2015

Barriers to Resolving Transboundary Pollution

Ian A. Thaure
Bard College

Barriers to Resolving Transboundary Pollution:

Lessons from Game Theory and Real-World Application

Senior Project submitted to

The Division of Social Studies

of Bard College

by

Ian Thaure

Annandale-on-Hudson, New York

May 2015

Acknowledgements

My thanks to Professor Kris Feder for her guidance throughout the senior project advising process and my undergraduate experience.

My thanks to Professor Aniruddha Mitra for introducing me to foundational game theory concepts and help with Chapter 2 of this project.

To my family and friends who have supported me throughout my undergraduate career at Bard College.

Abstract

Transboundary pollution challenges the conventional economic analysis of pollution as a negative externality to market activity. Different political entities cannot regulate pollution that flows among them. Developing institutions to regulate pollution is impossible at the international level due to the principle of state sovereignty. Game theory offers economic insight into the strategic interactions among nations that experience and contribute to transboundary pollution. Though game theory underestimates nations proclivities to remain cooperative, the conclusions of models align with real-world experience. Barriers to resolving transboundary pollution include the polluter pays principle and liability norms.

Contents:

Introduction:

Many complicated forms of environmental pollution involve transmission mechanisms that disperse damages over a vast geographical scale. These transmission mechanisms allow pollutants to travel and accumulate far from their points of origin. Some of these pollutants accumulate in environmental sinks held in common between nations, sinks such as rivers, forests, and other ecosystems. Other pollutants accumulate in global environmental sinks, like the planet's atmosphere and oceans. These sorts of pollutants display a complete disregard for geographical or political boundaries, hence the inspiration for the title *transboundary pollution*.

Transboundary pollution poses a unique conundrum to policymakers. Many of the most critical environmental problems the world faces today result from transboundary pollution between different political entities. International law and relations both operate on the principle of state sovereignty, whereby no nation may be subject to the authority of another nation without its explicit or tacit consent.[1] Maintenance of sovereignty prevents nations from regulating each other's contributions to transboundary pollution. So, the resolution of transboundary pollution disputes requires cooperation between nations. At the same time, the interests of different nations often compete and conflict with each other, preventing the cooperation required for resolution. In an international realm that holds sovereignty as a fundamental right to every nation, most international environmental problems that involve transboundary pollution simply remain unresolved until critical damage, or the threat of critical damage, manifests. Transboundary

[1] O'Connell. 1995.

pollution continues to spark debate and concern as the international community struggles to negotiate tangible solutions.

Most of the academic literature on transboundary pollution employs a very narrow definition of the concept. Commonly, transboundary pollution is defined as pollution that affects multiple nations, where many nations are responsible for emissions.[2] This definition suffers from two oversights. First, it implies that transboundary pollution only occurs at the international level. However, transboundary pollution may occur between any political entities that cannot regulate each other, including towns, cities, and states. Second, the conventional definition implies that multiple entities bear responsibility for transboundary pollution. However, many cases of transboundary pollution involve purely unidirectional flows: pollution simply flows from one entity to another with no element of reciprocity. In fact, the premier lawsuits concerning transboundary pollution in the United States were both interstate *and* unidirectional problems.[3] For a more robust perspective and analysis, this project utilizes its own, broader definition of transboundary pollution: pollution that crosses political boundaries, damaging jurisdictional territories (state, country, etc.) outside of its origin(s).

From this definition, transboundary pollution may be further divided between two transmission categories: unidirectional transboundary pollution and reciprocal transboundary pollution. Unidirectional pollution simply flows from source entities to victim entities, whereas reciprocal pollution flows between all parties involved. [4] For example, ozone-depleting substances pose a reciprocal problem. Any nation may emit ozone-depleting substances, putting other nations at risk of increased exposure to ultraviolet radiation. On the other hand, river

---

[2] Folmer et al. 1998.
[3] See *Merrill 1997* for an in-depth discussion of *Georgia v. Tennessee Copper Co. 1907* and *Missouri v. Illinois 1906*.
[4] Merrill. 1997.

pollution often poses a unidirectional problem. The physical flow of a river tends to result in a downstream accumulation of pollution, which allows upstream entities to transfer most of their pollution to downstream entities. Unidirectional transboundary pollution clearly divides involvement and repercussions between polluting entities and victim entities. With reciprocal transboundary pollution, on the other hand, entities each bear damages to the extent that all involved entities engage in polluting activities.

Conventionally, economists deal with pollution as an efficiency problem. Economics characterizes pollution as a negative externality to market activity: firms and consumers participate in a market that damages other entities,[5] either through the act of production or through the use of goods by consumers. In a free market, firms and consumers ignore negative externalities when making economic decisions, operating only on the basis of their own costs and benefits. So, a free market that involves pollution does nothing to abate it. The result is an inefficient outcome. The market produces dead weight loss through excess pollution damage.

There are many different regulatory schemes that correct for negative externalities. However, not all of them are efficient. Most often, policymakers attempt to control for pollution through a command and control approach that calls for universal abatement across all firms that engage in a polluting activity.[6] This approach, however, ignores differences in abatement costs across firms. Some firms have cheaper marginal abatement costs than others, yet command and control requires all firms to engage in the same level of abatement. The command and control approach simply trades one inefficiency, negative externalities, for another, needlessly expensive abatement programs.

[5] Baumol and Oates. 1988.
[6] Ibid.

In terms of efficient approaches, regulators may employ a Pigovian tax, a tax on firms or consumers in a market set at the marginal rate of pollution damage. Or, regulators may instead opt to restrict market activity through a quota system. A quota system may limit the production of goods, the number of firms participating in a market, or the number of transactions that take place in a market. Both Pigovian taxes and quota systems, when set at a level that fully internalizes the negative external costs at play, bring about an efficient outcome. In this new, efficient outcome, the price of a good reflects both its marginal cost of production *and* the marginal pollution damages involved in its use or production.

Transboundary pollution, however, complicates the conventional economic wisdom. The approaches listed so far all rely on a regulatory institution with jurisdiction over the entire scope of pollution damages. However, transboundary pollution flows among different political entities. For example, solid waste management in the United States takes place over a national market. Yet, through the Resource Recovery and Conservation Act, the federal government assigns the responsibility of regulating solid waste management to state governments. Often, in response to state attempts at stricter regulation, disposers simply ship waste to states that impose weaker regulations and controls. When polluting activities can simply be relocated outside of regulations, or when regulations are not enforced on the appropriate scale of a polluting activity, pollution leaks through the regulations, often completely.

Of course, solid waste management in the United States does not pose the same problem as a global pollutant, such as greenhouse gases. The federal government may, at any time it chooses, assume the role of national regulator and employ any of the strategies listed previously to correct transboundary pollution that occurs between states. Hence, most of the literature on transboundary pollution focuses on international environmental problems because there is no supranational entity that may impose binding resolutions between nations. However, simply the existence of an institution does not guarantee its capability nor willingness to manage transboundary pollution. At present, states continue to struggle with transboundary issues in solid waste management in the absence of significant federal initiative. Nonetheless, this project focuses on international cases of transboundary pollution only. Unlike at the national level, the principle of state sovereignty prevents the development of a binding regulatory institutions at the international level.

Despite the absence of binding institutions, many nations attempt to resolve international transboundary pollution problems. The United Nations Environment Programme lists 132 multilateral agreements concerning international environmental issues since 1991.[7] Clearly, most nations of the world realize that cooperating to resolve transboundary pollution can result in significant gains for everyone involved. Strategic interaction among different political entities does not necessarily lead to the destruction of resources held in common.[8] However, the lack of institutions to bind nations to commitments calls many aspects of environmental agreements into question: to what extent do these agreements resolve transboundary pollution? To what extent do countries comply with these agreements? What circumstances facilitate compliance? The conventional economic wisdom of pollution control offers no additional insight into these points.

This project offers a broad analysis of transboundary pollution with respect to real-world experience and economic theory. Chapter 1 examines a handful of international environmental agreements that attempt to resolve the world's most critical transboundary pollution problems. This section outlines agreements based on their organizational structure, enforcement bodies, noncompliance procedures, compliance incentives, and policy norms. Agreements are evaluated

[7] Barrett. 1997.
[8] Carraro and Siniscalco. 1993.

on their success in achieving observable abatement increases following their enactment. The purpose of this chapter is to provide a basis of real-world experience with international transboundary pollution.

Chapter 2 reviews contemporary game theory models of transboundary pollution interactions. Game theory provides a conceptual mode of analysis that directly acknowledges the lack of binding regulatory institutions. The chapter begins with a simple prisoner's dilemma and expands the model by introducing new options for players and relaxing restrictive assumptions. From this literature review comes a set of general conclusions on transboundary pollution disputes and their potential for resolution. Ultimately, the literature argues that self-enforcing environmental agreements are possible, though limited in their potential accomplishments beyond fundamentally uncooperative outcomes.

Chapter 3 reconciles the conclusions of game theory with real-world experience, identifying successes and limitations of the models in their conclusions on international environmental agreements. Though game theory models fail to acknowledge broader motivations for cooperation in the context of international relations, the models generally predict the success or failure of cooperation based on the extent of damages and costs involved in a dispute and the institutions (or lack thereof) employed in an agreement. In addition, real-world experience supports the conclusion that international agreements have little potential to improve upon the status quo of a pollution dispute.

Chapter 4 discusses two real-world barriers that prevent better resolution of transboundary pollution. The first barrier this chapter identifies is the polluter pays principle. This guiding policy principle acts as a barrier because it prevents nations from engaging in transfer payments to other nations in exchange for abatement. Such transfer payments promote better outcomes, as demonstrated by the game theory analysis presented in Chapter 2. The second barrier is international norms in liability regimes for transboundary pollution. This chapter critiques the manner in which international law applies the fundamentals of tort law to transboundary pollution disputes. On one hand, the norm of state liability has proved unworkable in application. On the other, the norm of strict liability aggravates bargaining incentives and induces extreme strategic behavior on the part of both victims and originators of transboundary pollution. Furthermore, the burden of establishing reasonable standards of care and evidence of harm renders most transboundary pollution problems unfit for legal resolution.

Finally, this project concludes with basic principles to promote better resolution of transboundary pollution. As a method to overcome real-world barriers, policymakers should aim to introduce linkages between different issues of transboundary pollution to encourage cooperation. Lastly, this section stresses the importance of establishing internal science and technology programs within international environmental agreements.

Chapter 1: International Environmental Agreements

This chapter discusses three famous international attempts at regulating transboundary pollutants: the Montreal Protocol on Substances that Deplete the Ozone Layer, The Convention on Long-Range Transboundary Air Pollution (LRTAP), and the Kyoto Protocol to the United Nations Framework Convention on Climate Change. All of these agreements cover reciprocal forms of transboundary pollution. that threaten nations with high levels of environmental damage. Each agreement experienced different degrees of success depending on the nature of pollutants they targeted as well as the policy norms and procedures they employed.

Ozone Depleting Substances and the Montreal Protocol

Stratospheric ozone depletion is perhaps the most serious form of global, reciprocal transboundary pollution that international coordination has reconciled. Though scientists had long been concerned about ozone depletion in the earth's atmosphere, an anthropogenic explanation did not surface until the 1970s with investigations into the role of chlorofluorocarbons (CFCs). CFCs are a wide variety of industrial chemicals used in aerosols and refrigerants. The chemical stability of CFCs allows them to persist into the stratosphere, whereby chlorine atoms break off and convert ozone into oxygen.[9] The reduction in ozone allows more ultraviolet radiation to enter the lower atmosphere, dramatically increasing the risk of skin cancer worldwide. Hence, atmospheric ozone functions as a public good that generates

[9] Parson. 1993.

shared benefits in the form of ozone protection and private costs in the form of abatement.[10] Abatement entails a reduction in the use of CFCs and other ozone-depleting substances.

Decades of scientific research and international negotiations culminated in 1989 with the ratification of the Montreal Protocol on Substances that Deplete the Ozone Layer. Originally, the agreement called for a 50% reduction in production and consumption of five principal CFCs by 1999 from a 1986 baseline, with interim freeze and elimination controls.[11] Since the Protocol has gone into effect, nations across the globe have dramatically cut CFC usage and production. In 2004, global consumption stood at about 7% of the Protocol's 1986 baseline.[12] The Montreal Protocol is also the only international treaty ever to reserve universal ratification.[13] Such results have led scholars and policymakers to consider the Montreal Protocol a hallmark of international coordination. Kofi Annan, former Secretary General of the United Nations, claimed in 2003 that the Montreal Protocol was the "single most successful international agreement to date."[14]

The Montreal Protocol is simple in terms of obligations. Essentially, signatories are required to follow the Protocol's reduction goals and freeze controls,[15] submit consumption and production data, and allow for technology transfers. Compliance monitoring relies on the consumption and production data sets that parties submit. Though this self-reporting mechanism contains the potential for signatories to falsify their data, review bodies ultimately compile all submitted data to evaluate trends. Along with implementation and governing committees, the Protocol establishes a number of expert advisory panels that oversee data collection and implementation.

[10] Epstein et al. 2014.
[11] Parson. 1993.
[12] Norman et al. 2008.
[13] United Nations Development Program. 2012.
[14] Green. 2009.
[15] Later revisions of the Protocol revised the original 50% goal upwards.

However, what makes the Montreal Protocol relatively unique among international environmental agreements is its commitment to engaged implementation. To begin with, the Montreal Protocol establishes a multi-lateral compliance fund, simply referred to as the multi-lateral fund. It draws from the contributions of developed nations to aid developing nations in the phase-out of ozone-depleting substances through compliance-inducing projects. The multi-lateral fund does not function as a pure monetary transfer program. Funding only covers incremental costs incurred by developing nations towards fulfilling protocol obligations.[16] Projects go through several stages of review and verification before being awarded funding. The World Bank and three United Nations programs provide implementation oversight for multi-lateral fund projects.[17] These organizations provide the necessary oversight and administrative background to ensure that funding is managed appropriately and that projects remain feasible. They also provide an alternative to weak government capacity in developing nations. So, the multi-lateral fund provides the financial and administrative means to ensure for global phase-out of ozone-depleting substances, allowing all nations, developed and developing, to participate in the Protocol.

Not only has the multi-lateral fund been instrumental in carrying out implementation, but also the Montreal Protocol's expert advisory committees. Part of the Montreal Protocol's institutional structure includes three main advisory committees: a scientific assessment panel, an environmental assessment panel, and a technology and economic assessment panel.[18] These review bodies provides a direct institutional link between expert scientific, engineering, and economic communities and the implementation process, ensuring detailed accounts of feasibility

[16] Green. 2009.
[17] Parson. 1993.
[18] Greene. 1998.

concerns and practical constraints.[19] The scientific assessment panel developed the standardized metric "ozone-depletion potential" to regulate emissions across targeted ozone-depleting substances, allowing regulators to focus on the most damaging substances first and signatories to reduce emissions through a basket approach across targeted substances.[20] The technology and economic assessment panel works with industries that utilize ozone-depleting substances to encourage new commercial substitutes and determine essential use exemptions.[21] The Montreal Protocol functions not simply as a mandate, but rather a procedure that its own institutions actively facilitate.

Another peculiar aspect of the Protocol is its reliance on informal procedures and soft enforcement approaches. The Protocol originally lacked a formal noncompliance procedure. Though signatories agreed to trade sanctions against noncompliant parties and nonsignatories, the noncompliance procedure was deferred to *after* the Protocol's ratification. As such, nations were not heavily involved with the design of the procedure.[22] Instead, the procedure developed organically through interactions between protocol institutions and parties in response to issues as they arose. As opposed to an accusatory and dispute resolution system, the Protocol's noncompliance procedures engages with offending parties informally to encourage compliance.[23] Informal procedures have facilitated smooth resolutions and transitions towards compliance. All original noncompliance issues the Protocol faced were volunteered for resolution by the offending parties themselves.[24] Through informal arrangements, nations have been more willing

---

[19] Greene. 1998.
[20] Epstein et al. 2014.
[21] Greene. 1998.
[22] Ibid.
[23] Ibid.
[24] Victor. 1998.

to address noncompliance than they may have been under the threat of harsh penalties and strict noncompliance procedures.[25]

However, the sticks and carrots of the Protocol also motivate parties towards addressing noncompliance. In terms of sticks, the Protocol calls for trade sanctions against noncompliant members and nonsignatories. Trade sanctions were not only decisive in motivating countries within the agreement to remain compliant, but also for non-parties that sought to avoid sanctions by unilaterally complying with the Protocol, as in the case of Taiwan and South Korea.[26] At the same time, the Montreal Protocol offers more in terms of carrots with the multi-lateral fund. Access to funding not only provides an incentive for developing nations to comply, but also the means with which to comply. Lack of institutional capacity prevented many developing nations from compiling production and consumption data.[27] Now, every developing nation with a long-standing multi-lateral fund program has reported baseline data.[28] Furthermore, the fund even motivated developing nations to agree to commitments *ahead* of the Protocol's schedule. Of the 60 countries eligible for the funding, 46 agreed to a complete phase-out sooner than the proposed schedule.[29] Thus, powerful and tangible incentives backed the informal noncompliance procedure.

Though the Montreal Protocol's institutional design proved successful the characteristics of ozone-depleting substances were relatively conducive to international negotiation. The concentration of ozone-depleting substances among few industrial actors in specific nation made reductions more manageable from a collective action standpoint.[30] Also, the chemical industry

---

[25] Greene. 1998.
[26] Parson. 1993.
[27] Ibid.
[28] Victor. 1998.
[29] Parson. 1993.
[30] Epstein et al. 2014.

had CFC substitutes in their research and development pipeline by the time of the Montreal Protocol's implementation.[31] Thus, industry opposition, which gridlocked regime development in the 80s, had petered out.[32] As the Montreal Protocol underwent implementation, most states and industrial producers of ozone-depleting substances had a strong interest in making the regime work.[33]

Nevertheless, despite beneficial circumstances, the Montreal Protocol remains unique among international environmental agreements. It functions with soft enforcement through several redundancies and non-dedicated review bodies while providing funding for compliance. Furthermore, it blends a management approach with enforcement to avoid unproductive antagonism while maintaining a credible threat of tough action.[34] The Montreal Protocol displays institutional synergies unlike any other international agreement while acknowledging practical constraints to its goals. It owes its success not just to beneficial circumstances, but also ingenious design.

Acid Rain, Europe, and LRTAP

The peculiar domestic and geographical elements of acid rain set it apart from other transboundary pollution problems. Acid rain entails both transboundary and domestic damages, often exacting a greater damage toll on the source nation itself than transboundary victims. Furthermore, acid rain blurs the unidirectional and reciprocal distinction. For example, transboundary flows of acid rain in Europe are fairly reciprocal between countries, whereas acid

[31] Parson. 1993.
[32] Ibid.
[33] Greene. 1998.
[34] Victor. 1998.

rain in North America flows mainly from the United States into Canada. Even within the United States, acid rain flows mainly from Midwestern territories to Northeastern territories. Hence, acid rain problems differ in transboundary characteristics depending on the geography of parties at play. Acid rain's fuzzy boundaries complicate resolution and the negotiation process, offering a primary insight into why acid rain controls have historically met with mixed success.

The first international attempt to regulate transboundary flows of acid rain began in Europe with the Convention on Long-Range Transboundary Air Pollution (LRTAP). Signed in 1979, almost a decade before the Montreal Protocol, it originally held no standards to regulate emissions among European nations. At the time of its ratification, international transboundary acid rain was a breakthrough discovery that lacked strong scientific consensus and understanding. Though Scandinavian scientists proved that acid deposition in their countries came mostly from other countries[35] and the OECD published findings in 1977 that found similar depositions in other European countries,[36] the relative novelty of acid rain put nations in uneasy negotiating positions. Not only were nations unsure of their contributions to transboundary pollution, but also of emission control costs an agreement would force them to accept. As such, LRTAP was designed as a framework agreement for scientific consensus building and future controls. Originally, it required parties to commit only to exchange information, consultation, and research.[37] It also established a European-wide scientific monitoring program for acid rain, the European Monitoring and Evaluation Program (EMEP). As scientists further investigated acid rain phenomena and damage across Europe, nations agreed to multiple control protocols on

[35] Munton et al. 1999.
[36] Levy. 1993.
[37] Wettestad. 1997.

key emissions: a sulfur dioxide protocol in 1985, a nitrogen oxide protocol in 1988, and a volatile organic compound protocol in 1991.[38]

The greatest strength of LRTAP was its ability to build scientific consensus. EMEP maintained an extensive monitoring program that harmonized data collection, evaluated competing explanations of damage, and assessed environmental threats.[39] EMEP's work later provided the basis for abatement commitments and critical-load evaluation. Without harmonization efforts, countries would have been unable to pool their results or make comparative assessments of environmental quality.[40] The development of critical-load evaluations eventually replaced flat-rate reduction commitments in later emission protocols. Critical-load evaluations overcame the arbitrary nature and fairness problems of flat-reductions, providing a scientific basis for emission controls.[41]

Furthermore, scientific research and consensus strengthened LRTAP's ability to monitor commitments and pressure recalcitrant parties. Though LRTAP relied on self-reported data from signatories, scientific assessment in the context of transport models provided a check against false or doctored data.[42] Apparently, the Czech government considered submitting false reports in the late 1980s, but was discouraged by possible detection through EMEP.[43] LRTAP allowed scientists from across Europe to bypass their national governments and submit data directly to international bodies, thwarting the interests of Eastern European governments that sought to suppress knowledge of environmental damages.[44]

---

[38] Ibid.
[39] Levy. 1993.
[40] Ibid.
[41] Ibid.
[42] Munton et al. 1999.
[43] Ibid.
[44] Levy. 1993.

Still, LRTAP floundered in developing European-wide abatement commitments. Instead of obtaining strong reduction commitments from each other, nations proposed different levels of abatement that they considered economical to achieve within their own borders. Most nations that committed to the emission protocols were already on the path to compliance, while key source nations either held out or bargained for weaker commitments. For example, many source nations held out from signing the sulfur dioxide protocol, including significant polluters like the United Kingdom and most of Eastern Europe.[45] Russia only agreed to the protocol on the condition that it could simply move polluting industries to the other side of the Ural Mountains while failing to address key sources of European pollution, like its nickel smelters on the Kola Peninsula.[46] Though high-compliant nations attempted to use their status to bully other nations into compliance, and even pushed the United Kingdom to accept the sulfur dioxide protocol, poorer nations of Western Europe and all of Eastern Europe escaped LRTAP commitments almost entirely.[47]

In addition, much of Europe's emission reductions had little to do with the LRTAP regime. Rather, many reductions were simply the byproducts of changing domestic energy and transportation policies. Both the European Commission's power plant directive and the vehicle emissions directives were motivated by German politicians who sought to level the European playing field with their domestic policy, not transboundary pollution concerns.[48] Many European nations were already seeking to expand nuclear power, which incidentally reduced acid rain emissions. Russia's energy sector restructuring was encouraged by their declining resource base

[45] See Table 4.2 in *Munton et al. 1999* for a detailed list of signatories for each of the emission protocols.
[46] Levy. 1995.
[47] Levy. 1993.
[48] Ibid.

in Europe and abundant fossil fuels in Siberia.[49] Britain's switch to expanded use of natural gas also had a huge effect on emission reduction.[50] As such, indirect concerns take most of the credit for emission reductions across Europe.

Furthermore, acid rain was so poorly understood at the time of LRTAP's construction and implementation that, as scientific consensus advanced, many nations underwent emission reductions not to reduce transboundary flows, but rather to resolve damages within their own borders. Once Austria, Finland, Switzerland, and the Netherlands became aware of damage within their countries, they began to pursue emission reductions.[51] Germany's reductions were stimulated by the discovery of large swathes of forest decay within its borders.[52] The United Kingdom, despite the significant transboundary flows it contributed to, discovered extensive damage within its own borders.[53] The fact that abatement programs also alleviated domestic damages motivated many states towards accepting LRTAP commitments.

So, LRTAP attempted to deal with an unwieldy environmental problem in an unwieldy fashion. The success it found in reducing emissions stems from regional and domestic concerns among signatories that were poorly aligned with LRTAP's international focus. Also, LRTAP's flat-rate reductions collided with the regional realities of acid rain. Acid rain was not internationally reciprocal across Europe. Rather, focal points and dynamics varied between regions. Flat-rate reduction did little to correct the behavior of major polluters or change pollution dynamics. In response to LRTAP's failure for relevant alleviation, many countries went outside of the agreement to pursue their own multilateral programs for regional concerns. Even

[49] Munton et al. 1999.
[50] Levy. 1995.
[51] Levy. 1993.
[52] Wettestad. 1997.
[53] Levy 1995.

during LRTAP, Sweden, Finland, and Russia formed regional air quality agreements to better suit their needs.[54]

Greenhouse Gases and the Kyoto Protocol

The Kyoto Protocol stands apart from other environmental agreements as a pariah. Unlike the other agreements highlighted so far, few parties involved in Kyoto had any desire to see it work. Belligerent negotiations rendered arbitrary and inefficient controls, haphazard implementation, and poorly planned institutions. Ultimately, Kyoto spent only half of its lifetime as a legally binding protocol. Even when finally implemented, Kyoto failed to regulate most of the world's key emitters of greenhouse gases. For those nations it did regulate, the Protocol achieved little more than a freeze in emissions along its baseline.

As with ozone-depleting substances, scientific investigation revealed the environmental problems posed by greenhouse gases. However, unlike with ozone-depleting substances, scientific developments took longer to motivate international negotiations on controls. By the mid 1970s, scientists had begun to make the connection between carbon dioxide and global temperature differences.[55] The first world climate conference, held in 1979, established a world climate program among scientists.[56] 1988 led to the formation of the International Panel on Climate Change.[57] It was not until 1992, with the Framework Convention on Climate Change, that nations responded to the issue. Under the framework, developed nations and the territories of the former Soviet Union voluntarily committed to limit their concentration of greenhouse gases

[54] Levy 1993.
[55] Thoms. 2003.
[56] Chichilnisky and Sheeran. 2009.
[57] Thoms. 2003.

according to a 1990 emissions baseline.[58] Moreover, the FCCC established the foundational policy norms for the Kyoto Protocol: that pollution should be controlled through inviolable targets and timetables[59] and that developing nations should be excluded from emission controls.

The initial Kyoto Protocol, signed in 1997 by 160 nations, arose out of the framework convention. Under the Protocol, industrialized countries were to cut their emissions of six key greenhouse gases—carbon dioxide, methane, nitrous oxide, sulphur hexafluoride, hydrofluorocarbons, and perfluorocarbons—by an average of 5.2 percent below their 1990 emission baseline by 2008-2012.[60] Most European countries and the United States were given a 7-8 percent reduction target while other countries were simply required to freeze emissions along their baseline.[61] To meet obligations, countries were expected to carry out most of their abatement efforts domestically.

However, the Protocol established three institutional mechanisms to allow for flexibility and transnational abatement: joint implementation, carbon markets, and the clean development mechanism (CDM). Joint implementation is a bilateral process between industrialized countries whereby one country invests in the abatement efforts of another for emission credit.[62] To incorporate a market mechanism into controls, Kyoto sought to establish carbon markets whereby nations exchange carbon permits allotted by emission caps; through carbon markets, nations that abate beyond their goals can sell permits to other nations, promoting cost effective global abatement. At the same time, Kyoto lacked any legal language that required the creation of carbon markets;[63] parties simply agreed to carbon markets in principle[64] and most permit

58 Nordhaus and Boyer. 1999.
59 Victor. 2001.
60 Chichilnisky and Sheeran. 2009.
61 See Table in *Victor 2001*.
62 Chichilnisky and Sheeran. 2009.
63 Victor. 2001.
64 Ibid.

trading that took place under implementation occurred only among European nations through the European Union Emissions Trading Scheme. Lastly, the clean development mechanism allowed developed nations to offset emissions by funding clean energy projects in developing nations.[65]

Despite the Protocol's initial approval, implementation faltered soon after. The exit of the United States in 2001 dealt a blow to the Protocol's potential. The United States accounted for nearly a third of the emissions that the Protocol sought to control.[66] Scrambling to save the Protocol and prevent further exit, signatories convened again in 2001 at Marraketch. The Marraketch negotiations relaxed Kyoto's controls, lowering targeted reductions among signatories from an average of 5.2 percent to 1.4 percent.[67] Though the Protocol retained the same regulatory structure and abatement channels, its controls lost most of their strength. Even after Marraketch, only with Russia's ratification in 2005 did the Kyoto Protocol finally become legally binding.[68] So, negotiations and exit triggers slashed the Protocol down to paltry commitments by the time it was implemented. Though, with Russia's entry, the 2005 Kyoto protocol represented approximately half of global emissions,[69] it placed low demands on its signatories and failed to include the United States, the greatest polluter among developed nations.

Furthermore, Kyoto excluded developing nations from its controls, creating a leakage channel that undermined abatement efforts. The Kyoto Protocol allowed developing nations to not only maintain their current pollution levels, but also increase their emissions freely. Even during Kyoto's discussion and implementation, many developing nations were among the most significant global emitters of greenhouse gases. Some even had higher emissions than developed

---

[65] Chichilnisky and Sheeran. 2009.
[66] See *Victor 2001*.
[67] Thoms. 2003.
[68] Chichilnisky and Sheeran. 2009.
[69] Thoms. 2003.

nations beholden to the Protocol. Around 2005, India and South Korea had higher emissions than Germany and France, respectively.[70] It is well known that China ranks right behind the United States as the world's greatest greenhouse gas emitter, yet China was excluded from any abatement controls. The Kyoto Protocol's exclusion of developing nations from any form of emission controls doomed its ability to taper global emissions.

Though signatories successfully weakened the Protocol, they did little to address the critical flaws of its structure. Firstly, the Protocol's emission targets lacked both a monitoring mechanisms and an account of full emission lifecycles. In terms of emission lifecycle, while Kyoto took the destruction of carbon sinks into account for emission levels, it did not fully account for their creation.[71] Kyoto's failure to credit the creation of carbon sinks meant that signatories could not reliably engage in sink creation projects, like reforestation, to offset their emissions. In addition, the Protocol had scant means to monitor its targeted greenhouse gases: except for carbon dioxide, emission sources for target greenhouse gases were not accurately understood.[72] Furthermore, Kyoto's strict control schedules conflicted with emission tendencies. Greenhouse gas emissions vary with economic growth and technological change, neither of which can be planned for by governments.[73] Attempting emission controls according to the Kyoto schedule created vast uncertainties for signatories that could only be mitigated through weak commitments.

Also problematic was the relationship between how emission caps were set and carbon markets. Essentially, in a carbon market, the price of a permit is the opportunity cost of reducing

[70] Sunstein. 2007.
[71] Green. 2009.
[72] Victor. 2001.
[73] Ibid.

carbon emission and caps are the measure of permit scarcity.[74] However, the Protocol's emission caps were capricious. As caps were set by a historic baseline, countries with inefficient energy systems received lenient emission caps that bestowed major permit windfalls.[75] Russia commitments serve as a prime example. During Kyoto's initial negotiation in 1997, Russia's emissions stood at 70% of its 1990 baseline. However, Russia bargained for an emission target equal to 100% its baseline.[76] So, under 1997 controls, assuming no increase in emissions, Russia could have safely sold emission permits equal to 30% of its cap without engaging in any abatement whatsoever.

Commentators dubbed these windfalls as "hot air:" permit endowments in excess of anticipated future emissions.[77] Hot air permits are costless to supply because they arise not from abatement efforts but institutional privilege. Their exchange only results in a financial transfer between buyer and seller. The buyer of the hot air permit covers part of their responsibility, while the seller engages in no additional abatement. Thus, hot air permits undermine environmental quality by crowding out permits supplied by actual abatement.[78]

Furthermore, permit trading implicitly relies upon a system of property rights that Kyoto did not attempt to reconcile. Similar pollution permit systems that Kyoto took inspiration from, like the United States' sulfur emission trading program, were organized under clear, national systems of property rights. The problem with permit trading between nations is differences in property law. Permits issued by different nations are not necessarily comparable nor perfect substitutes.[79] Furthermore, international law is poorly suited for adjudicating property rights

[74] Chichilnisky and Sheeran. 2009.
[75] Nordhaus and Boyer. 1999.
[76] Sunstein. 2007.
[77] Bohringer and Vogt. 2003.
[78] Victor. 2001
[79] Ibid.

between nations.[80] Hence, the main exchange that serviced Kyoto carbon permits was the European Union Emissions Trading scheme, which served regional trading under a regional regulatory structure.

The only institution under Kyoto that remotely functioned as intended was the Clean Development Mechanism. Many developed nations under the Kyoto protocol utilized the mechanism to ease the cost of their commitments. By 2009, nearly 1200 projects were registered under the mechanism worldwide, though 60 percent of those projects were implemented in China alone.[81] Also, unlike with carbon markets, Kyoto created an accreditation committee to approve of projects and determine offset credits.[82] Thus, CDM projects took place in a regulated environment. Though many commentators criticized CDM for the fact that most of its projects focused on China at the expense of other, poorer developing nations, CDM has nonetheless served its purpose as a flexibility mechanism for abatement.

In respect to the monumental task Kyoto attempted to address, climate change remains a problem that modern institutions and political negotiations are ill-equipped to deal with. The effects of climate change are relatively uncertain, intangible, and lack a distinct timeframe for expectations.[83] In addition, addressing climate change has a greater benefit for future generations while imposing enormous costs on the present generations. Most abatement measures undertaken today will not yield stabilizing effects until far into the future.[84]

Yet Kyoto's architects held poor regard to feasibility or efficiency. They set up controls that were neither steeped in science nor respectful of costs. They proposed institutions without following through to design them properly. Instead of finding common resolution, they

[80] Ibid.
[81] Chichilnisky and Sheeran. 2009.
[82] Ibid.
[83] Thoms. 2003.
[84] Bohringer and Vogt. 2003.

antagonized rich and poor nations against each other. Worse, instead of admitting the Protocol's faults and attempting to start fresh, Kyoto's architects simply surrendered all potential success in controlling global greenhouse gas emissions for the sake of implementation. Kyoto stands as a monumental failure of collective environmental action.

Conclusions

As these agreements demonstrate, the international community has experienced both extraordinary successes and failures with respect to resolving transboundary pollution. In addition, these agreements show that the threat of immense environmental damage does not necessarily guarantee resolution. Rather, resolution depends in part on the immediacy and clarity of environmental threats. Ozone depletion and acid rain both threatened the international community with tangible damages, increased skin cancer for the former and destruction of natural resources vital to economic needs for the latter, whereas the effects of climate change continue to remain mired in uncertainty. Success also depends upon the personal interests of all nations involved in a transboundary pollution problem. Acid rain controls met with success due to incidental changes in energy policy. The Montreal Protocol offered financial aid to encourage abatement from developing nations, giving them a financial incentive and the means to comply. Kyoto offered nothing and found no harmonization with incidental benefits, so it failed to generate much reform or abatement at all.

Chapter 2: Game Theory Models of Transboundary Pollution

From an economic perspective, transboundary pollution involves strategic interactions between nations. The benefits that a nation receives from engaging in abatement of transboundary pollution depend not only on its individual efforts, but also on the efforts of all other relevant nations. With no central institution to bind nations to commitments, each nation must consider the strategies of other nations in order to secure the best outcome for themselves. Hence, transboundary pollution may be modeled through game theory as a strategic interaction between nations. Game theory offers a mode of analysis that aligns with the central characteristics of transboundary pollution. Primarily, game theory models acknowledge the absence of binding control institutions. Secondly, game theory weighs the interests of groups against the potentially conflicting interests of individuals.

Scholars of game theory have shown interest in developing models that explain the cooperation dynamics of transboundary pollution problems. This chapter explores various models circulating throughout the literature to arrive at a set of conclusions to weigh against the practical record of real-world environmental agreements. The literature shows that nations can reach superior outcomes through the use of coalitions and transfer payments. However, the most efficient agreements require full cooperation. Free riding incentives abound in transboundary pollution disputes, which prevent nations from sustaining full cooperative outcomes.

The Prisoner's Dilemma

Let us begin with the simplest model of transboundary pollution. Two symmetrical nations, A and B, each have two action options: pollute or abate. Let us assume that pollution between them is reciprocal; the benefits that each nation receives depends upon the other's action. Furthermore, in keeping with general game theory assumptions, let us also assume that both nations reveal their actions simultaneously and that each nation's payoffs and preference are common knowledge. We may model this simple, reciprocal interaction as a classic prisoner's dilemma. **Figure 1.1** provides an example of a reciprocal transboundary pollution dilemma. Payoffs are listed in the form (a, b), where a denotes nation A's payoff and b denotes nation B's payoff.

Consider the preferences of nation A. If nation B decides to pollute, nation A's best response is to also pollute: $(3, 3) > (1,6)$. If nation B decides to abate, nation A's best response is still to pollute: $(6, 1) > (4,4)$. So, in both cases, nation A 's best response to pollute. By the rule of symmetry, nation B's best response is to pollute when nation A chooses to pollute or abate. For both nations, regardless of the choice the other player makes, choosing to pollute results in a higher payoff than choosing to abate.

Thus, the outcome that results from both nations' best responses is where both nations choose to pollute. This is the Nash equilibrium: the combination of best responses such that no nation has any incentive for a unilateral change in action.[85] For nation A to abate while nation B pollutes results in a worse outcome for A than the Nash equilibrium: $(1, 6) < (3,3)$. By the rule of symmetry, neither nation benefits from choosing to abate when the other chooses to pollute. So, the preferences of both nations reinforce the Nash equilibrium.

[85] My thanks to Professor Aniruddha Mitra for this definition.

		Nation B	
		Abate	Pollute
	Abate	4,4	1,6
Nation A	Pollute	6,1	3,3

Figure 1.1

However, the Nash equilibrium is inefficient. Both nations prefer the outcome where both choose to abate: $(4, 4) > (3,3)$ for both nation A and B. As such, the outcome where both nations choose to abate improves upon the Nash equilibrium. The improvement is also Pareto efficient: no nation is made worse off moving from the outcome $(3, 3)$ to $(4,4)$. In addition, the outcome $(4,4)$ maximizes aggregate payoffs. Let us refer to the outcome that maximizes aggregate payoffs as the *full cooperative outcome*.

Though the full cooperative outcome improves upon the Nash equilibrium, it does not constitute an equilibrium. Like most problems involving public goods and management of resources held in common, this model of reciprocal transboundary pollution contains free-riding incentives. Each nation prefers another outcome over the full cooperative outcome. For nation A, $(6, 1) > (4,4)$. Similarly, for nation B, $(1, 6) > (4,4)$. Either nation may reap additional benefits by choosing to pollute while the other abates, encouraging each to deviate from the full cooperative outcome. However, if both nations deviate from full cooperation by choosing to pollute, then their payoffs simply revert to the inefficient Nash equilibrium. Though the full cooperative outcome results in higher payoffs for both nations than the Nash equilibrium, free-riding provides an incentive to deviate and return to the Nash equilibrium. The incentive to break away from full cooperation prevents it from being an equilibrium outcome.[86]

[86] Missfeldt. 1999.

Of course, unidirectional transboundary pollution does not conform to the reciprocal model. Unidirectional pollution occurs when "transboundary pollution between two states moves in the same direction. State A always sends pollution in the direction of State B[.]" (Merrill. 1997.) Nations engaged in a problem of unidirectional transboundary pollution are not symmetric. The victim nation does not have the option to engage in the polluting activity itself; its action set remains empty. Instead, the victim nation's payoff rests entirely on the decision of the source nation. Assuming, as in the previous example, that the benefit to polluting is higher than the benefit to abatement, the source nation prefers to pollute. Negotiating unidirectional transboundary pollution, then, is a "cooperator's loss" game: cooperation makes the source state worse off than noncooperation.[87] There are no collective gains from abatement, since only the source nation has the option to undergo abatement. As such, pollution and noncooperation is the preferred strategy of source nations.

Does communication between nations play any role? Nations involved in a transboundary pollution problem that can communicate with each other have potential to develop institutional arrangements that change payoff parameters, thus changing preferences. Employing a transfer-of-gains mechanism, a institutional arrangement that redistributes payoffs, is one such method. For example, both nations in a prisoner's dilemma may agree to impose fines on any nation that free-rides by choosing to pollute while the other abates.[88] The fine can then be awarded to the nation that abates to encourage their commitment.

Figure 1.2 presents an adjustment of the earlier prisoner's dilemma example whereby *A* and *B* communicate with each other and agree to a transfer payment of 3 from any nation that pollutes while the other abates. The payment is then awarded to the nation that abates. This

[87] Ibid.
[88] Barrett. 1997.

		Nation B	
		Abate	Pollute
	Abate	4,4	4,3
Nation A	Pollute	3,4	3,3

Figure 1.2

transfer payment changes the payoff parameters of outcomes, along with the preferences of each nation. Consider the preferences of nation A under this arrangement. If nation B decides to pollute, the best response of nation A is to abate: $(4,3) > (3,3)$. If nation B decides to abate, the best response of nation A is also to abate: $(4,4) > (3,4)$. So, nation A's best strategy is to abate in either case. By the rule of symmetry, nation B's best strategy is to abate in either case that A chooses to pollute or abate. Thus, the Nash equilibrium is where both nations choose to abate. With this transfer arrangement, the Nash equilibrium constitutes the full cooperative outcome. The change in parameters eliminates free-rider incentives, prompting both nations to choose to abate for the sake of optimization. So, with a transfer-of-gains mechanism, the full cooperative outcome may be sustained.

Though institutional arrangements can change payoff parameters, they do no good unless arrangements are binding between nations.[89] Primarily, agreements need a mechanism that allows participants to monitor each other to determine when free-riding occurs. Secondly, nations must commit to carrying out the institutional arrangements they agree upon. Monitoring alone only allow nations to unilaterally respond to each other's actions. Nations need to go above monitoring and commit to fulfilling the agreement in order to sustain parameter changes. Without credible commitments or an enforcement mechanism, nations may simply enter into agreements they have no intention of honoring or refuse to honor commitments when they

[89] Ibid.

conflict with self-interest, ultimately nullifying the parameter changes an agreement aims to impose.

Problems of monitoring and enforcement call into question the extent to which nations comply with institutional arrangements. Nations stand to gain from either refusing to honor commitments or defecting from an agreement, seeking the original payoffs offered by free-riding. Though institutions may change the parameters of a prisoner's dilemma, any nation that seeks to free-ride also prefers the original parameters. A free-rider may reinstate those parameters by simply refusing to fulfill commitments. Even with the parameter changes in the earlier example, each nation prefers to free-ride and provide no transfer payment to the other and return the game to the original parameters. Thus, institutional arrangements are not self-enforcing. By consequence, nether is the full cooperative outcome they hope to enforce.

Nevertheless, even without monitoring or enforcement mechanisms, transfer payments still provide a useful incentive, especially for unidirectional transboundary pollution problems. The use of a transfer-of-gains mechanisms allows us to model unidirectional pollution as a prisoner's dilemma by expanding the action set of the victim nation. Let us assume that two nations, S and V, are involved in a unidirectional transboundary pollution problem. The source nation, S, has the action options: pollute or abate. The victim nation, V, has the action options: transfer payment, no payment. Essentially, V has the option to offer S a payment to encourage abatement. With the potential to receive a transfer payment that compensates for the costs of abatement, the source nation now has an incentive to cooperate.[90]

Figure 1.3 presents a prisoner's dilemma for unidirectional transboundary pollution with the nations and action sets discussed above. The general outcome of unidirectional pollution is

[90] The transfer payment needs to compensate the state for both the physical and opportunity cost of undergoing abatement. Otherwise, the transfer payment provides no incentive for the source state to engage in abatement, as pollution remains more profitable.

		Nation V	
		Transfer	No Transfer
	Pollute	4,-5	2,-3
Nation S	Abate	3,-2	1,0

Figure 1.3

where S pollutes and V offers no transfer payment. If the potential transfer payment is set to 2, the model operates like a prisoner's dilemma. Though V's payoff in most outcomes still results in a loss, the outcome where S abates and V provides a transfer is a Pareto improvement from the general outcome. Even though V still receives a negative payoff, both V and S's payoffs nonetheless improve.

However, the model does not overcome the fundamental problems of a prisoner's dilemma. S's best response is still to pollute in either case that V offers payment or no payment; the general outcome of a unidirectional game, pollute and no payment, is the Nash equilibrium. In addition, the transfer payment also contains free-riding incentives for both parties. S prefers to pollute while receiving transfers from V: $(4, -5) > (3, -2)$. Likewise, V prefers to offer no transfer payment while S undergoes abatement: $(1, 0) > (3, -2)$. So, similar free-rider problems that arise in reciprocal models also arise with a transfer of gains mechanism in the unidirectional model.

The prisoner's dilemma model offers a wealth of explanatory power concerning the fundamental conflicting interests involved in transboundary pollution. Nations benefit from cooperation on abatement, but self-optimization compels nations to pollute. In addition, nations have stronger incentives to free-ride on the abatement efforts of others rather than undergo abatement themselves.

The N-Nation Pollution Model

Though the prisoner's dilemma provides simple and intuitive insight, it lacks most of the complex dimensions that characterize transboundary pollution. The most glaring oversight of the prisoner's dilemma model is that most transboundary pollution involve more than two nations. Multiple nations can alter free-riding incentives. How does the analysis change when more nations are introduced?

The N-nation pollution game, as outlined in *Barrett 1997*, explores the same tensions involved in the prisoner's dilemma across more than two parties. Suppose that there are N symmetrical nations involved in a transboundary pollution problem. All nations have the action set: pollute or abate. The benefits of abatement depend on collective abatement effort. Any nation benefits not only from their own abatement efforts, if they engage in any, but also the abatement efforts of other nations. Let k be the number of nations that opt to abate ($k \subseteq N$), π_p be the payoff function for polluting nations, and π_a be the payoff function for abating nations. We may model the payoff structure presented in the introductory prisoner's dilemma example. To model the payoffs of **Figure 1.1**, the payoff functions are

$$\pi_p = 3 + 3k$$

$$\pi_a = -2 + 3k.$$

A nation that chooses to pollute receives the benefit of economic gains wrought by the polluting activity (3) plus any benefits from the abatement efforts of other countries $(3k)$. An abating nation bears the cost of abatement (-2) but reaps the gains from its own abatement efforts (3), as well as the abatement efforts of other nations $(3(k-1))$.

The same preference dynamics of the prisoner's dilemma operate in this N-nation example. For any nation i, regardless of the value of k, the payoff to polluting is always higher than the payoff to abatement. Because nation i's decision to abate or pollute affects k, nation i's payoff functions are

$$\pi_p^i = 3 + 3k$$

$$\pi_a^i = -2 + 3(k+1),$$

or

$$\pi_a^i = 1 + 3k.$$

Hence, any nation always receive a higher payoff from choosing to pollute, regardless of the choices of other nations. Thus, the Nash equilibrium of this example is for all countries to choose to pollute.

In the aggregate, though, payoffs are maximized across all nations when all nations engage in abatement. Aggregate net benefits, π, are simply the sum of payoffs wrought by polluting nations and abating nations:

$$\pi = k\pi_a + (N-k)\pi_p$$

We can demonstrate that the outcome where all nations abate is the full cooperative outcome. Simplifying π yields the following:

$$\pi = -5k + 3kN + 3N.$$

If all nations abate, then $k = N$ and $\pi = 3N^2 - 2N$. Any other combination of strategies that does not involve all nations choosing to abate involves at least one nation that chooses to pollute. In any case where at least one nation chooses to pollute, $k < N$. When $k < N$, the following holds true:

$$3N^2 - 2N > -5k + 3kN + 3N.$$

Hence, when all nations choose to abate, aggregate benefits are always larger than any other combination of strategies that involve at least one nation choosing to pollute. The full cooperative outcome, then, requires that all nations choose to abate.

On the other hand, each nation maximizes their individual payoff by choosing to pollute. Though full cooperation leads to higher payoffs across all nations, all nations prefer the payoff to polluting over abatement. Essentially, this N-nation pollution example demonstrates the same tensions as its corresponding prisoner's dilemma. Free-riding incentives do not allow full cooperation to constitute an equilibrium point. Without binding commitments to abatement, nations revert to the inefficient Nash equilibrium.

However, different parameter values in an N-nation pollution game can result in equilibria whereby either all countries commit to the full cooperative outcome or a handful of countries commit to abatement while others continue to pollute. Suppose that payoff functions are

$$\pi_p = 3 + 3k$$

$$\pi_a = 5 + 2k.$$ [91]

Figure 1.4 graphs these two functions together, with the x-axis measured by k and the y-axis measured by individual payoffs. The functions intersect at the point $N = 5$. The equilibrium of this example is for five nations to abate and the rest to pollute. Past the point where five nations abate, the payoff to polluting exceeds the payoff to abatement. So, any additional nations, in a situation where five nations commit to abatement, receive a higher benefit from polluting. Just as in the earlier example, the full cooperative outcome maximizes aggregate net benefits.

[91] These parameters have been chosen arbitrarily for heuristic purposes.

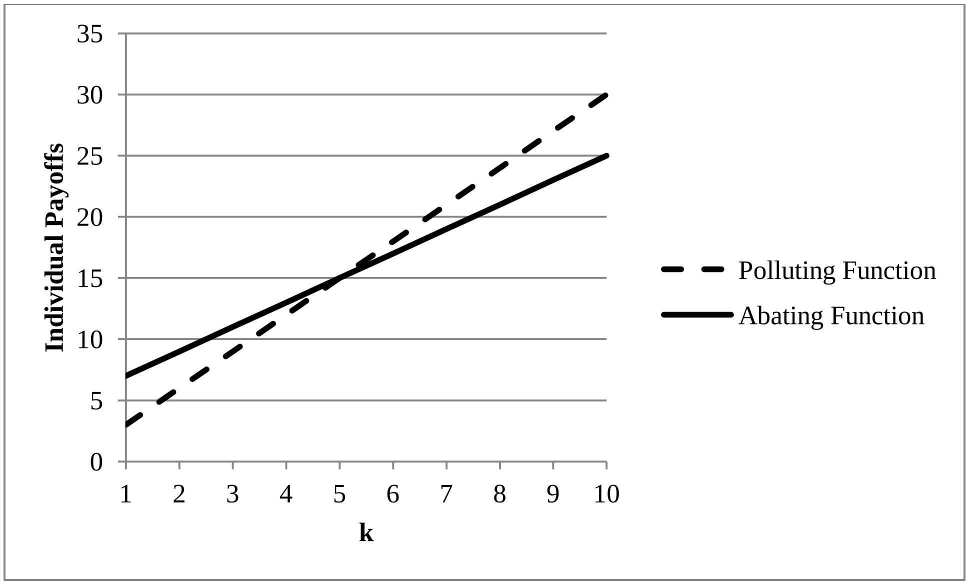

Figure 1.4

However, in this case, a handful of nations experience higher individual payoffs by choosing to abate. By these parameters, at most 5 prefer to abate. Thus, there are equilibria in N-nation pollution models that involve nations undergoing abatement.

In fact, given certain parameters, the full cooperative outcome can constitute an equilibrium. Suppose that the payoff functions from the earlier example are reversed:

$$\pi_p = 5 + 2k$$

$$\pi_a = 3 + 3k$$

As in the earlier example, these functions intersect at the point $N = 5$. But in this case, if there are less than 5 nations, all of them choose to pollute. If there are more than 5 nations, *all nations choose to abate***.** Thus, if $N > 5$, the full cooperative outcome is the equilibrium.

Equilibria in N-nation pollution models that involve some, or all, countries undergoing abatement are individually self-enforcing. For example, with parameters set forward in **Figure 1.4**, the equilibrium is for five nations to abate and for any additional nations to pollute. If any of the five polluting nations were to defect from abatement, choosing to pollute, that nation would receive a *lower* individual payoff. Similarly, if any polluting nation decided to abate while five nations were already committed to abatement, that nation would receive a lower individual payoff as well. Hence, there can be self-enforcing arrangements of abatement. Though the model does not indicate which nation will decide to abate, nations that choose to abate do so for the sake of self-optimization.

Unidirectional transboundary pollution, on the other hand, does not conform to the N-nation pollution model. Unidirectional transboundary pollution rarely involves more than a handful of nations. In addition, not all nations involved get to choose between polluting or abatement. However, with more victims comes the opportunity to provide higher transfers. Also, victims may choose to provide transfers to a handful of source states, receiving partial pollution relief. Nonetheless, the same free-rider problems occur regardless of the number of nations involved. The N-nation pollution model offers little additional insight into understanding unidirectional transboundary pollution.

Barrett's Model of Environmental Agreements

Restrictive assumptions still hold back the N-nation pollution model. For instance, most nations are not given a simple choice between pollution and abatement. Rather, nations select a level of abatement to engage in with respect to the actions of others and abatement costs. We can use the model developed in *Barrett 1997* to treat abatement as a continuous variable.[92] Assume that there are N symmetrical nations that take their actions simultaneously. All nations have the option to choose a level of abatement from a continuous spectrum of real numbers greater than

[92] *Barrett 1997* examines an adapted version of the model presented in *Barrett 1994*.

zero. Instead of two separate payoff functions, let π_i be the payoff function for nation i, where $i \in N$. Benefits, B_i, are a function of the aggregate abatement, Q, taken by all N nations. By definition, $Q = \sum_{i=1}^{N} q_i$. Costs, C_i, are a function of the quantity of abatement that nation i undergoes: q_i. The payoff function for any nation i is simply the difference between the benefits it receives from global abatement and the costs of its own abatement efforts:

$$\pi_i = B(Q) - C(q_i).$$

Optimization requires that nation i undergo a quantity of abatement q_i^* that maximizes π_i. Taking the first derivative of π_i with respect to q_i yields the following:

$$\frac{d\pi_i}{dq_i} = \frac{dB(Q)}{dQ} \cdot \frac{\partial Q}{\partial q_i} - \frac{dC(q_i)}{dq_i}.$$

Maximizing this equation, and taking into consideration that $\frac{\partial Q}{\partial q_i} = 1$, yields the following condition:

$$\frac{dB(Q)}{dQ} = \frac{dC(q_i)}{dq_i}.$$

Essentially, nations engage in abatement up to the point where *marginal benefits to abatement equal marginal costs*. Hence, nation i selects an abatement level q_i that maximizes its individual total benefits.

In the full cooperative outcome, each nation commits to a level of abatement that maximizes aggregate net benefits. Let π be the aggregate payoff function, where $\pi = \sum_{i=1}^{N} \pi_i$. The payoff function takes the following form:

$$\pi = NB(Q) - \sum_{i=1}^{N} C(q_i)$$

Taking the derivative of π with respect to q_i and setting that derivative equal to zero yields the following maximization condition:

$$\frac{NdB(Q)}{dQ} = \frac{dC(q_i)}{dq_i}.$$

Essentially, full cooperation requires that every nation engage in abatement up to the point where *aggregate marginal benefits equal individual marginal costs*.

Like the previous models explored so far, Barrett's model demonstrates the same tensions of cooperation. The full cooperative outcome results in the highest aggregate payoffs. However, under full cooperation, the marginal benefits to aggregate abatement are larger than individual marginal benefits (by the constant N). Hence, no nation is individually optimized under full cooperation. Each nation has an incentive to disregard the benefits of full cooperation in pursuit of higher individual benefits. Under a situation of full cooperation, each nation gains by defecting from cooperation, instead producing a quantity of abatement that maximizes their individual payoff function. So, full cooperation is unsustainable.

The benefit of Barrett's model is that it couches transboundary pollution strategies in the familiar economic logic of marginal costs and marginal benefits. It allows us to view transboundary pollution as a problem of economic decision-making, where, instead of discrete actions, nations make decisions along a continuum of action. The tensions inherent in resolving transboundary pollution problems, then, arise from differences between group marginal benefits and individual marginal benefits. The group always stands to benefit from cooperation, but the individual benefits more from pursuing their own rational optimization strategy.

Coalitions and Barrett's Self-Enforcing Agreement

In the full cooperative outcome, all nations commit to a level of abatement such that any deviation by any single nation makes all other nations worse off. However, the full cooperative outcome does not constitute an equilibrium. Any nation receives a higher payoff by defecting from full cooperation, producing instead a level of abatement that maximizes its individual payoff.

In Barrett's model of environmental agreements, nations may undergo any level of abatement they wish. In the full cooperative outcome, aggregate marginal benefits are higher than individual marginal benefits by the number of participating nations. Suppose that, under a situation of full cooperation, a single nation defects from the agreement and that the remaining nations wish to maintain a cooperative outcome among themselves. The remaining nations may commit to a level of abatement, q_c^*, that maximizes the new group's aggregate benefits. In this case, aggregate marginal benefits among the group members are *still* larger than individual marginal benefits (now by $(N - 1)$). So, the group still benefits from maintaining a cooperative outcome despite the free-riding behavior of the defector. Furthermore, any number of nations may choose to cooperate among themselves and still maintain aggregate marginal benefits that are higher than individual benefits. Hence, there is a rationale for nations to enter into coalitions, in the absence of full cooperation, that attempt to maximize aggregate benefits among its members.

Coalition theses have been introduced into transboundary pollution models to examine cooperative behavior in the absence of full cooperation. However, just as with the full cooperative outcome, coalitions encounter the same problem of free-riding. When individual marginal benefits are lower than the coalition's aggregate benefits, nations have an incentive to defect from the agreement. The introduction of coalitions does little to expand the analysis without some criteria for stability.

Coalitions need to satisfy two different conditions of stability: internal stability and external stability.[93] Coalitions are internally stable if no member of the coalition gains by unilaterally withdrawing. On the other hand, coalitions are externally stable if no nation outside of the coalition benefits by unilaterally conforming to the terms of the coalition, effectively inviting themselves in. *Barrett 1997* provides the conditions for internal and external stability in accordance with his model: Let K be a coalition of nations attempting to maximize aggregate benefits among themselves, where $K \subseteq N$. Let π_m be the payoff function for any country m that is a member of the coalition ($m \in K$). Let π_n be the payoff function for any country n that is not a member of the coalition ($n \in N \mid n \notin K$). A coalition K is stable if it satisfies the two different conditions of stability:

$$(i) Internal\ Stability: \pi_m(K) \geq \ \pi_n(K-1)$$

$$(ii) External\ Stability: \pi_n(K) \geq \ \pi_m(K+1).$$

Essentially, (i) requires that the payoff for coalition members must be greater than or equal to the payoff of being outside the coalition in the case that a single coalition member defects. That way, a defecting nation either makes themselves worse off or remains indifferent between defection and commitment. Also, (ii) requires that the payoff for nations outside of the coalition must be greater than or equal to the payoff of coalition members with the addition of another member. That way, a nation that unilaterally accedes to the coalition makes themselves worse off or remains indifferent.

With these two conditions, Barrett forms a strategy for self-enforcing environmental agreements, as shown in *Barrett 1997*, with the assumption that nations experience homogenous

[93] Tulkens. 1998.

cost and benefit functions. Nonmembers simply pursue a strategy of individual optimization, maximizing their payoff function:

$$\pi_n = B(Q) - C(q_n),$$

where q_n is the quantity that maximizes total individual benefits. Coalition members, on the other hand, seek to maximize aggregate benefits among themselves. The payoff function for a coalition member takes the following form:

$$\pi_m = KB(Q) - \sum_{m=1}^{K} C(q_m).$$

Differentiating with respect to q_m results in the following:

$$\frac{d\pi_m}{dq_m} = \frac{KdB(Q)}{dQ} \cdot \left[\frac{\partial Q}{\partial Q_K} \cdot \frac{\partial Q_K}{\partial q_m} + \frac{\partial Q}{\partial Q_n} \cdot \frac{\partial Q_n}{\partial Q_K} \cdot \frac{\partial Q_K}{\partial q_m}\right] - \frac{dC(q_m)}{dq_m}.$$

Maximization requires

$$\frac{KdB(Q)}{dQ} \cdot \left[\frac{\partial Q}{\partial Q_K} \cdot \frac{\partial Q_K}{\partial q_m} + \frac{\partial Q}{\partial Q_n} \cdot \frac{\partial Q_n}{\partial Q_K} \cdot \frac{\partial Q_K}{\partial q_m}\right] = \frac{dC(q_m)}{dq_m}.$$

The operators $\frac{\partial Q}{\partial Q_K}$, $\frac{\partial Q_K}{\partial q_m}$, and $\frac{\partial Q}{\partial Q_n}$ are each equal to one. Consider the value of $\frac{\partial Q_n}{\partial Q_K}$, how the quantity of abatement undertaken by nonmembers changes with respect to a change in abatement by members. Obviously, if all countries commit to the coalition ($K = N$), then $\frac{\partial Q_n}{\partial Q_K} = 0$ and the solution simply conforms to the full cooperative outcome. At the same time, Barrett's model is built on the assumption of simultaneous moves. Nonmembers lack any opportunity to react to the decision of the coalition members. So, taken as a reaction specification, $\frac{\partial Q_n}{\partial Q_K} = 0$.

Of course, if coalitions operate no differently than the full cooperative outcome, coalitions offer no expanded analysis of cooperation. Barrett drops the assumption of simultaneous moves to view $\frac{\partial Q_n}{\partial Q_K}$ as a specification that guides the decision of coalition members. Implicitly, this new assumption allows the coalition to move first in a Stackelberg fashion.[94] Through taking their action first, coalition members know the optimization strategy of nonmemebers—maximize π_n—and can choose $\frac{\partial Q_n}{\partial Q_K}$ such that they would find no benefit from revising their abatement decisions after observing nonmember abatement decisions.[95] Barrett defines $\frac{\partial Q_n}{\partial Q_K}$ as follows:

$$\frac{\partial Q_n}{\partial Q_K} = \frac{\frac{d^2B(Q)}{dQ^2} \cdot (N-K)}{\left[\frac{d^2C(q)}{dq^2} - \frac{d^2B(Q)}{dQ^2} \cdot (N-K)\right]}.$$

Substituting this back into the earlier equation results in the maximization condition for a self-enforcing environmental agreement:

$$\frac{KdB(Q)}{dQ} \cdot \left[1 + \frac{\frac{d^2B(Q)}{dQ^2} \cdot (N-K)}{\left[\frac{d^2C(q)}{dq^2} - \frac{d^2B(Q)}{dQ^2} \cdot (N-K)\right]}\right] = \frac{dC(q_m)}{dq_m}.$$

Essentially, coalition benefits depend not only on the number of nations participating in the coalition, but also on the number of nonmembers and the rates of marginal benefits and marginal costs. Whether marginal benefits and marginal costs are increasing, decreasing, or constant, plays a significant role in coalition benefits because second-order derivatives are involved in the maximization condition.

Is Barrett's assumption of a first-moving coalition justified? The assumption remains entirely exogenous to the model and argues for an asymmetry of information.[96] Nevertheless, nations must agree to form a coalition before taking their collective actions because the coalition

94 De Zeeuw. 1998. Finnus. 2001.
95 Barrett. 1997.
96 Finnus. 2001.

strategy depends on the number of nations joining; implicitly, there is some dimension of sequentially involved in coalition formation. In a game with sequential moves, the idea of a self-enforcing coalition only makes sense if the coalition moves first. A second-moving coalition would have to contend with nonmembers who anticipate the coalition's abatement contributions and adjust their strategies accordingly before the coalition moves. Nonmembers then take the expected abatement of the coalition as given and engage in less abatement themselves, defeating the logic of forming a coalition in the first place. In addition, a first-moving coalition does not necessarily conflict with practical experience. For example, Scandinavian countries jointly committed to high levels of acid rain abatement before the LRTAP's controls became legally binding.[97]

In order to determine the number of nations that comprise a self-enforcing environmental agreement, functional specifications for costs and benefits must be made.[98] However, with specifications, the model is no longer general. Benefits and costs depend in part upon the rate of environmental assimilation of a given pollutant, which varies drastically between different types of transboundary pollution. A choice of assimilation behavior changes the outcomes of a transboundary pollution model.[99] Using his own specifications, *Barrett 1994* conducts a series of simulations for games involving the formation of self-enforcing coalitions with the assumption that $N = 100$. He finds that the size of self-enforcing coalitions depends more on the ratio of costs and benefits than on their values themselves.[100] Thus, when the marginal benefits to abatement are high (marginal environmental damage is high) relative to abatement costs, self-enforcing agreements consist of many nations. However, though self-enforcing agreements

[97] Munton et al. 1999.
[98] Carraro and Siniscalco. 1993.
[99] Cesar and De Zeeuw. 1995.
[100] Barrett. 1994.

always improve upon level of abatement that takes place in a noncooperative outcome, they do not offer substantial improvements given most parameters.[101]

Barrett's findings align with the other papers that discuss self-enforcing agreements and coalitions. *Carraro and Siniscalco 1993* find that coalitions tend to be small regardless of the number of nations involved in a transboundary pollution game. *Hoel 1992*, instead of using the concept of a coalition, models an agreement where a group of countries commit to abatement under the condition that they are no worse off than under noncooperation. Under this constrained social optimum, he finds that few countries can be expected to participate and that coalitions only result in a modest increase in abatement. *Rubio and Ulph 2006* presents an analysis inspired by Barrett's model with the explicit additional assumption of non-negative emissions. Even with non-negative constraints, self-enforcing agreements may be as large as full cooperation or as small as two or three countries depending on benefits to abatement (environmental damages).

Though self-enforcing coalitions improve upon the non-cooperative outcome, they do not solve the commitment problem inherent with free-riding incentives; the success of a coalition *still* depends on commitment from its members. Though self-enforcing coalitions are stable in the sense that no nation inside the agreement gains from leaving and no nation outside the agreement gains from joining, all nations still have an incentive to remains uncooperative, since nonmember countries gain through free-riding on the abatement efforts of coalition members.[102] Consider a case where $(K^* - 1)$ countries form a coalition. The coalition becomes self-enforcing when one more nations joins in. However, the nations that decide against joining the coalition each benefit more than the nation that joins the coalition. The nations that abstain from joining the coalition gain free-riding benefits from the abatement efforts of the joining nation. All the

[101] Ibid.
[102] Ibid.

nations that decide against joining the coalition receive the benefits of the stable coalition while only the nation that joins bears the costs. Yet, if no nation joins, then all nonmembers are worse off. Hence, the decision for players to join a self-enforcing coalition takes the form of a game of chicken.[103] Furthermore, Barrett's model of self-enforcing environmental agreements explicitly assumes that compliance within the agreement is full, even though there is no mechanism inherent in the coalition to ensure that each member undergoes abatement exactly at the efficient level of q_m.[104]

Transfer Payments to Expand Coalitions

A coalition may attempt to expand its membership base by offering nonmembers a transfer payment in exchange for entry into the coalition. Essentially, a coalition may reap higher benefits by offering transfer payments to induce nonmembers to commit to q_m instead of q_n. Such transfer payments must be self-financed in the sense that coalition members gain more in abatement benefits than they lose in transfer payments.[105]

Depending on the types and levels of transfer payments involved, a coalition may expand its membership even up to achieving the full cooperative outcome in the form of a grand coalition. Assuming that a coalition simply makes self-financed, lump-sum payments to new members that cover their individual losses, full cooperation may be achieved if 60% of involved nations are part of the original coalition.[106] Or, a coalition can tie transfer payments directly to additional abatement efforts from nonmembers above what they produce in a Nash equilibrium.

[103] Ibid.
[104] Barrett. 1997.
[105] Carraro and Siniscalco. 1993.
[106] Ibid.

This is the grand stable coalition thesis, advanced by *Chander and Tulkens 1995* and *Tulkens 1997*. The grand stable coalition thesis views coalitions as a cooperative option given to all nations. A grand coalition that incorporates all nations provides a higher aggregate benefit than any other coalition; it is the full cooperative outcome. Members of the grand coalition, to ensure all nations adhere to the agreement, make transfer payments to cover the costs of additional abatement from nonmembers. Through a formulaic approach, the grand coalition can ensure that the costs of transfer payments and aggregate abatement costs are shared *collectively* among all parties.[107] Combined with the threat of coalition members disbanding the agreement entirely, committing to q_n if nonmembers do not change their abatement behavior, all countries enter into the grand coalition and the full cooperative outcome occurs.[108]

Just as with transfer payments in the unidirectional prisoner's dilemma, transfer payments by coalitions require commitment from both members and nonmembers. Though transfer payments preserve the profitability of coalitions, they create instability.[109] Without commitment, coalition members can attempt to back out of transfer payments, free-riding on payments financed by other coalition members. Alternatively, members outside of the coalition may simply take transfer payments without fulfilling additional abatement commitments. The grand stable coalition thesis especially relies on commitment between coalition members, not just on transfer payments but also the threat of playing q_n against defectors. Nations remaining committed to the coalition may simply renegotiate the agreement to maximize benefits among themselves instead of embracing an undesirable outcome from playing q_n.[110] All in all, transfer payments have no

[107] Tulkens. 1998.
[108] Ibid.
[109] Carraro and Siniscalco. 1993.
[110] Barrett. 1997.

self-enforcing mechanism and therefore lack stability despite their potential for encouraging cooperation.

Dynamic Games

So far, the game theory analysis presented here has only covered static games. Essentially, the models have been one-shot deals, where nations, though they can communicate beforehand and collude to form coalitions, ultimately settle on one path of action in a single round of decision-making. Noticeably, these games lack any dimension of time. Most real world environmental agreements and their negotiations play out over time and often involve changing control strategies in their later stages, like the Montreal Protocol. Furthermore, many agreements lose and gain members as time progresses, changing the aggregate benefits that an agreement results in and altering participation incentives.

Dynamic games introduce a time element that allows nations to consider strategies throughout different strategic periods. Games may be either divided into discrete stages, with a set number of rounds, or cover an infinite time horizon. To better capture the dynamics of real-world environmental agreements, this chapter will only focus on games with infinite time horizons. Environmental agreements rarely have a set expiration date and are frequently revised, re-enacted, and revived. Furthermore, transboundary pollution in the real world does not conform to discrete time horizons. Even after a discrete time horizon, nations can simply return to emitting individually rational levels of pollution. The tensions to engage in transboundary pollution remain eternal to the problem. So, real-world environmental agreements are better modeled through the application of an infinite time horizon.

Dynamic games bring the advantage of allowing for the incorporation of stock pollutant behavior: pollution damage that depends on the stock of pollutants in a given environment. Static games treat pollution damage as an instantaneous flow, though many of the world's most concerning pollutants display stock behavior.[111] All of the agreements discussed in Chapter 1 dealt with stock pollutants, where scientists advocated for abatement with respect to measures of critical environmental loads rather than general abatement. So, dynamic games that model stock pollutants offer a better characterization of modern transboundary pollution problems than static games.

However, the introduction of stock pollution damage changes the mode of analysis. With stock pollution, not only are pollution flows important but also the assimilative capacity of the environment. Barrett's model, though useful for general analysis, abstracts from pollution flows by deeming abatement as the variable choice given to nations, not emissions. As such, we require a model that allows nations variable control over their emissions levels and to consider pollution stock.

Dynamic games with stock pollutants employ a motion equation that incorporates pollution flow from all countries and the assimilative capacity of the environment, all as a function of time. *Missfeldt 1999* presents a general motion equation, replicated here:

$$S = \sum_{i=1}^{n} P_i(t) - \alpha f\big(S(t)\big),$$

where S is the stock of pollution, P_i the flow of pollution from nation i as a function of time, and α the assimilation capacity of the environment as a function of the stock of pollution. Hence, the

[111] De Zeeuw. 1998.

change in pollution stock at a given point t depends on total emitted pollutants across all nations minus natural assimilation.[112]

The benefits to any nation i depend not only on pollution flows among nations collectively, but also on the increase and decrease of pollution stock. We may employ our own dynamic payoff function that works with motion equation displayed earlier. Let the payoff function for any nation i take the following form:

$$\pi_i = B(P_i(t)) - D\big(S(t)\big),$$

where B is the benefits of emitting pollution as a function of personal emissions (P_i) and D is the damage caused by the total stock of pollution. Though the variable has changed from abatement to emissions in this model, the fundamental payoff logic remains. Individual payoff depends on the difference between benefits from personal emissions and damage from the stock of pollution to which all nations contribute. However, payoff depends not only on other nations immediate actions, but also on actions throughout time as the pollution stock increases and decreases.

Taking the first derivative of π_i with respect to P_i yields the following:

$$\frac{d\pi_i}{dP_i} = \frac{dB(P_i)}{dPi} - \frac{dD(S)}{dS} . \frac{\partial S}{\partial P_i}$$

which, by setting $\frac{d\pi_i}{dP_i} = 0$, can then be rearranged to reveal the maximization condition:

$$\frac{dB(P_i)}{dPi} = \frac{dD(S)}{dS} \cdot \frac{\partial S}{\partial P_i}.$$

In each round of time, nations pollute to the point where *individual marginal benefits of emissions equal marginal damages from the pollution stock*.

Just as in static games, full cooperation maximizes net benefits across all players. The aggregate payoff function takes the following form:

---

112 Missfedlt. 1999.

$$\pi = \sum_{i=1}^{N} B(P_i) - ND(S(t)),$$

where N is the number of nations. Taking the derivative of π with respect to P_i and solving for the maximization condition yields the following:

$$\frac{dB(P_i)}{dPi} = \frac{NdD(S)}{dS} \cdot \frac{\partial S}{\partial P_i}.$$

Under full cooperation, nations pollute in each round to the point where *individual marginal benefits of emissions equal aggregate marginal damages from the pollution stock.*

Yet, damages remain unclear without a specification for $\frac{\partial S}{\partial P_i}$: the change in pollution stock with respect to a change in individual emissions. Under both noncooperation and full cooperation, $\frac{\partial S}{\partial P_i}$ in turn depends on a specification for the assimilation function. Though nations directly contribute to pollution stock, their additions also affect environmental assimilation capacity. Commonly throughout most instances of transboundary pollution, higher levels of pollution stock reduce assimilation capacity. So, as nations continue to increase pollution stock, damages increase disproportionately because assimilation capacity reduces. Nonetheless, a functional specification for environmental assimilation capacity is needed to clarify the extent of damage.

However, providing a functional specification for environmental assimilation poses problems for this general analysis. The outcomes of stock pollution control models are sensitive to the choice of assimilation function.[113] For example, in the context of greenhouse gases, assimilation of pollutants cannot be approximated by a linear function, although many models attempt to do so.[114] Likewise, the assimilation of greenhouse gases differs from the assimilation

[113] Cesar and de Zeeuw. 1995.
[114] Missfeldt. 1999.

of ozone-depleting substances or pollutants that contribute to acid rain. Non-linear specifications lead to entirely different outcomes than linear specifications, with the possibility of multiple equilibria.[115] Thus, the fact that environmental assimilation differs between pollutants renders the construction of a general model counterintuitive. Attempting to construct a general model with a functional specification for environmental assimilation would only skew the analysis towards a certain characterization of pollutants.

Unlike static games, dynamic games reach different Nash equilibria depending on assumptions governing the availability of information. The most widely discussed modes of Nash analysis in the context of dynamic pollution games are open-loop information structures and feedback information structure.[116] Under an open-loop information structure, nations only know the initial value of pollution concentration and variables. No nation can adjust their behavior because each lacks the knowledge of current pollution levels. So, nations commit indefinitely to their initial strategies and take other nations' strategies as given.[117] Under a feedback information structure, nations observe the choices that other nations make through time and the changing pollution stock, but not the initial value of pollution concentration. Feedback Nash equilibria are also known as subgame perfect equilibria: they display strong time consistency throughout successive rounds and subgames of the overall model.[118] Technically, a third concept exists, closed-loop information structures, that combines the information available in both open-loop and feedback information structures. However, closed-loop analysis is not normally undertaken in game theory models concerning transboundary pollution because it

[115] Ibid.
[116] Ibid.
[117] Van de Ploeg and de Zeeuw. 1992.
[118] Finus. 2001.

requires too much complexity.[119] Nonetheless, Nash equilibria resulting from either open-loop or feedback Nash equilibria are not Pareto efficient: cooperation leads to better aggregate payoffs among players.[120]

Papers throughout the literature agree that feedback Nash equilibria result in higher pollution stock—and therefore higher damages—than open-loop Nash equilibria.[121] How does access to present information result in a worse outcome? *De Zeeuw 1998* offers a clear explanation of the intuition:

> Each country knows that in a feedback information structure, the other countries observe the stock of pollutants and react to higher stocks with lower output and pollution. Therefore, each country knows that an increase in output and pollution will then be partly offset by a decrease in all the other countries.[122]

As the stock of pollutants increases, damages increase and nations respond by lowering emissions to optimize their individual payoffs. At the same time, reduction among other nations creates an incentive for any nation to *increase* their own emissions, knowing that other nations offset their contributions. Alternatively, under a closed-loop information structure, nations cannot observe developments in the pollutant stock. Thus, nations do not know the strategies of other nations in response to higher pollution stocks.

Dynamic games also expand upon the theory of coalition formation by providing opportunities for trigger mechanisms. In the static context, coalitions are maintained according to internal and external stability. No member can gain by leaving the coalition, and nonmembers cannot gain by unilaterally acceding to the coalition. However, in a dynamic context, nations have the opportunity to react to defectors in later time stages. When a nation defects, coalition members have the opportunity to punish the defector in future rounds by ending the coalition. If

[119] Missfedlt. 1999.
[120] Van de Ploeg and de Zeeuw. 1992.
[121] Missfedlt. 1999.
[122] De Zeeuw. 1998.

the deviation of one nation leads to the coalition falling apart, defectors make themselves worse-off by ending the cooperative benefits that the coalition provides.[123] So, if all coalition members commit to a trigger strategy, following coalition controls when every member cooperates and reverting to noncooperative emission levels when any nation defects, larger coalitions may be sustained indefinitely.

However, there are three main issues with the use of trigger strategies. Firstly, the incentive for nations to sustain a coalition through trigger strategies depends on the future value discount rate of nations. In order to accurately represent economic preferences over an infinite time horizon, the incorporation of a positive time preference is needed.[124] In terms of dynamic models, nations' strategies need to account for a preference for present benefits over future benefits. Taking positive time preference into consideration, nations with a high discount rate gain less from the use of trigger strategies. Those nations have less preference for sacrificing present benefits for future gains. So, nations in a coalition only find trigger strategies desirable provided that their future value discount rate is low.[125]

Secondly, trigger mechanisms run into the same problem of enforcement as transfer mechanisms. While trigger mechanisms deter defection, they also hurt the coalition members that undergo them. All players have an incentive to renegotiate the coalition and attempt cooperation.[126] Commitment is required to make trigger strategies credible, otherwise nations may simply renegotiate a coalition when a member defects.

Furthermore, trigger mechanisms rely on a coalition's ability to monitor the actions of its members. With many transboundary pollutants, sources do not always remain clear, nor do

[123] De Zeeuw. 2008.
[124] Marini and Scaramozzino. 2000.
[125] Missfeldt. 1999.
[126] Barrett. 1997.

damages directly manifest after an increase in emissions. If deviations take time to detect, trigger mechanisms are not likely to work.[127] An effective monitoring mechanisms is required to keep trigger threats credible.

Conclusions

Game theory models demonstrate that the best outcome for all nations, full cooperation, is inherently unstable. Nations gain more from pursuing individual optimization strategies than respecting the commitments required by full cooperation. Though full cooperation remains unstable, opportunities to form coalitions and undergo transfer payments improve upon noncooperative outcome in nearly all models presented in this chapter. Coalitions and transfer payments, though, still require some basic level of commitment, as well as monitoring and enforcement abilities. Even with coalitions and transfer payments, environmental agreements cannot be expected to improve much upon what nations would already pursue in the absence of an agreement.

[127] De Zeeuw. 2008.

Chapter 3: Reconciling Game Theory with Practical Experience

Now that we have covered both game theory and real world environmental agreements, we can place the two chapters into conversation with each other. This chapter posits seven general arguments that support and critique conclusions from game theory, as well as point out fundamental shortcomings and characteristics of real-world environmental agreements. This chapter presents a unique contribution, as the literature on game theory and commentary on real-world environmental agreements rarely acknowledgement the views of each other.

The Behavior of Nations in Real-World Agreements Aligns Closer to Open-Loop Analysis than Feedback Analysis

Static game theory models provide useful insights into strategic behavior without the burden of complexity. However, static models also lack any sort of time dimension. Though static games may be expanded into games with discrete stages or time intervals, real-world environmental agreements operate under an indefinite time schedule. Though signatories may agree upon expiration dates for agreements, like with the Kyoto Protocol, transboundary pollution problems do not end with agreements or protocols. As long as free-rider incentives exist to benefit nations that engage in pollution, nations may simply return to polluting after the end of an agreement.

Thus, the tensions of cooperation remain ever-present with transboundary pollution problems in the real world. Agreements and control schedules often take decades to discuss, implement, and oversee. Also, signatories periodically revise agreements to account for changes in technology and pollution stock. For example, the Montreal Protocol has undergone multiple revisions over its lifetime and remains a binding international agreement even today.[128] Considering the lifespan of real-world environmental agreements and the perpetual problem posed by transboundary pollution, signatories to an environmental agreement need to formulate their strategies in a dynamic context with a near-infinite time horizon in order to achieve efficient outcomes.

However, as reviewed in the dynamic games section of Chapter 2, information assumptions change the analysis and outcomes of dynamic models. Though most models assume that nations know the costs and benefits of pollution and abatement, analyses differ with regard to available information on pollution stock and player strategies. This forms the crux of the dichotomy between open-loop and feedback information structures: open-loop analysis assumes that nations only know initial parameters and commit to initial strategies indefinitely, as they cannot view changes in other nations' behavior or the pollution stock, while feedback analysis assumes that nations only know present changes in pollution stock and others nations' responses to those changes.

Which one aligns closer to reality, open-loop or feedback analysis? Though feedback analysis remains theoretically useful because of its strong time consistency,[129] its assumption of access to present information remains unrealistic. Most environmental agreements and negotiations take place in a vacuum of present information. Both the Montreal Protocol and

[128] United Nations Development Program. 2012.
[129] De Zeeuw. 1998.

LRTAP created their own scientific institutions to rectify the lack of scientific development, consensus, and understanding at the time. Furthermore, access to present information requires a monitoring ability that has yet to be established for many transboundary pollutants. Most of the greenhouse gases that Kyoto attempted to control had no reliable monitoring potential at the time the agreement was put forward.[130] Even in situations where models can be developed, models nonetheless take time and resources to develop. These form transaction costs that feedback analysis often ignores, even though transaction costs present significant obstacles for access to present information in the real-world.

Secondly, feedback analysis assumes that nations easily adjust their emissions in response to new information. In reality, an abundance of uncontrollable factors affect pollution emissions that inhibit flexibility once a nation commits to an emission strategy. For example, greenhouse gas emissions vary with economic growth and technological change, neither of which can be accurately planned by governments.[131] Furthermore, energy infrastructure, often directly responsible for most of a nation's pollution emissions, does not lend itself to flexible adjustments. Energy infrastructure takes time to develop, plan, renew, or replace, often over decades. For example, four-fifths of the power plants expected to operate in 2010 were already in operation or under construction by the time of Kyoto's negotiation.[132] The United States had already committed to infrastructure that complicated the scheduled reductions imposed by Kyoto. Given these real-world inflexibilities, nations have little immediate ability to alter the course of their emission trends.

Given the inflexibilities and transaction costs associated with present information, real-world agreements align closer with open-loop analysis than feedback analysis. Most

[130] Victor. 2001.
[131] Ibid.
[132] Ibid.

environmental agreements operate on the basis of an initial set of commitments that nations must fulfill over a schedule. The Montreal Protocol, LRTAP, and Kyoto, all committed signatories to reduce pollution with respect to a past baseline. Signatories to real-world agreements are not given the option to alter commitments in response to the behavior of other signatories. Again, given inflexibilities and transaction costs, signatories bear little ability to alter their initial strategies anyway. Therefore, open-loop analysis bears more practical value than feedback analysis.

As reviewed in Chapter 2, there is general agreement throughout the game theory literature that open-loop Nash equilibria result in better outcomes (less pollution stock) than feedback Nash equilibria. Under a feedback information structure, nations observe how others reduce their emissions as pollution stock increases. So, all nations have an incentive to free-ride on the abatement of others and increase their own emissions. Without the ability to view present information or strategies under an open-loop information structure, nations cannot make the observations that leads to the free-riding behavior that occurs under feedback information structures. Instead, nations acting under an open-loop information structure commit to an initial strategy based only on initial information. So, even if real-world environmental agreements only result in noncooperative outcomes, they nonetheless achieve better outcomes than feedback analysis predicts.

Game Theory Models Lack Dimensions of Trust Linkage that are Critical to International Negotiations.

Game theory models assume that *homo economicus* forms the basis of modern statecraft: nations' primary motivation in negotiations is the maximization of self-gain, even at the cost of a breakdown in negotiations. However, *homo economicus* clashes with the delicacies of modern statecraft. Though certain agreements and negotiations may be one-shot games, behavior in a broad international context needs to account for trust linkages and diffuse reciprocity:

> Questions of treaty compliance arise in an environment of diffuse reciprocity, with manifold opportunities for subtle expressions of displeasure, suspicion, and reluctance to deal with treaty-violators in other contexts. A reputation for unreliability cannot be confined to the area of activity in which it is earned. It is inevitable that a state's defection from treaty rules will generate repercussions and linkages throughout the network of its relationships with others in the community.[133]

Diffuse reciprocity is the idea that nations expect other nations to conform to generally accepted standards of behavior in international relations and negotiations.[134] As *Chayes and Chayes 1991* argues, nations often approach treaty compliance as a generally accepted standard of international conduct. Nations that engage in noncompliance or defect from a treaty threaten their bargaining ability in other international agreements when members of the international community label and shun untrustworthy states. Though maximizing self-gain remains a fundamental motivation, nations must also consider the reputation they establish in the international sphere to maximize self-gain in the long run and across all international treaties and agreements they participate in. Therefore, game theory models underestimate nations' proclivity for cooperation in international contexts.

Of course, reputation concerns may also prompt states towards aggressive negotiation or stubborn refusals of commitments. It may pay for a country to remain incalcitrant and bargain down commitments in initial rounds of negotiations to gain a reputation as a tough negotiator.[135]

---

[133] Chayes and Chayes. 1991.
[134] Keohane. 1986.
[135] Maler. 1990.

Nevertheless, such behavior cannot escape the pressure of diffuse responsibility. Even though signatories may seek out tough negotiation positions, noncompliance and stubborn behavior has trust consequences. So, once an agreement becomes settled, diffuse responsibility pressures nations to commit.

The Centralization and Support of Scientific Research is Central to the Vitality of Agreements.

Another problematic assumption incorporated into game theory models is perfect information, that all involved parties have access to all information regarding the costs and benefits of pollution and abatement. None of the real-world agreements examined in this project were formed under circumstances of perfect information. In fact, many of these agreements were enacted during periods of weak scientific understanding. The early iterations of both the Montreal Protocol and LRTAP had no controls on pollution. Instead, they only committed parties to support scientific research and development. LRTAP created the European Monitoring and Evaluation Program (EMEP) and tasked it with harmonization and evaluation of emission data, as well as the development of common methodologies and pollutant modeling.[136] The Vienna Convention, the precursor to the Montreal Protocol, empowered UNEP to conduct working groups of scientific and engineering experts.[137] In turn, the Montreal Protocol developed its own dedicated scientific assessment panel.[138]

Dedicated scientific institutions within both regimes developed the information necessary to administer effective pollution controls. These institutions provided concerted scientific efforts to overcome coordination problems with data management. In addition, they advanced the

[136] Di Primio. 1998.
[137] Parson. 1993.
[138] Greene. 1998.

implementation of controls based on scientific understanding of pollutant behavior. Under the Montreal Protocol, scientists developed the standardized metric "ozone-depletion potential" to regulate emissions and focus early efforts on the most damaging targeted compounds.[139] Similarly, EMEP harmonized research methods and rationalized controls around critical-load analysis.[140] Thus, these regimes internally developed the information necessary to account for pollution damages and gradually adjusted control schemes to operate under a scientific basis.

Furthermore, scientific institutions enhanced these regimes by providing an effective monitoring capability. Harmonized data allowed regulatory bodies to make comparable environmental quality assessments between signatories,[141] as well as to ensure that signatories followed through with their commitments. In addition, though both the Montreal Protocol and LRTAP relied on self-reported data, scientific analysis proved capable of identifying and deterring bogus data submissions. *Munton et al. 1999* discusses evaluation capacity in the context of LRTAP:

> [The] collective emissions data, while relying on national reports, were considered highly reliable because they were regularly evaluated by scientists in the context of the creation of transport models. Erroneous national reports would have been detected in the course of modeling because of the anomalous results they would have generated. In fact, there are reports that the Czech government considered submitting false reports in the late 1980s, but opted not to precisely because it believed that the subterfuge would be detected through EMEP.[142]

EMEP's objective of creating a European-wide model of acid rain provided it with benchmarks against which to compare self-reported data. Thus, scientific development endowed EMEP with the ability to hold signatories accountable for the data they submitted. The Montreal Protocol's scientific assessment panel provided a similar function, though data conflicts were mainly

139 Epstein et al. 2014.
140 Levy. 1993.
141 Ibid.
142 Munton et al. 1999.

resolved by the Ozone Secretariat on an informal basis.[143] Dedicated scientific institutions provided a monitoring ability to ensure implementation of commitments.

Intriguingly, the support of scientific development spurred the transition of weak regimes into strong regimes by bypassing and eliminating political deadlock. Again, the early iterations of the Montreal Protocol and LRTAP contained no pollution controls: signatories only agreed to support scientific development. However, once the science developed and signatories were made aware of the extent of damages, signatories were pressured into accepting pollution controls in both regimes. In the case of LRTAP, research solidified consensus and led countries to discover domestic acid rain damage that had been previously unaccounted for.[144] Discoveries of domestic damage flipped many recalcitrant signatories into accepting LRTAP controls, including Austria, Finland, Switzerland, the Netherlands,[145] and Germany.[146] In addition, EMEP allowed Eastern European researchers to bypass their respective governments and submit data directly through LRTAP, refuting denials of involvement in acid rain.[147] Similarly, the United States shifted into a strong leadership position for CFC controls after the signing of the Vienna Convention once scientific consensus expanded.[148] Though the United States was not necessarily shifted by the Montreal Protocol's own scientific panels, those panels nonetheless drove international negotiations through their stamp of legitimacy in the regime.[149] Thus, scientific development helped curtail political reservations to abatement commitments.

---

[143] Victor. 1998.
[144] Levy. 1995.
[145] Levy. 1993.
[146] Munton et al. 1999.
[147] Ibid.
[148] Parson. 1993.
[149] Thoms. 2003.

Kyoto serves as a counterexample. While IPCC publications drove forward scientific consensus,[150] the institution was not *dedicated* to the Kyoto Protocol. The Kyoto Protocol suffered from incomplete and unharmonized data as countries ignored the proposed common framework and relied on their own nontransparent accounting methods.[151] Furthermore, Kyoto targeted pollutants that lacked practical scientific means of measurement or monitoring. When Kyoto incorporated methane and nitrous oxide into its targeted gases, no reliable proxy measures existed and emission factors, especially natural factors, varied widely.[152] In fact, the Kyoto Protocol had no reliable means to estimate or monitor greenhouse gases beyond carbon dioxide emitted from traded fossil fuels.[153] The lack of dedicated scientific input, review, and data harmonization crippled Kyoto from rationalizing its controls or accurately accounting for the emissions it sought to control.

On another note, the Montreal Protocol stands out among environmental regimes in its establishment of internal technical advising panels. Made up of engineering experts, these panels consulted with industries and services that relied on ozone-depleting substances, providing technical options for substitutes and elimination. For example, input from technical panels led to the development of essential-use exemptions for key industrial chemicals, avoiding complicated implementation problems that might have otherwise discouraged signatories.[154] Panels also addressed the practical concerns and constraints necessary to reduce friction in the compliance process.[155] These panels provided a key link between the regime and technical constraints.

[150] Ibid.
[151] Victor. 2001.
[152] Ibid.
[153] Ibid.
[154] Greene. 1998.
[155] Ibid.

Technical panels allowed the Montreal Protocol to promote abatement while avoiding painful economic disruptions.

Yet, despite the success and experiences of the Montreal Protocol, most environmental agreements are conducted in a complete vacuum of technical understanding. Neither LRTAP nor Kyoto established technical panels, nor most international environmental agreements. Understandably, regimes may remedy their ignorance of technical constraints and options through cooperation with other institutions. Though LRTAP benefitted from EC mandated reforms in automobile fuel efficiency and energy infrastructure, LRTAP failed to promote any reassessments of technical or economic constraints among its signatories.[156] Kyoto, on the other hand, struggled to address all practical considerations, technical or otherwise. Kyoto's architects reached no agreement on how flexibility instruments were to work, how sinks were to be measured (if at all), or how the agreement was to be enforced until the Marraketch accords,[157] after the infamous exit of the United States. The Protocol's arbitrary limitations on carbon sink creation and accounting, as well as emission trading, placed daunting and uncertain cost burdens on its signatories, especially the United States.[158] Furthermore, Kyoto failed to flesh out the procedures or instuitions for its own flexibility channels.[159] Only years after Kyoto's approval did signatories come to understand how mechanisms like the CDM or emission trading markets would work. Initially, signatories entered into the agreement under utter confusion regarding potential costs and implementation constraints.

So, environmental regimes stand to benefit from internal technical advisory bodies, as well as internal scientific ones. When incorporated, these advisory bodies engage with

[156] Levy. 1995.
[157] Thoms. 2003.
[158] Nordhaus and Boyer. 1999.
[159] Victor. 2001.

signatories of environmental regimes to indentify low-cost abatement measures. Technical advisory bodies also highlight infeasibilities in a regime's control measures, prompting internal reassessment. Besides technical advisory bodies, internal scientific advisory bodies have proved effect throughout international environmental regimes.[160] Internal scientific bodies endow regimes with the information necessary to rationalize abatement commitments as well as the key ability to monitor commitments. Through monitoring comes the ability for a regime to identify noncompliance and substantiate threats against it.

Transfer Payments Help Agreements Succeed, Though They are Seldom Applied.

Chapter 2 demonstrated how transfer payments can influence payoff parameters in both unidirectional and reciprocal games, as well as broaden coalitions in both static and dynamic games. Yet, the success of transfer payments rests on the assumption of commitment. Nations must commit to carrying out transfer payments and receiving nations must commit to undergoing abatement in exchange for those payments. Otherwise, nations on either side of the arrangement bear the inventive to free-ride on the actions of the other.

In real-world agreements, transfer payments are uncommon.[161] Most agreements rely on targeted percentage reductions across all parties, regardless of costs, or employ differentiated responsibilities for developing nations. Nonetheless, transfer payments have been successfully employed in the Montreal Protocol to the benefit of all signatories. In addition, the pollution dynamics of both LRTAP and Kyoto presented clear grounds for the implementation of transfer programs.

---

[160] Ibid.
[161] Maler. 1990.

To review, the Montreal Protocol carries out a transfer payment program through the multilateral fund. Through contributions from developed nations, the multi-lateral fund supplies funding to developing nations for projects that implement the freeze and elimination of ozone-depleting substances. The multi-lateral fund employs simple safeguards to prevent free-riding: funds are conditional upon project proposals that go through a review and approval process. The multi-lateral fund also subcontracts with the World Bank, UNEP, and UNDP for implementation needs. That subcontracting arrangement provides for redundancy and utilizes existing resources without creating new bureaucracy.[162]

The multi-lateral fund solved simple compliance problems primarily among developing nations. During its first years of implementation, the Montreal Protocol struggled to collect data from many of its signatories because developing nations lacked the institutional capacity to produce necessary data sets.[163] Hence, many multi-lateral fund projects ensured basic compliance, keeping developing nations in the agreements as opposed to harassing them through the noncompliance process. By keeping developing nations in the agreement, the multi-lateral fund prevented leakage that would have undermined abatement efforts.

Multi-lateral fund projects have been implemented across the world and seen successful results. From 1991 to 2012, nearly 2200 multi-lateral fund projects have been implemented across 105 countries, resulting in the phase out of nearly 67,000 tons of ozone-depleting substances.[164] Many of these projects have addressed direct technical concerns in key commodity industries in the developing world. UNDP champions the Montreal Protocol and multi-lateral fund projects as "an example of integration of environmental and equity concerns while promoting human development." (United Nations Development Program. 2012) The regime and

[162] Parson. 1993.
[163] Greene. 1998.
[164] United Nations Development Program. 2012.

all of its signatories have reaped immense gains towards the reduction of ozone-depleting substances from this transfer program.

LRTAP had no transfer program despite the obvious potential such a program presented at the time. Many countries with low abatement costs were directly upwind from sensitive European ecosystems. Poland and Czechoslovakia, countries whose abatement costs were 80% less than Western European countries, contributed a large percentage of Norway and Sweden's acidification.[165] Yet, LRTAP pursued uniform, self-financed reductions across Europe, which varied widely in terms of costs. These reduction ignored the regional nature of acid rain transmission and damages; some countries' emissions presented more of a transboundary threat than others. A transfer payment scheme could have been arranged precisely to fund cheap abatement among Europe's greatest contributors to acid rain.

Certainly, European nations, at the time of LRTAP's implementation, were aware of the great potential that transfer payments presented because many nations decided to arrange their own programs. European nations unilaterally and bilaterally engaged in a mix of uncoordinated loans, grants, and assistance for acid rain abatement.[166] Game theory scholars have also commented on acid rain abatement in Europe and advocated for transfer payments: *Kaitala et al. 1992* conducts a game theory model of acid rain between Finland and Russia, concluding that Finland should have offered payments to Russia in exchange for abatement. Even still, without centralized and concerted funding from LRTAP, many Eastern European nations struggled to meet commitment. Though nearly all European nations achieved 30% reductions in sulfur

[165] Levy. 1993.
[166] Levy. 1995.

dioxide by 1994, most Eastern European countries failed to meet the nitrogen oxide protocol or volatile organic compounds protocol.[167]

Kyoto, on the other hand, made the bizarre decision to both allow for transfers to occur through permit exchanges and excuse developing nations entirely from any pollution controls. Developed countries sought to purchase abatement from developing nations to ease the burden of their commitments. The United States determined its low compliance cost estimation on the condition that it purchased 75% of its abatement overseas.[168] Yet, despite the fact that they offered cheap abatement options, developing nations were excluded from controls and unable to participate in permit exchanges. The exclusion hampered the purchasing opportunities available to developed nations. Though the Clean Development Mechanism (CDM) allowed developed nations to engage in projects similar to those carried out by the Montreal Protocol's multi-lateral fund, credit for CDM projects was arbitrarily capped by Kyoto and no recognition was given for negative-carbon projects, such as the creation of new carbon sinks.[169] Furthermore, institutions and guidelines for the CDM and carbon-permit exchanges were not developed until long after the exit of the United States. For years, signatories had no ability to forecast the use of these institutions into their abatement plans.

The exclusion of developing nations dealt a critical blow to Kyoto. Whereas the Montreal Protocol fostered the participation of developing nations, extending its controls globally, Kyoto failed to regulate 40% of world emissions. Worse, the exclusion allowed developing nations to *increase* their emissions, creating leakage. Even more puzzling is the fact that Kyoto was inspired by the Montreal Protocol.[170] According to Kyoto's architects, the exclusion of

[167] See *Munton et al 1999.*
[168] Victor. 2001.
[169] Chichilinsky and Sheeran. 2009.
[170] See *Thoms 2003, Sunstein 2007*, and *Green 2009*.

developing nations embodied a principle of common but differentiated responsibilities based on the Montreal Protocol.[171] However, the Montreal Protocol did not exclude developing nations. It included developing nations while offering them financial relief through the multi-lateral fund. Somehow, Kyoto's architects interpreted the Montreal Protocol's precedent as one of *exclusive* responsibilities between developed and developing nations. That bastardized interpretation prevented Kyoto from subjecting developing nations to controls or instituting a transfer payment mechanism to encourage commitment.

The stakes of the Montreal Protocol perhaps explain why it remains one of the few agreements that operates a transfer program. Signatories realized that full participation from developing nations was necessary to prevent leakage.[172] Nonetheless, even at the time Kyoto settled on the exclusion of developing nations, it was clear that such exclusion doomed controls to a modest impact on global warming.[173] A transfer program could have rectified this obvious oversight, yet none was formed because of the policy norm of exclusive responsibilities.

As will be discussed in Chapter 4, other currently active policy and legal norms discourage the employment of transfer payments to resolve transboundary pollution. Even despite the clear economic potential of transfer payments, nations do not consider transfer payments viable when operating under these norms. Nevertheless, the Montreal Protocol provides successful precedent for the use of transfer payments. LRTAP demonstrates that transfer payments may even be employed outside of an agreement to generate regional benefits. Transfer payments remain a practical policy tool that future agreements should consider to make up for the failure of agreements like Kyoto.

[171] Chichilinsky and Sheeran. 2009.
[172] Parson. 1993.
[173] Nordhaus and Boyer. 1999.

Soft Enforcement Mechanisms Have Found More Success than Hard Enforcement Mechanisms.

Intuitively, transboundary pollution regimes require enforcement mechanisms to ensure that signatories carry out commitments. Otherwise, free-rider incentives motivate signatories to either under-provide towards their commitments or, if they cannot avoid detection by monitoring mechanisms, defect from the regime entirely. The success of a transboundary pollution regime hinges on enforcement mechanisms that combat free-riding behaviors.

However, many transboundary pollution regimes have found more success with soft enforcement mechanisms, like informal noncompliance procedures, non-dedicated review bodies, and active negotiations, than hard enforcement mechanisms, like formal noncompliance procedures and sanctions. In fact, few dispute resolution systems in environmental regimes are invested with any significant resolution powers.[174] Both the Montreal Protocol and LRTAP advanced their goals with soft enforcement mechanisms. Soft enforcement mechanisms, though weak, permit strong consensus-building powers, whereas strong mechanisms tend to generate hostility among signatories and threaten the integrity of an environmental regimes.[175]

The Montreal Protocol serves as a better example of effective soft enforcement mechanisms than LRTAP, since LRTAP owes much of its success to incidental changes and motivations. The Montreal Protocol directly brought about compliance through soft enforcement mechanisms. Though the Montreal Protocol included provisions for trade sanctions against outsiders and noncompliant members, no formal sanctions were technically incorporated into the compliance review process. Instead, the implementation committee sought to resolve noncompliance problems on a relatively informal basis with linkages to other Protocol

[174] Victor. 2001.
[175] Levy. 1993.

institutions. The implementation committee tackled problems of incomplete and missing data. Work between the implementation committee and the multilateral fund provided technical and financial assistance to developing nations for data reporting.[176] Technical advising panels also assisted Protocol signatories with data preparation, especially among nations of the former Soviet Union.[177] This facilitative approach worked with noncompliant nations to facilitate resolution. In fact, all original noncompliance issues were submitted for review by the affected parties themselves.[178] Instead of aggravating noncompliant nations and prompting withdrawal from the Protocol, informal processes addressed issues while ensuring for participation.

Of course, the Montreal Protocol's soft enforcement mechanisms did not achieve success merely by themselves. Rather, they thrived through institutional synergies. The Montreal Protocol had the advantage of multiple non-dedicated review bodies, such as its expert advisory bodies, that provided the advantage of redundancy.[179] In addition, the noncompliance process was backed by powerful sticks and carrots: trade sanctions and multi-lateral fund respectively.[180] Hence, the Montreal Protocol's soft enforcement mechanisms operated above a bedrock of formal hard measures. At the same time, soft enforcement mechanisms need not form the only enforcement mechanisms of an environmental regime. The Montreal Protocol's blend of soft and hard enforcement mechanisms avoids unproductive antagonism while retaining the credible threat of tough action to ensure cooperation.[181]

Nonetheless, many elements of soft enforcement mechanisms prove invaluable to international environmental agreements. Soft enforcement mechanisms work through diffuse

---

[176] Ibid.
[177] Greene. 1998.
[178] Ibid.
[179] Greene. 1998.
[180] Victor. 1998.
[181] Victor. 1998.

responses that lend themselves to cleaner calibration than formal sanctions.[182] Also, soft enforcement mechanisms harmonize with the many inherent uncertainties of treaty formation:

> treaties, like all legal instruments, contain ambiguities because of unforeseen circumstances, changing conditions or technology, or differences that were papered over in the negotiation process. Disputes about compliance with treaties are therefore not occasional departures from a steady state, but a continuing fact of life, as in any complex legal regime.[183]

Again, all of the environmental agreements reviewed by this project were enacted during times of underdeveloped scientific understanding. Without complete understanding, signatories enter into environmental agreements without full knowledge of potential costs or consequences of abatement procedures. Soft enforcement mechanisms ensure that regimes remain flexible enough to deal with new developments or expected difficulties without punishing signatories for factors beyond their control.

Therefore, considering the novelty of most transboundary pollution problems as they arise, environmental agreements find more success with soft enforcement mechanisms. Utilizing soft enforcement mechanisms as the first line of defense against noncompliance brings the advantage of engagement without hostility. Yet, soft enforcement mechanisms may encounter compliance problems too difficult to resolve. In that case, hard enforcement mechanisms may prove necessary. A blend of both soft and hard enforcement mechanisms combines the best of both approaches: the diffuse interactions of soft mechanisms and the tangible threats of hard mechanisms. The Montreal Protocol demonstrates the feasibility of such a dual approach.

Most Agreements Achieve No More Than What Countries are Already Comfortable Committing To.

---

[182] Chayes and Chayes. 1991.
[183] Ibid.

Throughout most international environmental agreements, signatories rarely commit to abatement above and beyond their own desires without support mechanisms. Many agreements enact weak controls that place signatories ahead of schedule. Some nations escape controls entirely while remaining in a regime, while others only comply due to incidental changes in energy policies, economic activity, or technology. Certainly, there are exceptions. The United States bore two-thirds of Kyoto's cost burden and the Montreal Protocol carried out a successful transfer payment program. However, as will be shown below, these exceptions prove the general rule.

With the Montreal Protocol, the threat of environmental damage was high enough to justify many of the bold maneuvers the Protocol took. Even still, nations were already ahead of the first control requirements by the time they were enacted.[184] Furthermore, previous ozone-control proposals had met with harsh industry backlash throughout the 1980s. The chemical industry refused to accept the Montreal Protocol until profitable substitutes for CFCs were near completion and its own scientists confirmed CFCs' threat to the ozone layer.[185] Much of the success of the Montreal Protocol stems from the fact that industry and powerful nations endorsed it once controls were deemed economically feasible.[186]

With LRTAP, incidental reforms allowed most signatories to meet controls, like European Commission directives:

> Neither the [European Commission's] power plant directive nor the vehicle emissions directives were motivated by transboundary concerns. In both cases, Germany adopted strict national standards out of concern for its own forests, and sought to extend those standards to other members to equalize the terms of economic competition. [187]

---

[184] Parson. 1993.
[185] Gareau. 2010.
[186] Sunstein. 2007.
[187] Levy. 1993.

Critical energy policy reforms were mainly initiated at the EU level, not among signatories. Neither the sulphur nor the nitrous oxide protocols motivate signatories to revise their domestic emission-reduction policies.[188] Furthermore, many European countries incidentally complied with LRTAP by tackling domestic pollution problems that the regime helped them identify. As covered previously, most of central Europe accepted controls once they discovered domestic damages that abatement could alleviate. As such, European countries were poised to undergo abatement, both deliberately and incidentally, regardless of LRTAP.

With Kyoto, all signatories, save for the United States, bargained for paltry reductions against their baselines:

> Indeed, many of the nations that accepted specified reductions actually promised to do little or nothing beyond what had already been done as a result of economic developments. Russia was given a target of 100% its 1990 emissions, but by 1997, its actual emissions had already dropped to a mere 70% of that amount due to economic difficulties...Germany appeared to accept a significant reduction—8% by 2012—but in 1997, its own emissions were already 10% lower than in 1990, as a result of reunification with the former East Germany...For the United Kingdom...the target reduction of 8% was less severe than it seemed, because in 1997, the United Kingdom was already at a level 5% below that of 1990.[189]

Economic difficulties in Europe and the break-up of the Soviet Union had already set up most European nations for compliance with Kyoto's commitments. Nonetheless, after the exit of the United States, signatories at the Marraketch accords bargained for weaker commitments.[190] While developed nations accepted little more than a freeze in emissions according to the 1990 baseline, developing nations won complete exemption from any controls whatsoever.

How can we interpret weak commitments in the context of game theory? Do weak commitments mean that agreements only result in self-interested commitments, or are there at

---

[188] Ibid.
[189] Sunstein. 2007.
[190] Thoms. 2003.

least self-enforcing coalitions at work within? The basic logic of coalition behavior is that some nations may engage in abatement above their self-interest through collective behavior. Certainly, that behavior manifests in real-world agreements. For example, Scandinavian nations unilaterally committed to higher levels of acid rain abatement than other LRTAP signatories. At the same time, those nations engaged in higher abatement due to technological advantage and higher potential benefits from protecting sensitive ecosystems. Nations in the real world are not homogenous, yet the game theory models in this project only considered homogenous nations.[191] Nonetheless, the weak starting point of most agreements demonstrates that most countries accede to agreements only on the condition that their commitments do not drastically conflict with their self-interest.

Conclusions

Game theory models elucidate the conventional problems of cooperation associated with transboundary pollution. However, their conclusions appear too dismal with respect to cooperation. With real-world agreements, countries normally cooperate and commit to agreements unless controls are wholly unfeasible. At the same time, the inability of real-world agreements to push signatories towards full cooperation, in the sense of maximizing aggregate benefits, aligns with the conclusions of game theory. Though countries cooperate with international agreements, agreements usually affirm the commitments that countries would have pursued anyway in the absence of an agreement.

---

[191] Models with heterogeneous nations exist in the game theory literature. However, those models are far too complex for the scope of this project and do not offer much additional insight beyond models with homogenous players. See *Barrett 1997* for a brief overview of models with heterogeneous nations.

In addition, despite the success story of the Montreal Protocol, its unique circumstances set it apart from other environmental regimes. The high benefits to CFC reduction propelled the Montreal Protocol to encourage the development of substitutes for ozone-depleting substances, hence its establishment of a technical advisory committee.[192] What's more, the Montreal Protocol's noncompliance mechanisms were never used to solve difficult compliance problems.[193] Though other environmental regimes struggle with data submission, basic institutional support proves an easy remedy. Serious compliance problem, like failure to meet control requirements or illicit trading of ozone-depleting substances, were never handled through this compliance mechanism.

Furthermore, production of ozone-depleting substances was concentrated among a few industrial actors in key nations.[194] In the 1970s, the United States produced half of the worlds CFCs, and DuPont alone produced half of the United State's total.[195] Only a handful of other production operations existed, most of them concentrated in Europe and Japan.[196] Hence, ozone-depleting substances were a relatively simple transboundary pollution problem in terms of pollution sources, whereas other transboundary pollution problems, like acid rain and greenhouse gases, have multiple complex sources for regulators and noncompliance procedures to consider.

Despite all its brilliant design and innovative arrangements, the Montreal Protocol nonetheless bore the unique advantage that both signatories and industrial producers supported the main principles and objectives of the regime.[197] All signatories had a vested interest in making the Montreal Protocol successful. Most other forms of transboundary pollution lack any

[192] Wettestad. 2001.
[193] Ibid.
[194] Epstein et al. 2014.
[195] Parson. 1993.
[196] Ibid.
[197] Greene. 1998.

such sort of consensus, let alone motivate abatement. The common expectation of environmental agreements is either incidental or scant success in promoting abatement.

Chapter 4: Barriers to Resolution

Though game theory models and real-world experience do not offer much hope for resolving transboundary pollution, two fundamental barriers exist in the real world that prevent agreements from reaching better outcomes. The first one is a policy norm: the polluter pays principle. The second one is a legal norm: tort law and liability regimes for transboundary pollution. This chapter discusses the problems that both norms pose for the resolution of transboundary pollution.

The Polluter Pays Principle

The polluter pays principle has served as a general principle of international environmental law and policy since its introduction by the OECD in 1972.[198] The general aim of the principle is simple: "Polluters should bear the expenses of carrying out measures...decided by public authorities to ensure that the environment is in an acceptable state." (OECD. 1972.) By assuming the costs of pollution abatement, polluters then have the incentive to economize on pollution, either through cutting down on output that results in pollution, implementing technological solutions, or finding non-polluting substitutes for materials or production processes. Though consumers of goods from polluting industries may bear some of the burden of abatement through higher prices, the polluter bears the ultimate responsibility for undertaking and financing abatement efforts. Many international agreements directly acknowledge and

[198] Gaines. 1991.

support the polluter pays principle, the most prominent of which include the Treaty of Maastricht[199] and the Rio Declaration on Environment and Development of 1992.[200]

The polluter pays principle aims to serve both efficiency and equity concerns, though the OECD denies that it constitutes an equity principle.[201] In terms of efficiency, the polluter pays principle prevents nations from subsidizing abatement of pollution. Abatement subsidies create distortions in a nation's domestic economy. Industries that benefit from abatement subsidies gain additional income, which may affect their choice of pollution level. Subsidies also require some method of financing. Taxes implemented to fund a subsidy can create their own distortions and inefficiencies as well. So, the polluter pays principle steers national policies away from creating the inefficiencies and distortions that result from subsidies.

Furthermore, the effects of an abatement subsidy spill over into international trade. Industries that receive subsidies for abatement gain additional income, allowing them to undercut competitors in the international market.[202] When capital moves across international borders, abatement subsidies encourage industries to move their production to nations that offer them. Thus, the polluter pays principle also prevents distortions in international trade. Polluting industries receive no special treatment regardless of where they locate when all nations adhere to the principle. Though the same effect occurs if all nations offer the same subsidy, the polluter pays principle prevents the race-to-the-bottom that would occur between nations as they compete to offer subsidies.

In terms of equity, the polluter pays principle places the burden of abatement on the actor with the most control over the polluting activity. Victims of pollution rarely exhibit any degree

[199] OECD. 1992.
[200] Luppi et al. 2012.
[201] See *OECD 1992*.
[202] Gaines. 1991.

of control over polluting activities. Specifically, in the case of unidirectional transboundary pollution, victims exhibit no direct control over polluting activities. Polluters, on the other hand, subject victims to pollution and willfully elect to engage in a polluting activity. Cases of reciprocal transboundary pollution complicate the distinction between victim and polluter. Each actor may engage in a polluting activity, but they also suffer from the pollution of other actors. Nonetheless, each actor exhibits no direct control over the polluting activity of other actors. In both reciprocal and unidirectional cases, victims lack control over the polluting activities of others whereas polluters exhibit full control of their own polluting activities. Therefore, placing the burden of abatement on polluters satisfies equity concerns by targeting actors with the most control over pollution.

However, the polluter pays principle warrants critique in relation to transboundary pollution. Before undergoing that critique, though, it is important to note that transboundary pollution remains outside the original premise of the principle. Most OECD documents concerning the polluter pays principle fail to explicitly address transboundary pollution in any context. Though the original 1972 recommendation lists "transfrontier pollution" among the exceptions to the polluter pays principle—briefly and without any elaboration on the matter—no such reference exists in subsequent recommendations in 1974 and 1992.[203] Instead, the OECD views trade distortions that arise from differences in cost allocation decisions between nations as the central concern of the polluter pays principle.[204] Nevertheless, the principle has become an accepted policy norm in the realm of transboundary pollution. Many international transboundary pollution agreements post-1990 make explicit reference to the polluter pays principle without

[203] See *OECD 1972, 1975,* and *1992*.
[204] Gaines. 1991.

defining it.[205] In addition, both the history of international case law and international conventions enshrine the polluter pays principle as a fundamental legal norm in transboundary pollution disputes.[206] Regardless of how the OECD views appropriate application of the polluter pays principle, the international community applies it to transboundary pollution.

Though the polluter pays principle appears reasonable in terms of a nation's domestic pollution policy, the principle conflicts with the realities and dynamics of transboundary pollution between different political entities. To reiterate from the introduction of this project, transboundary pollution normally takes place in an environment without a central regulatory authority. A polluter pays principle necessitates some sort of institution that can enforce property rights and legal obligations: the right of victims to a clean environment and the obligation of polluters to assume the burden of pollution abatement. Without such an enforcement institution, the principle meets with incentive problems.

Consider a case of unidirectional transboundary pollution, where a source nation emits pollution and a victim nation bears damage from that pollution. The source nation receives economic benefits from the output produced by polluting industries. Were the source nation to engage in abatement, it would suffer either from the costs of abatement or the loss of potential output, or both. Source nations have no economic incentive to embrace the polluter pays principle with regards to themselves, since they benefit more from the status quo. Only the victim gains from the application of the polluter pays principle. As the victim has no way to force the polluter pays principle upon the source nation, the source nation may simply refuse to acknowledge the principle.

---

[205] See *Vicha 1991*.
[206] Kettlewell. 1992.

Aside from the problems of enforceability, the polluter pays principle also creates a barrier to resolving transboundary pollution because it prevents the possibility of engaging in transfer payments in exchange for abatement. Transfer payments embody a victim pays principle. In the case of unidirectional pollution, the victim nation offers payment in exchange for abatement from the source state. In the case of reciprocal pollution, nation that engage in transfer payments assume the role of victims by paying for relief. Hence, transfer payments are incompatible with the polluter pays principle.

Yet, transfer payments offer a useful resolution mechanism for transboundary pollution in both theoretical and practical application. In terms of theory, the use of transfer payments expands the action set of victim nations in unidirectional games. In reciprocal games, transfer payments allow for the expansion of coalitions. Assuming cooperation and commitment, extra benefits generated by the coalition can finance payments that encourage other members to join. The polluter pays principle denies actors of any of these options, leaving victim states without recourse for alleviation and coalitions without a method to encourage expansion.

As an example of practical application, the Montreal Protocol clearly violates the polluter pays principle through the operation of its multi-lateral fund. The fund is a transfer program between developed and developing nations whereby funds from developed nations are exchanged for abatement in the developing world. Developed nations receive no such funding for their abatement efforts. Yet, the multi-lateral fund benefits all parties involved in the Montreal Protocol. Developing nations acquire the funds necessary to participate in the agreement and developed nations prevent leakage from the agreement.

In addition, the polluter pays principle encourages the alienation of developing nations from international environmental agreements. The need for developing nations to engage in economic development often outweighs their ability to abate any pollution that results from that development.[207] Thus, many developing nations refuse to acknowledge any responsibility for global consequences of the pollution they create.[208] In extreme cases, like the Kyoto protocol, developing nations completely exempt themselves from any controls whatsoever. Forcing a polluter pays principle on developing nations, in the context of international transboundary pollution, serves only to agitate and prompt developing nations to either remain outside of an agreement or seek special exemption.

The Montreal Protocol facilitated the inclusion of developing nations because it ignored the polluter pays principle. By including a transfer payment program, the Montreal Protocol kept developing nations under the agreement, holding them to abatement responsibilities while providing the means and incentives to comply. Kyoto, offering no compensation to developing nations, had no means to motivate an acceptance of abatement commitments. The only way Kyoto won the approval of developing nations was by excluding them entirely from pollution controls.

Arguably, the exceptions listed in OECD recommendations for the polluter pays principle cover the use of transfer payments. OECD recommendations exclude the use of the polluter pays principle in cases "where measures taken to promote a country's specific socio-economic objectives, such as the reduction of serious inter-regional imbalances, would have the incidental effect on constituting aid for pollution control purposes." (OECD. 1975.) Though the exception mainly applies to imbalances *within* nations, the vague wording may also apply to imbalances *between* nations. Nevertheless, the OECD does not explicit address transboundary pollution in its recommendations. Furthermore, many more exceptions riddle the OECD recommendations:

---

[207] Ibid.
[208] Ibid.

cases of stimulating experimentation with new technologies, cases of rapid implementation, and even cases where implementation is simply too difficult are all listed exceptions.[209] With all of its exceptions considered, the polluter pays principle constitutes only a guiding principle in international economic relations, not doctrine.

Despite that flexibility, the polluter pays principle remains a poor policy norm and threatens the success of international environmental agreements. It remains inconsistent with both theoretical application and practical experience. Even where international agreements themselves do not embrace transfer payments, nations often engage in transfer payments through existing foreign aid programs. Though the polluter pays principle may be sound policy for foreign trade and domestic environmental policy, it limits the success of transboundary pollution agreements.

Furthermore, international environmental agreements tend to find selective inspiration from the policy and legal norms of past agreements. Kyoto, for example, misinterpreted the differentiated responsibilities of the Montreal Protocol as exclusive responsibilities. Policymakers operating only on the basis of broad policy norms may ignore exceptions to the polluter pays principle entirely. Even a loose application or acknowledgement of the polluter pays principle may compel nations to refuse to participate in transfer programs. The international community needs to account for and clarify appropriate application of the polluter pays principle. Otherwise, the principle will remain open to strict interpretation. Until that clarification, the principle will continue to impede the resolution of transboundary pollution.

Legal Norms: Tort Law and Liability Regimes

---

[209] OECD. 1975.

So far, we have only viewed transboundary pollution through an economic lens: how do harm, costs, and benefits influence the behavior of economic actors involved in a transboundary pollution problem? Do participants have an incentive to cooperate towards resolution? However, international agreements, nations, and political actors are all creatures of law that respond to and operate under legal norms as well as economic motives. Many international environmental agreements are not pollution control regimes that hold states to abatement commitments at all. Instead, they are liability regimes that assign blame, standards of care and diligence, and recourse for damages. Liability regimes govern many cases of transboundary pollution, including nuclear power generation, oil spills, and international trade in hazardous waste.

Liability regimes and control regimes are compatible with each other. An international agreement can set pollution targets while also holding nations liable for transboundary damages that result from their pollution. However, agreements rarely combine the two in practice. Most pollution control agreements deliberately sidestep liability issues and many liability regimes lack any pollution controls.[210] So far, this project has only observed pollution control agreements, which begs the question: what does it mean to resolve transboundary pollution through a liability regime?

Surprisingly, international law lacks a consistent or broad body of law concerning transboundary pollution. Treatment of transboundary pollution remains poorly developed in practically all legal contexts.[211] Most presently employed legal precedents and norms actually stem from a handful of early 20th century cases from the Supreme Court of the United States.[212] In terms of international law, the *Trail Smelter Arbitration* stands alone as one of the *only*

[210] Brunnee. 2004.
[211] Merrill. 1997.
[212] Ibid.

examples of legal arbitration for transboundary pollution,[213] and its precedent essentially appropriated the American legal precedent of state liability for pollution between political bodies.[214] Needless to say, no thorough or mindful force crafted modern legal precedents for transboundary pollution. Rather, a haphazard case-by-case history forms the basis of common transboundary pollution law that guides current legal norms.

Nevertheless, modern legal norms and precedents for transboundary pollution find their roots in tort law, which provides for a simple legal logic for hazardous enterprises: the one who administers a dangerous activity is liable for the harm inflicted by the activity.[215] Tort law essentially affirms the polluter pays principle in a legal context. This precedent is established in both US domestic law, with *Missouri v. Illinois* and *Georgia v. Tennessee Copper Co.*,[216] and international law, with the *Trail Smelter Arbitration*.[217] However, instead of applying liability to "the one who administers a dangerous activity," (i.e. the firm or person) international law has historically applied it to the state, or nation, that houses a polluting activity. The *Trail Smelter Arbitration* held Canada liable for the pollution damages in the United States caused by a smelter within its borders.[218] Aside from legal precedent, international environmental agreements also establish state liability for transboundary pollution. Principle 21 of the Stockholm Declaration of 1972 declares that states have "the responsibility to ensure that activities within their jurisdiction or control do not cause damage to the environment of other states or of areas beyond the limits of national jurisdiction." (United Nations Environment Program. 1972) Principle 2 of the Rio Declaration of 1992 echoes Principle 21 of the Stockholm Declaration.[219] So, through both legal

[213] Ibid.
[214] Dinwoodie. 1972.
[215] Jones. 1992.
[216] See *Merrill 1997*.
[217] See Dinwoodie. 1972.
[218] Ibid.
[219] See *United Nations Environment Program 1992*.

precedent and international agreement, the international community ascribes liability to the nation that originates transboundary pollution. Though nations may not wholly administer or oversee polluting activities within their borders, they are, in principle, liable for any damages born outside their borders caused by those activities.

Nevertheless, international law does not offer a clear norm governing the *extent* of a nation's liability. Tort law distinguishes between three types of liability: strict liability, negligence, and partial liability. On one hand, strict liability entails that a nation bears liability for any and all pollution damage to which it subjects another nation. On the other hand, both negligence and partial liability hold nations liable for damages in accordance with some standard of reasonable care. In most legal contexts, that standard is left to courts to decide on a comparative basis with available active precautionary measures.[220] Negligence and partial liability differ in that negligence holds nations liable for all damages when they fail to exercise reasonable care, whereas partial liability holds nations liable for only those damages directly resulting from the failure to exercise reasonable care.

Generally, international law has favored the application of strict liability in cases of transboundary pollution. Many legal regimes apply strict liability to high-profile transboundary pollution problems, such as nuclear power generation, oil spills, and transportation of hazardous waste.[221] However, strict liability is not necessarily encapsulated in precedent or foundational legal treaties. The Stockholm and Rio declarations only call for state liability, not strict liability.

All degrees of liability have their trade-offs, advantages, and disadvantages. In practice, courts struggle to identify the standard of reasonable care necessary for the application of negligence or partial liability, whereas courts need not concern themselves with identifying a

---

[220] Schafer and Schonenberger. 1999.

[221] See *Brunnee 2004* for in-depth examples.

standard of reasonable care under strict liability.[222] On the other hand, the need to establish a standard of reasonable care under partial liability or negligence rations out claims of injury, discouraging petty lawsuits that may arise under a standard of strict liability.[223] At the same time, the burden of proving negligence rests with the victim, who usually lacks the ability deliver such proof.[224] So, strict liability offers the surest legal relief for victims at the risk of burdening the judicial system, whereas negligence and partial liability limit the leverage of victims based on available information.

Nevertheless, the legal norms of state liability and strict liability are not consistently applied throughout international liability regimes. Most liability regimes for transboundary pollution established post-1990 have largely abandoned the application of state liability in favor of civil liability.[225] Thus, modern liability regimes hold the owners and operators of hazardous enterprises liable for transboundary pollution damages, not the nations where those activities are carried out. Instead, nations are only expected to administer legal recourse through their own judicial system when a dispute arises. Though state liability has largely been abandoned, strict liability remains an active legal norm.[226]

Both state liability and strict liability act as barriers for resolution of transboundary pollution disputes. State liability prevents resolution of transboundary pollution disputes because pollution cannot easily be attributed to a nation as wrongful conduct on its behalf.[227] As with ozone-depleting substances, acid rain, and greenhouse gases, nations often lack any awareness of transboundary pollution they engage in or suffer from until clear scientific developments. There

[222] Ibid.
[223] Jones. 1992.
[224] Ibid.
[225] Brunnee. 2004.
[226] Ibid.
[227] Rao. 2003.

is little care or precaution that nations can take against transboundary pollution when they lack full knowledge of its effects or even existence. Moreover, victims are not guaranteed reparations because they must file suit in the courts of the very nation that harmed them.[228] Nonetheless, state liability no longer posses much of a barrier to resolving transboundary pollution because it has largely been abandoned.

Strict liability, on the other hand, remains an international legal norm. From a pure bargaining perspective, strict liability acts as a barrier because it aggravates bargaining dynamics between victims and polluters in a transboundary pollution dispute. Strict liability limits outcomes to either full reparation or no reparation. Thus, both victims and polluters have their entire interests at stake in a transboundary pollution lawsuit with strict liability.[229] Those limited outcomes do not allow for a partial settlement, such as transfer payments on condition of abatement. Consider the case of unidirectional transboundary pollution. Partial relief may be established through the use of transfer payments. However, if a victim operates under the norm of strict liability, they have the incentive to bypass a negotiation process entirely and seek full relief in a court of law. Confronted with strict liability, the polluter bears the incentive to avoid a lawsuit and deny the pollution problem entirely even if it found a partial agreement acceptable.[230] This is despite the fact that both the victim and polluter benefit from a partial settlement based on transfer payments: the victim receives less pollution and the polluting nation receives payment in exchange for abatement. Strict liability encourages nations to take extreme positions that clash with mutual resolution. Therefore, strict liability limits transboundary pollution resolution to extreme outcomes.

---

[228] Ibid.
[229] Merrill. 1997.
[230] Ibid.

In addition, strict liability often places a high burden of proof on the victim party.[231] Considering the lack of scientific knowledge often involved in a transboundary pollution dispute, victims often have little means to prove that they damage suffer from directly results from the actions of a source entity. Such was the case with *Missouri v. Illinois*. Missouri also dumped sewage into its own rivers, so the state had no way to prove that its suffering was caused by sewage from Illinois.[232] Most reciprocal transboundary pollution encounters the same problem. Consider the case of acid rain in Europe. Nations bore damages from a combination of chemical fluxes from surrounding nations without a clear method to determine which fluxes arose from which sources. Strict liability offers the best potential outcome for victims but requires an unreasonable burden of proof.

The only example of transboundary pollution that has found workable resolution through strict liability is international oil spills.[233] However, oil spills are the exception that proves the rule. With international oil spills, the origins and extent of damage are both clear and simple to prove in a court of law. In addition, as trans-national oil shipment is a well-established enterprise, a reasonable standard of care exists to weigh injuries against. Few other transboundary pollutants involve the same features of clarity and precedent.

Arguably, all forms of liability fail to harmonize with the realities of transboundary pollution. Even negligence and partial liability remain problematic because they require an establishment of reasonable care. Given the novelty of most transboundary pollution problems as they arise, no such standard may be established until well after damages have occurred. Hence, legal recourse offers even less hope for resolution of transboundary pollution than the economic bargaining reviewed throughout this project. Though it may be appropriate to establish liability

[231] Ibid.
[232] See Merrill. 1997.
[233] Brunnee. 2004.

guidelines for hazardous enterprises, liability regimes in and of themselves cannot hope to resolve transboundary pollution problems. Legal recourse and pure liability regimes should be limited only to cases where clear and capable judicial authority and exists and where reasonable care and sufficient proof of pollution sources may be established. Unfortunately, many crucial transboundary pollution problems lack either circumstance.

Conclusions

Both liability norms and the polluter pays principle conflict with the resolution of transboundary pollution. However, the removal of either norm brings with it further problems. Though liability norms rarely resolve transboundary pollution in and of themselves, they nonetheless communicate expectations for international enterprise and encourage preventative care through the power of diffuse reciprocity. In addition, liability and control regimes remain compatible with each other. Instead of abandoning liability concepts altogether, an environmental regime may set liability standards while resolving pollution primarily trough abatement controls.

The removal of the polluter pays principle, however, does not lend itself to a simple solution. While the removal of the polluter pays principle does not imply the establishment of a victim pays principle, it nonetheless allows for the victim pays principle to operate. With the potential to receive transfer payments under a victim pays principle, nations may find it attractive to substitute domestic pollution for transboundary pollution, thus leading to an increase in transboundary pollution.

Nonetheless, by linking policy issues together, nations may induce each other to cooperate against extreme outcomes in the absence of a polluter pays principle. Because international transboundary pollution takes place in a dynamic setting, nations may discourage substitution by linking substitution issues together. Essentially, when all nations wield the opportunity to substitute into transboundary pollution, each nation must anticipate that substitution on their behalf leads to further substitution by other nations, making all nations worse off. Thus, linkage mitigates the threat of substitution.

Nevertheless, the polluter pays principle bears similar value to liability norms in the sense that both communicate the importance of preventative action against transboundary pollution. Yet, the establishment of liability norms within an environmental control regime accomplishes the same goal without the burden of preventing transfer payments. As long as nations aim to resolve pollution primarily through abatement controls, liability norms in international environmental agreements may encapsulate the polluter pays principle without interfering with the resolution process.

Final Conclusions

Even without binding regulatory institutions, transboundary pollution does not result in a hopeless tragedy of the commons. Though some agreements have failed, international effort currently abates and manages many transboundary pollutants that threaten significant damage through innovative institutions. Nonetheless, transboundary pollution conflicts with the incentives of individual actors, discouraging the possibility of full resolution. Both game theory and real-world agreements demonstrate that nations remain unlikely to cooperate on transboundary pollution without key institutional arrangements, which in and of themselves require cooperation. Even removing the barriers caused by the polluter pays principle or liability regimes may not result in much improvement beyond the current status quo of international environmental negotiations. Transboundary pollution resolutions are not likely to achieve the most efficient outcome possible for all nations involved.

With the right mechanisms, practical experience demonstrates that international environmental cooperation responds better to immediate threats of significant and clear damage. Such was the case with ozone-depleting substances and acid rain. However, problems like climate change, where damages are unclear and expected to accrue over a wide timescales, clash with the interests of uncertainty-adverse, future-discounting nations. Truly, climate change is one of the greatest challenges this century faces precisely because it threatens future generations with relatively intangible damages. By the time present damages become severe enough to motivate action, international abatement efforts may not be able to halt the positive feedback cycles of climate change, let alone reverse them.

Nevertheless, this project has only viewed transboundary pollution as a singular policy issue. As single issues, transboundary pollution agreements are unstable due to free-riding incentives and payoff asymmetries. However, linking different transboundary pollution issues together in the same negotiations can reduce those problems. Issue linkage can stabilize an agreement in the same way that institutional arrangements can, leading to a more symmetric distribution of gains from cooperation.[234] The loss by one nation in a transboundary pollution dispute may simply be offset by a gain in another area of policy.[235] In addition, issue linkage can expand the action set of victim nations in a situation of unidirectional transboundary pollution,[236] providing a similar function to transfer payments. Issue linkage also increases the power of trigger strategies by allowing nations to respond to free-riding in one issue by severing other agreements.

Though many game theory models already consider issue linkage,[237] future research should consider limiting linkage with transboundary pollution to other transboundary pollution problems. Linking transboundary pollution with other unrelated international issues may create enormous transaction costs and prevent flexibility in negotiations. Limiting linkage to only other pollution problems allows nations to resolve multiple issues of transboundary pollution together, maintaining the benefits of more symmetric payoffs and powerful triggers, without spillovers into other critical international issues.

[234] Ibid.
[235] Botteon and Carraro. 1998.
[236] Finnus. 2001.
[237] See *Finnus 2001* and *Botteon and Carraro* 1998.

Finally, the development of scientific research and technological advancement remains fundamental to resolving transboundary pollution. The more information available to policymakers and nations, the more likely that environmental agreements will result in more equitable and efficient arrangements. Aside from encouraging scientific advancement and consensus, future agreements should also include technical and engineering experts into their implementation process in the spirit of the Montreal Protocol to reconcile policy commitments with practical constraints.

Bibliography

Barrett, Scott. "Towards a Theory of Environmental Cooperation." In *New Directions in The Economic Theory of The Environment*, edited by Carlo Carraro and Domenico Siniscalco, 239-80. Cambridge, UK: Cambridge University Press, 1997.

Barrett, Scott. "Self-Enforcing International Environmental Agreements." *Oxford Economic Papers*, Special Issue on Environmental Economics, 46 (October 1994): 878-94.

Baumol, William J., and Wallace E. Oates. *The Theory of Environmental Policy*. 2nd ed. Cambridge, UK: Cambridge University Press, 1988. First published 1975 by Prentice-Hall.

Bohringer, Christoph, and Carsten Vogt. "Economic and Environmental Impacts of the Kyoto Protocol." *The Canadian Journal of Economics* 36, no. 2 (May 2003): 475-94.

Botteon, Michelle and Carlo Carraro. "Strategies for Environmental Negotiations: Issue Linkage with Heterogeneous Counties." In *Game Theory and the Environment*, edited by Nick Hanley and Henk Folmer, 181-203. Cheltenham, UK: Edward Elgar Publishing, 1998.

Brunnee, Jutta. "Of Sense and Sensibility: Reflections on International Liability Regimes as Tools for Environmental Protection." *The International and Comparative Law Quarterly* 53, no. 2 (April 2004): 351-67.

Carraro, Carlo, and Domenico Siniscalco. "Strategies for the International Protection of the Environment." *Journal of Public Economics* 52 (1993): 309-28.

Chander, Parkash, and Henry Tulkens. "A Core-Theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution." *International Tax and Public Finance* 2, no. 2 (August 1995): 279-93.

Chayes, Abram, and Antonia Handler Chayes. "Compliance Without Enforcement: State

Behavior Under Regulatory Treaties." *Negotiation Journal* 7 (1991): 311-31.

Chichilnisky, Graciela, and Kristen A. Sheeran. *Saving Kyoto*. New Holland Publishers, 2009.

Cesar, Herman, and Aart De Zeeuw. "Sustainability and the Greenhouse Effect: Robustness Analysis of the Assimilation Function." In *Control and Game-Theoretic Models of the Environment*. Edited by Carlo Carraro and Jerzy A. Fliar, 25-46. Vol. 2 of *Annals of the International Society of Dynamic Games.* Boston: Birkhauser, 1995.

De Zeeuw, Aart. "Dynamic Effects on the Stability of International Environmental Agreements." *Journal of Environmental Economics and Management* 55 (2008): 163-74.

De Zeeuw, Aart. "International Dynamic Pollution Control." In *Game Theory and the Environment*, edited by Nick Hanley and Henk Folmer, 237-54. Cheltenham, UK: Edward Elgar Publishing, 1998.

Di Primio, Juan Carlos. "Data Quality and Compliance Control in the European Air Pollution Regime." In *The Implementation and Effectiveness of International Environmental Commitments: Theory and Practice*, edited by David G. Victor, Kal Raustiala, and Eugene B. Skolnikoff, 283-303. Global Environmental Accord. International Institute for Applied Systems Analysis, 1998.

Dinwoodie, D. H. "The Politics of International Pollution Control: The Trail Smelter Case." *International Journal* 27, no. 2 (Spring 1972): 219-35.

Epstein, Graham, Irene Perez, Micheal Schoon, and Chanda L. Meek. "Governing the Invisible Commons: Ozone Regulation and the Montreal Protocol." *International Journal of the Commons* 8, no. 2 (August 2014): 337-60.

Finus, Micheal. *Game Theory and International Environmental Cooperation*. New Horizons in Environmental Economics. Cheltenham, UK: Edward Elgar Publishing, 2001.

Folmer, Henk, Nick Hanley, and Fanny Missfeldt. "Game-Theoretic Modeling of Environmental and Resource Problems: an Introduction." Introduction to *Game Theory and the Environment*, edited by Nick Hanley and Henk Folmer, 1-29. Cheltenham, UK: Edward Elgar Publishing, 1998.

Gaines, Sanford E. "The Polluter-Pays Principle: From Economic Equity to Environmental Ethos." *Texas International Law Journal* 23 (1991): 463-95.

Gareau, Brian J. "A Critical Review of the Successful CFC Phase-out versus the Delayed Methyl Bromide Phase-out in the Montreal Protocol." *International Environmental Agreements: Politics, Law & Economics* 10, no. 3 (2010): 209-31.

Green, Bryan A. "Lessons From the Montreal Protocol: Guidance for the Next International Climate Change Agreement." *Environmental Law* 39, no. 1 (Winter 2009): 253-83.

Greene, Owen. "The System for Implementation Review in the Ozone Regime." In *The Implementation and Effectiveness of International Environmental Commitments: Theory and Practice*, edited by David G. Victor, Kal Raustiala, and Eugene B. Skolnikoff, 89-136. Global Environmental Accord. International Institute for Applied Systems Analysis, 1998.

Hoel, Michael. "International Environmental Conventions: The Case of Uniform Reductions of Emissions." *Environmental and Resource Economics* 2 (1992): 141-59.

Jones, William K. "Strict Liability for Hazardous Enterprise." *Columbia Law Review* 92, no. 7 (November 1992): 1705-79.

Kaitala, Veijo, Matti Pohjola, and Olli Tahvonen. "An Economic Analysis of Transboundary Air Pollution between Finland and the Former Soviet Union." *The Scandinavian Journal of Economics* 94, no. 3 (September 1992): 409-24.

Keohane, Robert O. "Reciprocity in International Relations." *International Organization* 40, no.

1 (Winter 1986): 1-27.

Kettlewell, Ursula. "The Answer to Global Pollution? A Critical Examination of the Problems and Potential of the Polluter-Pays Principle." *Colorado Journal of International Environmental Law and Policy* 3, no. 2 (1992): 430-55.

Levy, Marc A. "International Cooperation to Combat Acid Rain." *Green Globe Yearbook of International Cooperation on Environment and Development*, 1995b, 59-68.

Levy, Marc A. "European Acid Rain: The Power of Tote-Board Diplomacy." In *Institutions for the Earth*, edited by Peter M. Haas, Robert O. Keohane, and Marc A. Levy, 75-132. Global Environmental Accord. 1993. Reprint, Cambridge, MA: MIT Press, 1995.

Luppi, Barbara, Francesco Parisi, and Shruti Rajagopalan. "The Rise and Fall of the Polluter-Pays Principle in Developing Nations." *International Review of Law and Economics* 32 (2012): 135-44.

Maler, Karl-Goran. "International Environmental Problems." *Oxford Review of Economic Policy*, Economic Policy Towards the Environment, 6, no. 1 (Spring 1990): 80-108.

Marini, Giancarlo, and Pasquale Scaramozzino. "Social Time Preference." *Social Time Preference* 13, no. 4 (December 2000): 639-45.

Missfeldt, Fanny. "Game-Theoretic Modeling of Transboundary Pollution." *Journal of Economic Surveys* 13, no. 3 (1999): 287-321.

Merrill, Thomas W. "Golden Rules for Transboundary Pollution." *Duke Law Journal* 46, no. 5 (March 1997): 931-1019.

Munton, Don, Marvin Soroos, Elena Nikitina, and Marc A. Levy. "Acid Rain in Europe and North America." In *The Effectiveness of International Environmental Regimes*, edited by Oran R. Young, 155-248. Cambridge, MA: MIT Press, 1999.

Nordhaus, William D., and Joseph G. Boyer. "Requiem for Kyoto: An Economic Analysis of the Kyoto Protocol." *The Energy Journal*, Special Issue: The Costs of the Kyoto Protocol: A Multi-Model Evaluation, 20 (1999): 93-130.

Norman, Catherine S., Stephen J. DeCanio, and Lin Fan. "The Montreal Protocol at 20: Ongoing Opportunities for Integration with Climate Protection." *Global Environmental Change* 18 (2008): 330-40.

O'Connell, Mary Ellen. "Enforcement and the Success of International Environmental Law." *Indiana Journal of Global Legal Studies* 3, no. 1 (Fall 1995): 47-64.

OECD. The Polluter-Pays Principle: OECD Analyses and Recommendations. Environment Directorate no. OCDE/GD(92)81. Paris, 1992.

OECD. "Council Recommendation on the Implementation of the Polluter-Pays Principle." *International Legal Materials* 14, no. 1 (January 1975): 234-41.

OECD. Recommendation of the Council on Guiding Principles Concerning International Economic Aspects of Environmental Policies. Recommendation no. C(72)128. Paris, 1972.

Parson, Edward A. "Protecting the Ozone Layer." In *Institutions for the Earth*, edited by Peter M. Haas, Robert O. Keohane, and Marc A. Levy, 27-74. Global Environmental Accord. 1993. Reprint, Cambridge, MA: MIT Press, 1995.

Rao, Pemmaraju Sreenivasa. *International Liability for Injurious Consequences Arising Out of Acts Not Prohibited by International Law*. UN Doc. no. A/CN.4/531. International Law Commission, 2003.

Rubio, Santiago J., and Alistair Ulph. "Self-Enforcing International Environmental Agreements Revisited." *Oxford Economic Papers*, 58, no. 2 (April 2006): 233-63.

Schafer, Hans-Bernd and Andreas Schonenberger. "Strict Liability versus Negligence." In Vol. 2

of *Encyclopedia of Law and Economics*, 597-624. Cheltenham, UK: Edward Elgar Publishing, 1999.

Sunstein, Cass R. "Of Montreal and Kyoto: A Tale of Two Protocols." *Havard Environmental Law Review* 31 (2007): 1-65.

Thoms, Laura. "A Comparative Analysis of International Regimes on Ozone and Climate Change with Implications for Regime Design." *Columbia Journal of Transnational Law* 41 (2003): 795-859.

Tulkens, Henry. "Cooperation Versus Free-Riding in International Environmental Affairs: Two Approaches." In *Game Theory and the Environment*, edited by Nick Hanley and Henk Folmer, 30-44. Cheltenham, UK: Edward Elgar Publishing, 1998.

United Nations Development Program. *25 Years: Montreal Protocol*. New York: Bureau for Development Policy, 2012.

United Nations Environment Program. "Rio Declaration on Environment and Development," June 1992.

United Nations Environment Program. "Declaration of the United Nations Conference on the Human Environment," June 1972.

Van Der Ploeg, Frederick, and Aart J. De Zeeuw. "International Aspects of Pollution Control." *Environmental and Resource Economics* 2 (1992): 117-39.

Vicha, Ondrej. "The Polluter-Pays Principle in OECD Recommendations and its Application in International and EC/EU Law." *Czech Yearbook of Public and Private International Law* 2 (2011): 57-67.

Victor, David G. *The Collapse of the Kyoto Protocol and the Struggle to Slow Global Warming*. Princeton, NJ: Princeton University Press, 2001.

Victor, David G. "The Operation and Effectiveness of the Montreal Protocol's Non-Compliance Procedure." In *The Implementation and Effectiveness of International Environmental Commitments: Theory and Practice*, edited by David G. Victor, Kal Raustiala, and Eugene B. Skolnikoff, 137-176. Global Environmental Accord. International Institute for Applied Systems Analysis, 1998.

Wettestad, Jorgen. "Designing Effective Environmental Regimes: The Conditional Keys." *Global Governance* 7, no. 3 (2001): 317-41.

Wettestad, Jorgen. "Acid Lessons? LRTAP Implementation and Effectiveness." *Global Environmental Change* 7, no. 3 (1997): 235-49.