When Do (Explicit) Threats Of Sanctions Work?

Stine Aakre
Master Thesis in Political Science
University of Oslo
Department of Political Science
Spring 2006

Acknowledgements

I wish to express my deepest appreciation to my thesis advisor Professor Jon Hovi, whose instruction and advice throughout this project proved invaluable. I would also like to thank Professor Geir B. Asheim for reviewing parts of this thesis, and for his helpful comments and suggestions. Many thanks are also due to Dr. Hilde J. Skorpen for initial support and encouragement. Finally, I am indebted to Sunniva F. Meyer for reading and commenting on drafts of this thesis as they developed. Responsibility for any faults is entirely my own.

Word count: 20 533

Stine Aakre
Oslo, May 2006

Table of Contents

1. Introduction

Are economic sanctions an effective foreign policy tool, and if so, under what conditions? This question has been the topic of several empirical and theoretical studies in the past decades. Quite a few of these have concluded that economic sanctions are largely ineffective (Morgan and Schwebach 1997; Pape 1997; Pape 1998). In one of the most influential quantitative studies of economic sanctions, Gary Hufbauer, Jeffrey Schott and Kimberly Eliott (1990: 93) report a success rate of only 34 percent in sanctions episodes spanning the period 1914 to 1990. After a careful reconsideration of the cases contained in this study, Robert A. Pape (1997:103) concludes that only 5 of these sanctions episodes were clearly successful.

Despite the seemingly poor empirical record, economic sanctions are nevertheless a prominent tool of statecraft (Drezner 2003: 643; Hufbauer 1999; Tsebelis 1990: 3). Three interpretations of this empirical anomaly have been offered. The first interpretation holds that sender states overestimate the prospects for success or that they have not yet learned to recognize instances in which sanctions are likely to be successful (Morgan and Schwebach 1997; Tsebelis 1990). A second interpretation offered is that sanctions fail internationally because they are employed for reasons other than extracting concessions from the targeted state, such as satisfying domestic interest groups or demonstrating a willingness to "do something" (Baldwin 1999/2000:102; Hufbauer, Schott and Elliott 1990: 105; Morgan and Schwebach 1997:28; Pape 1997:109). A third interpretation suggests that the conclusion that economic sanctions are largely inconsequential might be due to selection effects (Drezner 2003; Hovi 2001: 515 - 517; Lacy and Niou 2004; Morgan and Schwebach 1997: 46). More specifically, the abovementioned empirical studies have focused primarily on cases in which sanctions were actually imposed. However, as Lacy and Niou (2004: 25) comment, economic sanctions, "like punishments more generally, often work as threats even if they do not succeed when carried out. Sanctions are observed only when the threat of sanctions has failed."

1.1 Research Question

This study aims to contribute towards answering the question of *when* sanctions are likely to be effective in inducing policy changes in the targeted government. More specifically, the purpose of this thesis is to attempt to clarify the conditions under which disputes involving sanctions are likely to be settled at the threat stage, prior to the actual imposition of sanctions. Drawing on previous work by Jon Hovi, Robert Huseby and Detlef Sprinz (2005) in particular, but also on work by Dean Lacy and Emerson Niou (2004), I develop a game-theoretic model of economic sanctions in order to derive hypotheses about the conditions under which *explicit* threats of sanctions should succeed in extracting concessions from the targeted state. By distinguishing between implicit and explicit threats of sanctions, and more importantly by incorporating both types of threats, the model developed and analyzed here differs from the model presented by Hovi, Huseby and Sprinz and from that presented by Lacy and Niou.

Most conflicts involving sanctions, or threats thereof, are the result of the target state having violated some norm or standard of importance to the sender state, for example by forcefully acquiring disputed territory, illegally acquiring nuclear weapons, engaging in human rights violations, or by supporting terrorism (Hovi 2001; Hovi, Huseby and Sprinz 2005; Martin 1992; Schott 1998; Tsebelis 1990). What are the conditions under which one should expect explicit threats of sanctions to succeed in restoring compliance in such cases?

The purpose of this thesis is to attempt to establish a set of conditions under which explicit threats of sanctions work in cases where the target state has a desire to violate a norm or standard of importance to the sender state. In such cases, it seems reasonable to assume that the target state can foresee upon making a decision to violate an international norm that it could risk becoming the object of economic sanctions, or explicit threats thereof (Hovi 2001: 517). Accordingly, one should expect to observe states violating international norms only insofar as they are prepared to suffer the consequences. My research question is as follows:

What are the conditions under which an explicit *threat of sanctions could make the target state wish to reconsider?*

1.2 Method

This thesis aims to clarify the conditions under which threats of economic sanctions are likely to be effective in extracting concessions from the targeted state. The question of effectiveness will be approached using noncooperative game theory, and the purpose is to construct a formal model of economic sanctions and to draw on this to derive a set of empirical predictions with regards to the conditions under which explicit threats of sanctions can be effective.[1] Below, I discuss the basic assumptions of game theory in more detail. I conclude this section by briefly considering the implications of these assumptions as they are applied in this thesis.

Game theory can be defined as the "theory of rational behavior by two or more interacting rational individuals, each of them determined to maximize his own interests, whether selfish or unselfish, as specified by his own utility function" (Harsanyi 1986: 89). As indicated by Harsanyi's definition, game theoretic modeling focuses on two aspects of behavior in determining the outcome of a given game; rational choice and strategic interaction.

Game theory constitutes one type of rational choice theory, in the sense that game theory as a way to model strategic interaction relies on the assumption of rational choice. Game theory relies on this assumption when specifying the players' preferences for the different outcomes. The rationality assumption links the choices of decision makers to their goals and beliefs; decision makers have goals and pursue these goals efficiently by choosing among the actions available to them (Elster 1986:

---

[1] A noncooperative game is a game in which the players cannot make binding commitments, as opposed to a cooperative game, in which they can (Rasmusen 2001: 21).

3). The theory of rational behavior is a normative theory; "it tells us what we ought to do in order to achieve our aims as well as possible" (Elster 1986: 1). In my analysis of different sanctions scenarios, I will mainly be basing my discussion on Jon Elster's "thin theory" of rationality. Rationality in this sense involves three criteria: Preferences are transitive, beliefs are non-contradictory, and actions are consistent with beliefs and preferences (ibid.).

Second, Harsanyi's definition indicates that game theory is concerned with the interaction of two or more rational decision makers. As opposed to models of individual rationality, game theoretic models consider decision making in a strategic environment, in the sense that the actors in a game are seen to make (rational) choices while taking into account the range of possible choices of other players. In other words, game theory is concerned with the actions of decision makers who are aware that their actions affect each other (Rasmusen 2001: 11). With regards to this strategic decision making environment, game theory further assumes that the actors possess common knowledge about the rules of the game, including the rationality of other actors: "Common knowledge plays a fundamental role in the manner in which players' expectations take shape. It is a beginning point for modeling the strategic interest that is the essence of game theory." (Gates and Humes 1997: 9) In any strategic interaction, "information is common knowledge if it is known to all the players, if each player knows that all the players know it, if each player knows that all the players know that all the players know it, and so forth ad infinitum" (Rasmusen 2001: 47). The rules of the game consist of a list of players, their possible actions, the players' payoffs associated with each possible outcome, and the nature of the

information available to each player (ibid: 12).[2]

---

[2] With regards to the information available to the players, a distinction can be made between games of complete information and games of incomplete information. In the former, all the players' payoffs are common knowledge, while in the latter at least one player has private information about its payoffs (Morrow 1994: 63).

Given the assumptions of rationality and strategic interaction, the purpose of game theoretic modeling is to arrive at predictions with regards to which outcomes are stable, in the sense that no actor will have any incentives to deviate from them (Morrow 1994: 8). As Scott Gates and Brian D. Humes (1997: 10) summarize, "[r]ational choice provides a way of understanding an individual's preference for one outcome over another. Strategic interaction then shapes the action that is selected since it is the interaction of choice that leads to different outcomes associated with different payoffs. Game theoretic models then explain the structure and rules for how individuals' decisions and actions are interrelated and how different social outcomes come to be."

Given the aim of this thesis – i.e., to provide a theoretical treatment of the conditions under which explicit threats of sanctions can be expected to succeed in attaining states' foreign policy goals – the primary advantage of formal modeling as a research strategy is the rigor and precision of argument that such a method requires (Gates and Humes 1997: 6 – 7; Morrow 1994: 6 – 8). In developing a general model of economic sanctions, formalization requires that the assumptions with regards to whom the relevant actors are, what their payoffs and preferences are, what actions are available to them, and what information and beliefs they possess be stated explicitly (Milner 2004: 273; Watson 2002: 7). As Gates and Humes (1997: 6) comment when relating game theoretic modeling to King, Keohane and Verba's (1994) four characteristics of scientific research, this explicit statement of assumptions indeed makes the procedures public; "[t]he logic of the analysis is presented for all to see and evaluate. With formal models there is no glossing over the details."

In order to derive hypotheses about the conditions under which explicit threats of sanctions are likely to be effective, I conceptualize a typical dispute involving threats of sanctions by means of a simple two-player, one-shot game. Obviously, this model necessarily fails to capture many of the complexities of real-world cases involving economic sanctions. By allowing only two actors, and by assuming that the relevant actors are states, for instance, the role of domestic groups and of secondary sender

states in determining the outcome of a given sanctions scenario has been left out in the model developed here[3]. As Duncan Snidal (2004: 247) notes, the assumption that states are the relevant actors "already assumes that a state-centric perspective is a useful way to analyze international politics." Moreover, such an approach also requires the assumption that states' decision making processes "can reasonably be approximated as a coherent and consistent decision-making arrangement." (ibid.) By assuming a one-shot interaction, I necessarily ignore the role that reputation might play in interactions where one state threatens to impose sanctions on another. While the model constructed here is obviously simplistic, however, empirical examples will be used to assess the substantive relevance of its premises and predictions. Moreover, as Gates and Humes note (1997: 8), the necessary process of imposing simplifying assumptions in formal models is by no means unique to game theory – as with all social science modeling efforts, "[t]he objective should be to provide the best explanation with the simplest model". Accordingly, even if the model developed here necessarily abstracts from many important features of actual cases of economic sanctions, the model presented here seems an appropriate, and manageable, starting point.

1.3 Disposition

This thesis is divided into seven chapters. In Chapter 2, I review the diverse literature on economic sanctions and define key concepts relevant in the assessment of the extent to which sanctions are a successful foreign policy tool. In chapter 3, I present a general model of economic sanctions and discuss its assumptions. In chapters 4 and 5 this general model will be analyzed under different assumptions with regards to the strategic decision-making environment. In chapter 4, I analyze a sanctions game in which the disputants are assumed to possess complete information about all relevant

---

[3] However, as will be discussed in Chapter 3, I have chosen to incorporate the possibility of domestic pressure into my model.

aspects of the conflict. In chapter 5, I relax the assumption of symmetric information and determine the outcome of a sanctions game in which the target state is assumed to lack information about the type of sender with which it is interacting. In chapter 6, I discuss the main findings of this thesis, while Chapter 7 concludes this thesis.

2. The Sanctions Debate

While it seems a rather dominant opinion of the academic community that economic sanctions are largely ineffective, this thesis aims to contribute towards answering the question of *when* sanctions are likely to be effective, and more specifically, the conditions under which explicit threats of sanctions can succeed in inducing target compliance. In this chapter, I define key concepts and review the diverse empirical and theoretical literature on sanctions, as well as some of the main issues of contention with regards to determining the extent to which sanctions are a successful foreign policy tool.

2.1 Defining Economic Sanctions

According to David A. Baldwin (1985: 32), the concept of economic sanctions should be understood as encompassing the following three basic components:

1) *"Type of policy instrument used in the influence attempt, i.e., economic.*

2) *Domain of the influence attempt, i.e., other international actor(s).*

3) *Scope of the influence attempt, i.e., some dimension(s) of the target(s)' behavior (including beliefs, attitudes, opinions, expectations, emotions, and/or propensities to act)"*

With regards to the first component of Baldwin's definition - the instrument employed in a state's influence attempt - a distinction can be made between financial sanctions and trade sanctions. Examples of the former would be the blocking of financial flows and the freezing or expropriation of assets, while trade suspensions and tariff increases would constitute examples of the latter (Pape 1997: 93). Each type of economic pressure can be applied with varying intensity and scope. First of

all, the entire economy might be targeted, or just one critical sector (ibid.). Second, sanctions can be applied either by individual states, or by a coalition of states (Hovi 2001: 511)[4]. Trade sanctions can be further divided into embargoes and boycotts, where the latter can be either voluntary or imposed by law (ibid.).

The second component of Baldwin's definition of economic sanctions refers to the purpose of employing economic tools; to attain foreign policy goals by influencing the behavior of the targeted international actor(s). In the following, I employ the standard terminology whereby the term "target" is used to denote the state becoming the object of economic sanctions, or a threat thereof. Similarly, I employ the term "sender" to denote the state whose foreign policy goals are being pursued through the threat or imposition of sanctions (Hufbauer, Schott and Elliott 1990: 1).

With regards to the third component of the abovementioned definition of economic sanctions, Baldwin (1985: 32 – 33) argues that the scope of an influence attempt may be any dimension of the targeted state's behavior. Thus, according to Baldwin, what defines economic sanctions is not the type of foreign policy goal pursued, but rather the policy instruments employed in attaining them. Accordingly, the concept of economic sanctions should be understood as encompassing all aspects of economic statecraft, whether the foreign policy goal pursued is changing the capabilities, tariff policies, the level of economic welfare, or domestic or foreign policies in the targeted state (ibid.: 40).

Another definition of economic sanctions has been offered by Robert A. Pape (1997: 93 – 98). While Pape agrees that the first two components in Baldwin's definition should constitute part of any definition of economic sanctions, he prefers limiting the scope of the influence attempts significantly. While he agrees with Baldwin that states use economic tools against other states for a variety of purposes, he disagrees with the contention that economic sanctions should be understood as encompassing

---

[4] That is, unilateral and multilateral economic sanctions, respectively.

all goals pursued by means of economic pressure (ibid.). More specifically, Pape argues that there are three main types of economic pressure: economic sanctions, trade wars, and economic warfare[5]. The former concept is defined as economic pressure that "seek[s] to lower the aggregate economic welfare of a target state by reducing international trade in order to coerce the target government to change its political behavior" (ibid.). Given this definition, the following foreign policy goals sought by means of economic coercion would constitute examples of economic sanctions; the protection human rights, restoration of democracy and discouraging terrorist activities. Employing economic instruments in the attempt to weaken the targeted state's military capabilities, or in the attempt to affect the targeted state's international economic policies, however, do not constitute examples of economic sanctions according to Pape's definition. Given Baldwin's understanding of the concept of economic sanctions, the latter two types of influence attempts would constitute examples of sanctions, whereas Pape argues that such attempts should be considered instances of economic warfare and trade wars, respectively.

There are two main reasons why Pape prefers to limit the concept of economic sanctions to include only economic pressure with the goal of influencing a targeted state's political behavior. First, he argues that the different types of goals sought by the sender state are likely to have different determinants of success, and thus require separate theoretical investigations; "[f]or example, knowing whether a certain type of economic sanction often helps the coercer government's standing in the polls tells us little about whether the same sanctions or other instruments would be likely to succeed in coercing target states to change their political behavior." (Pape 1997: 95) Second, he argues that Baldwin's understanding of the concept creates problems

---

[5] A "trade war" is used to denote incidences in which "a state threatens to inflict economic harm or actually inflicts it in order to persuade the target state to agree to terms of trade more favorable to the coercing state". (Pape 1997: 94) The concept "economic warfare" refers to incidences in which a state "seeks to weaken an adversary's aggregate economic potential in order to weaken its military capabilities, either in a peacetime arms race or in an ongoing war." (ibid.)

when determining the extent to which sanctions are a successful foreign policy tool:

> *"beyond a certain point, excessively loose operationalization of dependent variables not only hinders theory building but departs from science altogether. Baldwin argues that the mere imposition of economic sanctions should automatically qualify as a success: "to make the target of an influence attempt pay a price for non-compliance is to be at least partially successful." If failure is defined to be impossible, the dependent variable cannot vary and the theory cannot be falsified." (Pape 1997: 95)*

As Pape points out, and as will be discussed below, how one defines economic sanctions also has implications for determining the extent to which sanctions are a successful foreign policy tool. Do economic sanctions 'work'?

2.2 Are Economic Sanctions a Successful Foreign Policy Tool?

Despite the prevalence of economic sanctions as a foreign policy tool, sanctions scholars have failed to reach consensus on whether sanctions 'work'. The problems inherent in extracting concessions from the target state by means of economic punishment are well-documented in the literature on sanctions (Drezner 2000; Elliott 1998; Hufbauer, Schott and Elliott 1990; Pape 1997). As Pape (1997: 93) summarizes:

"Pervasive nationalism often makes states and societies willing to endure considerable punishment rather than abandon what are seen as the interests of the nation, making even weak or disorganized states unwilling to bend to the demands of foreigners. In addition, states that have modern administrative capabilities can usually mitigate the economic damage of sanctions through substitution and other techniques. Finally, even when such capabilities are lacking and ruling elites are unpopular, they can still often protect themselves and their supporters by shifting the economic burden of sanctions onto opponents or disenfranchised groups."

What are the main issues of contention with regards to assessing the extent to which economic sanctions are a successful foreign policy tool?

2.2.1 Defining Success

In one of the most influential quantitative studies of economic sanctions, Hufbauer, Schott and Eliott (1990: 12) measure success by ''the extent to which the policy outcome sought by the sender country was in fact achieved''. This understanding of the concept of success corresponds to what David A. Baldwin (1999/2000: 90) refers to as the *effectiveness* of an undertaking involving sanctions. Some scholars would contend, however, that such a concept of success is excessively narrow. Part of this contention can be traced back to disagreements over what is to be included in the concept of economic sanctions, and more specifically, the purposes for which they are applied. As Baldwin (1985: 132) comments; "To view the use of economic statecraft strictly in terms of securing compliance with explicit and publicly stated demands is to load the dice in favor of failure. Third parties, secondary goals, implicit and unstated goals are all likely to be significant components of such undertakings". In other words, success should not be equated with effectiveness. More specifically, Baldwin (1999/2000: 89 – 92) argues that there are altogether five dimensions on which to measure the success of an undertaking involving economic sanctions: (1) effectiveness in achieving the sender's goals; (2) costs to the sender; (3) costs to the

target(s); (4) the type of foreign policy goal pursued by the sender; and (5) stakes for the target(s).[6] Accordingly, even if economic sanctions do not succeed in extracting concessions from the targeted state, they might still meet other important policy goals, and as such, be considered successful (Baldwin 1985: 371).

Other scholars have argued that the concept of success as applied by Hufbauer, Schott and Elliott is not rigorous enough. While Hufbauer, Schott and Elliott largely agree with Pape's definition of the concept of economic sanctions, they have also chosen to include what Pape refers to as instances of economic warfare when assessing the effectiveness of economic sanctions as a foreign policy tool (Hufbauer, Schott and Elliott 1990: 12; Pape 1997: 96). For instance, whereas Hufbauer, Schott and Elliott count as success the sanctions imposed by the United Kingdom against Argentina in 1982, in the sense that "sanctions made a useful contribution by rendering the target's military capability less effective than otherwise", Pape argues that "force was both sufficient and necessary to cause the outcome" (Elliott 1998: 53; Pape 1997: 99).

In the following, and in subsequent chapters, Pape's understanding of the concepts of economic sanctions and success will be employed in the attempt to provide an answer to the conditions under which explicit threats of sanctions can be expected to 'work'. The reason for this choice is that this thesis aims at providing an answer to the question of the conditions under which explicit threats of sanctions can be expected to succeed in restoring target compliance with a given norm or standard. Accordingly, economic sanctions will be considered successful to the extent that they succeed in extracting political concessions from the targeted government. (Hovi, Huseby and Sprinz 2005: 483). It should be noted, however, that whereas Baldwin's success criteria allows a judgment of the *utility* of employing economic tools of

---

[6] Ceteris paribus, economic sanctions are more successful "the greater the effectiveness, the lower the costs for the user, the higher the costs of non-compliance for the target, the higher the stakes, and the more difficult the undertaking."(ibid.: 92)

coercion, my approach only allows for an assessment of the conditions under which sanctions are *effective*. Baldwin's approach provides for a more comprehensive and policy-relevant assessment of sanctions, in the sense that one can make inferences about when sanctions *should* be employed, as opposed to other policy alternatives, such as military force; "[f]rom the standpoint of the logic of choice, any discussion of economic sanctions that fails to compare their likely cost-effectiveness with that of alternative courses of actions provides no useful policy guidance with respect to deciding whether sanctions should be used in a given situation" (Baldwin 1999/2000: 85-86).

2.2.2 Empirical Record of the Effectiveness of Economic Sanctions

In their study, Hufbauer, Schott and Elliott (1990: 93) report a success rate of only 34% in sanctions episodes spanning the period 1914 to 1990.[7] Based on the lessons drawn from the 115 cases considered, they provide a list of suggestions for increasing the prospects of success. Economic sanctions are more likely to be effective if: high policy goals are avoided; few sender countries are required to implement the measures; the target is already experiencing economic and political difficulties; the relations between sender and target are generally friendly; incremental application of sanctions are avoided; the costs inflicted on the target are heavy; the costs to sender(s) are modest; and sanctions are not accompanied by companion policies such as military operations or covert actions (ibid.: 94 – 105).

After a careful reconsideration of the cases contained in the abovementioned study Pape concludes that only about 5 percent of past sanctions attempts can properly be coined as successful, leading him to assert that "economic sanctions have little

---

[7] By their standards, successful cases of sanctions are those with an overall success score of 9 or higher. The score is arrived at by multiplying the policy result score by the sanctions contribution score, where 4 is the maximum score for each (Hufbauer, Schott and Elliott 1990: 92).

independent usefulness for pursuit of noneconomic goals" (Pape 1997: 93). According to this view, there is "no sound basis for even qualified optimism about the effects of sanctions" (ibid.). Moreover, these results lead Pape to raise the question of "whether decision makers who impose sanctions systematically overestimate the prospects of coercive success of sanctions".

Based on the abovementioned findings, and Pape's reexamination in particular, one would seem justified in contending that policymakers are 'fools' to consider such a tool of foreign policy (Morgan and Schwebach 1997). However, a few objections to this assertion deserve attention. First, as David Baldwin points out, even if sanctions are arguably often ineffective in extracting political concessions from the targeted government, it would be wrong to infer from this that it is always unwise to consider such a tool; "Even when the expectation of success is very low, the use of sanctions is justified if there is no policy alternative with a higher expectation of success. Sometimes policymakers must choose from a set of dismal alternatives".

Second, it should be noted that the abovementioned findings are based almost entirely on cases in which sanctions were actually *imposed*. However, economic sanctions are usually threatened before they are imposed (Hovi 2001; Hovi, Huseby and Sprinz 2005; Lacy and Niou 2004). Moreover, in most cases economic sanctions are imposed only if the targeted state refuses to make concessions. Accordingly, it might be that economic sanctions 'work' more often than the abovementioned studies suggest (Lacy and Niou 2004). The main findings of Daniel W. Drezner's (2003) study of U.S. economic coercion employed in trade, environmental, and labor disputes lends support to this assumption; disputes that are ended at the threat stage yield significantly larger concessions when compared to instances in which sanctions are actually imposed.

2.3 Distinguishing Between Threats of Sanctions and Imposed Sanctions

Game-theoretic approaches to studying economic sanctions that treat the actors as rational utility maximizers share a common prediction: to the extent that sanctions are effective at all, they tend to be so at the threat stage (Drezner 2003; Lacy and Niou 2004). Because the imposition of sanctions represents a deadweight loss of utility for both sender and target in the form of disrupted economic exchange, each has an incentive to reach an agreement before the actual imposition of sanctions (Drezner 2003: 645). If the sender prefers the status quo to imposing sanctions, then there should be no coercion attempt. If the target prefers conceding to incurring the cost of sanctions, it has an incentive to acquiesce before the imposition of sanctions (ibid.).

A threat can be defined as a "contingent assertion signaling an intention to hurt somebody – physically, economically or otherwise – unless that somebody acts in the way prescribed by the threatener." (Hovi 1998: 11) A threat is effective if it "causes the target to change its behavior in accordance with the threatener's desires" (ibid: 13). In order to be effective, a threat must be: (1) relevant, in the sense that the target prefers to act contrary to the sender's desires in the first place, and that the target has the necessary freedom of action to change its policies in compliance with the demands being made; (2) sufficiently clear in conveying what measures are required of the target in order for it to avoid sanctions being imposed; (3) complete, in the sense that the imposition of sanctions is made contingent on the target's actions; (4) credible, which implies that the sender state must prefer to impose sanctions, should the target government fail to concede; and (5) sufficiently severe, that is, the target must prefer to back down, rather than to stand firm and suffer the consequences (ibid.). A threat is only effective insofar as it causes a target to acquiesce to the sender's demands, *without* sanctions being imposed (ibid.).

In this thesis, a distinction is made between implicit and explicit threats of sanctions (Hovi 2001: 10 – 11). Most conflicts involving sanctions are the result of the target state having violated some norm or standard of importance to the sender state.

Accordingly, it seems reasonable to assume that in most cases where the sanctions tool is employed, the targeted government was able to foresee when implementing a given policy that it could risk having sanctions imposed or explicitly threatened (Hovi 2001: 517). Moreover, it seems reasonable to assume that a potential target government will only violate a given international norm insofar as it is prepared to suffer the expected consequences of implementing such policies (ibid.).[8] In other words, sanctions are usually imposed or explicitly threatened as a result of the targeted government having ignored an implicit threat of sanctions (ibid.).

An implicit threat of sanctions can be considered effective to the extent that the fear that sanctions would otherwise be imposed or explicitly threatened, deters a potential target government from transgressing a given norm or standard of importance to the sender state. If the implicit threat of sanctions has proven ineffective in deterring the target from transgressing a given norm or standard of importance to the sender, the latter may choose to *explicitly* threaten sanctions. Explicit threats can be considered effective to the extent that the fear that sanctions would otherwise be imposed compels the targeted state into changing its policies in compliance with the stated demands. If such a threat fails in extracting target concessions, the sender may choose to impose sanctions.

From the above discussion, economic sanctions are successful to the extent that: (1) implicit threats of economic sanctions are effective; (2) explicit threats of sanctions are effective; or (3) imposed sanctions are effective.

Using a game-theoretic model of two-sided incomplete information, Lacy and Niou (2004: 36) identify the conditions for (explicit) threats of sanctions to work: threats of sanctions will be effective if "the target is compliant, the cost of complying is low, and the costs of sanctions is high". A compliant target is defined as a target who

---

[8] It should be noted, however, that norm transgressions are not necessarily always intentional (Hovi, Huseby and Sprinz 2005: 18).

"would rather concede on issue X than suffer economic sanctions (ibid.: 30). Moreover, as Lacy and Niou's (ibid.: 27) findings suggest, "states that ignore the threat of sanctions are unlikely to change their behavior after sanctions are imposed". That is, if a (credible) threat of sanctions has failed to deter the target state, this is usually a good indication that it is not prepared to yield even if sanctions are imposed (Hovi, Huseby and Sprinz 2005: 482).

Yet, as Hufbauer, Schott and Elliott conclude after having assessed the universe of sanctions cases in the twentieth century, imposed sanctions sometimes *do* succeed in bringing about desired changes in policies in the targeted state. In their 2005 article *When Do (Imposed) Economic Sanctions Work*, Hovi, Huseby and Sprinz identify the conditions under which imposed sanctions work. Using a game-theoretical model of incomplete information, they identify two conditions for sanctions to work after they have been imposed. The first condition is that Target must initially (a) underestimate the impact of sanctions; or (b) miscalculate Sender's determination to impose sanctions; or (c) wrongly believe that sanctions will be imposed and maintained whether it yields or not. Second, Target's misperceptions must be corrected after sanctions are imposed. In other words, Target lacks information about aspects relevant to its decision-making at the time of deciding to violate a norm, and such information is only revealed through the process of sanctions imposition. However, as this information is revealed, Target discovers that it is not prepared to stand firm and suffer the consequences.

Hovi, Huseby and Sprinz also identify the conditions under which *implicit* threats of sanctions work; whenever the target state is sufficiently convinced that the sender would otherwise prefer to impose potent sanctions, it prefers to not violate the norm. Yet, sometimes such implicit threats of sanctions fail to deter targets from violating a given norm or standard. What are the conditions under which an explicit threat could make the target wish to reconsider?

Drawing on previous work by Hovi, Huseby and Sprinz (2005) and by Lacy and Niou (2004), this thesis aims at clarifying the conditions under which *explicit* threats

of sanctions can work in situations where the implicit threat of sanctions has failed to deter the target. While the model developed and analyzed here to a large extent resembles the model presented by Hovi, Huseby and Sprinz in particular, but also that of Lacy and Niou, there are some important differences. More specifically, whereas Hovi, Huseby and Sprinz only consider implicit threats of sanctions, I have chosen to introduce the opportunity to explicitly threaten sanctions. The latter opportunity is in keeping with the model presented by Lacy and Niou. Contrary to the model developed by Hovi, Huseby, and Sprinz, however, Lacy and Niou do not address the conditions under which implicit threats of sanctions will be effective. Thus, an important difference between the model developed here and the model presented by Lacy and Niou, is that I allow target the opportunity to avoid becoming the object of an explicit threat of sanctions. That is, whenever the target is not prepared to suffer the expected consequences, it can choose to not violate the norm in the first place. Thus, while their model predicts that explicit threats of sanctions will be effective if the target is compliant, I allow compliant targets the opportunity to not provoke an explicit threat of sanctions.

To sum up, whereas Hovi, Huseby and Sprinz only explore the conditions under which implicit threats of sanctions and imposed sanctions are likely to be effective, and Lacy and Niou only explore the conditions under which explicit threats of sanctions and imposed sanctions are likely to be effective, the model analyzed here incorporates all three types or sanctions effectiveness. Compared to Hovi, Huseby and Sprinz, I consider the implications of adding a choice of whether to launch an explicit threat of sanctions on the part of the sender state. And compared to Lacy and Niou, I consider the implications of assuming that prior to a threat, a potential target state can avoid becoming the object of an explicit threat of sanctions.

In the model developed here, the following conditions must be met in order for an explicit to be effective: (1) the target state must prefer to violate an international norm; (2) the sender state must prefer to threaten potent sanctions; and (3) the target state must prefer to back down given an explicit threat of potent sanctions, rather than

stand firm and risk having sanctions imposed. In order for the target state to prefer to violate an international norm, it must be prepared to suffer the expected consequences. In order for the target state to prefer to back down to an explicit threat of sanctions, it must somehow decide that it is *not* prepared to suffer the expected consequences. Given that the target has already made a (rational) decision to violate an international norm, what are the conditions under which an explicit threat of sanctions could make the target state wish to reconsider?

3. A Formal Model of Economic Sanctions

Cases of sanctions that end at the threat stage exhibit a significantly higher success rate than cases in which sanctions are actually imposed (Drezner 2003). In addition, if the sender state can attain target acquiescence without incurring the costs of imposing sanctions, such an outcome should be preferred to an outcome in which sanctions are actually imposed (Drezner 2003; Hovi 2001; Hovi, Huseby and Sprinz 2005; Lacy and Niou 2004). Accordingly, threatening sanctions may be as important as the actual imposition of sanctions as a strategy in disputes where the sender state wishes to attain concessions from the targeted state (Lacy and Niou 2004).

In the following chapter, I present a general game-theoretic model of economic sanctions. In subsequent chapters, this model will be analyzed under different assumptions about the nature of the information available to the disputants and about their preferences over the different possible outcomes of the conflict. The purpose of this modeling exercise is to attempt to clarify the conditions under which one should expect explicit threats of economic sanctions to have an impact on the policies of the targeted government.

3.1 The Model

As discussed, in order for a threat of sanctions to be effective, it must be credible, sufficiently severe, relevant, clear and complete (Hovi 1998). In keeping with Hovi, Huseby and Sprinz (2005) and with Lacy and Niou (2004), I treat the latter three requirements as given, in the sense that: (1) the target is assumed to derive a benefit from acting contrary to the sender's wishes; (2) the target is assumed to know that, in order to avoid having sanctions imposed, it must concede on the disputed issue; and (3) the target is assumed to know that sanctions will not be imposed if it backs down to a threat of sanctions.

The model of economic sanctions developed by Hovi, Huseby and Sprinz states the conditions under which the target can be expected to comply with the sender's demands only after sanctions are imposed. In their model, there are two other possible outcomes; target compliance without sanctions being imposed, and sanctions imposition resulting in stalemate. The concern of this thesis is with the former of these two outcomes, in the sense that an attempt will be made to establish the conditions under which one should expect compliance prior to the actual imposition of sanctions. However, whereas Hovi, Huseby and Sprinz only incorporate an implicit threat of sanctions in their model, I assume that the sender state has an opportunity to *explicitly* threaten economic sanctions prior to their actual imposition.[9] In this respect, the model bears resemblance to the model developed by Lacy and Niou in their 2004 article *A Theory of Economic Sanctions and Issue Linkage: The Roles of Preferences, Information and Threats*. Yet, there are some important differences between the model developed here and the model presented by Lacy and Niou.

First of all, in the model developed here it is assumed that economic sanctions will only be explicitly threatened insofar as a potential target government has transgressed an international norm that the sender state wants it to respect (Hovi 2001; Hovi, Huseby and Sprinz 2005; Martin 1992: 16; Schott 1998; Tsebelis 1990). While Lacy and Niou agree with the assumption that "any incidence of sanctions involves a dispute on at least one issue as well as a conflict over the sanctions" their model is based on the assumption that most conflicts begin with a threat by the coercer (Lacy and Niou 2004: 26 – 28). Thus, implicitly, they are assuming that the targeted state is

---

[9] With regards to the three main types of sanctions success discussed in the previous chapter, this means that Hovi, Huseby and Sprinz only focus on the first and third types (i.e., instances in which implicit threats of sanctions are effective, and instances in which imposed sanctions are effective). Thus, in their model, if the target decides to violate a given norm, the sender state is assumed to have a choice of either imposing or not imposing sanctions. In my model, however, I assume that prior to such a choice by the sender, it has the option of explicitly threatening (or not threatening) sanctions.

already acting contrary to the sender's wishes. The latter assumption is not explicitly discussed in their article, but it follows from the preference rankings of the sender state; the outcomes in which (a) the target state backs down to an explicit threat of sanctions and (b) the target state yields to imposed sanctions, are both preferred to the status quo (ibid.).[10] In the model developed here, however, I allow target the opportunity of avoiding a conflict with the sender if it is not prepared to suffer the (expected) consequences of such a conflict. Thus, whereas the sequence of strategic moves in their model begins with a decision by the sender of whether to initiate a conflict by threatening economic sanctions, I assume instead that it is the target state who initiates a conflict by violating a norm or standard that sender wants it to respect.

One implication of this additional assumption is that the model analyzed here also allows for sanctions effectiveness in the sense that a potential target government might be deterred from implementing policies contrary to the sender's wishes, whereas in the model developed by Lacy and Niou, sanctions are only effective to the extent that: (a) an explicit threat of sanctions succeeds in extracting target concessions; and (b) imposed sanctions succeed in extracting target concessions.

Second, while Lacy and Niou only allow for a binary choice at the threat stage; that of either threatening or not threatening sanctions, I have chosen to expand the policy options for the sender by assuming that it can choose to not threaten economic sanctions, threaten lenient sanctions or threaten potent sanctions (Hovi, Huseby and Sprinz 2005: 491). The reason why I have chosen to include three policy options instead of two is that this enables me to take into account the possibility that the target's response to a threat of sanctions might be contingent on the exact nature of the measures threatened. As previously discussed, a threat of sanctions will only be effective insofar as it is credible and sufficiently severe. More specifically, the target

---

[10] Moreover, it follows from the assumption that, for the target, the only difference between the status quo outcome and the outcome that results if sanctions are threatened but not imposed, is "a reputation gain for the target due to resisting the coercer's demands." (ibid.: 30)

will only back down if the costs that would otherwise be incurred outweigh the benefits associated with violating the norm. Thus, allowing more than one type of sanctions threat makes it possible to take into consideration that the target might be prepared to back down to some threats, while preferring to stand firm to others, and that what measures the sender can credibly threaten determines the response of the target (ibid.: 491). Lacy and Niou, on the other hand, approach the possibility that a potential target government might not always be prepared to back down to a threat of sanctions by assuming that the target could be one of two possible types; one which is never prepared to back down to a threat of sanctions, and one which is.

In terms of the sequence of decisions and payoffs, the only difference between the model presented here and that presented by Hovi, Huseby and Sprinz (2005) is that I allow Target the opportunity to explicitly threaten sanctions prior to imposing them. Accordingly, the following presentation draws heavily on their article.

3.2 The Sequence of Decisions

Figure 1 illustrates the sanctions game between two players, Target and Sender. Target begins the game by deciding whether to violate a norm or standard of importance to Sender, or to remain in status quo by not violating the norm. If Target violates the norm, Sender must choose whether to threaten economic sanctions. If Sender does nothing, the game ends and Target continues to act in violation of the norm. Alternatively, Sender may threaten to impose either lenient or potent sanctions.

If Sender threatens sanctions, Target faces two options; that of either standing firm or backing down. If Target backs down, the game ends and compliance is restored. If Target stands firm to a threat of sanctions, Sender must either carry out its threat out or renege.

If Sender does not impose sanctions, the game ends and Target continues its noncompliance. If sanctions are imposed, Target is faced with the options of either yielding or not yielding.
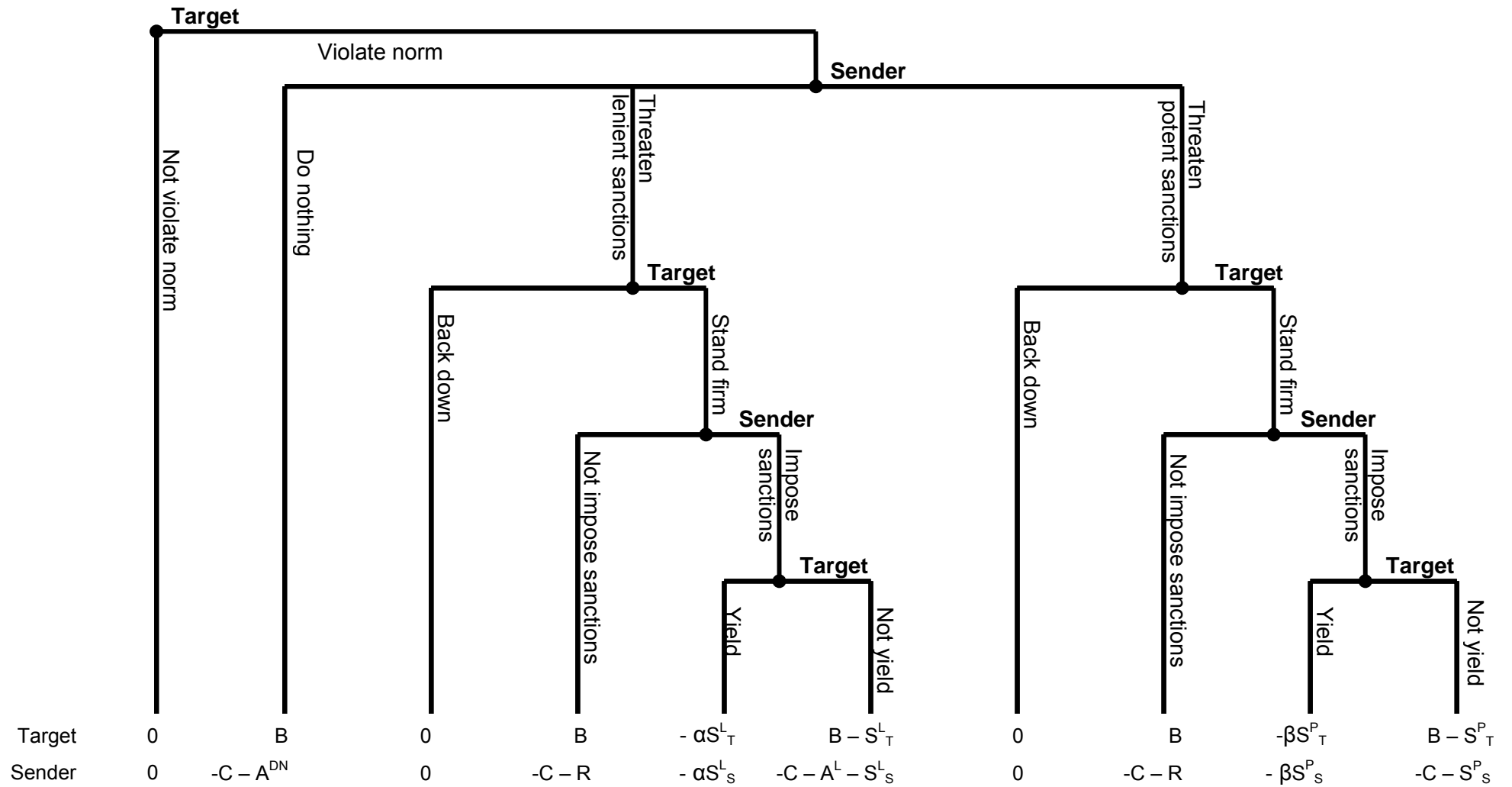
Figure 1 A Model of Economic Sanctions

Model adapted from Hovi, Huseby and Sprinz (2005: 491)

The Sanctions Game – Key

B	Target's benefits associated with violating the norm
C	Violation cost incurred by Sender if Target violates the norm and does not subsequently acquiesce
A	Audience costs incurred by Sender if it chooses action a after Target has violated the norm ($a = dn, L$)
R	Costs to Sender associated with reneging on a threat
S	Sanctions costs incurred by state j if sanctions are of type t ($j= S, T; t= L, P$)
α, β	Constants ($0 < (\alpha, \beta) < 1$)

3.3 Discussion of the Model's Payoffs

In the following paragraphs, I consider the different types of costs and benefits associated with the different possible outcomes of the game in more detail.

3.3.1 Benefits

It will be assumed that Target derives some benefit from violating a given norm or standard; $B > 0$. For example, Target might gain certain advantages by acquiring a weapon of mass destruction or some contested territory, or by violating human rights in its dealings with domestic groups. In keeping with Hovi, Huseby and Sprinz (2005: 492), it will be assumed that Target gets to enjoy this benefit only insofar as it does not subsequently acquiesce to Sender's demands. It could be argued, however, that this assumption is not always reasonable. For instance, one could imagine that under certain conditions, states might derive at least some benefits by violating a norm, even if it is only short term. Despite this objection, the abovementioned

assumption to a large extent seems an excusable one – the purpose of this modeling exercise is to attempt to establish a set of general conditions for explicit threats of sanctions to be effective.

3.3.2 Costs

In this model, four types of costs are assumed to be of relevance in determining the outcome of a given sanctions episode. First of all, I assume that a norm violation by Target will inflict certain costs on Sender, i.e. the norm in question is assumed to be of either moral or material importance to Sender. For instance, if the norm violation in question is the forceful acquisition of disputed territory, the cost incurred might be in the form of increased insecurity or loss of natural resources. C will be used to denote such costs, and it is assumed to always be greater than zero (C > 0). Like Hovi, Huseby and Sprinz (2005: 492), I assume that this cost is incurred only if Target does not subsequently acquiesce.

Second, given a norm violation, there might exist international or domestic pressure on Sender to take a stance on the matter (Baldwin 1985: 96; Kaempfer and Lowenberg 1988: 786). Thus, if Target fails to concede on the disputed issue, Sender suffers audience costs, A, if its actions are perceived as inadequate.[11] More specifically, it is assumed that audience costs will be incurred unless Sender imposes potent sanctions, given that Target is acting in violation of a norm and refuses to acquiesce to Sender's demands. By including such a component in the model, one will also be able to take into account situations in which the economic costs of sanctions are outweighed by the political gains of pacifying Sender audiences. Furthermore, I assume that in the case of a norm violation, doing nothing is

---

[11] It could be argued, however, that audience costs might be also incurred if Sender's actions are considered excessively harsh. However, in the following analysis, such a possibility will not be discussed further, nor has it been taken into account in the model's payoffs.

considered worse than imposing lenient sanctions. Thus, to Sender, I assume that the costs are defined by the following inequality: $A^{DN} > A^L > 0$.

Unlike Hovi, Huseby and Sprinz, I have chosen to include Sender costs of reneging, R, into the model. The reason is that, whereas they only include implicit threats of sanctions, I have chosen to introduce the opportunity of explicitly threatening sanctions into the sequence of events. Thus, third, it will be assumed that there are costs involved if Sender reneges on an explicit threat of economic sanctions, should Target fail to back down; $R > 0$. It seems reasonable to assume that in such cases, the future credibility of the government will be undermined. As such, R can be said to reflect any damages to the Sender's reputation associated with explicitly threatening sanctions, and subsequently not imposing sanctions if Target fails to make concessions. This type of cost might also serve the purpose of adding to the credibility of a threat launched. As Schelling comments, insofar as a country's reputation for action is at stake, "this kind of face is one of the few things worth fighting over" (Schelling, quoted in Baldwin 1985: 99).

Finally, I assume that sanctions impose costs on both Sender and Target relative to the status quo.[12] First of all, the economic relationship that has existed between the contending parties will be severed. It seems rather unproblematic to assume that the economic relationship that is disrupted by sanctions should initially have been beneficial to both parties. The imposition of sanctions might also entail monitoring costs to Sender. Costs incurred by Target and Sender will be represented by S_T and S_S, respectively, where $S > 0$. Moreover, it is assumed that lenient sanctions are less

---

[12] As Hufbauer, Schott and Elliott (1990: 76) note, it could be argued that not all sanctions entail costs for the sender state. More specifically, if the sanctions in question involve aid suspension, aid reduction or the cutting of official credits, the sender might actually enjoy benefits due to reductions in budget expenditures. However, even in these instances, "the corollary loss of trade contacts may entail an economic burden, in the form of lost sales and jobs, on the sender country" (ibid.) In the following, I assume that Sender prefers to avoid the costs of sanctions if Target complies with the norm.

costly than potent sanctions for both Sender and Target. Given that Target does not yield, Sender's costs of lenient and potent sanctions are S^L_S and S^P_S, respectively ($0 < S^L_S < S^P_S$). The corresponding costs for Target are S^L_T and S^P_T ($0 < S^L_T < S^P_T$). If Target subsequently yields, however, it is assumed that the costs incurred will be smaller than if it were to stand firm (indefinitely). In this case, the costs incurred by Target is αS^L_T given that sanctions are lenient, and βS^P_T given potent sanctions. The factors α and β are introduced in order to take into account that the costs of imposed sanctions are assumed to be lower if Target yields than if it were to not yield; $0 < (\alpha, \beta) < 1$. For Sender, the corresponding costs are αS^L_S and βS^P_S.

3.4 Discussion of Assumptions With Regards to the Players' Payoff Rankings

What are the actors' preferences over the different possible outcomes of the model presented in Figure 1? Based on the above discussion, one can impose some reasonable restrictions with regards to the orderings of payoffs in the sanctions game. These assumptions will be retained throughout. Note that all payoffs are relative to the status quo outcome.

For Target, the best possible outcomes are those in which it violates the norm without sanctions being imposed. The outcomes that satisfy this condition are if Target violates the norm, and: (a) Sender does nothing; or (b) Sender threatens, but does not subsequently impose sanctions. In order for these outcomes to be comparable, it must be assumed that Target derives no additional benefits or costs by becoming the object of an explicit threat of sanctions.

Like Hovi, Huseby and Sprinz, I assume that Target is prepared to yield to potent sanctions, but not to lenient sanctions, implying that $S^L_T (1 - \alpha) < B < S^P_T (1 - \beta)$. While the benefits Target derives from violating a norm outweigh the costs of lenient sanctions, potent sanctions are considered so severe that Target would rather: (a) yield than not yield if potent sanctions are imposed; and (b) back down than stand

firm if potent sanctions are explicitly (and credibly) threatened. In addition, it will be assumed that Target will only violate the norm if it expects to benefit from doing so. This implies that if it is common knowledge that potent sanctions would otherwise be explicitly (and credibly) threatened, Target prefers not violating the norm to violating the norm.

For Sender, Target acquiescence without sanctions being imposed are the best possible outcomes in the conflict. There are two ways in which Sender might attain its most favored outcome: (a) if Target does not violate the norm (i.e., if an implicit threat of sanctions is effective); or (b) if Target violates the norm, but backs down prior to the imposition of sanctions (i.e., if an explicit threat of sanctions is effective). In order for these two outcomes to be comparable, it must be assumed that Sender derives no additional benefits or costs by explicitly threatening sanctions – effective, implicit threats and effective, explicit threats are considered equally successful.

In chapters 4 and 5, the model presented in Figure 1 will be analyzed under different assumptions with regards to the nature of the information available to the players, and with regards to their preferences.

4. The Sanctions Game Under Complete Information

The purpose of this and the subsequent chapter is to attempt to establish a set of conditions under which explicit threats of sanctions work in cases where the target has a desire to violate a norm or standard that the sender wants it to respect. In such cases, it seems reasonable to assume that a potential target government is able to assess the likely consequences of its actions, and more specifically, that a norm violation could result in retaliatory measures (Hovi 2001: 517). Accordingly, one should expect to observe states violating international norms only insofar as they are prepared to suffer the consequences. What are the conditions under which an explicit threat of sanctions could make the target wish to reconsider?

Given the assumptions of the model, a threat of sanctions would need to be credible *and* potent in order to be effective. How might the credibility and potency of an explicit threat of economic sanctions affect Target's choice of whether to acquiesce, and how might Target's information about these parameters affect its decision?

In this chapter, I analyze the sanctions game under the assumption of complete information.

4.1 Equilibria Under Complete Information

With regards to the potency aspect, it is assumed that Target is prepared to yield to potent sanctions but not to lenient sanctions ($S^L_T (1 - \alpha) < B < S^P_T (1 - \beta)$). Given this, the following conditions must hold in order for threats of sanctions to be considered credible:

$$R > \beta S^P_S - C \tag{1}$$

$$R > A^L + S^L_S \tag{2}$$

The first inequality defines the condition under which Sender is prepared to impose potent sanctions, while the second relates to the condition under which Sender is prepared to impose lenient sanctions. Conversely, empty threats of sanctions are defined by the following inequalities:

$$R < \beta S^P_S - C \tag{3}$$

$$R < A^L + S^L_S \tag{4}$$

Inequality 3 defines the condition under which Sender cannot credibly threaten to impose potent sanctions, while inequality 4 defines the condition under which Sender cannot credibly threaten lenient economic sanctions.

Given the above sets of conditions, there are altogether four different possible sanctions scenarios to consider:

A. Both types of sanctions threats are credible (conditions 1 and 2 hold);

B. Only a threat of potent sanctions is credible (conditions 1 and 4 hold);

C. Only a threat of lenient sanctions is credible (conditions 2 and 3 hold);

D. Neither type of sanctions threat is credible (conditions 3 and 4 hold).

The outcomes associated with each sanctions scenario will be determined below.

4.1.1 Sanctions Scenario A: Both types of sanctions threats are credible

If $R > \text{Max } [(\beta S^P_S - C), (A^L + S^L_S)]$, the costs of reneging are so significant that Sender would rather impose sanctions if a threat of sanctions were to prove ineffective. Given a threat of potent sanctions, Target prefers backing down to standing firm and having sanctions imposed, since $\beta S^P_T > 0$. If Sender threatens lenient sanctions, however, such measures will not be considered sufficiently severe by Target, since $B > S^L_T$. Accordingly, Target will not back down if lenient sanctions are threatened.

Given that Target is only prepared to back down to a threat of potent sanctions, Sender prefers launching such a threat to the other two alternatives of action, since costs are otherwise incurred.[13]

---

[13] That is, if Sender threatens lenient sanctions, it receives a payoff of $(-C - A^L - S^L_S)$, which is less than the payoff obtained if Sender threatens potent sanctions; zero. Similarly, if Sender does nothing,

Given the assumption that Target will only violate the norm if it expects to benefit from doing so, the subgame perfect equilibrium of a sanctions scenario in which conditions 1 and 2 hold is that Target does not violate the norm, and that Sender would otherwise credibly threaten to impose potent economic sanctions.

Given conditions 1 and 2, Sender can credibly threaten sanctions which would also be considered sufficiently severe by Target. An explicit threat of sanctions would be effective in the sense that, if Sender were to announce the imposition of (potent) sanctions, Target would prefer to back down. However, given complete information, Target can foresee that a credible threat of potent sanctions would be forthcoming if it should decide to violate the norm. Accordingly, when conditions 1 and 2 hold, Target is effectively deterred from transgressing the norm, and no explicit threat of sanctions will be observed. [14]

4.1.2 Sanctions Scenario B: Only a threat of potent sanctions is credible

If $A^L + S^L_S > R > \beta S^P_S - C$, the costs of reneging are still so great that Sender would rather impose sanctions if a threat of potent sanctions proves ineffective. However, Sender would rather incur the costs of reneging than impose lenient sanctions, as lenient sanctions would not work and would remain in place indefinitely.

---

it receives a payoff of $(-C - A^{DN})$, which is less than can be obtained if Sender threatens potent sanctions.

[14] Compared to cases where sanctions are explicitly threatened or imposed, and the latter in particular, this type of sanctions effectiveness would be generally more difficult to establish empirically. More specifically, what must be established is that a potential target government is acting in compliance with a given norm or standard *because* sanctions would otherwise likely be imposed or explicitly threatened (Achen and Snidal 1989: 162-163). As Hovi, Huseby and Sprinz (2005: 490) point out, states act in compliance with international norms "for a variety of reasons, many of which have nothing to do with sanctions."

Again, a threat of potent sanctions is preferred to the other alternatives of action, since this would produce the best possible outcome for Sender (C, A^L, A^{DN}, $S^L_S > 0$). The subgame perfect equilibrium of a sanctions game in which conditions 1 and 4 hold is that Target does not violate the norm and that Sender would otherwise explicitly threaten potent sanctions. Again, Target is effectively deterred from violating the norm, and no explicit threat of sanctions will be observed.

The first and second scenarios are similar, in that what prevents a norm transgression is that Sender would otherwise credibly threaten to impose potent economic sanctions. As shown, as long as Sender's preferences are defined by condition 1, its optimal response to a norm transgression would be to threaten potent economic sanctions. Condition 1 holds as long as the costs of imposing potent economic sanctions are outweighed by the sum of the violation cost and the cost of reneging on a threat of sanctions.

4.1.3 Sanctions Scenario C: Only a threat of lenient sanctions is credible

If $A^L + S^L_S < R < \beta S^P_S - C$, only a threat of lenient sanctions is credible. Potent sanctions would be sufficiently severe to induce Target to yield if imposed. However, because the costs associated with imposing potent sanctions exceed the sum of the costs which would be incurred if potent sanctions are not imposed, Sender cannot credibly launch such a threat. Target would thus stand firm to a threat of potent economic sanctions. While a threat of potent sanctions would be considered sufficiently severe, it is not credible, and hence would not induce Target to comply with Sender's demands.

Given a threat of lenient sanctions, Target would also prefer to stand firm. While a threat of lenient sanctions would be credible, such a threat would lack the severity required in order to be effective, since $B > S^L_T (1 - x)$. While both threats would be ineffective in this scenario, a threat of lenient sanctions is preferred by Sender to a

threat of potent sanctions. This follows from the fact that the latter policy choice yields a payoff of (-C – R), which, by the definition of a threat being credible, does not exceed the payoff associated with the imposition of lenient sanctions. Accordingly, when Sender's preferences are defined by conditions 2 and 3, Sender prefers threatening lenient sanctions to threatening potent sanctions.

The equilibrium play in the sanctions scenario where conditions 2 and 3 hold depends on additional assumptions with regards to Sender's preferences over the different outcomes associated with either threatening lenient sanctions or doing nothing. Announcing the imposition of lenient sanctions is a more attractive policy option than doing nothing whenever: the audience costs associated with doing nothing are significant compared to the audience costs incurred when imposing lenient sanctions; and the costs of imposing lenient sanctions are not too great. Put differently, Sender prefers threatening lenient sanctions to doing nothing as long as $A^{DN} > A^L + S^L_S$. In the following, I assume that this inequality holds, thereby disregarding the possibility that Sender might be prepared to impose lenient sanctions but yet prefer to do nothing. In other words, I assume that Sender would rather impose lenient sanctions when its interests are violated than do nothing and allow Target to transgress the norm without incurring any costs.

Given the abovementioned assumption, the subgame perfect equilibrium is that Target violates the norm, and that Sender threatens, and subsequently imposes, lenient sanctions. Since lenient sanctions are not sufficiently severe to compel Target into compliance with Sender's demands, the result is a stalemate in which Target does not yield and sanctions remain in place indefinitely. In equilibrium, Target would suffer the costs of lenient sanctions, but these costs are outweighed by the benefits associated with violating the norm, and the net result is still beneficial

compared to the status quo. Accordingly, Target will not be prepared to make concessions. [15]

4.1.4 Sanctions Scenario D: Neither type of sanctions threat is credible

If $R < Min [(\beta S^P_S - C), (A^L + S^L_S)]$, Sender cannot credibly threaten sanctions. Whether Sender prefers launching an empty threat of sanctions to doing nothing depends on the relationship between the costs of reneging and the audience cost incurred by Sender if it responds to a norm violation by doing nothing.

Like Lacy and Niou (2004: 31), I assume that Sender prefers the outcome where sanctions are not threatened (i.e., where Sender does nothing) to the outcome where sanctions are threatened but not imposed. In terms of the parameters of the model, this implies that the costs associated with reneging on a threat to impose sanctions are sufficiently great to outweigh the costs incurred when sanctions are not threatened in the first place: $R > A^{DN}$.

Given the abovementioned assumption, the subgame perfect equilibrium is that Target violates the norm, and that Sender ends the game by doing nothing. In this

---

[15] The conflict that erupted in 1998 between the United States and Pakistan over the latter's decision to conduct nuclear tests might serve to illustrate this point. In June 1998, the United States announced the imposition of economic sanctions, as provided for by the Glenn Amendment (Elliott et al. forthcoming). The fact that sanctions had recently been imposed against India, as required by the same amendment, would lend support to the assumption that the threat was considered credible. While the economic impact of the imposed measures was quite significant, the option of abandoning the nuclear weapons program was considered prohibitively costly, as indicated by Prime Minister Nawaz Sharif: "Pakistan has been obliged to exercise the nuclear option because of the weaponization of India's nuclear program. This had led to the collapse of existing deterrence and had radically altered the strategic balance in our region" (ibid.).

scenario, one would observe a norm violation that is allowed to go unpunished, in the sense that Target is allowed to proceed with its behavior without costs whatsoever. [16]

4.2 Concluding Remarks

None of the scenarios analyzed above involve Target backing down to an explicit threat of sanctions. The following conditions must be met in order for an explicit threat of sanctions to be effective: (i) Target must prefer to violate an international norm; (ii) Sender must prefer to threaten potent sanctions; and (iii) Target must prefer to back down given an explicit threat of potent sanctions, rather than stand firm and risk having sanctions imposed.

In order for Target to prefer backing down to an explicit threat of sanctions, the threat must be potent and credible. In other words, Sender must be prepared to impose potent economic sanctions, implying that its preferences are defined either by conditions 1 and 2 *or* by conditions 1 and 4. The scenarios analyzed above reveal that whenever Sender is prepared to *impose* potent sanctions, it also prefers *threatening* potent sanctions to the other alternatives of action, since only a threat of potent sanctions can be effective in extracting Target concessions. However, if Sender can credibly threaten to impose potent sanctions, the outcome predicted by the model is that Target is deterred from transgressing the norm in the fist place.

---

[16] The lack of response to North Korea's decision to resume efforts to develop a nuclear weapons program might serve as an example (Elliott 2003: 2). While the U.S. has expressed a desire to take the matter to the United Nations Security Council, and even if the costs incurred by the sender states from economic disruption would probably not be very substantial (given North Korea's limited financial and trade relations with the outside world), any plans for multilaterally backed economic sanctions have yet to materialize (Elliott 2003: 5; Sanger 2005, April 25). North Korea's neighbors in particular have expressed concerns about employing the sanctions tool, as economic disruption could increase the risk of either a military response or of economic and political instability (Elliott 2003: 4).

If Sender cannot credibly threaten to impose potent sanctions, on the other hand, there are conditions under which an explicit threat of sanctions would be observed, but where such a threat fails in extracting Target concessions. The reason for this ineffectiveness is that the threat is not sufficiently severe (i.e., if conditions 2 and 3 hold). The third type of outcome predicted by the various equilibria in the complete information version of the sanctions game is that Target violates the norm and that Sender ends the conflict by doing nothing. This happens if Sender's preferences are defined by conditions 3 and 4.

With complete information about Sender's preferences, Target would never violate a norm and subsequently back down to an explicit threat of sanctions. The reason is that, whenever Sender can credibly threaten to impose potent sanctions, Target can anticipate that an explicit threat of potent sanctions would be forthcoming if it were to violate the norm. Thus, whenever it is common knowledge that Sender can credibly threaten potent sanctions, the outcome predicted is that Target will be deterred from violating the norm in the first place.

In the next chapter, I analyze a game in which Target is assumed to lack information about the type of sender with which it is interacting.

5. The Sanctions Game Under Incomplete Information

In this chapter, I determine the equilibria of the sanctions game under incomplete information. Assume that Target faces uncertainty with regards to the type of opponent with which it is interacting. More specifically, assume that Target faces uncertainty with regards to how Sender might respond to a norm violation. Does Target's uncertainty about Sender's type affect its behavior in the game, in the sense that one can expect to observe instances in which Target will back down to an explicit threat of sanctions?

As demonstrated in the previous chapter, the effectiveness of a threat of economic sanctions is determined by its potency and its credibility, in the sense that both requirements must be met in order for Target to prefer to back down to a threat of sanctions. In a game of incomplete information, Target still knows whether or not it considers a threat to be sufficiently severe, since this is determined by its own preferences. Given the assumptions of the model, only a threat of potent sanctions is sufficiently severe, implying that $S^L_T (1 - \alpha) < B < S^P_T (1 - \beta)$. However, Target is assumed to be unable to predict with certainty what measures Sender is actually prepared to impose. As a consequence, Target cannot judge the credibility of a threat with certainty. Thus, what is of importance to Target's decision of whether or not to make concessions, is its subjective assessment of what measures Sender would otherwise be prepared to impose.

5.1 Defining the Types of Sender

In the previous chapter, it was shown that Target's decision of whether or not to comply with the norm is affected by whether Sender can credibly threaten potent sanctions. Based on what measures Sender could be assumed to be prepared to impose, a total of four different scenarios were considered: one in which both threats

of sanctions are credible; one in which only a threat of potent sanctions is credible; one in which only a threat of lenient sanctions is credible; and finally, one in which neither threat is credible.

Under the assumption of complete information, we saw that if Sender is prepared to *impose* potent sanctions (i.e., when its preferences are defined by conditions 1 and 2 *or* by conditions 1 and 4), it also prefers *threatening* potent sanctions to the other alternatives of action, since costs would otherwise be incurred. In order for the same to hold under incomplete information, additional assumptions with regards to Sender's preferences over the different possible outcomes are required. The reason is that, under incomplete information, Target cannot judge the credibility of a threat with certainty, and hence, might under certain conditions choose to stand firm *even if* potent sanctions are threatened. Accordingly, we must make assumptions with regards to how Sender might rank the following outcomes: Target stands firm to a threat of potent sanctions, but yields as potent sanctions are imposed (i.e., the outcome in which the threat of sanctions proves ineffective, but where imposed sanctions are effective); Target stands firm to a threat of lenient sanctions, and does not subsequently yield (i.e., the outcome in which both the threat of sanctions and imposed sanctions prove ineffective); and Sender does not threaten sanctions. The assumption that the first outcome is preferred to the second requires that $A^{DN} > \beta S^P_S - C$. [17] The assumption that the first outcome is preferred to the third requires that $A^L > \beta S^P_S - S^L_S - C$.[18]

---

[17] If the abovementioned inequality does not hold, Sender prefers to do nothing, rather than (credibly) threaten of potent sanctions. This possibility will be ignored in the following.

[18] If the abovementioned inequality does not hold, Sender prefers to (credibly) threaten lenient sanctions, rather than (credibly) threaten potent sanctions. This possibility will be ignored in the following.

Given the above discussion, assume that Sender may be one of three possible types:

Table 5.1 Sender Types

Type I	• A *Type I* Sender is not prepared to impose sanctions, implying that conditions 3 and 4 hold. • Sender is *Type I* with probability *(1 – p – q)*
Type II	• A *Type II* Sender is only prepared to impose lenient sanctions, implying that conditions 2 and 3 hold. • Sender is *Type II* with probability *q*
Type III	• A *Type III* Sender is prepared to impose potent sanctions, implying that either conditions 1 and 2 *or* conditions 1 and 4 hold. • A *Type III* Sender prefers threatening potent sanctions if Target violates the norm. • Sender is *Type III* with probability *p*

5.2 Equilibria Under Incomplete Information

In the sanctions game of incomplete information, it is assumed that Sender knows its type while Target does not. This one-sided incomplete information is modeled by letting Nature make the first move, deciding with probability *p* that Sender is *Type III*, with probability *q* that Sender is *Type II*, and with probability *1- p- q* that Sender is *Type I*, where $p \geq 0$, $q \geq 0$, and $p + q \leq 1$. While Sender is assumed to be able to observe Nature's move, Target does not know Sender's type when making its decision whether to violate a norm or standard of importance to Sender. However, the probability distribution according to which Nature's decision has been made is assumed to be common knowledge. Moreover, Target is also assumed to know the preferences of each type of Sender.

For Target, yielding is preferred to not yielding if sanctions are potent, but not otherwise: $S^L_T (1 - \alpha) < B < S^P_T (1 - \beta)$.

If Sender is *Type I*, it is not prepared to impose sanctions, since $R < Min [(\beta S^P_S - C)$, $(A^L + S^L_S)]$. A Sender of *Type II* would be prepared to impose lenient sanctions, but not potent sanctions, as $\beta S^P_T - C < R < A^L + S^L_S$. A *Type III* Sender is prepared to impose potent sanctions, implying that $R > \beta S^P_S - C$.

For Target, the expected utility of standing firm given a threat of lenient sanctions exceeds the utility of backing down, given the assumption $B > S^L_T > 0$. Given a threat of potent sanctions, Target's choice of action depends on its belief that it is interacting with a *Type III* Sender, as only a *Type III* Sender would be prepared to impose potent sanctions. If Target knows that Sender is *Type III*, backing down is preferred to standing firm, since potent sanctions would otherwise be imposed. Conversely, if Target knows with certainty that Sender is either *Type I* or *Type II*, it will stand firm to a threat of potent sanctions, since such a threat would be empty. Given a threat of potent sanctions, let *u* denote Target's updated belief that Sender is *Type III*, *v* the belief that Sender is *Type II* and $(1 - u - v)$ the belief that Sender is *Type I*. For Target, the expected utility of standing firm given a threat of potent sanctions is determined as follows:

$$EU_{TARGET} \text{ (stand firm} \mid \text{ threat of potent sanctions)} = u (-\beta S^P_T) + vB + (1 - u - v) B$$

Backing down to a threat of sanctions yields a payoff of zero for Target. Thus, given a threat of potent sanctions, backing down is preferred to standing firm as long as:

$$u > \frac{B}{\beta S^P_T + B} \tag{5}$$

The higher the probability that Target attaches to Sender being *Type III*, the less likely that it will stand firm. The right-hand side of the inequality is always a number between zero and one, since $S^P_T > B > 0$ and $0 < \beta < 1$. This inequality thus puts a meaningful restraint on u.

Conversely, if:

$$u < \frac{B}{\beta S^P_T + B} \qquad (6)$$

Target prefers to stand firm given a threat of potent sanctions.

Limiting the remainder of this analysis to *pure strategy* equilibrium candidates, and disregarding dominated strategies, there are four remaining possible equilibrium candidates that need to be considered in more detail in the sanctions game of incomplete information.[19] All four involve threatening potent sanctions if Sender is *Type III*. The choice of policy if Sender is either *Type I* or *Type II*, however, is contingent on Target's belief that it is interacting with a *Type III* Sender. More specifically, a *Type I* Sender might either do nothing or threaten potent sanctions, while a *Type II* Sender might either threaten lenient or potent sanctions. Accordingly, the following combinations of Sender policy choices at the threat stage could potentially constitute part of a complete equilibrium strategy for Sender in the

---

[19] A pure strategy can be defined as a "strategy that does not include any probabilistic moves" (Morrow 1994: 66). A mixed strategy can be defined as "a probability distribution on the set of [a player's] pure strategies." (ibid.: 88)

incomplete information version of the model:

    A. Sender threatens potent sanctions regardless of type;

    B. A *Type I* Sender does nothing, a *Type II* Sender threatens lenient sanctions, while a *Type III* Sender threatens potent sanctions;

    C. A *Type I* Sender does nothing, while senders of *Type II* and *III* threaten potent sanctions;

    D. A *Type II* Sender threatens lenient sanctions, while senders of *Type I* and *Type III* threaten potent sanctions.

Below, I analyze each in order to determine the conditions under which, if any, such behavior could be considered rational in the sanctions game of incomplete information. Again, it should be noted that the following analysis is restricted to possible equilibrium candidates in pure strategies. A more detailed discussion of the model's mixed strategy equilibria is provided in Appendix 1 and 2.

5.2.1 Sanctions Scenario A: Sender threatens potent sanctions regardless of type

Consider a possible equilibrium candidate in which Sender follows a strategy which prescribes threatening potent economic sanctions regardless of type. If all Sender types respond to a norm violation by threatening potent sanctions, the updated probability, u, that Sender is *Type III* would simply equal the corresponding prior probability, p. Put differently, no additional information about Sender's type would be revealed through its choice of policy.

For Target, backing down is an optimal response to a threat of potent sanctions if $u > B/(\beta S^P_T + B)$. Thus, as long as Target's belief that Sender is actually prepared to impose potent sanctions satisfies this condition, announcing a policy of potent

sanctions would yield the best possible payoff for Sender regardless of its type. If Nature determines Sender to be *Type I*, it would receive a payoff of - A^{DN} – C if it does nothing, and by definition, this is less than could be achieved if it were to threaten potent sanctions; zero. Similarly, if Nature determines Sender to be *Type II*, announcing a policy of lenient sanctions would yield -A^L – S^L_S – C, whereas launching a threat of potent sanctions would yield zero. Accordingly, announcing a policy of potent sanctions is a best response given that Target will back down.

Given the above, Target's expected utility of violating the norm equals zero. Thus, since it has been assumed that Target will only violate the norm if it expects to benefit from doing so, the model has a pooling perfect Bayesian equilibrium in which Target does not violate the norm, and in which all Sender types would otherwise threaten potent sanctions. This is an equilibrium which holds as long as $u = p$ and $p > B/ (B + \beta S^P_T)$.

If $u < B/ (\beta S^P_T + B)$, a Sender strategy consisting of launching of a threat of potent sanctions regardless of type cannot constitute part of an equilibrium. As demonstrated, Target's optimal response given such a belief would be to stand firm to a threat of potent sanctions. Given that Target prefers to stand firm to a threat of potent sanctions, however, Senders of *Type I* and *Type II* would wish to announce policies other than the imposition of potent sanctions. More specifically, a *Type I* Sender would rather do nothing, while a *Type II* Sender would rather threaten lenient sanctions. This yields our next possible equilibrium candidate.

5.2.2 Sanctions Scenario B: A *Type I* Sender does nothing, a *Type II* Sender threatens lenient sanctions, and a *Type III* Sender threatens potent sanctions

Consider the second potential equilibrium candidate, in which Sender *does* reveal its type through its choice of policy at the threat stage. More specifically, assume that Sender follows a strategy which prescribes for it to do nothing if it is *Type I*, to threaten lenient sanctions if it is *Type II*, and threaten potent sanctions if it is *Type III*.

If Sender were to follow such a strategy, Target's updated belief that Sender is *Type III* given a threat of potent economic sanctions would equal one, which by definition exceeds the right-hand side of inequality 5. For Target, the optimal response in such a scenario would be to back down given a threat of potent sanctions, but not otherwise.

Facing a strategy which prescribes for Target to back down to a threat of potent sanctions, however, Senders of *Type I* and *Type II* would have an incentive to alter their strategies. As demonstrated in the previous scenario, as long as Target prefers to back down to a threat of potent sanctions, any strategy which prescribes for Senders of *Type I* and *II* to announce policies other than potent sanctions cannot occur in equilibrium. Accordingly, there are no pure strategy perfect Bayesian equilibria in which Sender does nothing if it is *Type I*, threatens lenient sanctions if it is *Type II* and threatens potent sanctions if it is *Type III*. Such a strategy would constitute optimal behavior given that Target will stand firm to a threat of potent sanctions; however, given such a strategy choice on the part of Sender, standing firm would not be optimal for Target.

5.2.3 Sanctions Scenario C: A *Type I* Sender does nothing, while senders of *Type II* and *III* threaten potent sanctions

In this scenario, *some* new information would be revealed through Sender's choice of policy. More specifically, given a threat of potent sanctions, the updated probability that Sender is *Type I* would equal zero. However, Target would still be unable to determine whether Sender is *Type II* or *Type III*, as they would both threaten potent sanctions.

As shown above, Target will stand firm to a threat of potent sanctions if the updated probability, *u*, that Sender is *Type III* is less than $B/ (\beta S^P_T + B)$. However, when this condition holds, *Type II* Sender would prefer to launch a (credible) threat of lenient sanctions rather than an empty threat of potent sanctions, since $R > A^L + S^L_S$. Conversely, if condition 5 holds, Target will back down rather than stand firm and risk having potent sanctions imposed. However, given this, a *Type I* Sender has an incentive to alter its choice of policy, since $A^{DN} + C > 0$. Accordingly, there are no pure strategy perfect Bayesian equilibria in which Sender does nothing if it is *Type I*, and threatens potent sanctions otherwise.

5.2.4 Sanctions Scenario D: A *Type II* Sender threatens lenient sanctions, while senders of *Type I* and *III* threaten potent sanctions

A strategy which consists of launching a threat of potent sanctions if *Type I* or *Type III*, and threatening lenient sanctions if *Type II*, is similar to the one just considered. Furthermore, by similar reasoning, following such a strategy cannot constitute equilibrium behavior on the part of Sender.

The only pure-strategy equilibrium identified thus far is a pooling perfect Bayesian equilibrium in which Target does not violate the norm, and in which all Sender types would threaten potent sanctions otherwise. While an explicit threat of sanctions would be effective if made, such a threat does not occur in equilibrium, since the anticipation that potent sanctions would otherwise be threatened is sufficient to deter Target from violating the norm in the first place.

In the game of incomplete information, Target does not need to be absolutely convinced that it is in fact interacting with a type of Sender who is prepared to impose potent sanctions in order to comply with the norm. Thus, whereas in the sanctions game of complete information, a threat of sanctions would have to be both sufficiently severe and credible, the latter requirement is somewhat relaxed when Sender is assumed to have private information about its type. More specifically, as long Target's belief, p, that it is interacting with a *Type III* Sender exceeds $B/(\beta S^P_T + B)$, Target is effectively deterred from violating the norm.[20]

When $p < B/(\beta S^P_T + B)$, the model does not have any perfect Bayesian equilibria in pure strategies. However, there are equilibria in mixed strategies, as described in Appendix 1 and 2. In equilibrium, Target violates the norm (with certainty), Sender may threaten potent sanctions, and Target may back down to an explicit threat of sanctions, given a belief that $u = B/(\beta S^P_T + B)$. If Target stands firm to an explicit threat of potent sanctions, only a *Type III* Sender will actually carry out the threat to impose sanctions. If Sender imposes (potent) sanctions, Target will yield (with certainty).

---

[20] Since the equilibrium just described is a pooling perfect Bayesian equilibrium, Bayes' rule requires that $u = p$.

5.3 Concluding Remarks

In order for an explicit threat of sanctions to be effective, Target must prefer to: (1) violate the international norm; (2) back down *after* the threat is issued; but (3) *before* sanctions are imposed. Under incomplete information, this requires that a threat of sanctions be considered both sufficiently severe and sufficiently credible. In order for the former requirement to be met, Sender would need to threaten potent sanctions. In order for the latter requirement to be met, Target must be sufficiently convinced that it is interacting with a *Type III* Sender. We have seen that backing down to a threat of potent sanctions is optimal as long as $u > B/(\beta S^P_T + B)$. However, given the assumptions of the model, Target will only violate the norm if it expects to benefit from doing so. Thus, if it is common knowledge that an explicit threat of sanctions would be effective, the outcome predicted by the model is that Target does not violate the norm in the first place.

6. Summary and Discussion of Findings

In order to explore the conditions under which explicit threats of sanctions are effective, the model developed in Chapter 3 was analyzed under different assumptions with regards to the nature of the information available to the disputants. I begin this discussion by briefly recapitulating the equilibrium results of this analysis.

Target will only violate the norm if the benefits associated with doing so exceed the costs incurred. Assuming complete information, this means that Target will violate the norm whenever: (1) Sender is not prepared to impose sanctions; or (2) Sender is prepared to impose sanctions, but the costs incurred by standing firm to a (credible) threat of sanctions do not exceed the benefits associated with violating the norm. Thus, as long as Sender cannot credibly threaten potent sanctions, Target will violate the norm. Moreover, to the extent that sanctions are explicitly threatened, they do not succeed in extracting concessions. If Sender can credibly threaten potent sanctions, the outcome predicted is that Target is effectively deterred from violating the norm.

Assuming that Sender has private information about its preferences, we found that if Target is sufficiently convinced that Sender can credibly threaten potent sanctions, the outcome predicted is that Target does not violate the norm. This is the only pure-strategy perfect Bayesian equilibrium in the model of incomplete information.

Given the assumptions of the model, economic sanctions, and threats thereof, must be potent in order to be effective in extracting Target concessions. In addition, a threat must be considered credible. Assuming complete information of all relevant aspects of the interaction, these requirements mean that Sender must be prepared to impose potent sanctions. Assuming incomplete information, threats of sanctions are effective to the extent that Target attaches a sufficiently high probability to Sender being prepared to impose potent sanctions. This means that as long as Target is sufficiently convinced that the sender with which it is interacting is in fact prepared to impose

potent sanctions, even empty threats of sanctions can be effective in extracting concessions. Thus, what distinguishes the sanctions game of complete information from the game of incomplete information is that, in the latter Sender will under certain conditions prefer to threaten potent sanctions *even if* it is not prepared to carry out the threat. Under complete information, a sender who cannot credibly threaten potent sanctions, would rather choose to not threaten sanctions, or threaten lenient sanctions.

Accordingly, with regards to the conditions under which explicit threats would be effective, we have that Sender must be prepared to impose potent sanctions, or Target must be sufficiently convinced that Sender is prepared to impose potent sanctions. Yet, common to both models is that they predict that, whenever an explicit threat of sanctions would be effective in inducing compliance, such a threat is not required. Conversely, whenever explicit threats of sanctions are observed, they will be ineffective in extracting Target concessions. The only exception is when the players employ mixed strategies. Below, I discuss the reasons why the model predicts that effective, explicit threats of sanctions should only be observed if the players adopt mixed strategies.

6.1 Are Explicit Threats of Sanctions Effecitve? If So, When?

In order for a threat of sanctions to be effective, it must be credible, sufficiently severe, relevant, clear and complete. Treating the latter three requirements as given, I have chosen to consider how the potency and credibility aspects might interact in determining the outcome of a given dispute by allowing three policy options for Sender (Hovi, Huseby and Sprinz 2005). As Hovi, Huseby and Sprinz (2005: 491) comment, this assumption allows for the possibility that "Target might violate an international norm and yet reconsider when it learns the true costs of sanctions". What are the conditions under which an explicit threat of sanctions can make Target wish to reconsider?

In the model developed here, the following conditions must be met in order for an explicit to be effective: (1) Target must prefer to violate an international norm; (2) Sender must prefer to threaten potent sanctions; and (3) Target must prefer to back down given an explicit threat of potent sanctions, rather than stand firm and risk having sanctions imposed.

Under complete information, an explicit threat of sanctions would be effective insofar as Sender is prepared to impose potent sanctions. However, since Target can foresee this, and since it has been assumed that Target will only violate an international norm if it expects to benefit from doing so, the model predicts that Target does not violate the norm. Put differently, whenever an explicit threat of sanctions would be effective in compelling Target to back down, the implicit threat that potent sanctions would otherwise be explicitly, and effectively, threatened is sufficient to deter Target from violating the norm in the first place. Conversely, the model predicts that to the extent that explicit threats of sanctions are observed, Target will not be prepared to make concessions, since such threats will not be considered sufficiently severe. Furthermore, whenever an explicit threat of sanctions proves ineffective, imposed sanctions will also fail in extracting Target concessions.

Similarly, in the sanctions game of incomplete information, it was demonstrated that whenever it is optimal for Target to follow a strategy which involves backing down to a threat of potent sanctions, the outcome predicted is that Target will not violate the norm in the first place. In order to prefer backing down to an explicit threat of potent sanctions in the game of incomplete information, Target must be sufficiently convinced that the threat is credible. However, since Target compliance is the best possible outcome for Sender, and since even an empty threat of sanctions would be effective given that Target is sufficiently convinced that Sender is in fact prepared to impose potent sanctions, Sender's best response is to threaten potent sanctions regardless of type if Target violates the norm. Accordingly, no new information about Sender's preferences can be revealed through its choice of policy. Moreover, since Target can foresee that a threat of potent sanctions would otherwise be

forthcoming, and that it is not prepared to stand firm and risk having potent sanctions imposed, Target chooses to not violate the norm. Again, given that $u = p$ and $p > B/(\beta S^P_T + B)$, an explicit threat of sanctions would be effective, but is not required in order to induce Target to comply with Sender's demands.

The abovementioned findings suggest that, whenever it is common knowledge that an explicit threat of sanctions would be effective (with certainty) – either in the sense that (a) Sender is prepared to impose potent sanctions, or in the sense that (b) Target *believes* that Sender is prepared to impose potent sanctions – Target will be effectively deterred from violating the norm, since potent sanctions would otherwise be explicitly, and effectively, threatened. In terms of the three abovementioned conditions for explicit threats to be effective, this means that whenever the second and third conditions are met, and this is common knowledge, Target prefers to *not* violate the norm. This raises an interesting question – what would happen if this is not common knowledge?

One possibility explored in this thesis, is whether the predictions derived from the model developed by Hovi, Huseby and Sprinz – namely that Target might violate an international norm, but reconsider as it learns the true costs of sanctions – might also apply to the threat stage. More specifically, could there be conditions under which Target might violate a norm because it mistakenly believed that Sender would not (credibly) threaten potent sanctions, but where it backs down to an explicit threat as it learns that Sender is in fact prepared to impose potent sanctions? If Target initially believed that Sender might either threaten lenient sanctions or not threaten sanctions at all, but comes to learn that it was wrong, an explicit threat of sanctions would be effective, provided that Target somehow becomes sufficiently convinced that Sender is prepared to carry out the threat. However, we have seen that in order for Sender to reveal information about its preferences at the threat stage which were not previously known to Target, Sender's response to a norm violation must be made contingent on its type. In other words, it must be that Sender only threatens potent sanctions if it is in fact prepared to impose such measures. Facing such a strategy, backing down to a

threat of potent sanctions would be optimal for Target. However, as demonstrated, as soon as it is optimal for Target to follow a strategy which prescribes backing down to a threat of potent sanctions, it is not optimal for Sender to follow a strategy which prescribes actions other than announcing the imposition of potent sanctions at the threat stage. Again, as long as it is common knowledge that Target will back down to a threat of potent sanctions (with certainty), announcing the imposition of potent sanctions is costless to senders of *Type I* and *II*, and indeed, is preferred to the other alternatives of action, since doing so ensures the best possible payoff in the sanctions game.

A second possibility could be if Target, instead of following a strategy which prescribes backing down (or standing firm) with certainty, plays a mixed strategy of standing firm or backing down to a threat of potent sanctions. Similarly, if Sender follows a mixed strategy, could there be instances in which Target violates a norm but backs down to an explicit threat of sanctions? While the model's mixed strategy equilibria have not received much attention in this analysis, the findings in Appendix 1 and 2 suggest that if Target assigns some probability to standing firm to a threat of potent sanctions, then Sender will sometimes be deterred from threatening potent sanctions if it is not prepared to carry out the threat. In equilibrium, Target violates the norm.[21] Moreover, in equilibrium, Target *may* back down to an explicit threat of potent sanctions.[22]

Thus far, it has been established that, if Target decides that it is in its best interest to violate the norm, it should only reconsider if it is somehow learns that Sender is more likely to be prepared to impose potent sanctions than it had originally estimated.

[21] Given a belief that *p,* the prior probability that Sender is *Type III,* is less than $B/(\beta S^P_T + B)$.

[22] Given a belief that *u,* the updated probability that Sender is *Type III,* equals $B/(\beta S^P_T + B)$. On becoming the object of an explicit threat of potent sanctions, Target upgrades the probability that Sender is *Type III* (conversely, if Sender does nothing or threatens lenient sanctions, *u* would equal zero). (Watson 2002: 287 – 288).

Moreover, in *pure strategies*, this can only happen if Sender follows a strategy which prescribes threatening potent sanctions only if it is in fact prepared to impose them. However, if Target follows a strategy which prescribes backing down to a potent threat of sanctions, but not otherwise, the optimal response for Sender is to threaten potent sanctions regardless of whether it is actually prepared to carry out the threat. Put differently, whenever it is common knowledge that Target will back down given a threat of potent sanctions, Target cannot be "surprised" at the threat stage. How, then, might a strategy combination consisting of violating the norm and backing down given a threat of potent sanctions on the part of Target, and threatening potent sanctions only if Sender is prepared to impose potent sanctions, constitute equilibrium play? The only way in which following a strategy which prescribes for Sender to threaten potent sanctions only if it is *Type III* at the threat stage is optimal, is when it is too costly for Sender to bluff about its intentions to carry out the threat. We have seen that whenever Target follows a strategy which prescribes standing firm to a threat of potent sanctions, Sender prefers to not threaten sanctions, or threaten lenient sanctions, if it is not prepared to carry out a threat of potent sanctions. This suggests that in order to observe an outcome in which Target violates an international norm and backs down to an explicit threat of sanctions, Sender must be deterred from bluffing about its intentions to carry out a threat of potent sanctions. The latter could happen if Sender is sufficiently (but mistakenly) convinced that Target will stand firm to a threat of potent sanctions. This possibility has not been explored in this thesis, as it would require the introduction of asymmetric information about Target's preferences. Moreover, since it has been established that Target will not violate the norm in the first place if it *knows* that Sender is prepared to impose potent sanctions, what is required is not only asymmetric information about Target's preferences, but two-sided incomplete information. However, the results produced here suggest that such a modelling effort might provide important contributions towards answering the question of the conditions under which explicit threats of sanctions can be expected to be effective in extracting Target concessions.

6.2 Theoretical Contribution

While the model developed and analyzed here resembles the model presented by Hovi, Huseby and Sprinz in particular, but also that presented by Lacy and Niou, there are some important differences. More specifically, whereas Hovi, Huseby and Sprinz only explore the conditions under which *implicit* threats of sanctions and *imposed* sanctions are likely to be effective, and Lacy and Niou only explore the conditions under which *explicit* threats of sanctions and *imposed* sanctions are likely to be effective, the model analyzed here incorporates all three types or sanctions effectiveness .

The predictions derived from the model analyzed in this thesis suggest that empirical studies that concentrate on cases in which sanctions were actually imposed when assessing the question of whether or not economic sanctions 'work', might potentially miss the real impact of sanctions as a foreign policy tool (Lacy and Niou: 38). More specifically, the model presented here, like that presented by Hovi, Huseby and Sprinz and Lacy and Niou, makes clear predictions about the conditions under which disputes are likely to be settled *without* sanctions being imposed – that is, the conditions under which the threat of sanctions will be sufficient in extracting concessions from the targeted state.

In the model developed by Lacy and Niou, the conditions under which (explicit) threats of sanctions will be effective are that "the target is compliant, the cost of complying is low, and the costs of sanctions is high" (Lacy and Niou: 36). A compliant target is defined as a target who "would rather concede on issue X than suffer economic sanctions (Lacy and Niou 2004: 30). The findings in the model developed here, however, suggest that, when explicit threats of sanctions are likely to be effective, the implicit threat that sanctions would otherwise be effectively threatened is likely to be sufficient in deterring Target from violating the norm in the first place. Conversely, when explicit threats of sanctions are observed, they are unlikely to succeed in extracting concessions from the targeted state. More specifically, the model predicts that the outcome in which Target backs down to an

explicit threat of sanctions, should only be observed if the disputants employ mixed strategies.

One reason why the predictions derived from the model developed here differs from the predictions derived from the model developed by Lacy and Niou, is that, like Hovi, Huseby and Sprinz, I have chosen to allow Target the opportunity of avoiding a conflict with Sender in the first place if it is not prepared to suffer the expected consequences of violating a given norm. That is, whenever Target is not prepared to suffer the consequences - in the sense that it attaches a sufficiently high probability to Sender being *Type III* – it can choose to *avoid* becoming the object of an explicit threat of sanctions. Moreover, this also means that Target will only become the object of an explicit threat of sanctions if it has already made a decision that it *is* prepared to suffer the expected consequences. This difference of assumptions probably imposes some restrictions on the conditions under which explicit threats of sanctions are likely to be effective in extracting Target concessions. While their model predicts that explicit threats of sanctions will be effective if the target is compliant, I allow compliant targets the opportunity to not provoke an explicit threat of sanctions.

The model developed by Hovi, Huseby and Sprinz (2005: 495) makes clear predictions about the conditions under which implicit threats of sanctions will be effective; norm adherence is predicted if there is a high probability that a norm violation would cause Sender to impose potent sanctions. When the implicit threat of sanctions is sufficient to deter Target from violating the norm in the first place, it does not make a difference whether or not Sender is assumed to have the opportunity to explicitly threaten sanctions prior to the actual imposition, since neither the explicit threat nor the actual imposition of sanctions is required in order to extract Target concessions.

In keeping with the findings of Hovi, Huseby and Sprinz, the model presented here supports the conclusion that with complete information about all relevant aspects of the interaction, imposed sanctions cannot be effective. Moreover, the predictions of the model developed here suggests that this also extends to explicit threats of

sanctions; with complete information, explicit threats of sanctions cannot succeed in extracting concessions from the targeted state. More generally, with complete information, adding an opportunity to explicitly threaten sanctions prior to actually imposing them does not affect the outcomes predicted.

Limited to pure strategy equilibrium outcomes, Target will concede only at the threat stage in the model analyzed by Lacy and Niou. Similarly, in the model presented here, to the extent that Target prefers to make concessions, it will do so without sanctions being imposed. However, this prediction only applies to *implicit* threats of sanctions, while Lacy and Niou's predictions relate to *explicit* threats of sanctions. The model analyzed here does permit outcomes where sanctions are explicitly threatened and Target backs down, but only if mixed strategies are involved. Like Lacy and Niou's findings, but unlike the findings of Hovi, Huseby and Sprinz, the model analyzed here only predicts instances of successful, *imposed* sanctions if the players employ mixed strategies. In the model developed by Hovi, Huseby and Sprinz, successful instances of imposed sanctions are predicted whenever Target initially believes that Sender is not prepared to impose potent sanctions, and where Sender is in fact prepared to impose potent sanctions. The model presented here generates only mixed strategy equilibria whenever Target considers it unlikely that Sender is prepared to impose potent sanctions. In equilibrium, Target will violate the norm and might sometimes back down to an explicit threat of sanctions, and sometimes stand firm. However, *if* Target stands firm to an explicit threat of potent sanctions *and if* Sender is prepared to carry out the threat, imposed sanctions will be effective (with certainty). In the model presented by Lacy and Niou, on the other hand, imposed sanctions may work, but they may also prove ineffective.

7. Conclusion

Empirical studies that focus exclusively on cases in which sanctions were actually imposed when assessing whether or not economic sanctions 'work', might potentially miss the real impact of sanctions as a foreign policy tool, since sanctions often work as threats, and since sanctions are usually only implemented if the threat of sanctions fails. With this as my point of departure, in this thesis I set out to explore the conditions under which threats of sanctions are likely to be effective in extracting concessions from the targeted state. Drawing on previous work of Hovi, Huseby and Sprinz (2005) and Lacy and Niou (2004), a game-theoretic model incorporating the three main stages of sanctions effectiveness was developed and analyzed. The predictions derived from this model lend support to the selection effects argument, in the sense that it identifies conditions under which disputes involving sanctions are likely to be settled *without* sanctions being imposed.

In order to be effective, a threat of sanctions must be potent. Second, the threat must be credible. Assuming complete information, this requires that the sender state must be prepared to impose potent sanctions. Assuming incomplete information, a threat of sanctions is effective insofar as the target is sufficiently convinced that the sender is prepared to impose potent sanctions

With regards to the conditions under which implicit threats of sanctions are successful, the model predicts that a potential target state will be effectively deterred from violating a given norm or standard of importance to the sender state if the former is sufficiently convinced that the latter would otherwise be prepared to impose potent economic sanctions. Yet, sometimes implicit threats of sanctions prove ineffective. What are the conditions under which an explicit threat of sanctions could make the target state wish to reconsider?

With regards to the conditions under which explicit threats of sanctions are likely to succeed in restoring compliance with international norms, the model does not predict

this outcome in pure strategies. In chapters 4 and 5, it was demonstrated that whenever explicit threats of sanctions would be effective in extracting concessions, the implicit threat of sanctions is adequate to deter the target from implementing policies in violation of a given norm or standard. Put differently, whenever an explicit threat of sanctions would be effective in compelling the target to back down, the implicit threat that potent sanctions would otherwise be explicitly, and effectively, threatened is sufficient to deter the target from violating the norm in the first place. Conversely, limited to pure strategy equilibrium outcomes, the model predicts that to the extent that explicit threats of sanctions are observed, the target will not be prepared to make concessions, since such threats will not be considered sufficiently severe. Furthermore, whenever an explicit threat of sanctions proves ineffective, imposed sanctions will also fail in extracting concessions.

The model developed in this thesis predicts that successful instances of explicit threats of sanctions, as well as successful instances of imposed sanctions, should only be observed if the disputants employ mixed strategies.

In chapter 6, I summarized and discussed the main findings of this thesis, particularly as they relate to the findings of Hovi, Huseby and Sprinz (2005) and Lacy and Niou (2004). Moreover, based on these findings, I also suggested ways in which future modelling efforts might provide important contributions towards answering the question of the conditions under which explicit threats of sanctions can succeed in restoring compliance with international norms.

Appendix 1 The Sanctions Game under Incomplete Information – a Mixed Strategy Perfect Bayesian Equilibrium (with $q = 0$)

For simplicity, assume that $q = 0$.

1. *Find the probability,* u, *of Sender being* Type III *that makes Target indifferent between standing firm and backing down to a threat of potent sanctions:*

   EU_{Target} (stand firm | threat of potent sanctions) $= u \, (-\beta S^P_T) + (1 - u) \, B$

   Backing down to a threat of potent sanctions yields a payoff of zero, with certainty. Target is indifferent between standing firm and backing down when:

   $0 = u \, (-\beta S^P_T) + (1 - u) \, B$

   $$u \quad = \quad \frac{B}{\beta S^P_T + B}$$

   The right-hand side of this equality is always a number between zero and one, since $S^P_T > B > 0$ and $0 < \beta < 1$. This equality thus puts a meaningful restraint on u.

2. *Find the probability,* w, *of Target standing firm to a threat of potent sanctions that makes a* Type I *Sender indifferent between doing nothing and threatening potent sanctions:*

   $EU_{Sender \; Type \; I}$ (threaten potent sanctions) $= w(-C - R) + (1 - w)0$

Doing nothing yields a payoff of $-C - A^{DN}$, with certainty. A *Type I* Sender is indifferent between doing nothing and threatening potent sanctions when:

$$-C - A^{DN} = w(-C - R) + (1 - w)0$$

$$w = \frac{C + A^{DN}}{C + R}$$

The right-hand side of this equality is always a number between zero and one, since $(R, A^{DN}, C) > 0$ and $R > A^{DN}$. This equality thus puts a meaningful restraint on w.

3. *Find the probability, z, of a* Type I *Sender threatening potent sanctions that could justify Target's belief that* $u = B/(\beta S^P_T + B)$

$$\mu_T(\textit{Type III} \mid \text{potent threat}) = \frac{\mu_T(\textit{Type III})\, \mu_T(\text{pot. threat} \mid \textit{Type III})}{\mu_T(\textit{Type III})\, \mu_T(\text{pot. threat} \mid \textit{Type III}) + \mu_T(\textit{Type I})\, \mu_T(\text{pot. threat} \mid \textit{Type I})}$$

$$\frac{B}{\beta S^P_T + B} = \frac{p\,(1)}{p\,(1) + (1 - p)z}$$

$$z = \frac{p \beta S^P_T}{(1 - p)B}$$

The right-hand side of this equality is only a number between zero and one *for certain values of p.* When $p > B/(\beta S^P_T + B)$ the probability of a *Type I* Sender threatening potent sanctions is greater than one, since Target will always back down to a threat of potent sanctions. This mixed strategy cannot be supported when $p > B/(\beta S^P_T + B)$. When $p < B/(\beta S^P_T + B)$, *z* is a number between zero

and one since $(S^P_T, B) > 0$ and $0 < \beta < 1$. This equality thus puts a meaningful restraint on z, but only for certain values of p.

4. *Find the conditions under which Target prefers violating the norm to not violating the norm, given the above:*

$EU_{Target}(VN) = p[w(-\beta S^P_T) + (1 - w)0] + (1 - p)[z(w(B) + (1 - w)0) + (1 - z)B]$

Not violating the norm yields a payoff of zero, with certainty. Target prefers violating the norm to not violating the norm when:

$0 < -pw\beta S^P_T + (1 - p)[zwB + B - zB]$

Substituting for z and w yields:

$$0 < \frac{-p\beta S^P_T (C+A^{DN})}{(C+R)} \quad \frac{+ (1-p)\, p\beta S^P_T B \,(C+A^{DN})}{(1-p)(C+R)B} \quad + (1-p)B \quad \frac{- (1-p)\, p\beta S^P_T B}{(1-p)B}$$

$$0 < \quad + (1 - p)B + p\beta S^P_T$$

$$\boldsymbol{p < \frac{B}{\beta S^P_T + B}}$$

When $p < B/(\beta S^P_T + B)$ and $q = 0$, there is a mixed strategy perfect Bayesian equilibrium in which Target violates the norm with certainty, and in which it stands firm to a threat of potent sanctions with probability w and backs down to a threat of potent sanctions with probability $(1 - w)$ with $u = B/(\beta S^P_T + B)$. Target yields to imposed potent sanctions with certainty. A *Type III* Sender (credibly) threatens potent sanctions with certainty, and a *Type I* Sender threatens potent sanctions with probability z and does nothing with probability $(1 - z)$. A *Type I* Sender does not impose potent sanctions (with certainty).

Appendix 2 The Sanctions Game under Incomplete Information – a Mixed Strategy Perfect Bayesian Equilibrium (with $p + q = 1$)

For simplicity, assume that $p + q = 1$.

1. *Find the probability,* u, *of Sender being* Type III *that makes Target indifferent between standing firm and backing down to a threat of potent sanctions:*

    EU_{Target} (stand firm | threat of potent sanctions) $= u(-\beta S^P_T) + (1- u)B$

    Backing down to a threat of potent sanctions yields a payoff of zero, with certainty. Target is indifferent between standing firm and backing down when:

    $0 = u(-\beta S^P_T) + (1- u)B$

$$u \quad = \quad \frac{B}{\beta S^P_T + B}$$

    See Appendix 1 for a discussion of this condition.

2. *Find the probability,* w, *of Target standing firm to a threat of potent sanctions that makes a* Type II *Sender indifferent between doing nothing and threatening potent sanctions:*

    $EU_{Sender\ Type\ II}$ (threaten potent sanctions) $= w(-C - R) + (1- w)0$

Threatening lenient sanctions yields a payoff of $-C - A^L - S^L_S$, with certainty. A *Type II* Sender is indifferent between threatening lenient sanctions and threatening potent sanctions when:

$$-C - A^L - S^L_S = w(-C - R) + (1 - w)0$$

$$w = \frac{C + A^L + S^L_S}{C + R}$$

The right-hand side of this equality is always a number between zero and one, since $(R, A^L, C, S^L_S) > 0$ and $R > A^{DN} + S^L_S$. This equality thus puts a meaningful restraint on w.

3. *Find the probability, z, of a* Type II *Sender threatening potent sanctions that could justify Target's belief that* $u = B/(\beta S^P_T + B)$

$$\mu_T(\textit{Type III potent threat}) = \frac{\mu_T(\textit{Type III}) \, \mu_T(\text{pot. threat} | \textit{Type III})}{\mu_T(\textit{Type III}) \, \mu_T(\text{pot. threat} | \textit{Type III}) + \mu_T(\textit{Type II}) \, \mu_T(\text{pot. threat} | \textit{Type II})}$$

$$\frac{B}{\beta S^P_T + B} = \frac{p(1)}{p(1) + (1 - p)z}$$

$$z = \frac{p \beta S^P_T}{(1 - p)B}$$

See Appendix 1 for a discussion of this condition.

4. *Find the conditions under which Target prefers violating the norm to not violating the norm, given the above:*

$$EU_{Target}(VN) = p[w(-\beta S^P_T) + (1 - w)0] + (1 - p)[z\,(w(B) + (1 - w)0) + (1 - z)(B - S^L_T)]$$

Not violating the norm yields a payoff of zero, with certainty. Target prefers violating the norm to not violating the norm when:

$$0 < -pw\beta S^P_T + (1 - p)[zwB + B - S^L_T - zB + zS^L_T]$$

Substituting for z and w yields:

$$0 < \frac{-p\beta S^P_T (C+A^L+S^L_S)}{(C+R)} + \frac{(1-p)\,p\beta S^P_T B\,(C+A^L+S^L_S)}{(1-p)(C+R)B} + (1-p)B - (1-p)S^L_T - \frac{(1-p)\,p\beta S^P_T B}{(1-p)B} + \frac{(1-p)p\beta S^P_T S^L_T}{(1-p)B}$$

$$0 < + (1 - p)B^2 - (1 - p)S^L_T B - p\beta S^P_T B + p\beta S^P_T S^L_T$$

$$0 < + B - pB^2 - S^L_T B + pS^L_T B - p\beta S^P_T B + p\beta S^P_T S^L_T$$

$$pB^2 - p\,S^L_T B + p\beta S^P_T B - p\beta S^P_T S^L_T < + B^2 - S^L_T B$$

$$p < \frac{B^2 - S^L_T B}{B^2 - S^L_T B + \beta S^P_T B - \beta S^P_T S^L_T}$$

$$p < \frac{(B - S^L_T)B}{(B - S^L_T)(B + \beta S^P_T)}$$

$$\boldsymbol{p} < \frac{\mathbf{B}}{\boldsymbol{\beta} \mathbf{S}^P_T + \mathbf{B}}$$

When $p < B/(\beta S^P_T + B)$ and $p + q = 1$, there is a mixed strategy perfect Bayesian equilibrium in which Target violates the norm with certainty, and in which it stands

firm to a threat of potent sanctions with probability w and backs down to a threat of potent sanctions with probability $(1 - w)$ with $u = B/(\beta S^P_T + B)$. Target yields to potent sanctions, and stands firm and does not yield to lenient sanctions, with certainty. A *Type III* Sender (credibly) threatens potent sanctions with certainty, and a *Type II* Sender threatens potent sanctions with probability z and (credibly) threatens lenient sanctions with probability $(1 - z)$. A *Type II* Sender does not impose potent sanctions, with certainty.

References

Achen, Christopher H. and Duncan Snidal (1989). "Rational Deterrence Theory and Comparative Case Studies", *World Politics* 41(2): 143-169.

Baldwin, David A. (1985). *Economic Statecraft.* Princeton, New Jersey: Princeton University Press.

Baldwin, David A. (1999/2000). "The Sanctions Debate and the Logic of Choice", *International Security* 24(3): 80-107.

Drezner, Daniel W. (2000). "Bargaining, Enforcement, and Multilateral Sanctions: When is Cooperation Counterproductive?", *International Organization* 54(1): 73-102.

Drezner, Daniel W. (2003). "The Hidden Hand of Economic Coercion", *International Organization* 57 (2003): 643-659.

Elliott, Kimberly Ann (1998). "The Sanctions Glass. Half Full or Completely Empty?", *International Security* 23(1): 50-65.

Elliott, Kimberly Ann (2003). "Economic Leverage and the North Korean Nuclear Crisis", International Economics Policy Brief 03-3. Washington: Institute for International Economics.

Elliott, Kimberly Ann, Jeffrey J. Schott, Gary Clyde Hufbauer and Barbara Oegg (forthcoming). *Economic Sanctions Reconsidered: Case Histories.* 3rd Edition. Washington: Institute for International Economics.

Elster, Jon (1986). "Introduction", pp. 1-33 in Jon Elster (ed.): *Rational Choice.* Oxford: Blackwell.

Gates, Scott and Brian D. Humes (1997). *Games, Information and Politics. Applying Game Theoretic Models to Political Science.* Ann Arbor: The University of Michigan Press.

Harsanyi, John C. (1986). "Advances in Understanding Rational Behavior", pp. 82-107 in Jon Elster (ed.): *Rational Choice.* Oxford: Blackwell.

Hovi, Jon (1998). *Games, Threats and Treaties. Understanding Commitments in International Relations.* London: Pinter.

Hovi, Jon (2001). "Hvorfor økonomiske sanksjoner virker – tross alt", *Internasjonal politikk* 59(4) 2001: 509-524.

Hovi, Jon, Robert Huseby and Detlef Sprinz (2005). "When do (Imposed) Economic Sanctions Work?", *World Politics* (in press).

Hubauer, Gary Clyde (1999). "Trade as a Weapon". Paper for the Fred J. Hansen Institute for World Peace, World Peace Week, San Diego State University. San Diego, California April 12-18 1999. Conference Paper.

Hufbauer, Gary C., Jeffrey J. Schott, and Kimberly Ann Elliott (1990). *Economic Sanctions Reconsidered: History and Current Policy*. 2nd Edition. Washington: Institute for International Economics.

Kaempfer, William H., and Anton D. Lowenberg (1992). "The Theory of International Economic Sanctions: A Public Choice Approach", *American Economic Review* 78(4): 786–93.

Lacy, Dean, and Emerson M. S. Niou (2004). "A Theory of Economic Sanctions and Issue Linkage: The Roles of Preferences, Information, and Threats", *The Journal of Politics* 66(1): 25-42.

Martin, Lisa L. (1992). *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton, New Jersey: Princeton University Press.

Milner, Helen V. (2004). "Formal Methods and International Political Economy", pp. 265-289 in Detlef F. Sprinz and Yael Wolinsky-Nahmias (eds.): *Models, Numbers & Cases. Methods for Studying International Relations*. Ann Arbor: The University of Michigan Press.

Morgan, T. Clifton, and Valerie L. Schwebach (1997). "Fools Suffer Gladly: The Use of Economic Sanctions in International Crises", *International Studies Quarterly* 41 (1): 27-50.

Morrow, James D. (1994). *Game Theory for Political Scientists*. Princeton, New Jersey: Princeton University Press.

Pape, Robert A. (1997). "Why Economic Sanctions Do Not Work", *International Security* 22(2): 90–136.

Pape, Robert A. (1998). "Why Economic Sanctions Still Do Not Work", *International Security* 23(1): 66–77.

Rasmusen, Eric (2001). *Games & Information. An Introduction to Game Theory*. 3rd Edition. Oxford: Blackwell Publishing.

Sanger, David E. (2005, April 25). "White House May go to U.N. Over North Korean Shipments". New York Times [online]. Available via Global Policy Forum, http://www.globalpolicy.org/security/sanction/nkorea/2005/0425quarantine.htm.

Schott, Jeffrey J. (1998). "US Economic Sanctions: Good Intentions, Bad Execution", Statement before the Committee on International Relations, United States House of Representatives, Washington, DC, June 3 1998.

Snidal, Duncan (2004). "Formal Models of International Politics", pp. 227-264 in Detlef F. Sprinz and Yael Wolinsky-Nahmias (eds.): *Models, Numbers & Cases. Methods for Studying International Relations*. Ann Arbor: The University of Michigan Press.

Tsebelis, George (1990). "Are Sanctions Effective? A Game-Theoretic Analysis", *Journal of Conflict Resolution* 34(1): 3–28.

Watson, Joel (2002). *Strategy. An Introduction to Game Theory*. New York: W. W. Norton & Company.