breakdown of cooperation.

The LSGT provides voters with an effective means to restore cooperation. Once the leader responsible for defection is replaced then the two nations return to a cooperative state of affairs. For mathematical simplicity and to retain a parallel to our baseline case, GT, the mechanism we consider is a trigger strategy. Yet, unlike the Grim Trigger strategy, a single incidence of defection in the Leader Specific Grim Trigger does not imply a permanent collapse of cooperation.

In many ways electoral competition is modeled unrealistically. The voters always reelect their leader providing she continues to support cooperation. A more realistic electoral model might assume that leaders produced a variety of goods and policies in addition to international cooperation. The electorate's decision might be a function of performance on all these issues. Yet, even in this more complex scenario, the incentives to cooperate remain. If the voters retain a leader who has unilaterally defected in the past then they should not expect significant gains from international cooperation. In contrast, if they remove such a leader, international cooperation can be restored immediately. While the electorate might still retain such a leader for her superior performance on other issues, but, *ceteris paribus*, defection makes it harder for her to stay in office. While a more realistic model of electoral competition makes the reelection decision less deterministic, leaders still want to avoid defection because it harms their prospects for retaining power. Electoral accountability makes defection an unattractive policy option. This result is true despite the popularity of defection with the voters. Indeed, if their leader exploits the cooperative behavior of other nations then the voters are better off; they receive the temptation payoff. The voters do not suffer from the withdrawal of future cooperation providing they are prepared

to drop the incumbent and choose a new leader.

Domestically the model predicts that leaders who fail to cooperate internationally suffer electorally, even if the policies they pursued were popular and in the national interest. We can find anecdotal support for these prediction in the fortunes of the John Major's British Conservative government. In the 1997 general election, the Labour party, under the leadership of Tony Blair, decisively defeated the incumbent Conservative government. On all measures of pocket book voting, there was no reason for such a defeat. Indeed, quite to the contrary, the economy was roaring. Most political pundits have pointed to two factors in the Conservatives' demise: scandal and Europe. Prior to the 1997 election, Major's government had consistently failed to cooperate with its fellow European governments in the creation of a single Europe. The Conservatives broke numerous European union agreements. For example, they refused to sign the social charter, they dropped out of the Exchange Rate Mechanism, and they threaten not to join European Monetary Union. The social charter is an agreement on union and labor practices throughout the European union. By refusing to sign or enact such legislation, the Conservatives were able to keep marginal labor rates lower in Britain than in its rival European partners. Such differences helped encourage firms to build new plants in Britain rather than in other parts of Europe. By refusing to sign the social charter, the Conservatives reduced unemployment in Britain at the expense of the rest of Europe.

In contrast to the Conservative's position, the opposition Labour party's manifesto fully endorsed European Integration.<sup>4</sup> Surprisingly, given that claims that Europe was, at least partly, responsible for the Conservatives problems, the Conservatives' policy position on Europe was the more popular. In a retrospective voting framework this appears somewhat contradictory. Yet, these facts fit exactly our model of

---

<sup>4</sup>This is a complete reversal of party position from those held during Britain's early years in the European Economic Community.

electoral competition and international cooperation. The British voters got to have their cake and eat it too. By not signing the social charter the British government exploited tighter European labor practices. However, British voters paid few of the consequences. The voters restored European cooperation by dumping the incumbent and electing the pro-European Labour party.<sup>5</sup>

In general, we think that our model offers a more realistic explanation of international cooperation than theories relying purely on external factors. For instance, it is obviously inappropriate to believe that nations use Grim Trigger type strategies { cooperation would have ended long ago. Similarly, there is little evidence of periods of punishment that should be associated with tit-for-tat type strategies. For example, in the mid and late 1960's France demanded that the US convert France's holdings of dollars into gold. This move helped precipitate Nixon's withdrawal of the gold standard.<sup>6</sup> Yet, following this incident France remained an active, and welcome, member of the international community, however, President Charles DeGaul was forced to resign when the people failed to endorse his referendum on constitutional reform.

Here, as elsewhere in the literature, international cooperation is modeled as a simple binary choice: cooperate or defect. In reality the situation is more complex. As Fearon (1998) points out, there are numerous agreements through which nations can cooperate (Milner 1997; Putnam 1988; Rheinhardt 1996). Given the incentives to defect, nations must be suitably patient in order to abide by the terms of the agreement. Yet, the more patient nations are, the greater the importance of striking a good deal (McGillivray 1998; Smith and Hayes 1997). As Fearon demonstrates, when nations

---

<sup>5</sup>Indeed in the context of our theory, the only part of this narrative that is difficult to explain is why the incumbent Conservatives adopted the popular policy of being obstinate over European integration.

<sup>6</sup>Interestingly, the French had made a similar move against Britain during the interwar years. By demanding gold, when the circulation of paper money far exceeded the gold that backed it, the French helped contribute to the monetary crises of the late 1920's and early 1930's.

value the future highly, they may be prepared to delay the start of cooperation to ensure that when cooperation starts it is a more advantageous arrangement. In general, most of the literature on international cooperation fails to specify the course of international agreements. Is international cooperation abiding by the terms of an existing arrangement? Is it negotiating a new arrangement? Or is it both? Our model suggests that leaders that fail to cooperate are punished domestically. However, such a statement ignores Fearon's critique about negotiating agreements. Prospectively, voters have little incentive to retain leaders that fail to demonstrate an ability to achieve advantageous agreements. Hence leaders must walk a tightrope. If seen as refusing to cooperate, then the removal of future cooperation makes them unelectable. Yet, on the other hand, too readily accepting poor agreements reveals leaders are weak, which also harms their electoral prospects. Major's Conservative government clearly failed on the former criteria. Yet, playing hardball and still reaching an agreement is a recipe for success. Major's predecessor, Margaret Thatcher withdrew British cooperation from the European union in order to renegotiate Britain's role in Europe. In particular, she was able to reduce Britain's contributions to the European Union which provides funds for such programs as the Common Agricultural Program.

### REGIME COMPARISONS: WHO COOPERATES?

So far we have outlined two mechanisms that nations can use to cooperate when faced with prisoners' dilemma type interactions. In the first, the grim trigger strategy, nations resist temptation today in order to preserve the long run stream of cooperative benefits. We might categorize the incentives to cooperate as an external or international reputation. By contrast, in the second mechanism, the Leader Specific Grim Trigger strategy, leaders cooperate to avoid domestic punishment. This ability of the electorate to punish their leaders creates audience costs. To avoid these costs, leaders cooperate. However, the extent to which leaders can use this internal mech-

anism depends upon their vulnerability to domestic removal. If leader B knows that leader A can not be punished by her electorate then she knows that leader A has no incentive to remain honest and will try to cheat. In anticipation of this, leader B will also defect. This implies that cooperation can only be maintained via the LSGT if both leaders are domestically vulnerable.

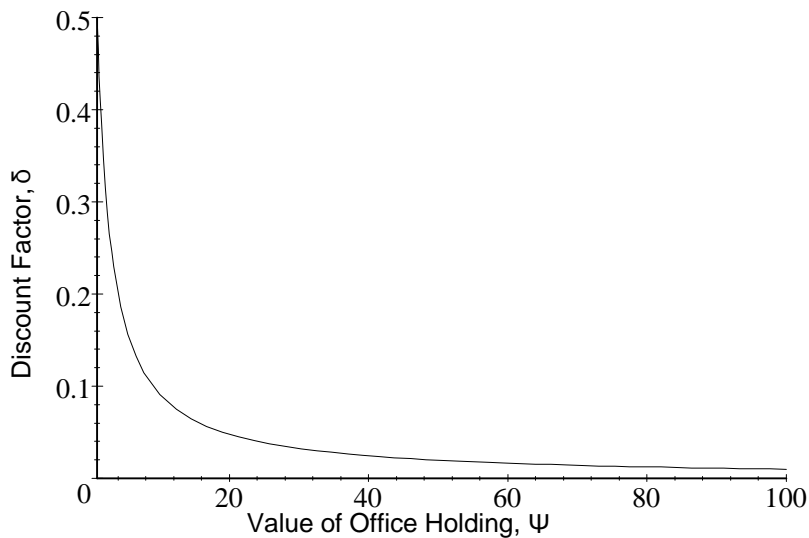
External, or international, reputation requires that  $\delta \geq \frac{T_i R_i}{T_i P_i}$  to ensure cooperation. Internal, or domestic, enforcement requires that  $\delta \geq \frac{T_i R}{(T + a_i R)}$ . We should remember that cooperation requires the belief that both sides, not just a single nation, will continue to cooperate. Table 1 show the level of patience required for different pairs of nations to cooperate. All pairs of nations, democratic or not, can support cooperation through external mechanisms. If both nations are democratic, then cooperation can also be supported through the domestic imposition of audience costs.

		Nation 2	
		Democratic	Autocratic
Nation 1	Democratic	either $\delta \geq \frac{T_i R}{T_i P}$ or $\delta \geq \frac{T_i R}{(T + a_i R)}$	$\delta \geq \frac{T_i R}{T_i P}$
	Autocratic	$\delta \geq \frac{T_i R}{T_i P}$	$\delta \geq \frac{T_i R}{T_i P}$

Table 1 suggests that democratic dyads never find it harder to cooperate than other pairs of nations and under many circumstances find cooperation easier. There is growing empirical support for this prediction (Eyerman and Hart 1996; Gaubatz 1996; Leeds 1998; Mansfield 1998; Mansfield et. al. 1998; Martin 1993; Milner 1997; Oneal and Russett 1997; Russett et. al. 1998.). Providing that  $a > R_i P$  then domestic enforcement makes cooperation possible under conditions where external mechanisms fail. If leaders value office holding highly relative to their evaluation of the international stakes then cooperation can occur under almost any conditions. As

a guide to the magnitude of these effects, we return to our earlier numerical example,  $T = 4$ ,  $R = 3$ ,  $P = 2$  and  $S = 1$ . For this example, external reputation requires  $\delta \geq 0.5$  to support cooperation. Figure 2 shows how the discount factor required to ensure cooperation via the LSGT strategy varies as a function of office holding. As the value of office holding rises, leaders require only the minimal level of patience to support cooperation. For example, if  $\alpha = 2$  then cooperation can be supported provided that  $\delta \geq 1/3$ . If, as we believe, politicians are primarily driven by office holding motives, then we should consider the office holding benefits much larger. If  $\alpha$  is 10 or 100 then the discount factor need only be  $1/11$  or  $1/101$ , respectively, to support cooperation via the LSGT.

Figure 2: The relationship between the value of office holding and the minimum discount factor required to support cooperation under the LSGT strategy.



## CONCLUSION

By simultaneously modeling domestic reselection and international cooperation we provide a mechanism for the production of domestic audience costs in democratic

systems. These audience costs allow democratic dyads to cooperate under conditions where other pairs of states find it impossible. Such predictions are supported by empirical studies (Eyerman and Hart 1996; Gaubatz 1996; Leeds 1998; Mansfield 1998; Mansfield et. al. 1998; Martin 1993; Milner 1997; Oneal and Russett 1997; Remmer 1998; Russett et. al. 1998.) In addition, such predictions contribute more generally to the growing literature that democratic dyads are different (Bremer 1992; Bueno de Mesquita et. al. 1998; Dixon 1994; Levy 1988; Maoz and Abdolai 1989; Maoz and Russett 1993; Ray 1995). However, unlike much of the democratic peace literature, the arguments here are deductive rather than descriptive.

We offer a theoretical explanation that integrates the standard model of international cooperation {the prisoners' dilemma} with a model of domestic reelection. We show that there exists an equilibrium in which it is incentive compatible for the electorate to remove leaders that fail to support cooperation. The key factor in the generation of these domestic audience costs is that punishment for failing to cooperate is directed at the leader responsible rather than the nation that she represents. We believe that the leader specific punishment schedules that we propose are more realistic than the nation based punishment schedules required for external enforcement of cooperation. Indeed, the start of cooperation appears to be strongly tied to regime changes. Extreme examples such as the end of World War II or the Cold War serve as illustrations.<sup>7</sup> Once the old regimes fell and new leaders took over cooperative arrangements rapidly developed. Few expect that UN sanctions against Iraq will continue following the demise of Saddam Hussein. Unfortunately, the Iraqi citizens

---

<sup>7</sup>The treatment of Germany following the two World wars provides an interesting comparison. Following the first World war, the allies appeared to insist that the German nation as a whole apologize in the form of disarmament and reparations. Little cooperation followed. The end of the second world war stands in stark contrast. At least on a humanitarian level, German behavior was far worse during the second world war. Yet, despite this, following the removal of the Nazi regime, the allies invested vast resources into rebuilding Germany and establishing cooperative arrangements..

continue to suffer because they can not remove their leader.<sup>8</sup>

Our model provides an explanation as to why democratic states cooperate at higher levels. Yet equally important in terms of theory building, we generate additional testable hypotheses. Our predictions include the domestic consequences of the success or failure of international cooperation and the relationship between leadership change and the restoration of cooperative arrangements.

---

<sup>8</sup>The citizens of autocratic countries are doubly disadvantaged. First, since autocratic leaders can not be removed, they are less likely to continue cooperation. Second, autocratic leaders typically survive in office much longer than democrats. This means that it is, on average, much longer before cooperation is restored.

## REFERENCES

- Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Harper Collins.
- Axelrod, Robert and Robert O. Keohane. 1986. "Achieving Cooperation Under Anarchy: Strategies and Institutions." in Kenneth A. Oye, ed. *Cooperation Under Anarchy*. Princeton NJ; Princeton University Press.
- Bremer, Stuart . 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965" *Journal of Conflict Resolution* 26 (June 1992) 309-41.
- Bueno de Mesquita, Bruce and David Lalman. 1992. *War and Reason: Domestic and International Imperatives* (Yale University Press: New Haven, Ct 1992).
- Bueno de Mesquita, Bruce and Randy Siverson. 1995. "War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability." *American Political Science Review* 89(December) pp 841-55.
- Bueno de Mesquita, Bruce., James D. Morrow, Randolph M. Siverson and Alastair Smith. 1998. "An Institutional Explanation of the Democratic Peace" Manuscript.
- Cowhey, Peter F. 1993. "Domestic Institutions and the Credibility of International Commitments: Japan and the United States." *International Organization* 47: 299-326.
- Dixon, William. 1994. "Democracy and the Peaceful Settlement of International Conflict" *American Political Science Review* 88 (March 1994): 14-32.
- Eyerman, Joe and Robert A. Hart. 1996. "Audience Cost: Democracy Speaks Louder than Words" *Journal of Conflict Resolution* 40(4 Dec.):597-616.
- Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review*. 88: 577-592.
- Fearon, James D. 1998. "Bargaining, Enforcement, and International Cooperation." *International Organization* 52: 269-305.
- Fudenberg, Drew and Eric Maskin. 1986. "The Folk Theorem in Repeated Games

with Discounting and with Incomplete Information," *Econometrica* 54: 533-545.

Gaubatz, Kurt Taylor. 1996. Democratic States and Commitment in International Relations. *International Organization* 50: 109-139.

Leeds, Brett Ashley. 1998. "Domestic Political Institutions, Credible Commitments and International Cooperation." Manuscript, Florida State University.

Levy, Jack. 1988 "Domestic Politics and War" *Journal of Interdisciplinary History* 18 (Spring 1988).

Mansfeld, Edward D. 1998. "The Proliferation of Preferential Trading Agreements." *Journal of Conflict Resolution* 42: 523-43.

Mansfeld, Edward D., Helen Milner and B. Peter Rosendorff. 1998. "Why Democracies Cooperate More: Electoral Control and International Trade Agreements." Paper delivered at the annual meeting of the American Political Science Association, Boston MA., Sept. 3-6, 1998.

Maoz, Zeev and Nazrin Abdolali. 1989. "Regime Types and International Conflict, 1816-1976" *Journal of Conflict Resolution* 33 (March 1989): 3-36.

Maoz, Zeev and Bruce Russett, "Normative and Structural Causes of the Democratic Peace," *American Political Science Review* 87 (September 1993):624-38.

Martin, Lisa L. 1993. Credibility, Costs and Institutions: Cooperation on Economic Sanctions." *World Politics* 45: 406-432.

McGillivray, Fiona. 1998. "How Voters Shape the Institutional Framework of International Agreements" in *Strategic Politicians, Institutions, and Foreign Policy*, edited by Randolph M. Siverson. p.79-96.

Milner, Helen. 1992. "International Theories of Cooperation Among Nations: Strengths and Weaknesses." *World Politics* 44: 466-496.

Milner, Helen V. 1997. *Interests, institutions, and information : domestic politics and international relations*. Princeton, NJ : Princeton University Press.

Milner, Helen and Peter Rosendorff. 1997. "Democratic Politics and International

Trade Negotiations." *Journal of Conflict Resolution* 41: 117-146.

Morrow, James D., 1994. Modeling the forms of international cooperation: distribution versus information *International Organization*. v. 48 Summer '94 p. 387-423.

Oneal, John R., and Bruce Russett. 1997. "The classical liberals were right: democracy, interdependence, and conflict, 1950-1985" *International Studies Quarterly*. 41(June): 267-93.

Putnam, Robert. 1988. "Diplomacy and Domestic Politics; the Logic of Two Level Games." *International Organization* 42: 427-460.

Ray, James Lee 1995. *Democracies in International Conflict* University of South Carolina Press, Columbia, SC.

Remmer, Karen. 1998. "Does Democracy Promote Interstate Cooperation?" *International Studies Quarterly* 42: 25-52.

Rheinhardt, Eric. 1996. "Posturing Parliaments: Ratification, Uncertainty, and International Bargaining." Ph.D. Dissertation, Columbia University.

Russett, Bruce M., John R. Oneal and David Davis 1998. "Third Leg of the Kantian Tripod for Peace." *International Organization* 52(30):441-68.

Smith, Alastair. 1995. "Alliance Formation and War," *International Studies Quarterly* 39(4): 405-425.

Smith, Alastair. 1998a. "International Crises and Domestic Politics" *American Political Science Review*. 92(Sept. 3): 623-638.

Smith, Alastair. 1998b. "The Effect of Foreign Policy Statements on Foreign Nations and Domestic Electorates." in *Strategic Politicians, Institutions and Foreign Policy*, edited by Randy Siverson, Ann Arbor MI: University of Michigan Press. pp. 221-254.

Smith, Alastair and David Hayes. 1997. "The Shadow of the Polls: Electoral Effects On International Agreements" *International Interactions* 23(1):79-108.