Four Essays in Behavioral Economics

by Haydar Doruk İriş

A dissertation submitted to

Faculdade de Economia
Universidade Nova de Lisboa

in partial fulfilment of the requirement for the degree of

Doctor of Philosophy
in
Economics

March 2011

Financial support from POCI 2010 and FSE
within the III Quadro Comunitário de Apoio
SFRH/BD/32928/2006 & PTDC/ECO/65284/2006

Ciência.Inovação 2010 Programa Operacional Ciência e Inovação 2010
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Four Essays in Behavioral Economics

Doruk İriş

Faculdade de Economia
Universidade Nova de Lisboa

March 2011

Contents

List of Figures

Acknowledgments

My Ph.D. period has been a process of personal and scientific growth, influenced by many people. I take the opportunity now to thank all these people who have supported and guided me in this process.

First I would like to thank my dear friend, hard-working co-author and patient supervisor, Chrysostomos Tabakis, who has played so many important roles in my Ph.D. life. With his keen eye for detail and his crucial supervision in regards to my dissertation, he has also introduced me to new areas of interest, such as international trade and environmental economics, which have both piqued my interest and resolve in applying my ideas to them. Futhermore, our intellectual and social talks were an essential part of my social life.

If I hadn't knocked on Luís Santos-Pinto's door four years ago, the title of my thesis would not be related to behavioral economics. I am very thankful to him for introducing me to this field which became the field of my interest. I am also grateful for all the encouragement I received in my research and for his reference, which provided me with the opportunity to visit and work at the University of California, San Diego (UCSD) for one year.

I also had the opportunity to work with Costas Hadjiyiannis from the University of Cyprus. Besides his rather nice company, I am grateful for his sharp intellect, which improved my dissertation significantly, and for introducing me to numerical analysis, which is something I will be utilizing in my future work. I look forward to our continuing Skype talks that are always full of interesting subjects, especially football.

Another person I would like to thank is Joel Sobel, for having me as a visiting researcher at UCSD. In addition to his and Vincent Crawford's wonderful behavioral economics course I took when I was in San Diego, I owe him a debt of gratitude for helping me refine my ability to validate new research ideas. I now take the opportunity to also thank the other people who I met in the department of economics at UCSD, especially Joel Watson and Nageeb Ali for insightful comments on my ideas.

Some people have really big hearts and I am lucky to have so many of these people around me. I would like to thank three great friends of mine, Burak Can, Danilo Lopomo Beteto, and Tahir Öztürk for all the delightful talks and discussions regarding our research ideas and economic literature. You made my research ideas your own and helped me discover the correct answers. I am grateful for having such friends, you are my brothers.

I would like to thank Susana Peralta for being available whenever i needed to talk and for her encouragement on my research. I am sad that we did not have the opportunity to work together in the end. I would also like to thank her and Ana Balcão Reis for choosing me as the research fellow for the "Institutions and Technology" project. I am especially appreciative to Ana Balcão Reis for her assistance with any problem that I had.

I am grateful to Guilherme Carmona for taking time out of his busy schedule to personally teach me the advanced game theory course, especially since this was not originally a part of his teaching responsibilities. I am also grateful to my dear friend, Luis Carvalho, for being the teaching assistant on this course and helping me improve my mathematical skills. We also had some very interesting and intellectual talks.

I am indebted to the Nova School of Business and Economics at the Universidade Nova de Lisboa, its faculty, program coordinators, and the administrative staff for the encouraging research environment they created that resulted in this dissertation. I would also like to thank the participants of the conferences and seminars I attended. Financial supports from *Fundação para a Ciência e a Tecnologia*, Ph.D. fellowships under POCI 2010, *SFRH/BD/32928/2006* and *PTDC/ECO/65284/2006*, are kindly acknowledged.

My colleagues from Ph.D.: Bernardo (Pimentel), Erica, European Gang at UCSD, Fazeer, Filipe (Lacerda), Gisele, Guilherme (Oliveira), Isaac, João (Filipe), Luisa, Luiz, Mafalda, Margarida, Mark, Miguel, Nina, Nuno (Sobreira), Rodrigo, Rui, Sonia, Tiago (Botelho), Tiago (Pires), Tymur, and Vladimir... I thank you all for all your help, support, company, and fun times.

I would further like to thank my colleagues and flatmates Renato Rosa and João Farinha, as well as my friends Bernardo Fazendeiro, Nuno Paixão, and Pedro Chaves for being such kind, interesting, intellectual, and fun people. Moreover, thank you to my friend Petula for being such great company not only in Lisbon, but also on our trips to Spain, Turkey and in Portugal. I am also grateful for your help with my english.

A special thanks to my dearest friend Kathy for just being yourself.

I feel very lucky to have met my mentor, brother Remzi Sanver and my co-author İpek Özkal-Sanver. You own a big share of my initiation on the challenging paths I took and which have ended in success. I am deeply thankful for everything you have done for me.

Finally, I am grateful for all the help I have received from my family: my mother, my father, my sister, nephew and niece, my brother- and parents-in-law, my aunts, my uncles, my cousins and my brothers in Eküri. You are the main reason for me being myself.

Introduction

Despite their centrol role in economic analysis, the assumptions that individuals behave rationally and maximize their self-interested payoffs does not match with a large body of evidence from psychology and experimental economics. Economic agents often pursue objectives other than self-interested payoff maximization. Many observed departures from such behavior arise through actions that consider other-regarding or social preferences.

This dissertation is composed of four papers, all related to other-regarding preferences, more specifically, reciprocity and inequity aversion. If a player has such preferences, then she does not only care about her payoff, but also the other players' payoffs. Furthermore, a reciprocal player places a positive weight on the payoff of another player if she expects the latter to take a further positive action than the one she perceives as fair (positive reciprocity); and a negative weight if she expects the reverse (negative reciprocity). Our reciprocity models thus follow Segal and Sobel (2007). Moreover, the reciprocal player's perception on the fair action of the other serves as a reference level, which is another widely used behavioral concept. A key aspect of reciprocal preferences is that players do not only care about the outcomes of their actions, but also the actions of other players.

An inequity averse player, on the other hand, maximizes her payoff, but also tries to reduce the difference between her payoffs and those of the rivals. She dislikes advantages inequity: she feels compassion towards others when the average payoffs of others is less than her own payoff. At the same time, she dislikes disadvantage inequity: she feels envious towards others when the average payoffs of others is more than her own payoff. Our inequity aversion model thus follows Fehr and Schmidt (1999).

All of this has led to the general goal of this dissertation, which is to understand the implications of such behavioral concepts in environments, where players with such non-standard preferences compete or cooperate.

In the first three chapters the implications of fairness and reciprocity is studied in different environments. In particular, in **Chapter 1** we study the implications of fairness and reciprocity for self-enforcing international environmental agreements on pollution abatement. To this end, we develop a dynamic game, in which countries with reciprocal preferences attempt to maintain international environmental cooperation. In our context, reciprocal countries reward kind behavior, but retaliate against countries behaving unfairly. Moreover, we maintain the assumption that binding commitments cannot be made at the international level and countries are therefore limited to cooperative environmental agreements that are self-enforcing. In such a setting, a country will choose to adhere to the cooperative path as long as the onetime gains it could achieve by unilaterally deviating from its agreed-upon environmental policies do not outweigh the discounted future welfare losses due to the ensuing breakdown of multilateral cooperation and the emergence of a noncooperative equilibrium characterized by inefficiently low abatement standards.

We demonstrate that reciprocal countries that have moderate expectations from each other with respect to their national abatement strategies can support a greater degree of environmental cooperation than self-interested ones. However, when only very high abatement standards are deemed fair, then reciprocity could have a detrimental effect on international environmental cooperation. Our model therefore provides a novel perspective on the failure and the recent success of the 2009 and 2010 U.N. Climate Change Conferences in Copenhagen and Cancun, respectively. Finally, we show that these results are robust to endogenizing fair-abatement-standard perceptions.

Chapter 2 is in the same line with Chapter 1, but in this Chapter we explore the impact of fairness and reciprocity on multilateral tariff cooperation. We demonstrate that reciprocal countries that are moderately demanding from their trading partners regarding their commercial policy can support a greater degree of cooperation than self-interested ones. However, when only very liberal import policies are considered fair, then reciprocity could have a detrimental effect on multilateral tariff cooperation. Thus, our model provides a novel reason for the occasional failure of trade negotiations. Finally, our results are robust to endogenizing fair-tariff perceptions.

In **Chapter 3** we study the firm behavior with managers that are motivated in part by personal animosity–or respect–towards their rivals. A reciprocal manager responds to unkind behavior of rivals with unkind actions, while at the same time, it responds to kind behavior of rivals with kind actions. We find that if fairness payoffs are small by comparison with monetary payoffs, then collusive action profiles (prices or quantities) are easier to sustain when firms have reciprocal managers. Thus, fairness concerns among firms with reciprocal managers can have adverse welfare consequences for consumers.

In **Chapter 4** we explore the role of inequity aversion as an explanation for observed behavior in experimental Cournot oligopoly. We show that inequity aversion can change the nature of the strategic interaction: quantities are strategic substitutes for sufficiently asymmetric output levels but strategic complements otherwise. We find that inequity aversion can explain why: (i) some experiments result in higher than Cournot-Nash production levels while others result in lower, (ii) collusion often occurs with only two players whereas with three or more players market outcomes are very close to Cournot-Nash, and (iii) players often achieve equal profits in asymmetric Cournot oligopoly.

Chapter 1

International Environmental Cooperation under Fairness and Reciprocity

This paper explores the implications of fairness and reciprocity for self-enforcing international environmental agreements on pollution abatement. Reciprocal countries reward kind behavior (positive reciprocity), but retaliate against countries behaving unfairly (negative reciprocity). We demonstrate that reciprocal countries that have moderate expectations from each other with respect to their national abatement strategies can support a greater degree of environmental cooperation than self-interested ones. However, when only very high abatement standards are deemed fair, then reciprocity could have a detrimental effect on international environmental cooperation. Our model therefore provides a novel perspective on the failure of the Copenhagen summit and the recent success of the Cancun one. Finally, we show that these results are robust to endogenizing fair-abatement-standard perceptions.

Keywords: Reciprocity; Environmental agreements; Abatement standards; Repeated games

JEL Classification: Q50; Q58; D63

1.1 Introduction

In this paper, we examine international environmental agreements (IEAs) on pollution abatement among governments with reciprocal preferences for environmental policy. Governments with reciprocal preferences reward "kind" or "fair" actions (positive reciprocity), whereas they punish "unkind" or "unfair" behavior (negative reciprocity). The significance of this is twofold. First, governments seem to exhibit such preferences, at least with regard to environmental policy. Second, our analysis provides important insights into the successes and failures of international environmental negotiations.

Many environmental problems are transboundary in nature and often have a global scope (e.g., climate change or marine pollution). Countries have therefore been seeking to sign IEAs in order to coordinate their environmental policies. Typically, governments stress the importance of fairness in sharing the burden of environmental protection. In fact, it is clearly stated in the Copenhagen Accord, the outcome of the 2009 United Nations (UN) Climate Change Conference in Copenhagen, that its endorsers "shall...on the basis of equity...enhance [their] long-term cooperative action to combat climate change."[1] In a similar spirit, France at the onset of the Copenhagen summit, in response to India's commitment to reduce emissions, expressed "her determination to work with India and all her partners to put together an ambitious, just and balanced agreement in Copenhagen," while President Obama at a press conference at the end of the summit agreed that it is not fair to expect developing countries like China and India to be bound by the same set of legal obligations as developed countries in the fight against

[1] See http://unfccc.int/resource/docs/2009/cop15/eng/11a01.pdf.

climate change.[2] Moreover, fairness considerations are central to the UN Environment Programme Medium-term Strategy 2010–2013 and to the official environmental policy agendas of most countries, including the European Union (EU) and the US.[3]

At the same time, the emphasis governments place on reciprocity is equally strong. As a matter of fact, reciprocity considerations did play a pivotal role in the disappointing outcome of the Copenhagen summit. In particular, most analysts agree that the principal reason for the failure of the summit was the disagreement between countries—especially between the US and the BASIC countries (i.e., Brazil, South Africa, India, and China) led by China—on how to share the burden of reducing greenhouse gas emissions. Essentially, the negotiations in Copenhagen revolved around fairness and reciprocity, with countries rejecting the various deals proposed as unfair or one-sided; and when the talks were concluded, participants started accusing each other of a total lack of willingness to compromise or reciprocate. For instance, early in the negotiations, one of India's top negotiators expressed "concern...that [India has] been offering unilateral concessions, without obtaining any reciprocity."[4] Furthermore, after the summit, the G77 group of 130 developing nations blamed President Obama for "locking the poor into

[2] See, respectively, http://www.ambafrance-uk.org/France-welcomes-India-s-climate.html and http://www.whitehouse.gov/the-press-office/remarks-president-during-press-availability-copenhagen.

[3] For the UN Environment Programme Medium-term Strategy 2010–2013, see http://www.unep.org/PDF/FinalMTSGCSS-X-8.pdf. See, respectively, http://ec.europa.eu/environment/integration/development_en.htm and http://www.usaid.gov/our_work/environment/climate/index.html for the EU and US environmental policy agendas.

[4] See http://indiatoday.intoday.in/site/Story/73850/Top%20Stories/Copenhagen+summit+braces+for+a+'no+deal'.html.

permanent poverty by refusing to reduce US emissions further," while the UK's Prime Minister Gordon Brown argued that, "If America and China were able to show they were doing more and I believe that they can, then all countries—Australia, Brazil, Japan, Korea—all those countries that have ranges [of emission cuts] would be prepared to go to their highest level of ambition."[5] In brief, in the international environmental arena, governments seem, to some extent, to not be simply maximizers of their own self-interested welfare, but to rather be exhibiting some preferences for fairness and reciprocity.[6]

On a more theoretical level, there is an extensive and rapidly growing literature providing evidence of both positive and negative reciprocity in individual decision making.[7] Political-economy models of environmental policy then suggest that if individuals have such preferences, these preferences will be reflected in the objective functions of governments. The first model we can invoke here is the median-voter model, where the government chooses policies that reflect majority opinion on the issue (in order to remain popular and stay

[5] See http://www.guardian.co.uk/environment/2009/dec/19/copenhagen-blame-game and http://news.bbc.co.uk/2/hi/uk_news/politics/8423831.stm, respectively.

[6] It is important to underscore here that fairness in this context is entirely subjective. It depends only on governments' perceptions and should not be confused with what is objectively or ethically fair. For the purposes of this paper, fairness simply relates to the reference level of governments regarding other countries' environmental policies (see also footnote 12).

[7] For instance, experiments asking individuals to contribute to public goods typically find that their contributions far exceed what self-interested utility maximization would entail (e.g., Andreoni, 1988; Palfrey and Prisbrey, 1997; Croson, 2007). This is usually interpreted as evidence of positive reciprocity. Analogous results arise from trust or gift-exchange experiments (e.g., Berg et al., 1995; Fehr et al., 1997; Fehr et al., 1998). On the other hand, evidence for negative reciprocity is found in ultimatum-game experiments with the typical result being that people reject offers that would be accepted under the self-interested hypothesis (e.g., Güth et al., 1982; Roth et al., 1991). Moreover, using survey data, Dohmen et al. (2009) provide evidence of both positive and negative reciprocity.

in office).[8] In such a setting, if the median voter has reciprocal preferences, then the government's actions will mirror these preferences. Alternatively, we could look at interest-group models. The framework that currently occupies center stage in the literature is that of Grossman and Helpman (1994), who focus on trade-policy issues. More specifically, in their paper, the incumbent government maximizes a weighted average of aggregate social welfare and political contributions by lobbies that wish to influence trade policy. Their framework has subsequently been applied to environmental policy by Fredriksson (1997) and Aidt (1998), among others. In these lobbying models, if individuals have preferences for reciprocity, these preferences will enter into the government's objective function with some weight.[9]

A question that naturally arises at this point is whether voters exhibit reciprocal preferences towards environmental policy. There exists considerable evidence that seems to suggest so. For example, in an October 1998 Program on International Policy Attitudes (PIPA) poll, 44% of Americans agreed that "the US should refuse to sign the Kyoto treaty until all the less-developed countries commit to limits," providing clear evidence of negative reciprocity.[10] In another PIPA poll in September-October 2009, 68% of respondents in 15 (developed and developing) countries endorsed the view that if their own country took measures to deal with the problem of climate change, then other countries would be more willing to act, reflecting

[8] See, for example, Eriksson and Persson (2003) and McAusland (2003).

[9] In fact, one could argue that governments are not genuinely motivated by fairness considerations, but only push for fairer environmental policies in order to increase their political support, since voters have such considerations. In any case, our results depend solely on governments exhibiting, to some degree, reciprocal preferences towards environmental policy, and not on why this is the case.

[10] See http://www.americans-world.org/digest/global_issues/global_warming/gw3.cfm.

positive-reciprocity thinking among individuals.[11] Another important conclusion that could be drawn from these surveys is that people have an explicit notion of what the "appropriate" environmental policies entail, i.e., they have a reference level regarding environmental policy against which the policies implemented by their own government or by the rest of the world can be evaluated.[12] For instance, in a 25-nation PIPA poll conducted November 2006 through January 2007, people around the globe were asked to evaluate the US handling of global warming or climate change. The percentage of respondents who were not sure/declined to answer was on average only 18%.[13] Analogous results were obtained in a 2009 23-nation PIPA survey regarding the handling of climate change by the US and China.[14] Similarly, in a special Eurobarometer survey on Europeans' attitudes towards climate change conducted in late August and September 2009, Europeans were asked whether the EU and their national government are doing enough to fight climate change. Only 12% of the respondents in the former case and 8% in the latter were unsure/declined to answer.[15] Finally, there exist numerous nongovernmental/nonprofit organizations devoted solely to the promotion of environmental protection and sustainable development, in which large numbers of people become organized. These organizations most often openly advocate

[11] See http://siteresources.worldbank.org/INTWDR2010/Resources/Background-report.pdf.

[12] The "reference level" is a concept widely studied in the behavioral economics literature. For more on this, see, for instance, Helson (1964) and Tversky and Kahneman (1991).

[13] See http://www.worldpublicopinion.org/pipa/pdf/jan07/BBC_USRole_Jan07_quaire.pdf.

[14] See http://www.worldpublicopinion.org/pipa/pdf/nov09/WPO_China_Nov09_quaire.pdf.

[15] See http://ec.europa.eu/public_opinion/archives/ebs/ebs_322_en.pdf.

fairness and reciprocity in international environmental cooperation.[16]

To address the implications of reciprocity and fairness for IEAs on pollution abatement, we develop a dynamic game in which reciprocal countries facing a free-riding Prisoner's Dilemma problem in their dealings with one another attempt to maintain cooperation in their national abatement strategies, where pollution is assumed to be transboundary in nature. Given the lack of a supranational authority with effective enforcement mechanisms regarding environmental policy, we restrict our attention to IEAs that are self-enforcing, as in Ferrara et al. (2009). In this context, a country will choose today to adhere to the cooperative path as long as the onetime gain it could achieve by unilaterally deviating from its agreed-upon abatement policies does not outweigh the discounted future welfare losses due to the ensuing breakdown in international environmental cooperation. It is important to stress here that we completely abstract from participation considerations in this paper. We instead look for the most cooperative equilibrium that can be supported by reciprocal countries within the context of a self-enforcing international agreement involving full participation.[17] This is in line with recent experiences with the Copenhagen Accord and the Cancun Agreements, to which 140 and 193 countries, respectively, have agreed, including all the major (polluting) economies. On the other hand, to model reciprocity, we follow Segal and Sobel (2007). More specifically, we assume that a country

[16] See, for example, http://www.greenpeace.org/international/Global/international/planet-2/report/2009/5/greenpeace-climate-vision.pdf for the climate vision of Greenpeace.

[17] For further elaboration on these points, see Barrett (1994), Wagner (2001), and Barrett (2005), among others. Note also that self-enforcing agreements involving full participation are commonly employed in the literature on multilateral trade negotiations (e.g., Dam, 1970; Dixit, 1987; Bagwell and Staiger, 2002).

attaches a positive (negative) weight to the self-interested welfare of another country if it expects the latter to behave kindly (unkindly) by setting a higher (lower) abatement standard than the one it perceives as fair. In other words, countries are assumed to have preferences over both outcomes and strategies.[18]

We find that reciprocal countries that have moderate expectations from each other with respect to environmental policy (i.e., when the abatement standards considered fair are not too high, but at the same time, too little pollution abatement is considered unfair) can support a higher degree of environmental cooperation and thus achieve higher welfare than can self-interested countries. The intuition underlying this result is straightforward. For such fair-abatement-standard perceptions, in the reciprocal game (i) the punitive Nash abatement standard is lower than in the self-interested game, which acts to make the punishment phase costlier in the former game; and (ii) countries are in a positive-reciprocity state. As a result, under the scenario in question, reciprocal countries are faced with both a weaker incentive to cheat and a stronger incentive to cooperate than self-interested ones, allowing them to support a "greener" equilibrium.

However, when reciprocal countries are highly demanding of each other regarding their environmental policies (i.e., when only very high abatement levels are perceived as fair), then the impact of reciprocity on international environmental cooperation is ambiguous. Intuitively, in such a case, reciprocal countries are in a negative-reciprocity state, meaning that they face

[18] Lange and Vogt (2003) also examine IEAs among countries with "nonstandard" preferences. However, their focus is on preferences for equity. Moreover, unlike here, they look at coalition formation.

a stronger incentive to defect than self-interested ones. On the other hand, their incentive to cooperate remains stronger (due to the lower punitive Nash abatement standard), but there exist cases in which the stronger-incentive-to-cheat effect dominates, leading to less pollution abatement in the reciprocal equilibrium as compared with the self-interested one. Finally, we show that these results are robust to allowing for fair-abatement-standard perceptions that are endogenously determined during the course of the game.

At a more general level, our findings provide a novel insight into the successes or failures of international environmental negotiations. In particular, assuming countries have (some) preferences for fairness and reciprocity, our results suggest that if they arrive at the negotiations table with expectations that are greatly elevated (for whatever reason), this could prove to be counterproductive, in the sense that they might no longer be able to support very "green" policies. It could then be argued that this might be one of the plausible explanations for the failure of the Copenhagen summit. There is little doubt that the pre-Copenhagen expectations for what could have been accomplished there were very high. For instance, according to the official UN website, more than 100 world leaders met in New York in September 2009 in order to "mobilize political will and strengthen momentum for a fair, effective, and ambitious climate deal in Copenhagen."[19] The same level of ambition is expressed in the quote from the French government above. Our findings, however, demonstrate that such "ambitious" expectations, represented in our model by high fair-abatement-standard perceptions, could end up hindering the efforts for deeper international environmental cooperation.

[19] See http://un.org/wcm/content/site/climatechange/lang/en/pages/2009summit.

In fact, this possibility had been anticipated by a number of analysts and negotiators before the summit. For example, in November 2009, Susanne Dröge from the German Institute for International and Security Affairs said that high expectations for the summit had to be adjusted, and noted that it was important that the expectations were adjusted before the summit, because otherwise the outcome would be bad.[20]

By contrast, at the 2010 UN Climate Change Conference in Cancun, expectations were much more modest. As a matter of fact, the UN Secretary General Ban Ki-moon, in his address to the opening ceremony of the high-level segment of the Cancun talks, urged nations to not be too demanding. He said, "We don't need final agreement on all the issues, but we do need progress on all the fronts. We cannot let the perfect be the enemy of the good."[21] In the light, then, of our model, it could be argued that the modest expectations in place at the Cancun summit could in part account for its success. As *The Economist* concludes, "So why did Cancun succeed in making progress within the UN process where Copenhagen so spectacularly failed? One reason is low expectations. Copenhagen was meant to produce an all-encompassing agreement; Cancun was expected to embarrass itself."[22] In summary, expectations emerge as a key factor in our analysis, having a significant effect on what can be achieved in the international environmental arena. Actually, to the best of our knowledge, our paper is the first to identify such a role for expectations in multilateral environmental negotiations. The policy implications are then immediate. The careful management of expec-

[20] "Can Copenhagen Still Be Saved?" *The Economist*, November 17, 2009.

[21] See http://www.guardian.co.uk/environment/2010/dec/08/ban-ki-moon-cancun-climate-deal.

[22] See http://www.economist.com/blogs/newsbook/2010/12/climate_change?page=1.

tations is critical for the success of international environmental negotiations, and most importantly, the creation of a pre-negotiations high-expectations environment should be avoided.

The remainder of the paper is organized as follows. Section 2 sets out the basics. Section 3 characterizes the static Nash equilibrium of our model, whereas Section 4 analyzes the dynamic game. Section 5 presents a simplified model in order to better illustrate the main insights from our analysis, while Section 6 endogenizes countries' fair-abatement-standard perceptions. Finally, Section 7 offers some concluding remarks. All the proofs are relegated to the Appendix.

1.2 The Model

We assume that the world consists of two countries, A and B, which trade two goods, a and b.[23] This focus on a two-country world is nonrestrictive, as our findings readily extend to $N > 2$ countries.[24] Country J is endowed with 1 unit of good $-j$ and zero units of good j, where $J \in \{A, B\}$ and $j \in \{a, b\}$.[25] Countries can either consume or export their endowment, or alternatively use it for pollution abatement. On the consumption side, we assume that demand functions are symmetric across countries and goods, and that the demand for good j is independent of the price of good $-j$. More specifically, the demand in country J for good j is given by $D\left(P_j^J\right)$, where P_j^J

[23] Our framework is inspired by Bagwell and Staiger (1999b).

[24] Upon request, a tecnical appendix is available from the authors in which an N-country model is presented.

[25] We choose to ignore the production process in the two countries for expositional simplicity. In any case, this assumption does not affect the qualitative nature of our findings.

is good j's price in J. We maintain the standard assumptions that $D\left(P_j^J\right)$ is strictly positive on some bounded interval $\left[0, \overline{P}_j^J\right)$, that $D\left(P_j^J\right) = 0$ for $P_j^J \geq \overline{P}_j^J$, and that $D\left(P_j^J\right)$ is twice continuously differentiable in P_j^J with $D'\left(P_j^J\right) < 0$ for $P_j^J \in \left[0, \overline{P}_j^J\right)$. Given this setup, good j $(-j)$ is country J's natural import (export) good.

In each period, country J unilaterally selects its abatement standard $q^J \in \left[0, q_O^J\right]$ so as to maximize its individual welfare, where $q_O^J < 1$ is the abatement level country J would choose under full international cooperation.[26] Pollution abatement is assumed to consume a fraction of a country's endowment. In particular, country J has a post-abatement endowment of good $-j$ of $1 - q^J$ units available for domestic consumption or export. The aggregate environmental damage country J, then, faces is a function of the level of its own emissions and those of country $-J$:

$$\Psi^J\left(q^J, q^{-J}\right) = \frac{1}{2}\left[1 - q^J + s\left(1 - q^{-J}\right)\right]^2, \tag{1.1}$$

where $s \in (\underline{s}, 1]$ is the degree of transboundary pollution, with $\underline{s} > 0$.[27] As s converges to $\underline{s}$, pollution becomes more local in nature, whereas as s converges to 1, pollution becomes a pure public bad affecting both countries equally.

We assume that the countries engage in free trade, implying $P_j^J = P_j^{-J}$,

[26] As will become evident below, country J would never find it optimal to select an abatement standard higher than q_O^J. We can therefore restrict the range of q^J to $\left[0, q_O^J\right]$ without loss of generality.

[27] As we discuss below, we assume s to be sufficiently high so that in the absence of an IEA, the countries would underinvest in pollution abatement from the point of view of global efficiency. See also footnote 33.

$\forall j$.[28] Letting $X_j^{-J}\left(P_j^{-J}\right) = 1 - q^{-J} - D\left(P_j^{-J}\right)$ denote country $-J$'s export supply function, the equilibrium prices can then be obtained from the usual market-clearing conditions:

$$D\left(P_j^J\right) = X_j^{-J}\left(P_j^{-J}\right). \tag{1.2}$$

As expected, the equilibrium price of good j is increasing in country $-J$'s abatement efforts (or, equivalently, decreasing in country $-J$'s post-abatement endowment of j).

The countries have preferences for fairness and reciprocity. More precisely, the welfare of country J is given by:

$$RW^J\left(q^J, q^{-J}, q_f^{-J}\right) = SW^J\left(q^J, q^{-J}\right) + \gamma w^J(q^{-J}, q_f^{-J}) SW^{-J}\left(q^J, q^{-J}\right). \tag{1.3}$$

The first term, SW^J, is the self-interested (or "standard") welfare function, i.e., the sum of consumer and producer surplus minus environmental damage:

$$\begin{aligned} SW^J\left(q^J, q^{-J}\right) = \int_{P_j^J(q^{-J})}^{\overline{P}_j^J} D(P)\, dP + \int_{P_{-j}^J(q^J)}^{\overline{P}_{-j}^J} D(P)\, dP \\ + \left(1 - q^J\right) P_{-j}^J\left(q^J\right) - \Psi^J\left(q^J, q^{-J}\right). \end{aligned} \tag{1.4}$$

The second term, $\gamma w^J(q^{-J}, q_f^{-J}) SW^{-J}$, captures the fairness payoff for country J, where $\gamma > 0$ is a scaling factor, and $w^J(q^{-J}, q_f^{-J})$ determines the (scaled) weight country J places on $-J$'s self-interested welfare SW^{-J} and

[28] It is direct to show that our basic findings are robust to the introduction of (optimally set) import tariffs. We therefore abstract from trade-protection considerations for analytical convenience. Upon request, a technical appendix is available from the authors in which import tariffs are introduced into our framework.

is of the following form:

$$w^J(q^{-J}, q_f^{-J}) \begin{cases} > 0 \text{ if } q^{-J} > q_f^{-J} \\ = 0 \text{ if } q^{-J} = q_f^{-J} \\ < 0 \text{ otherwise} \end{cases}, \tag{1.5}$$

with q_f^{-J} being the q^{-J} country J deems "fair." We maintain the assumptions that country J's weight function $w^J(q^{-J}, q_f^{-J})$ is twice continuously differentiable in both arguments, is strictly decreasing in its own fair-abatement-level perception, and is strictly increasing in country $-J$'s abatement level. We also assume that fair-abatement-standard perceptions are common knowledge.

To gain some insight into (1.5), note first that an increase in q^{-J} has two offsetting effects on the self-interested welfare of country J. On the one hand, it has a positive environmental effect, as it results in lower aggregate environmental damage for country J. On the other hand, it has a negative terms-of-trade effect, as it leads to a higher price of good j. Observe further that the first effect is a function of the degree of transboundary pollution s, whereas the latter effect is independent of s. It follows that as long as s is sufficiently high, which is our working assumption throughout the paper, the environmental effect of an increase in q^{-J} on country J's self-interested welfare outweighs its terms-of-trade effect, i.e., country J's self-interested welfare rises as country $-J$ implements a higher abatement standard.

The interpretation of the weight function w^J is then straightforward upon recalling that a reciprocal country cares about the other country's intentions. More specifically, if country J expects country $-J$ to set an abatement standard higher than the one it perceives as fair, then it is willing to sacrifice some of its own self-interested welfare in order to reward $-J$, exhibiting positive reciprocity. If instead country J expects country $-J$ to behave unfairly by selecting an abatement standard below the one it considers fair, then it is willing to sacrifice some of its self-interested welfare in order to punish $-J$, exhibiting negative reciprocity. Finally, if q^{-J} is exactly equal to q_f^{-J}, then RW^J collapses to SW^J, i.e., the reciprocal and self-interested welfare functions coincide for country J. In sum, equation (1.5) signifies that from country J's perspective, any q^{-J} in excess of q_f^{-J} is a fair (or kind) action that should be rewarded, whereas any q^{-J} below q_f^{-J} is an unfair (or unkind) action that should be punished.

1.3 Static Game

Our aim in this section is to characterize the static Nash equilibrium of our model, and compare it with the one that would emerge in a game with self-interested countries. This equilibrium will serve as a credible punishment in the dynamic game explored in the next section, the threat of which can support international environmental cooperation in a repeated setting.[29] To this end, let the static game with self-interested countries be represented by $\Gamma^S(SW)$, while $\Gamma^R(RW, w, \overrightarrow{q}_f)$ denotes the static game with reciprocal countries, where $\overrightarrow{q}_f \equiv \left(q_f^J, q_f^{-J}\right)$ is the fair-abatement-standard vector. We henceforth assume that $q_f^J = q_f^{-J} \equiv q_f$, i.e., the countries share a common

[29] Note that the static Nash equilibrium would be the unique equilibrium for the dynamic game as well if an IEA were not feasible (e.g., due to exogenous, political reasons or because the countries were highly impatient and did not value the future at all).

fair-abatement-level perception. The reason for this assumption is twofold. First, it considerably simplifies our analysis. Second, asymmetries of a not too high degree in fair-abatement-level perceptions between the (otherwise symmetric) countries would not affect the qualitative nature of our findings.[30] In addition, in all that follows, we maintain the assumption that γ is sufficiently small, meaning that the relative weight of the fairness payoff in the countries' objective function (or, equivalently, the relative weight the countries attach to each other's self-interested welfare) is not too high.[31]

It is direct to show that given that γ is sufficiently small, the cross-partial derivative of the welfare function of reciprocal country J with respect to its own abatement level and country J's abatement standard is strictly negative (i.e., $\partial RW^J/\partial q^J \partial q^{-J} < 0$). In other words, the choice variables are (strict) strategic substitutes, reflecting the free-riding incentives the countries face in their dealings with one another. Furthermore, given that s is "high" in our setting, the cross-partial derivative of country J's welfare function with respect to its abatement level and its fair-abatement-standard perception is strictly negative (i.e., $\partial RW^J/\partial q^J \partial q_f < 0$). To understand the sign of the latter derivative, simply recall that (i) given our focus on "high" degrees of transboundary pollution, the positive environmental effect of an increase in q^J on country $-J$'s self-interested welfare outweighs its negative terms-of-trade effect, i.e., pollution abatement by country J exerts a positive exter-

[30] Upon request, a technical appendix is available from the authors in which we reproduce our analysis allowing for asymmetries in fair-abatement-level perceptions between the countries that are not too high. See also footnote 36.

[31] The derivation of a closed-form solution for the upper bound of γ has proved elusive. In the simulations in Section 5, where we present a simplified model, γ is set less than or equal to 0.1.

nality on country $-J$; and (ii) a lower q_f results ceteris paribus in a larger w^J.

As we show in the Appendix, both $\Gamma^R(RW, w, \overrightarrow{q}_f)$ and $\Gamma^S(SW)$ admit largest and smallest pure symmetric Nash equilibria. In fact, if the countries' best-response functions have a slope strictly greater than -1, which we henceforth assume, then $\Gamma^R(RW, w, \overrightarrow{q}_f)$ and $\Gamma^S(SW)$ have unique equilibria: $\overrightarrow{q}_{NR} = (q_{NR}, q_{NR})$ and $\overrightarrow{q}_{NS} = (q_{NS}, q_{NS})$, respectively.[32] Moreover, it can be readily shown that the Nash equilibria of $\Gamma^R(RW, w, \overrightarrow{q}_f)$ and $\Gamma^S(SW)$ are characterized by an inefficiently low level of pollution abatement, i.e., $q_{OR} > q_{NR}$ and $q_{OS} > q_{NS}$, with $q_{OR} > q_{OS}$, where q_{OR} (q_{OS}) maximizes the countries' joint welfare, $RW^J + RW^{-J}$ ($SW^J + SW^{-J}$).[33] The latter result stems from the fact that given pollution is relatively transboundary, there exist significant pollution-abatement spillover effects, resulting in a suboptimal level of pollution abatement in the absence of an environmental agreement. However, under full international cooperation, the spillover effects would be internalized by both countries and therefore, higher abatement standards would be implemented.

We next show how countries' fair-abatement-level perception affects q_{NR}.

Lemma 1 *The (pure) Nash equilibrium of* $\Gamma^R(RW, w, \overrightarrow{q}_f)$, $\overrightarrow{q}_{NR} = (q_{NR}, q_{NR})$*, is strictly decreasing in the fair abatement standard* q_f*, i.e.,* $\partial q_{NR}/\partial q_f < 0$.

[32] This assumption is not critical for our results. It is just made for expositional simplicity.

[33] The inefficiency of the static Nash equilibria is shown in a technical appendix available from the authors upon request. We also demonstrate there that $\underline{s} = \max\{\underline{s}_S, \underline{s}_R\}$, where $\underline{s}_S$ satisfies $-D\left(P^J_{-j}(q_{NS})\right)\left(\partial P^J_{-j}(q_{NS})/\partial q^J\right) + \underline{s}_S(1 - q_{NS})(1 + \underline{s}_S) = 0$ and $\underline{s}_R$ satisfies $-D\left(P^J_{-j}(q_{NR})\right)\left(\partial P^J_{-j}(q_{NR})/\partial q^J\right) + \underline{s}_R(1 - q_{NR})(1 + \underline{s}_R) + \gamma w^J(q_{NR}, q_f)\left(\partial SW^{-J}/\partial q^J\right) + \gamma\left(\partial w^J(q_{NR}, q_f)/\partial q^J\right) SW^{-J} = 0$.

Intuitively, as we argued above, for a given q^{-J}, a higher q_f leads to a smaller w^J. Consequently, as q_f rises, country J chooses its environmental policy with less of country $-J$'s interests in mind, resulting in a Nash equilibrium characterized by less pollution abatement worldwide.

The following (nonrestrictive) assumption is now introduced: $q_f \geq q_{NS}$, i.e., too little pollution abatement is considered unfair, which is a reasonable assumption given our focus on environmental cooperation among countries. We finally compare the reciprocal static Nash abatement standard, q_{NR}, with the self-interested one, q_{NS}, as well as the welfare obtained in $\Gamma^R\left(RW, w, \overrightarrow{q}_f\right)$ and $\Gamma^S(SW)$.

Proposition 1 *Under our model's assumptions, (i) $q_{NR} \leq q_{NS}$; and (ii) for any country J, $RW^J\left(\overrightarrow{q}_{NR}, q_f\right) \leq SW^J(\overrightarrow{q}_{NS})$, with equalities holding in either case if and only if $q_f = q_{NS}$.*

Proposition 1 demonstrates that as compared with self-interested countries, reciprocal ones select lower Nash abatement standards, and thus, end up with lower welfare in Nash equilibrium. The intuition underlying Proposition 1 is direct. At q_{NS}, reciprocal countries are in a negative-reciprocity state wishing to punish each other by lowering their abatement standards (since $q_f \geq q_{NS}$). Therefore, the Nash equilibrium abatement standard of $\Gamma^R(RW, w, \overrightarrow{q}_f)$ is lower than that of $\Gamma^S(SW)$.

1.4 Dynamic Game

We now study repeated interaction between the countries. More specifically, the dynamic game we consider is simply the static one analyzed above infinitely repeated. We assume that countries cannot make binding international commitments, but are instead limited to environmental agreements that are self-enforcing. In such a setting, countries can still maintain international environmental cooperation, whose degree depends critically on how severely they can credibly punish an offender. Our aim in this section is to investigate the ramifications of fairness and reciprocity for the most cooperative abatement-standard equilibrium that can be supported within the context of an IEA involving full participation. In other words, we totally abstract from participation considerations, which is in line with recent experiences with the Copenhagen Accord and the Cancun Agreements, to which participation has been almost universal.

To this end, denote the dynamic game with reciprocal countries by $\Gamma_{\infty}^{R}\left(RW, w, \overrightarrow{q}_{f}\right)$, and the one with self-interested countries by $\Gamma_{\infty}^{S}\left(SW\right)$. Moreover, let $\delta \in (0,1)$ denote the discount factor between periods. Given the overall symmetry of our framework, for both games we focus on symmetric cooperative subgame-perfect equilibria in which (i) along the equilibrium path, the countries implement a common cooperative abatement standard $q_C \in (q_{NS}, q_{OS}]$ in each period; and (ii) if at any point in the game a defection occurs, both countries revert from the following period onwards to the noncooperative Nash abatement standard of the (relevant) stage game.[34] In other words, to enforce environmental cooperation, the countries employ grim-trigger strategies.

Let us begin our analysis with the dynamic game with self-interested

[34] Note that for both games we restrict our attention to cooperative abatement standards above q_{NS} but below q_{OS}. This enables us to better compare $\Gamma_{\infty}^{R}\left(RW, w, \overrightarrow{q}_{f}\right)$ with $\Gamma_{\infty}^{S}\left(SW\right)$, which is our main goal in this paper.

countries, $\Gamma_{\infty}^{S}(SW)$. To derive the incentive-compatibility constraint for self-interested country J, we first look at its static incentive to cheat, Ω_S^J, which simply equals the onetime increase it achieves in welfare when it optimally cheats by choosing an abatement standard on its reaction curve, while country $-J$ still cooperates with q_C:

$$\Omega_S^J(q_C) \equiv SW^J\left(BR_S^J(q_C), q_C\right) - SW^J(\overrightarrow{q}_C) \equiv SW_D^J - SW_C^J, \tag{1.6}$$

where $\overrightarrow{q}_C \equiv (q_C, q_C)$ and $BR_S^J(q_C)$ is country J's best-response abatement standard to q_C.

However, defection by any country leads to a permanent breakdown in international cooperation. Therefore, the discounted future welfare cost a defector faces is the discounted difference between the welfare under cooperation and the welfare in the punishment phase, given by:

$$\frac{\delta}{1-\delta}\left(SW^J(\overrightarrow{q}_C) - SW^J(\overrightarrow{q}_{NS})\right) \equiv \frac{\delta}{1-\delta}\left(SW_C^J - SW_N^J\right) \equiv \frac{\delta}{1-\delta}\omega_S^J(q_C), \tag{1.7}$$

where $\overrightarrow{q}_{NS} \equiv (q_{NS}, q_{NS})$ and ω_S^J is the per-period value of cooperation for country J.

Thus, the incentive-compatibility condition for self-interested country J to adhere to the cooperative path in $\Gamma_{\infty}^{S}(SW)$ is that the onetime gain from defection, Ω_S^J, does not exceed the discounted future value of cooperation, $(\delta/1-\delta)\,\omega_S^J$:

$$\Omega_S^J(q_C) \leq \frac{\delta}{1-\delta}\omega_S^J(q_C). \tag{1.8}$$

It follows from (1.8) that a given cooperative abatement standard q_C can be supported as a subgame-perfect equilibrium of $\Gamma^S_\infty(SW)$ as long as the countries are patient enough, or:

$$\delta \geq \delta^S_{q_C} \equiv \frac{SW^J_D - SW^J_C}{SW^J_D - SW^J_N}. \tag{1.9}$$

Analogous relationships hold for reciprocal countries. More specifically, the incentive-compatibility constraint for reciprocal country J to uphold international environmental cooperation in $\Gamma^R_\infty(RW, w, \overrightarrow{q}_f)$ is given by:

$$\Omega^J_R(q_C) \leq \frac{\delta}{1-\delta}\omega^J_R(q_C). \tag{1.10}$$

Furthermore, reciprocal countries can support a given cooperative abatement standard q_C as long as they sufficiently value the future, or:

$$\delta \geq \delta^R_{q_C} \equiv \frac{RW^J_D - RW^J_C}{RW^J_D - RW^J_N}. \tag{1.11}$$

Let $\overrightarrow{q}_{CS} \equiv (q_{CS}, q_{CS})$ denote the most cooperative equilibrium abatement-standard vector for $\Gamma^S_\infty(SW)$, i.e., q_{CS} is the highest pollution-abatement standard that does not invite cheating in the dynamic game with self-interested countries. Similarly, let $\overrightarrow{q}_{CR} \equiv (q_{CR}, q_{CR})$ represent the most cooperative equilibrium abatement-standard vector of $\Gamma^R_\infty(RW, w, \overrightarrow{q}_f)$.[35] Clearly, q_{CS} (q_{CR}) is the most cooperative equilibrium abatement standard of $\Gamma^S_\infty(SW)$ ($\Gamma^R_\infty(RW, w, \overrightarrow{q}_f)$) when $\delta = \delta^S_{q_{CS}}$ ($\delta = \delta^R_{q_{CR}}$). Furthermore, let us assume in the remainder of this section that $\delta \in [\underline{\delta}, \overline{\delta}]$ so that both

[35] Observe that the most cooperative abatement-standard equilibrium is the most natural focal point for either game as (i) it is the only equilibrium of the desired class that is not Pareto dominated; and (ii) nothing precludes preplay communication between the countries.

self-interested and reciprocal countries can maintain some environmental cooperation, but q_{OS} is infeasible for either of them. The following proposition compares q_{CR} against q_{CS} assuming that the countries are moderately demanding from each other regarding their environmental policy (i.e., assuming the fair abatement standard is not too high).

Proposition 2 *Let $q_f \leq q_{CS}$. Then the most cooperative equilibrium abatement standard of $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$ is higher than the one of $\Gamma_\infty^S(SW)$, i.e., $q_{CR} > q_{CS}$.*

To understand Proposition 2, recall that for any cooperative abatement standard q_C above the fair abatement level q_f, reciprocal countries attach a positive weight to each other's self-interested welfare, i.e., they are in a positive-reciprocity state. Two reinforcing forces are at work here. On the one hand, for any country J, the value of cooperation at q_C is higher in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$ than in $\Gamma_\infty^S(SW)$ because in the former game (i) the noncooperative (punitive) Nash abatement standard is lower, increasing the severity of punishments; and (ii) infinite Nash reversion would also be costly for country $-J$, which acts to heighten the cost of the punishment phase for country J itself (see Figure 1.1). On the other hand, the static incentive country J has to deviate from q_C is weaker in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$ than in $\Gamma_\infty^S(SW)$ since in the former game (i) the defect abatement standard is higher, as the countries are in a positive-reciprocity state; and (ii) defection would hurt $-J$, mitigating J's potential onetime gains from cheating (see Figure 1.2). It then follows that reciprocal countries can more easily support any given cooperative abatement standard above the fair one than can self-interested countries. As a result, when the fair abatement standard is not too high, reciprocity has a positive impact on international environmental cooperation (see Figure 1.3).

However, as the following proposition establishes, this is no longer necessarily true when countries are highly demanding of each other regarding pollution abatement.

Proposition 3 *Let $q_f > q_{CS}$. Then the effect of fairness and reciprocity on the most cooperative abatement-standard equilibrium of the dynamic game is ambiguous.*

To gain some insight into Proposition 3, recall that for any cooperative abatement standard q_C below the fair abatement level q_f, reciprocal countries attach a negative weight to each other's self-interested welfare, i.e., they are in a negative-reciprocity state. Two observations can then be readily made for any such $q_C < q_f$. On the one hand, for any country J, the value of cooperation at q_C is higher in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$ than in $\Gamma_\infty^S(SW)$ since the punitive Nash abatement standard is lower in the former game. Of course, infinite Nash reversion would be costly for country $-J$ as well, which acts to lower the cost of the punishment phase for country J in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$. However, the latter effect is relatively weak for a sufficiently small γ (see Figure 1.1). On the other hand, country J has a stronger incentive to defect in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$ than in $\Gamma_\infty^S(SW)$ since in the former game (i) the defect abatement standard is lower, because the countries are in a negative-reciprocity state, willing to incur some welfare cost in order to punish each other; and (ii) defection would hurt country $-J$, raising the gains from cheating for country J (see Figure 1.2). It is therefore ambiguous whether reciprocal or self-interested countries can more easily sustain any given cooperative abatement standard below the fair abatement standard. As a result, when only very "green" environmental policies are considered fair, the overall effect of reciprocity on international environmental cooperation could be negative (see Figure 1.3). Actually, this is more likely to happen when δ is relatively low, i.e., when the countries are relatively impatient. This is due to the fact that a lower δ weakens the relative significance of the stronger-incentive-to-cooperate force, while it leaves the stronger-incentive-to-cheat force unaffected. These results are more clearly illustrated in the next section within the context of a simplified model.[36]

At a more general level, Propositions 2 and 3 demonstrate that if, for whatever reason, countries become more demanding of each other with respect to their environmental policy (i.e., if the fair abatement standard increases), a given cooperative equilibrium that could have been otherwise supported, might no longer be feasible. This then suggests that if countries enter a round of multilateral environmental negotiations with elevated expectations due to domestic and/or global political pressure, they might fail to reach an agreement on deeper international environmental cooperation, even though such an agreement might have been attainable in the absence of these high expectations. Therefore, Propositions 2 and 3 provide a novel perspective on the successes and failures of international environmental negotiations, as at the Cancun and the Copenhagen summits, respectively.

[36] At this point, we should note that asymmetries in fair-abatement-standard perceptions would not affect the qualitative nature of our findings as long as the countries remained *symmetrically* demanding of each other with respect to their environmental policy. In particular, under asymmetric fair-abatement-standard perceptions, Proposition 2 would still hold as long as $q_f^{-J} \leq q_{CS}^{-J}$ for any country J, whereas Proposition 3 would still be valid as long as $q_f^{-J} > q_{CS}^{-J}$ for all J.

1.5 Simplified Model

In this section we reproduce the results of the paper within a simple setup with linear demand curves and a specific functional form for $w^J(q^{-J}, q_f)$. This enables us to better illustrate the insights from our model. To this end, let the demand for good j in country J be given by:

$$D\left(P_j^J\right) = \alpha - \beta P_j^J, \tag{1.12}$$

where $\alpha > 1/2, \beta > 0$ are constants. Moreover, let us assume that the weight function w^J is of the following form:

$$w^J(q^{-J}, q_f) = \frac{q^{-J} - q_f}{q^{-J} + q_f} \in [-1, 1). \tag{1.13}$$

We first look at the static game, and in particular at $\Gamma^S(SW)$. The abatement standard that would prevail under full international cooperation equals:

$$q_{OS} = 1 - \frac{2\alpha}{1 + 2\beta\left(1+s\right)^2}. \tag{1.14}$$

We next derive the best-response abatement standard of country J:

$$BR_S^J(q^{-J}) = \frac{3 + 4\beta\left(1 + \left(1 - q^{-J}\right)s\right) - 4\alpha}{3 + 4\beta}. \tag{1.15}$$

From equation (1.15), the self-interested static Nash abatement standard can then be readily obtained:

$$q_{NS} = 1 - \frac{4\alpha}{3 + 4\beta\left(1+s\right)}. \tag{1.16}$$

As discussed above, the Nash equilibrium of $\Gamma^S(SW)$ is characterized by an inefficiently low level of pollution abatement (i.e., $q_{OS} > q_{NS}$) as long as the degree of transboundary pollution s is large enough, or:

$$\beta > \frac{1}{4s(1+s)}. \tag{1.17}$$

The analysis for $\Gamma^R(RW, w, \overrightarrow{q}_f)$, however, is more involved. The best-response function for reciprocal country J is given by:

$$\begin{aligned} BR_R^J\left(q^{-J}\right) &= \Lambda\{-4\alpha\left(q^{-J}+q_f\right)+\left(q^{-J}+q_f\right)\left[3+4\beta\left(1+\left(1-q^{-J}\right)s\right)\right] \\ &\quad -\left(q^{-J}-q_f\right)\left[1-4\beta s\left(1-q^{-J}+s\right)\right]\gamma\}, \end{aligned} \tag{1.18}$$

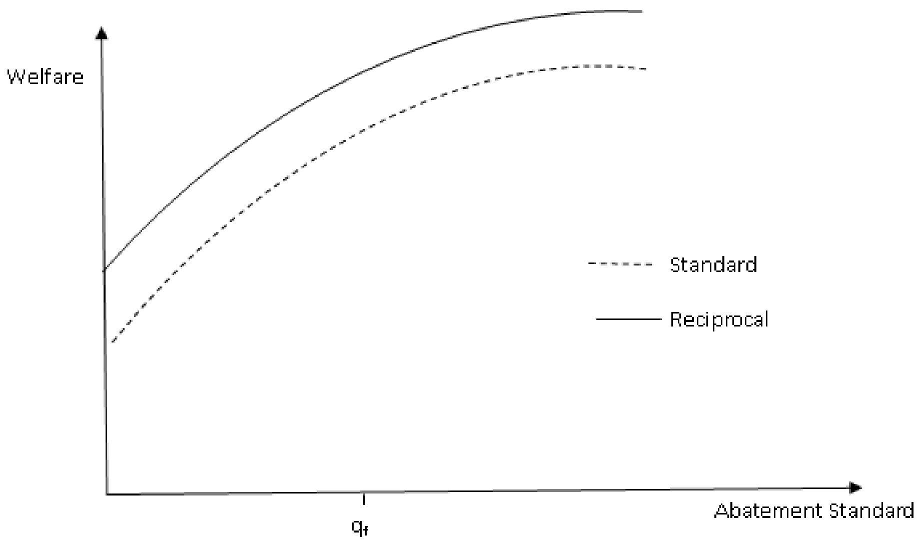
where $\Lambda = 1/\left[(3+4\beta)\left(q^{-J}+q_f\right)+\left(q^{-J}-q_f\right)\left(4\beta s^2-1\right)\gamma\right]$. For the remainder of this section, we resort to numerical/graphical analysis since the functional forms are too cumbersome for analytical results.[37] Numerical analysis shows that $q_{NR} < q_{NS}$ when $q_f > q_{NS}$, $q_{NR} = q_{NS}$ when $q_f = q_{NS}$, and $\partial q_{NR}/\partial q_f < 0$, confirming Proposition 1 and Lemma 1.

We now turn to the dynamic game and compare q_{CS} with q_{CR}, while maintaining the (nonrestrictive) assumptions that $q_f > q_{NS}$ and $q_C \in (q_{NS}, q_{OS}]$. Let us consider first the per-period value of cooperation for country J in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$ and $\Gamma_\infty^S(SW)$: ω_R^J and ω_S^J, respectively. Figure 1.1 depicts the relationship between the two: $\omega_R^J > \omega_S^J$ for any $q_C \in (q_{NS}, q_{OS}]$. Intuitively, two forces are at work here. First, the punitive Nash abatement standard is lower in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$ than in $\Gamma_\infty^S(SW)$, i.e., $q_{NR} < q_{NS}$. Sec-

[37] The analysis was carried out using Mathematica. The file is available from the authors upon request.

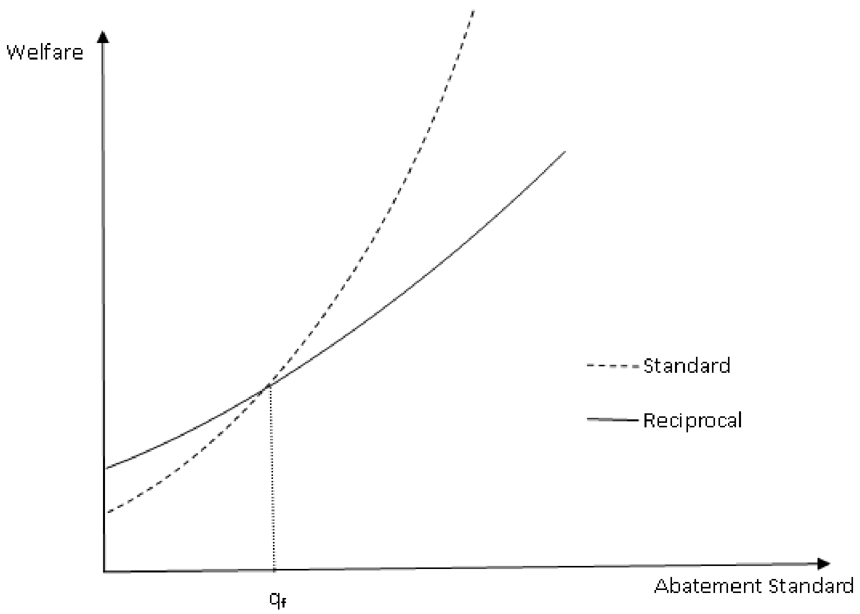
ond, infinite Nash reversion would also hurt J's partner. This could raise or lower the cost of the punishment phase for country J itself in $\Gamma_{\infty}^{R}(RW, w, \overrightarrow{q}_f)$, depending on whether the countries are in a positive- or negative-reciprocity state, i.e., depending on whether q_C is above or below q_f. In any case, for sufficiently low γ, this effect is relatively weak and is always dominated by the first effect.

Figure 1.1: Incentive to Cooperate



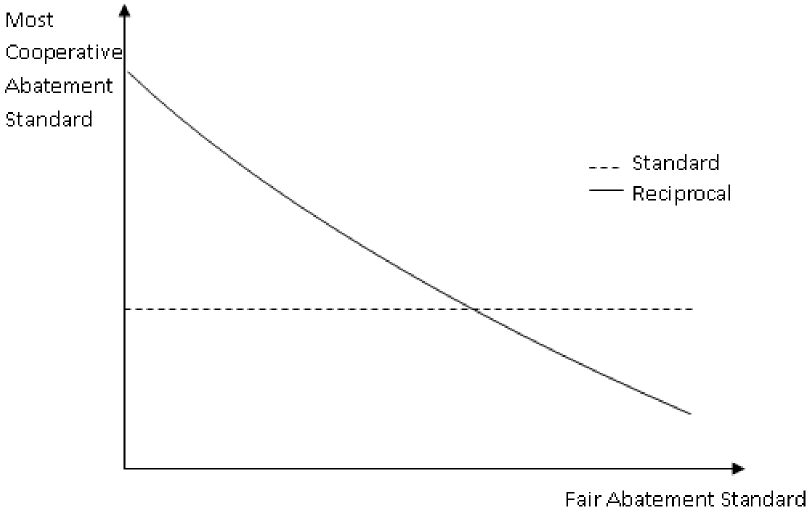
We next examine the static incentive country J has to cheat in $\Gamma_{\infty}^{R}(RW, w, \overrightarrow{q}_f)$ and $\Gamma_{\infty}^{S}(SW)$: Ω_R^J and Ω_S^J, correspondingly. As Figure 1.2 reveals, the former is weaker if and only if $q_C > q_f$. The intuition is direct. For any given q_C above q_f, the countries are in a positive-reciprocity state. This has a dampening effect on the defect abatement standard. Moreover, defection would be costly for J's partner, which acts to mitigate the potential onetime gains from cheating for J in $\Gamma_\infty^R(RW, w, \overrightarrow{q}_f)$. As a result, $\Omega_R^J < \Omega_S^J$ for any such q_C. Of course, the reverse is true for cooperative abatement standards below the fair abatement level, i.e., $\Omega_R^J > \Omega_S^J$ for $q_C < q_f$.

Figure 1.2: Incentive to Deviate



Therefore, for $q_f \in (q_{NS}, q_{CS}]$, reciprocal countries have a stronger incentive to cooperate and a weaker incentive to defect than self-interested countries around q_{CS}, implying that the former can support "greener" policies than the latter, or $q_{CR} > q_{CS}$. However, for $q_f > q_{CS}$, reciprocal countries have around q_{CS} both a stronger incentive to cheat and a stronger incentive to cooperate than self-interested ones. In other words, there are two offsetting forces at play for high fair-abatement-standard perceptions, making the comparison between q_{CS} and q_{CR} less clear-cut. Our simulations do confirm that for very high fair abatement standards, q_{CR} does indeed lie below q_{CS}, as we depict in Figure 1.3. To summarize, when countries are highly demanding from each other regarding their environmental policies, reciprocity could have a detrimental effect on international environmental cooperation.

Figure 1.3: Most Cooperative Abatement Standard



1.6 Endogenizing Fairness

We have hitherto assumed that perceptions of fairness are exogenous and constant between periods. This is consistent with the experimental work of Fehr and Falk (1999), who in a wage-setting context find virtually no change in either behavior or perceptions of fairness over time. However, one could argue that countries' perceptions of a fair abatement standard might adjust during the course of the game. As Kahneman et al. (1986, pp.730–1) write: "Psychological studies of adaptation suggest that any stable state of affairs tends to become accepted eventually, at least in the sense that alternatives to it no longer readily come to mind. Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction...[people] adapt their views of fairness to the norms of actual behavior."[38] In this section, we extend our analysis by allowing for endogenous formation of fair-abatement-standard perceptions, and investigate whether our main predictions so far continue to hold.

To this end, we adapt to our discrete-time framework an equation widely used in the habit formation literature (e.g., Ryder and Heal, 1973; Carroll et al., 2000; Fuhrer, 2000), assuming that current abatement standards affect future fair-abatement-standard perceptions as follows:

$$q_f^{-J,t} = \alpha q^{-J,t-1} + (1-\alpha)\, q_f^{-J,t-1}, \text{ for any } J, \tag{1.19}$$

where $\alpha \in (0,1)$. Equation (1.19) indicates that country J's fair-abatement-standard perception today is a linear combination of its last period's fair-abatement-standard perception and of the abatement standard country $-J$ actually implemented in that period. In other words, if country $-J$ sets an abatement standard today above the one country J deems fair, then country J will be more demanding in the next period (i.e., q_f^{-J} increases), which acts to lower its fairness payoff for any given q^{-J}. On the other hand, if country

[38] See Franciosi et al. (1995) for experimental support of these ideas in a price-setting context.

$-J$ selects an abatement standard today below q_f^{-J}, then country J will be less demanding in the next period (i.e., q_f^{-J} decreases), which acts to raise its fairness payoff given a q^{-J}.[39] It follows that as the game unfolds, country J's fair-abatement-level perception converges to $-J$'s actual environmental policy. More formally:

$$\left| q_f^{-J,t} - q^{-J,t} \right| \underset{t \to \infty}{\to} 0. \tag{1.20}$$

We are now prepared to examine whether our main conclusions heretofore are affected in any fundamental way by (1.19). Let us make the assumption that α is not too big (i.e., countries do not adjust their reference levels too quickly). Clearly, endogenizing fairness has no impact on the static game. At the same time, in the dynamic game, the major difference equation (1.19) introduces is that both q_{NR} and q_{CR} vary over time. However, given an α and an initial fair-abatement-level perception, we can readily derive the future fair-abatement-standard perceptions, and hence q_{NR}^t and q_{CR}^t for all t.

It is direct to show that if the fair abatement level is initially below the most cooperative equilibrium abatement standard of the self-interested game, then q_{CR} remains above q_{CS} along the equilibrium path, which is along the lines of Proposition 2. Intuitively, under this scenario, the countries start with a q_{CR} above q_{CS}. As the game progresses, q_f converges to q_{CR}, and thus, q_{CR} converges to q_{CS} (since w^J converges to zero). But for an α that is not too high, q_f never exceeds q_{CS} (implying that q_{CS} never exceeds q_{CR}

[39] It is only reasonable to assume that as international environmental cooperation becomes stronger, countries become more demanding with respect to environmental policy. For example, it is logical to expect that the abatement standards deemed fair nowadays are substantially higher compared with those in the 1970s or 1980s, when global warming did not occupy center stage in the policy arena and international environmental cooperation was substantially weaker.

by Proposition 2).

Moreover, a result analogous to Proposition 3 is obtained: If the fair abatement standard is initially higher than q_{CS}, then the effect of fairness and reciprocity on multilateral tariff cooperation is ambiguous, since the countries might start with a q_{CR} either below or above q_{CS}. Eventually though, under this scenario as well, the reciprocal game (slowly) converges to the self-interested one. In summary, allowing for endogenously formed fair-abatement-standard perceptions does not affect the qualitative nature of our findings (as long as α is not too high).

1.7 Conclusion

In this paper we examined the impact of fairness and reciprocity on international environmental cooperation in pollution abatement, where pollution was assumed to be transboundary in nature. More specifically, we investigated whether in the context of self-enforcing IEAs (involving full participation) reciprocal countries can support a higher degree of pollution abatement than self-interested ones. In our setting, a reciprocal country is willing to reward another country by raising its own abatement standard and therefore reducing transboundary pollution if it expects the latter to behave kindly by setting a higher abatement standard than the one deemed fair; nevertheless, the reverse is true when the latter is expected to behave unkindly by implementing an unfairly low level of pollution abatement. This is an important question for two reasons. First, governments and individuals seem to exhibit reciprocal preferences towards environmental policy. Second, our analysis provides a novel perspective on the successes and failures of international environmental negotiations.

We have established that reciprocal countries that have moderate expectations from each other regarding their national abatement policies (i.e., when their fair-abatement-standard perceptions are not too high) can sustain higher cooperative abatement levels than self-interested ones. On the other hand, when countries are highly demanding from each other with respect to environmental policy (i.e., when only very high abatement standards are perceived as fair), then reciprocity could have a detrimental effect on international environmental cooperation. Our findings therefore suggest a plausible explanation for the failure of the 2009 Copenhagen summit and the success of the 2010 Cancun one. In particular, it is now evident that countries entered the Copenhagen negotiations with overly ambitious expectations. Our analysis demonstrates that such high expectations (represented in our framework by high fair-abatement-standard perceptions) could prove to be counterproductive, hindering the efforts for deeper international environmental cooperation. The opposite is true for the Cancun summit. Expectations were much more modest and the outcome was surprisingly positive.

In concluding, a remark is in order. We focused here on symmetric countries, and argued that introducing asymmetries of a not too high degree in their fair-abatement-standard perceptions would not affect the qualitative nature of our findings. It would be interesting, though, to incorporate into our framework further (and larger) asymmetries among countries (e.g., in their size or their production structure). Such a model would provide us with valuable insights into the implications of fairness and reciprocity for

North-South environmental cooperation. We leave this avenue for future research.

1.8 Acknowledgements

We would like to thank Luca Opromolla, Luis Santos-Pinto, José Tavares, Anastasios Xepapadeas, and participants at CRETE 2010, ETSG 2010, IAREP/SABE/ICABEEP 2010, QED Jamboree 2010, UECE Lisbon Meetings 2010, and seminars at Cardiff Business School, ISEG, Istanbul Bilgi University, Izmir University of Economics, and Universidade Nova de Lisboa for very helpful comments and suggestions. A previous version of the paper was written while İriş was visiting UC San Diego, and Tabakis was an International Faculty Fellow at the MIT Sloan School of Management. Tabakis gratefully acknowledges financial support from the Fundação para a Ciência e a Tecnologia in Portugal. Any errors are ours.

Appendix 1.A Proofs

Lemma 2: For the static game with reciprocal countries $\Gamma^R(RW, w, \overrightarrow{q}_f)$, there exist largest and smallest pure symmetric Nash equilibria, $\overrightarrow{\overline{q}}_{NR} \equiv (\overline{q}_{NR}, \overline{q}_{NR})$ and $\overrightarrow{\underline{q}}_{NR} \equiv \left(\underline{q}_{NR}, \underline{q}_{NR}\right)$. Similarly, for the static game with self-interested countries $\Gamma^S(SW)$, there also exist largest and smallest pure symmetric Nash equilibria, $\overrightarrow{\overline{q}}_{NS} \equiv (\overline{q}_{NS}, \overline{q}_{NS})$ and $\overrightarrow{\underline{q}}_{NS} \equiv \left(\underline{q}_{NS}, \underline{q}_{NS}\right)$. Moreover, *if the countries' best-response functions have a slope strictly greater than* -1, *then* $\Gamma^R(RW, w, \overrightarrow{q}_f)$ and $\Gamma^S(SW)$ have unique equilibria.

Proof of Lemma 2: We first consider $\Gamma^R(RW, w, \overrightarrow{q}_f)$. Let us define new strategies $a^J = q^J$ and $a^{-J} = -q^{-J}$, reversing the natural order in country $-J$'s strategy set. Then, $\frac{\partial RW^J}{\partial a^J \partial a^{-J}} > 0$. Given now that the number of countries is two and that for any country J (i) $[0, q^J_{OR}]$ is a compact interval in $\mathcal{R}^+$; (ii) RW^J is twice continuously differentiable on $[0, q^J_{OR}]$; and (iii) $\frac{\partial RW^J}{\partial a^J \partial a^{-J}} > 0$, we know from Theorem 4 in Milgrom and Roberts (1990) that $\Gamma^R(RW, w, \overrightarrow{q}_f)$ is a (smooth strictly) supermodular game. It then follows from Theorem 5 in Milgrom and Roberts (1990) that (i) there exist largest and smallest serially undominated strategies for each country J, $\overline{q}^J$ and $\underline{q}^J$; and (ii) the strategy profiles $\overrightarrow{\overline{q}} \equiv \left(\overline{q}^J, \overline{q}^{-J}\right)$ and $\overrightarrow{\underline{q}} \equiv \left(\underline{q}^J, \underline{q}^{-J}\right)$ are pure Nash equilibrium profiles. Finally, given the overall symmetry of our setup, we have that $\overline{q}^J_{NR} = q^{-J}_{NR} \equiv \overline{q}_{NR}$ and $\underline{q}^J_{NR} = \underline{q}^{-J}_{NR} \equiv \underline{q}_{NR}$.

The second part of the lemma is straightforward upon recalling that $\Gamma^S(SW)$ can be obtained from $\Gamma^R(RW, w, \overrightarrow{q}_f)$ by setting $\gamma = 0$, meaning that $\Gamma^S(SW)$ is also a (smooth strictly) supermodular game. Last, the uniqueness of the equilibria when the countries' best-response functions have a slope strictly greater than -1 follows directly from Theorem 2.8 in Vives

(1999). *Q.E.D.*

Proof of Lemma 1: Given that (i) $\Gamma^R(RW, w, \overrightarrow{q}_f)$ is a supermodular game; and (ii) $\frac{\partial RW^J}{\partial q^J \partial q_f} < 0$ for any J, the lemma follows immediately from Theorem 6 in Milgrom and Roberts (1990). *Q.E.D.*

Proof of Proposition 1: If $q_f = q_{NS}$, then trivially $q_{NR} = q_{NS} \equiv q_N$, implying that for any J, $RW^J(\overrightarrow{q}_N, e_f) = SW^J(\overrightarrow{q}_N)$ since $w^J(q_N, q_f) = 0$ by (1.5). On the other hand, if $q_f > q_{NS}$, then $q_{NR} < q_{NS}$ by Lemma 1. These two inequalities imply that $q_{NR} < q_f$, and thus for any J, $w^J(q_{NR}, q_f) < 0$ from (1.5). Moreover, for $q_{NR} < q_{NS}$, $SW^J(\overrightarrow{q}_{NR}) < SW^J(\overrightarrow{q}_{NS})$ for all J. But then it follows that for all J, $RW^J(\overrightarrow{q}_{NR}, q_f) < SW^J(\overrightarrow{q}_{NS})$. *Q.E.D.*

Proof of Proposition 2: We first prove the following auxiliary lemma:

Lemma 3: *Let $q_C \geq q_f$ be a cooperative abatement standard. The critical discount factor above which international environmental cooperation can be maintained at q_C is lower in $\Gamma^R_\infty(RW, w, \overrightarrow{q}_f)$ than in $\Gamma^S_\infty(SW)$, i.e., $\delta^R_{q_C} < \delta^S_{q_C}$.*

Proof of Lemma 3: We want to show that $q_C \geq q_f$ implies that $\delta^R_{q_C} = \frac{RW^J_D - RW^J_C}{RW^J_D - RW^J_N} < \frac{SW^J_D - SW^J_C}{SW^J_D - SW^J_N} = \delta^S_{q_C}$. To do so, we will prove:

(i) If $q_C \geq q_f \Rightarrow RW^J_D - RW^J_C \leq SW^J_D - SW^J_C$ for any J.

(ii) If $q_C \geq q_f \Rightarrow RW^J_D - RW^J_N > SW^J_D - SW^J_N$ for any J.

Let us start with (i). We have that for any J:

$$RW^J_C = SW^J(\overrightarrow{q}_C) + \gamma w^J(q_C, q_f) SW^{-J}(\overrightarrow{q}_C) \text{ and}$$

$$RW_D^J = SW^J(BR_R^J(q_C), q_C) + \gamma w^J(q_C, q_f) SW^{-J}(BR_R^J(q_C), q_C).$$

Therefore:

$$\begin{aligned} RW_D^J - RW_C^J &= SW^J(BR_R^J(q_C), q_C) - SW^J(\overrightarrow{q}_C) \\ &+\gamma w^J(q_C, q_f)\left(SW^{-J}(BR_R^J(q_C), q_C) - SW^{-J}(\overrightarrow{q}_C)\right) \\ &\leq SW^J(BR_R^J(q_C), q_C) - SW^J(\overrightarrow{q}_C) \\ &\leq SW^J(BR_S^J(q_C), q_C) - SW^J(\overrightarrow{q}_C) = SW_D^J - SW_C^J. \end{aligned} \tag{1.21}$$

We know from (1.5) that $w^J(q_C, q_f) \geq 0$ if $q_C \geq q_f$. Furthermore, the welfare of self-interested country $-J$ is (weakly) lower when country J deviates while it still cooperates than when both countries cooperate, i.e., $SW^{-J}(BR_R^J(q_C), q_C) - SW^{-J}(\overrightarrow{q}_C) \leq 0$. The first inequality then follows. The second inequality stems from the fact that $BR_S^J(q_C)$ is the best reply of the self-interested country J. This concludes the proof of (i).

We now turn to (ii). Let us rewrite the result we want to show:

$$q_C \geq q_f \Rightarrow \left(RW_D^J - SW_D^J\right) - \left(RW_N^J - SW_N^J\right) > 0 \text{ for any } J.$$

By Proposition 1 we know that the Nash equilibrium abatement level of $\Gamma^S(SW)$ is (weakly) higher than that of $\Gamma^R(RW, w, \overrightarrow{q}_f)$, i.e., $q_{NR} \leq q_{NS}$. Thus, $q_f \geq q_{NS} \geq q_{NR}$, implying that $w^J(q_{NR}, q_f) \leq 0$ by (1.5). Therefore, the following inequality holds for any J:

$$\begin{aligned} RW_N^J &= SW^J(\overrightarrow{q}_{NR}) + \gamma w^J(q_{NR}, q_f) SW^{-J}(\overrightarrow{q}_{NR}) \\ &\leq SW^J(\overrightarrow{q}_{NR}) \leq SW^J(\overrightarrow{q}_{NS}) = SW_N^J. \end{aligned} \tag{1.22}$$

Next we will show that $RW_D^J - SW_D^J \geq 0$ for any J. Remember that γ is assumed to be sufficiently small. Taking a first-order Taylor series expansion of $RW^J(BR_R^J(q_C), q_C, q_f)$ around $\gamma = 0$, we obtain:

$$\begin{gathered} RW^J(BR_R^J(q_C), q_C, q_f) \approx SW^J(BR_S^J(q_C), q_C) + \\ +\gamma w^J(q_C, q_f) SW^{-J}(BR_S^J(q_C), q_C) \Leftrightarrow \\ \Leftrightarrow RW^J(BR_R^J(q_C), q_C, q_f) - SW^J(BR_S^J(q_C), q_C) \approx \\ \approx \gamma w^J(q_C, q_f) SW^{-J}(BR_S^J(q_C), q_C) \geq 0 \Leftrightarrow \\ \Leftrightarrow RW_D^J - SW_D^J \geq 0. \end{gathered} \tag{1.23}$$

The inequality holds due to $w^J(q_C, q_f) \geq 0$. By assumption, we have that $q_C > q_{NS}$, and thus q_f cannot be equal to both q_C and q_{NS} at the same time. Hence, at least one of the inequalities in (1.22) and (1.23) must be strict. This conclude the proof of part (ii). Therefore, by (i) and (ii), we finally have $\delta_{q_C}^R < \delta_{q_C}^S$. *Q.E.D.*

We are now ready to prove Proposition 2. We know from Lemma 3 that for any cooperative abatement standard $q_C \geq q_f$, $\delta_{q_C}^R < \delta_{q_C}^S$. So, given the assumption of Proposition 2 (i.e., $q_{CS} \geq q_f$), this is also true for the most cooperative equilibrium abatement standard of the repeated game with self-interested countries, q_{CS}: $\delta_{q_{CS}}^R < \delta_{q_{CS}}^S$. Note that both self-interested and reciprocal countries can sustain q_{CS} at the discount factor $\delta_{q_{CS}}^S$, but only reciprocal countries can support q_{CS} at $\delta_{q_{CS}}^R$. From (1.9) and (1.11), we have:

$$\begin{aligned} SW_D^J - SW_C^J &= \delta_{q_{CS}}^S \left(SW_D^J - SW_N^J\right) \text{ and} \\ RW_D^J - RW_C^J &= \delta_{q_{CS}}^R \left(RW_D^J - RW_N^J\right). \end{aligned}$$

Since $\delta^R_{q_{CS}} < \delta^S_{q_{CS}}$:

$$\begin{aligned} RW^J_D - RW^J_C &< \delta^S_{q_{CS}}\left(RW^J_D - RW^J_N\right) \Leftrightarrow \\ \Leftrightarrow \left(1-\delta^S_{q_{CS}}\right) &RW^J(BR^J_R(q_{CS}), q_{CS}, q_f) \\ < RW^J(\overrightarrow{q}_{CS}, q_f) &- \delta^S_{q_{CS}} RW^J\left(\overrightarrow{q}_{NR}, q_f\right), \end{aligned} \tag{1.24}$$

meaning that $\Omega^J_R(q_{CS}) < \frac{\delta^S_{q_{CS}}}{1-\delta^S_{q_{CS}}}\omega^J_R(q_{CS})$, or that the incentive-compatibility condition is not binding for a reciprocal country J at the pair $(q_{CS}, \delta^S_{q_{CS}})$.

Note here that RW^J_N does not depend on the cooperative abatement level. Moreover, for any cooperative abatement standard q_C higher than the most cooperative equilibrium abatement level of $\Gamma^s_\infty(SW)$, q_{CS}, the welfare for reciprocal country J under defection from q_C is higher than the welfare under deviation from q_{CS}:

$$RW^J(BR^J_R(q_C), q_C, q_f) > RW^J(BR^J_R(q_{CS}), q_{CS}, q_f).$$

At the same time, for such a $q_C > q_{CS}$, country J's welfare under cooperation is also higher at q_C than at q_{CS}:

$$RW^J(\overrightarrow{q}_C, q_f) > RW^J(\overrightarrow{q}_{CS}, q_f).$$

By the continuity of $RW^J(\bullet)$, there exists a cooperative abatement $\widehat{q}_C > q_{CS}$ such that (1.24) still holds, or $\Omega^J_R(\widehat{q}_C) < \frac{\delta^S_{q_{CS}}}{1-\delta^S_{q_{CS}}}\omega^J_R(\widehat{q}_C)$. Since the same analysis applies to any $(q_{CS}, \delta^S_{q_{CS}})$ pair for $\delta^S_{q_{CS}} \in [\underline{\delta}, \overline{\delta}]$, we have that for any $\delta \in [\underline{\delta}, \overline{\delta}]$, $q_{CS} < q_{CR}$. *Q.E.D.*

Proof of Proposition 3: Lemma 3 holds for any cooperative abatement standard q_C (weakly) higher than the fair abatement level q_f. However, for any $q_C < q_f$, it is ambiguous by (1.21) and (1.23) whether $\delta^R_{q_C}$ or $\delta^S_{q_C}$ is higher, since the weight function is negative at q_C. Hence, it is possible that the minimum discount factor required for countries with reciprocal preferences to sustain cooperation at q_C is higher than that for self-interested countries, i.e., $\delta^R_{q_C} > \delta^S_{q_C}$. Let us consider this case first, and focus on the most cooperative equilibrium abatement standard of $\Gamma^s_\infty(SW)$, q_{CS}. Under the scenario in question, both types of countries could sustain cooperation at q_{CS} only with a level of discount factor equal to $\delta^R_{q_{CS}}$ or above. Moreover, let us make the assumption that $\Omega^J_R(\bullet)$ is a strictly convex function whereas $\omega^J_R(\bullet)$ is a strictly concave one.[40] From (1.11) and (1.9), we have:

$$\begin{aligned} RW^J_D - RW^J_C &= \delta^R_{q_{CS}}\left(RW^J_D - RW^J_N\right) \text{ and} \\ SW^J_D - SW^J_C &= \delta^S_{q_{CS}}\left(SW^J_D - SW^J_N\right). \end{aligned}$$

Since $\delta^R_{q_{CS}} > \delta^S_{q_{CS}}$:

$$\begin{gathered} RW^J_D - RW^J_C > \delta^S_{q_{CS}}\left(RW^J_D - RW^J_N\right) \Leftrightarrow \\ \Leftrightarrow \left(1-\delta^S_{q_{CS}}\right) RW^J(BR^J_R(q_{CS}), q_{CS}, q_f) \\ > RW^J(\overrightarrow{q}_{CS}, q_f) - \delta^S_{q_{CS}} RW^J\left(\overrightarrow{q}_{NR}, q_f\right), \end{gathered}$$

meaning that $\Omega^J_R(q_{CS}) > \frac{\delta^S_{q_{CS}}}{1-\delta^S_{q_{CS}}}\omega^J_R(q_{CS})$, or that the incentive-compatibility condition is violated for a reciprocal country J at the pair $(q_{CS}, \delta^S_{q_{CS}})$.

[40] This assumption is clearly not restrictive given the type of result we are seeking here.

For any cooperative abatement standard q_C lower (higher) than the most cooperative equilibrium abatement standard of $\Gamma_\infty^s(SW)$, q_{CS}, the onetime gain for reciprocal country J under defection from q_C is lower (higher) than the static gain under deviation from q_{CS}:

$$\Omega_R^J(q_C) < (>)\Omega_R^J(q_{CS}).$$

At the same time, for such a $q_C < (>)q_{CS}$, country J's per-period gain from cooperation is also lower (higher) at q_C than at q_{CS}:

$$\omega_R^J(q_C) < (>)\omega_R^J(q_{CS}).$$

Given the strict convexity of $\Omega_R^J(\bullet)$ and the strict concavity of $\omega_R^J(\bullet)$, it follows that the incentive-compatibility condition for reciprocal country J can only be restored at a cooperative abatement standard $\widehat{q}_C < q_{CS}$. Since the same analysis applies to any $(q_{CS}, \delta_{q_{CS}}^S)$ pair for $\delta_{q_{CS}}^S \in [\underline{\delta}, \overline{\delta}]$, we have that for any $\delta \in [\underline{\delta}, \overline{\delta}]$, $q_{CS} > q_{CR}$.

Nevertheless, $\delta_{q_{CS}}^R < \delta_{q_{CS}}^S$ is also possible by (1.21) and (1.23). In this case, as we showed in the proof of Proposition 2, $q_{CS} < q_{CR}$. Therefore, when $q_f > q_{CS}$, it is ambiguous whether q_{CR} or q_{CS} is higher due to the ambiguity of whether $\delta_{q_{CS}}^R$ or $\delta_{q_{CS}}^S$ is higher. *Q.E.D.*

Chapter 2

Multilateral Tariff Cooperation under Fairness and Reciprocity

This paper explores the impact of fairness and reciprocity on multilateral tariff cooperation. Reciprocal countries reward kind behavior (positive reciprocity), but retaliate against countries behaving unkindly (negative reciprocity). We demonstrate that reciprocal countries that are moderately demanding from their trading partners regarding their commercial policy can support a greater degree of cooperation than self-interested ones. However, when only very liberal import policies are considered fair, then reciprocity could have a detrimental effect on multilateral tariff cooperation. Thus, our model provides a novel reason for the occasional failure of trade negotiations. Finally, we show that these results are robust to endogenizing fair-tariff perceptions.

Keywords: Reciprocity; Trade agreements; Trade policy; Repeated games
JEL Classification: F13; D63

2.1 Introduction

> "*We wish to do it [promote commerce] by throwing open all the doors of commerce and knocking off all its shackles. But as this cannot be done for others, unless they will do it for us, and there is no probability that Europe will do this, I suppose we may obliged to adopt a system which may shackle them in our ports, as they do to us in theirs.*" – Thomas Jefferson, 1785[1]

Reciprocity is an old theme in international trade negotiations. In this paper, we set out to explore the implications of reciprocal preferences for commercial policy and multilateral trade agreements. Governments with reciprocal preferences reward "kind" or "fair" actions (positive reciprocity), whereas they punish "unkind" or "unfair" behavior (negative reciprocity). This is an important question for two reasons. First, governments seem to exhibit such preferences, at least with respect to trade policy. Second, our analysis provides a novel perspective on the successes and the occasional failures of multilateral trade negotiations. We should stress here that our definition of "reciprocity" differs substantially from the standard one in the trade literature (e.g., Bagwell and Staiger, 1999a; Freund, 2003; Krishna and Mitra, 2005). In these papers, the term "reciprocity" refers broadly to mutual changes in trade policy by *self-interested* countries which bring about changes in each country's import volume that are of equal value to the changes in its export volume. Instead, in our framework, reciprocity characterizes the *preferences* of countries.

[1] PTJ 8: 633.

Most governments specifically state that one of their major goals when sitting at the negotiations table is to promote fair trade through reciprocity, with "fair trade" typically entailing that (i) domestic producers (and workers) are faced with fair trade policies worldwide; and (ii) the gains from trade are fairly distributed among trading partners. For instance, President Obama's 2009 Trade Policy Agenda Report states that "If we work together, free and fair trade...will be a powerful contributor to the national and global well being."[2] Analogous goals and concerns characterize the commercial policy of the European Union (EU). On its official website on external trade, it is written that "The EU has evolved during the process of globalization by aiming for the harmonious development of world trade and fostering fairness..."[3] At the same time, the emphasis placed on reciprocity is equally strong. For example, President Sarkozy and Chancellor Merkel in a joint letter to the President of the EU Council in 2007 argue that "Open markets can only develop their full potential if transparent rules facilitate fair competition in a spirit of reciprocity."[4] Moreover, a 2006 Communication of the EU Commission warns that "If necessary, targeted restrictions will be maintained [on behalf of the EU] for uncooperative countries with the aim of encouraging them towards a mutual opening up of markets."[5] Another such example of negative reciprocity is the extensive employment of anti-dumping and countervailing measures within the context of the World Trade Organization (WTO), aimed directly at punishing unfair trade practices. In brief,

[2] See http://www.ustr.gov/node/4442.

[3] See http://europa.eu/legislation_summaries/external_trade/index_en.htm.

[4] "Reinforcing the Lisbon Strategy through External Economic Measures," French Presidency/German Chancellor's Office, September 10, 2007.

[5] See http://europa.eu/legislation_summaries/external_trade/r11022_en.htm.

in the international trade arena, governments seem, to some extent, to not be simply maximizers of their own self-interested welfare, but to rather be exhibiting some preferences for fairness and reciprocity.

It could also be argued that it is reasonable to expect governments to have reciprocal preferences towards commercial policy since such preferences are exhibited by voters.[6] In experiments, there is ample evidence of both positive and negative reciprocity in individual decision making.[7] Public opinion polls suggest that a significant proportion of people do exhibit reciprocal preferences with respect to trade policy as well. For example, in a January 2004 PIPA poll, 67% of Americans agreed that "in general, if another country is willing to lower its barriers to products from the US if we [the US] will lower our [its] barriers to their products," the US should do so, whereas only 24% disagreed with this statement. More importantly, almost 75% of the ones endorsing this statement agreed that "the US should only lower its barriers if other countries do, because that is the only way to pressure them to open their markets," while just 24% of them thought that "the US

[6] In fact, one could argue that governments are not genuinely motivated by fairness considerations, but only push for fairer trade policies in order to increase their political support, since voters have such considerations. In any case, our results depend solely on governments exhibiting, to some degree, reciprocal preferences towards commercial policy, and not on why this is the case.

[7] For instance, experiments asking individuals to contribute to public goods typically find that their contributions far exceed what self-interested utility maximization would entail (e.g., Andreoni, 1988; Palfrey and Prisbrey, 1997; Croson, 2007). This is usually interpreted as evidence of positive reciprocity. Analogous results arise from trust or gift-exchange experiments (e.g., Berg et al., 1995; Fehr et al., 1997; Fehr et al., 1998). On the other hand, evidence for negative reciprocity is found in ultimatum-game experiments with the typical result being that people reject offers that would be accepted under the self-interested hypothesis (e.g., Güth et al., 1982; Roth et al., 1991). Moreover, in a recent paper, Dohmen et al. (2009) provide evidence of both positive and negative reciprocity using survey data.

should lower its barriers even if other countries do not, because consumers can buy cheaper imports and foreign competition spurs American companies to be more efficient."[8] Similar preferences towards commercial policy are exhibited by Europeans. In a spring 2001 Eurobarometer survey, more than 74% of EU-15 citizens did endorse reciprocity in international trade agreements, whereas merely 7% of them did not.[9] Another important conclusion that could be drawn from these polls is that people have an explicit notion of what fair trade involves, i.e., they have a reference level regarding fair commercial policy against which the policies implemented by their own government or by the rest of the world can be evaluated.[10] For instance, in a July 2004 CCFR poll, Americans were asked whether the US practices fair trade with various other countries, and whether the countries in question have fair trade policies towards the US. The percentage of respondents who were not sure/declined to answer was very low overall, ranging from 11% to 14% in the former case and from 11% to 15% in the latter.[11] Likewise, in a 19-nation poll conducted November 2003 through February 2004, people around the globe were asked whether rich countries are playing fair in trade negotiations with poor countries. In this poll, the percentage of respondents who were not sure/declined to answer was just 10% or lower in 14 out of the 19 countries.[12]

[8] See http://americans-world.org/digest/global_issues/intertrade/tradepolicy.cfm.

[9] See (in French) http://ec.europa.eu/public_opinion/archives/ebs/ebs_152_fr.pdf.

[10] The "reference level" is a concept widely studied in the behavioral economics literature. For more on this, see, for instance, Helson (1964) and Tversky and Kahneman (1991).

[11] See http://americans-world.org/digest/global_issues/intertrade/reservations_trade.cfm.

[12] See http://www.worldpublicopinion.org/pipa/pdf/jun03/GlobalIss_Jun04_quaire.pdf.

On a more theoretical level, if individuals have reciprocal preferences, political-economy models of trade policy suggest that these preferences will be reflected in the government's objective function. The first model we can invoke here is the median-voter model, where the government chooses policies that reflect majority opinion on the issue (in order to remain popular and stay elected).[13] In such a setting, if the median voter has reciprocal preferences, then the government's actions are going to mirror these preferences. Instead, we could look at interest-group models. The framework that currently occupies center stage in the literature is due to Grossman and Helpman (1994). In their paper, the incumbent government maximizes a weighted average of aggregate social welfare and political contributions by lobbies that wish to influence trade policy. Alternatively, political influences could be readily represented, as Baldwin (1987) has demonstrated, by a parameter that attaches additional weight to producer surplus in the government's objective function. In either case, if individuals have preferences for fairness and reciprocity, these preferences will enter into the government's objective function with some weight.[14]

To address the implications of reciprocity and fairness for commercial policy, we develop a dynamic game in which reciprocal countries facing a terms-of-trade Prisoner's Dilemma problem in their dealings with one another attempt to maintain tariff cooperation. To model reciprocity, we follow Segal and Sobel (2007). In particular, we assume that a country attaches a positive (negative) weight to the self-interested welfare of a trading partner

[13] See, for example, Mayer (1984).

[14] See Goldberg and Maggi (1999) and Gawande and Bandyopadhyay (2000) for US empirical evidence that social welfare does in fact receive a high weight in the government's objective function.

if it expects the latter to behave kindly (unkindly) by imposing an import tariff that is lower (higher) than the one it perceives as fair. In other words, countries are assumed to have preferences over both outcomes and strategies. Of course, fairer tariff policies correspond to a fairer distribution of the gains from trade among countries (or interest groups in different countries), meaning that our results could be readily reinterpreted in terms of the distribution of the trade gains among trading partners. Besides the countries' reciprocal preferences, we model trade agreements in a standard fashion. More specifically, we maintain the assumption that binding commitments cannot be made at the international level and countries are therefore limited to cooperative multilateral tariff agreements that are self-enforcing. This assumption reflects the lack of a strong mechanism within the WTO for enforcing the trade policies agreed upon under its auspices. In this context, a country will choose today to adhere to the cooperative path as long as the onetime gain it could achieve by unilaterally deviating from its agreed-upon trade policies does not outweigh the discounted welfare cost of the future trade war its defection would ignite.[15]

We find that reciprocal countries that are moderately demanding from their trading partners with respect to their commercial policy (i.e., when the tariffs considered fair are not too low, but at the same time, overly restrictive import policies are perceived as unfair) can support a greater degree of multilateral tariff cooperation and thus achieve higher welfare than self-interested ones. The intuition is straightforward. For such fair-tariff perceptions, in the reciprocal game (i) the punitive Nash tariffs are higher than in the self-

[15] See, for example, Dam (1970), Dixit (1987), and Bagwell and Staiger (2002) for further elaboration on these points.

interested one; and (ii) the countries are in a positive-reciprocity state. As a result, under the scenario in question, reciprocal countries face both a weaker incentive to defect and a stronger incentive to cooperate than self-interested ones, allowing them to maintain more liberal trade policies.

However, when reciprocal countries are highly demanding from their trading partners regarding their import policy (i.e., when only very liberal import policies are considered fair), then the effect of reciprocity on multilateral trade cooperation is ambiguous. Intuitively, in such a case, reciprocal countries are in a negative-reciprocity state, meaning that they face a stronger incentive to cheat than self-interested ones, even though their incentive to cooperate is still relatively stronger due to the harsher punishment a defector faces in the reciprocal game. Our simulations do confirm that for very low fair tariffs, there are indeed cases where self-interested countries can support lower cooperative tariffs in equilibrium than reciprocal ones. Finally, we demonstrate that our results are robust to allowing for fair-tariff perceptions that are endogenously determined during the course of the game.

At this point, it is important to note that our findings suggest a novel reason for the occasional failure of trade negotiations: Assuming countries have (some) preferences for fairness and reciprocity, if they arrive at the negotiating table with expectations that are highly elevated (i.e., they have very low fair-tariff perceptions), this could prove counterproductive, in the sense that they might no longer be able to support very liberal trade policies. As a matter of fact, Mr. Renato Ruggiero, the WTO Director-General in 1995-1999, referred to this possibility in one of his speeches in 1995: "I have heard it said that unrealistically high expectations could pose a threat to the success of the negotiations. I have also heard it suggested that failure to meet such expectations could make a multilateral solution impossible."[16] It could then be argued that this might be one of the possible explanations for the problems plaguing the Doha Round since its launch in 2001. More specifically, the success of the 1986-94 Uruguay Round as well as the deepening of globalization in general in the 1990s might have led to countries entering the Doha negotiations with expectations that were too high, hindering the efforts for further multilateral trade liberalization. Put differently, if countries had entered the Doha Round with lower expectations, its outcome might have been more favorable. Another plausible explanation, still along the lines of our model, is that developing countries might have arrived at the negotiations being too demanding from developed countries regarding their trade policies (especially with respect to agricultural goods), partly due to their feeling that the previous round was lopsided or unfair.[17] In any case, as Dani Rodrik writes: "In the end, it may well be the atmospherics – psychology and expectations – rather than the actual economic results on the ground that will determine the outcomes [of the Doha Round]."[18] In summary, expectations emerge as a key factor in our analysis, having a significant effect on what can be achieved in the international trade arena. In fact, to the best of our knowledge, our paper is the first to identify such a role for expectations in international trade negotiations. The policy implications of our model are then straightforward. The careful management of expectations is criti-

[16] See http://www.wto.org/english/news_e/pres95_e/pr9512_e.htm.

[17] For more on the latter explanation, see "The Doha Round...and Round...and Round," *The Economist* (print edition), July 31, 2008.

[18] See http://www.guardian.co.uk/commentisfree/2008/aug/08/wto.internationalaidanddevelopment.

cal for the success of multilateral trade negotiations, and most importantly, the creation of a pre-negotiations high-expectations environment should be avoided.

The remainder of the paper is organized as follows. Section 2 sets out the basics. Section 3 characterizes the static Nash equilibrium of our model, whereas Section 4 analyzes the dynamic game. Section 5 presents a simplified model in order to better illustrate the main insights from our analysis. Section 6 endogenizes countries' fair-tariff perceptions. Finally, Section 7 identifies some promising avenues for future research and concludes. All the proofs are relegated to the Appendix.

2.2 The Model

We assume the world consists of two countries, A and B, that trade two goods, a and b.[19] This focus on a 2-country world is nonrestrictive, as our findings readily extend to N countries.[20] Country J is endowed with 1 unit of good $-j$ and zero units of good j, where $J \in \{A, B\}$ and $j \in \{a, b\}$.[21] On the consumption side, we maintain the assumptions that demand functions are symmetric across countries and goods, and that the demand for good j is independent of the price of good $-j$. More specifically, the demand for good j in country J is given by $D\left(P_j^J\right)$, where P_j^J is good j's price in country J. We make the standard assumptions that $D\left(P_j^J\right)$ is strictly positive on some

[19] Our framework is inspired by Bagwell and Staiger (1999b).

[20] Upon request, a technical appendix is available from the authors in which an N-country model is presented.

[21] We choose to ignore the production process in the two countries for expositional simplicity. In any case, this assumption does not affect the qualitative nature of our findings.

bounded interval $\left[0, \overline{P}_j^J\right)$, that $D\left(P_j^J\right) = 0$ for $P_j^J \geq \overline{P}_j^J$, and that $D\left(P_j^J\right)$ is twice continuously differentiable in P_j^J with $D'\left(P_j^J\right) < 0$ for $P_j^J \in \left[0, \overline{P}_j^J\right)$. Given our setup, country J imports good j, whereas it exports good $-j$ in accordance with the following export supply function:

$$X_{-j}^J\left(P_{-j}^J\right) = 1 - D\left(P_{-j}^J\right). \tag{2.1}$$

In each period, the countries simultaneously select specific (nonprohibitive) import tariffs so as to maximize their individual welfare. The tariffs are picked with perfect information as to all past tariff choices. Let country J's import tariff be $\tau^J \in \Theta^J \subset \mathcal{R}^+$, where Θ^J is a compact interval. The no-arbitrage condition for good j yields:

$$P_j^J = P_j^{-J} + \tau^J. \tag{2.2}$$

The equilibrium prices can then be obtained from the usual market-clearing conditions:

$$D\left(P_j^J\left(\tau^J\right)\right) = X_j^{-J}\left(P_j^{-J}\left(\tau^J\right)\right). \tag{2.3}$$

The countries are assumed to have preferences for fairness and reciprocity.

In particular, the welfare of country J is given by:[22]

$$RW^{J}\left(\tau^{J},\tau^{-J},\tau_{f}^{-J}\right)=SW^{J}\left(\tau^{J},\tau^{-J}\right)+\gamma w^{J}(\tau^{-J},\tau_{f}^{-J})SW^{-J}\left(\tau^{J},\tau^{-J}\right). \tag{2.4}$$

The first term, SW^{J}, is the self-interested (or "standard") welfare function, i.e., the sum of consumer surplus, producer surplus, and tariff revenue:

$$\begin{aligned} SW^{J}\left(\tau^{J},\tau^{-J}\right) &= \int_{P_{j}^{J}(\tau^{J})}^{\overline{P}_{j}^{J}} D\left(P\right)dP+\int_{P_{-j}^{J}(\tau^{-J})}^{\overline{P}_{-j}^{J}} D\left(P\right)dP+ \\ &\quad + P_{-j}^{J}\tau^{-J}+\tau^{J}X_{j}^{-J}\left(\tau^{J}\right). \end{aligned} \tag{2.5}$$

The second term, $\gamma w^{J}(\tau^{-J},\tau_{f}^{-J})SW^{-J}$, captures the fairness payoff for country J, where (i) its relative significance is specified by the scaling factor $\gamma\geq 0$; and (ii) $w^{J}(\tau^{-J},\tau_{f}^{-J})$ determines the weight country J places on its trading partner's self-interested welfare SW^{-J}, and is of the following form:

$$w^{J}(\tau^{-J},\tau_{f}^{-J})\begin{cases} >0 \text{ if } \tau_{f}^{-J}>\tau^{-J} \\ =0 \text{ if } \tau^{-J}=\tau_{f}^{-J} \\ <0 \text{ otherwise} \end{cases}, \tag{2.6}$$

with τ_{f}^{-J} being the τ^{-J} country J deems "fair." We maintain the assumptions that country J's weight function $w^{J}(\tau^{-J},\tau_{f}^{-J})$ is twice continuously differentiable in both arguments, and is nondecreasing in its own fair-tariff

[22] An alternative way of modeling reciprocity in a dynamic setup is due to Dufwenberg and Kirchsteiger (2004). In their paper, they develop a theory of reciprocity for extensive form games, and introduce a new solution concept – sequential reciprocity equilibrium – where players update their beliefs about their co-players' intentions as the game unfolds and choose their actions accordingly. However, their framework would be highly intractable for the purposes of this paper.

perception and nonincreasing in country $-J$'s tariff. Furthermore, we assume that fair-tariff perceptions are common knowledge.

Intuitively, the function w^J reflects the fact that a reciprocal country cares about the intentions of its trading partner. More specifically, the first condition in (2.6) expresses positive reciprocity: If country J expects the tariff of country $-J$ to be smaller than its own perception of a fair tariff, then it is willing to sacrifice some of its self-interested welfare to reward its trading partner. On the other hand, the third condition in (2.6) expresses negative reciprocity: When country J expects country $-J$ to impose a tariff that exceeds the one it perceives as fair, then it wishes to punish its trading partner and is willing to sacrifice some of its own self-interested welfare in order to do so. Moreover, if τ^{-J} and τ_f^{-J} are equal, then the self-interested and the reciprocal welfare functions coincide for country J. In brief, equation (2.6) signifies that from country J's standpoint, any τ^{-J} below τ_f^{-J} is a fair (or kind) action on the part of country $-J$ that should be rewarded, whereas any τ^{-J} in excess of τ_f^{-J} is an unfair (or unkind) action that should be punished.

2.3 Static Game

Our aim in this section is to characterize the static Nash equilibrium of our model, and compare it with the one that would emerge in a game with self-interested countries. This equilibrium will serve as a credible punishment in the dynamic game considered in the next section, the threat of which can support multilateral tariff cooperation in a repeated setting.[23] To this end, let the static game with self-interested countries be denoted by $\Gamma^S(SW)$, while $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ represents the static game with reciprocal countries, where $\overrightarrow{\tau}_f \equiv \left(\tau_f^J, \tau_f^{-J}\right)$ is the fair-tariff vector. We henceforth assume that $\tau_f^J = \tau_f^{-J} \equiv \tau_f$, i.e., the countries have a common fair-tariff perception. The reason for this assumption is twofold. First, it considerably simplifies the analysis. Second, asymmetries of a not too high degree in fair-tariff perceptions between the (otherwise symmetric) countries would not affect the qualitative nature of our findings.[24]

It is direct to show that the cross-partial derivative of the welfare function of reciprocal country J with respect to its own tariff and country $-J$'s tariff is nonnegative (i.e., $\frac{\partial RW^J}{\partial \tau^J \partial \tau^{-J}} \geq 0$). This means country J's incremental returns from raising its own tariff are nondecreasing in its partner's tariff, i.e., the choice variables are strategic complements. On the other hand, the cross-partial derivative of country J's welfare function with respect to its tariff and its fair-tariff perception is nonpositive (i.e., $\frac{\partial RW^J}{\partial \tau^J \partial \tau_f} \leq 0$). To gain some insight for the sign of these derivatives, simply recall that (i) increasing τ^J inflicts a negative terms-of-trade externality on country $-J$ (given our countries are "large"); and (ii) a higher τ^{-J} (τ_f) results ceteris paribus in a smaller (larger) w^J.

[23] In fact, the static Nash equilibrium would be the unique equilibrium for the dynamic game as well if a multilateral trade agreement were not feasible (e.g., due to exogenous, political reasons or because the countries were highly impatient and did not value the future at all).

[24] Upon request, a technical appendix is available from the authors in which we reproduce our analysis allowing for asymmetries in fair-tariff perceptions between the countries that are not too high. See also footnote 29.

Our first result establishes the existence of a pure symmetric Nash equilibrium for both $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ and $\Gamma^S(SW)$.

Lemma 1 *For the static game with reciprocal countries $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$, there exist largest and smallest pure symmetric Nash equilibria, $\overrightarrow{\overline{\tau}}_{NR} \equiv (\overline{\tau}_{NR}, \overline{\tau}_{NR})$ and $\overrightarrow{\underline{\tau}}_{NR} \equiv (\underline{\tau}_{NR}, \underline{\tau}_{NR})$. Moreover, for the static game with self-interested countries $\Gamma^S(SW)$, there also exist largest and smallest pure symmetric Nash equilibria, $\overrightarrow{\overline{\tau}}_{NS} \equiv (\overline{\tau}_{NS}, \overline{\tau}_{NS})$ and $\overrightarrow{\underline{\tau}}_{NS} \equiv (\underline{\tau}_{NS}, \underline{\tau}_{NS})$.*

We next show how countries' fair-tariff perception affects the extremal equilibrium tariffs of $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$.

Lemma 2 *The largest and the smallest pure Nash equilibria of $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$, i.e., $\overrightarrow{\overline{\tau}}_{NR} \equiv (\overline{\tau}_{NR}, \overline{\tau}_{NR})$ and $\overrightarrow{\underline{\tau}}_{NR} \equiv (\underline{\tau}_{NR}, \underline{\tau}_{NR})$, are non-increasing in τ_f.*

Intuitively, as we argued above, for a given τ^{-J}, a higher τ_f leads to a larger w^J. Consequently, country J now wishes to reduce the terms-of-trade negative externality of its tariff on its trading partner when choosing its import policy, resulting in more liberal Nash tariff equilibria.

It turns out that all our results hold independently of whether we consider the largest or the smallest pure Nash equilibria of $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ and $\Gamma^S(SW)$. Therefore, without loss of generality, we drop from now on the "bar" notation and simply write $\overrightarrow{\tau}_{NR} \equiv (\tau_{NR}, \tau_{NR})$ and $\overrightarrow{\tau}_{NS} \equiv (\tau_{NS}, \tau_{NS})$, referring to either $\overrightarrow{\overline{\tau}}_{NR}$ and $\overrightarrow{\overline{\tau}}_{NS}$, or $\overrightarrow{\underline{\tau}}_{NR}$ and $\overrightarrow{\underline{\tau}}_{NS}$, respectively. In addition, we hereafter assume that (i) $\gamma \neq 0$ so that country J's fairness payoff is nonzero for $\tau^{-J} \neq \tau_f$; and (ii) $\tau_f \leq \tau_{NS}$, i.e., overly restrictive import policies are considered unfair, which is a reasonable assumption given our focus on trade cooperation among countries. We next compare τ_{NR} with τ_{NS} as well as the welfare obtained in $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ and $\Gamma^S(SW)$.

Proposition 1 *Under our model's assumptions, (i) $\tau_{NR} \geq \tau_{NS}$; and (ii) for any J, $RW^J(\overrightarrow{\tau}_{NR}, \tau_f) \leq SW^J(\overrightarrow{\tau}_{NS})$, with equality only holding for $\tau_f = \tau_{NS}$.*

Proposition 1 demonstrates that reciprocal countries end up with a more protectionist and thus welfare-inferior Nash equilibrium than self-interested countries. The intuition is straightforward. At τ_{NS}, reciprocal countries are in a negative-reciprocity state wishing to punish each other (since $\tau_f \leq \tau_{NS}$), implying that the Nash equilibrium tariff of $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ must exceed the one of $\Gamma^S(SW)$.[25]

2.4 Dynamic Game

We now study repeated interaction between the countries. In particular, the dynamic game we consider is the stage game analyzed above infinitely repeated. We assume that countries cannot make binding international commitments but are instead limited to self-enforcing trade agreements. In such a setting, countries can still maintain multilateral trade cooperation, whose degree depends critically on how severely they can credibly punish an offender. Our aim in this section is to evaluate the effect of fairness and reciprocity on the ability of countries to cooperate with low import tariffs.

[25] At this point, a technical note is in order. Our results do not require differentiability of $RW^J(\bullet)$. Rather, our findings would still hold under the significantly weaker assumptions that $RW^J(\bullet)$ has increasing differences in (τ^J, τ^{-J}) and decreasing differences in (τ^J, τ_f). Nevertheless, for expositional simplicity, we have chosen to work with differentiable welfare functions.

To this end, denote the infinitely repeated game with reciprocal countries by $\Gamma_{\infty}^{R}(RW, w, \overrightarrow{\tau}_f)$, and the one with self-interested countries by $\Gamma_{\infty}^{S}(SW)$. The discount factor between periods is $\delta \in (0,1)$. For both games, we focus on symmetric cooperative subgame-perfect equilibria in which (i) along the equilibrium path, the countries set a common cooperative tariff $\tau_C < \tau_{NS}$ in each period; and (ii) if at any point in the game a defection occurs, both countries revert from the following period onwards to the noncooperative Nash tariff of the (relevant) stage game.[26] In other words, to enforce cooperation, the countries employ a grim-trigger strategy.

We begin our analysis with $\Gamma_{\infty}^{S}(SW)$. The static incentive self-interested country J has to cheat is defined as:

$$SW^J(BR_S^J(\tau_C), \tau_C) - SW^J(\overrightarrow{\tau}_C) \equiv SW_D^J - SW_C^J \equiv \Omega_S^J(\tau_C)\,, \tag{2.7}$$

where $BR_S^J(\tau_C)$ is country J's best-response tariff to τ_C and $\overrightarrow{\tau}_C \equiv (\tau_C, \tau_C)$. Ω_S^J equals simply the onetime increase in welfare country J achieves when it optimally chooses a tariff on its reaction curve while its trading partner still cooperates with τ_C. On the other hand, violating multilateral cooperation also bears consequences as a trade war ensues. The discounted future welfare

[26] Given the overall symmetry of our framework, it is only natural to focus on symmetric equilibria, which imply an equal split between the countries of the gains from cooperation. Actually, it can be readily shown that such a split supports the highest degree of multilateral trade cooperation in our setting. On a different note, observe that for both games we restrict our attention to cooperative tariffs lower than τ_{NS}. This enables us to better compare $\Gamma_{\infty}^{R}(RW, w, \overrightarrow{\tau}_f)$ with $\Gamma_{\infty}^{S}(SW)$, which is our main goal in this paper.

cost a defector faces equals:

$$\frac{\delta}{1-\delta}\left(SW^J(\overrightarrow{\tau}_C) - SW^J(\overrightarrow{\tau}_{NS})\right) \equiv \frac{\delta}{1-\delta}\left(SW_C^J - SW_N^J\right) \equiv \frac{\delta}{1-\delta}\omega_S^J\left(\tau_C\right), \tag{2.8}$$

where ω_S^J is the per-period value of cooperation for country J, i.e., the per-period increase in country J's welfare under multilateral cooperation as compared with its welfare during a tariff war. Therefore, the incentive-compatibility condition for a self-interested country J to adhere to the cooperative path in $\Gamma_\infty^S(SW)$ is that the onetime gain from defection, Ω_S^J, does not outweigh the discounted value of future cooperation, $\frac{\delta}{1-\delta}\omega_S^J$:

$$\Omega_S^J\left(\tau_C\right) \leq \frac{\delta}{1-\delta}\omega_S^J\left(\tau_C\right). \tag{2.9}$$

From (2.9), it is direct to show that a given cooperative tariff τ_C can be supported as a subgame-perfect equilibrium of the dynamic game as long as countries are patient enough, or:

$$\delta \geq \delta_{\tau_C}^S \equiv \frac{SW_D^J - SW_C^J}{SW_D^J - SW_N^J}. \tag{2.10}$$

Analogous relationships hold for countries with reciprocal preferences. In particular, the incentive-compatibility condition for a reciprocal country J to uphold multilateral cooperation is given by:

$$\Omega_R^J\left(\tau_C\right) \leq \frac{\delta}{1-\delta}\omega_R^J\left(\tau_C\right). \tag{2.11}$$

Moreover, for a given cooperative tariff τ_C, the minimum discount factor required so that the tariff in question can be multilaterally sustained equals:

$$\delta^{R}_{\tau_C} \equiv \frac{RW^J_D - RW^J_C}{RW^J_D - RW^J_N}. \tag{2.12}$$

Our next lemma demonstrates that reciprocal countries can support a fairly liberal cooperative tariff as long as they are sufficiently patient.

Lemma 3 *Let $\tau_C \leq \tau_f$ be a cooperative tariff. Then a sufficiently high discount factor exists such that $\overrightarrow{\tau}_C$ is a subgame-perfect Nash equilibrium for $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$.*

Let us, in all that follows, maintain the (nonrestrictive) assumption that γ is sufficiently small, meaning that the relative weight of the fairness payoff in the countries' objective function (or, equivalently, the relative weight the countries place on their trading partner's self-interested welfare) is not too high.[27] We are at this point ready to state our first result about the impact of fairness and reciprocity on multilateral trade cooperation. Using (2.10) and (2.12), we now compare $\delta^S_{\tau_C}$ against $\delta^R_{\tau_C}$, where $\tau_C \leq \tau_f$. Remember that $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ and $\Gamma^S_\infty(SW)$ are identical in all respects except for the fairness payoff in the countries' welfare function.

Proposition 2 *Let $\tau_C \leq \tau_f$ be a cooperative tariff. The critical discount factor above which multilateral cooperation can be maintained at τ_C is lower in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ than in $\Gamma^S_\infty(SW)$, i.e., $\delta^R_{\tau_C} < \delta^S_{\tau_C}$.*

The intuition underlying Proposition 2 is straightforward once it is recalled that for any cooperative tariff τ_C lower than the fair tariff τ_f, the countries

[27] The derivation of a closed-form solution for the upper bound of γ has proved elusive. In the simulations in the next section where a simplified model is presented, γ is set less than or equal to 0.1.

attach a positive weight to their partner's self-interested welfare, i.e., they are in a positive-reciprocity state. Two reinforcing forces are at work here. On the one hand, for any country J, the value of cooperation at τ_C is higher in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ than in $\Gamma^S_\infty(SW)$ since in the former game (i) the noncooperative (punitive) Nash tariff is higher; and (ii) infinite Nash reversion would also be costly for J's trading partner, which acts to heighten the cost of the punishment phase for country J itself. On the other hand, the static incentive country J has to deviate from τ_C is weaker in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ than in $\Gamma^S_\infty(SW)$ because in the former game (i) the defect tariff is lower, since the countries are in a positive-reciprocity state; and (ii) defection would hurt J's partner, mitigating J's potential onetime gains from cheating. It then follows that reciprocal countries can more easily support any given cooperative tariff below the fair tariff than self-interested ones.

Let now $\overrightarrow{\tau}_{CS} \equiv (\tau_{CS}, \tau_{CS})$ be the most cooperative equilibrium tariff vector of $\Gamma^s_\infty(SW)$, i.e., τ_{CS} is the smallest nonnegative tariff that does not invite cheating in the dynamic game with self-interested countries. Similarly, $\overrightarrow{\tau}_{CR} \equiv (\tau_{CR}, \tau_{CR})$ is the most cooperative equilibrium tariff vector of $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$.[28] Clearly, τ_{CS} (τ_{CR}) is the most cooperative equilibrium tariff of $\Gamma^s_\infty(SW)$ ($\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$) when $\delta = \delta^S_{\tau_{CS}}$ ($\delta = \delta^R_{\tau_{CR}}$). Moreover, we assume in the remainder of this section that $\delta \in [\underline{\delta}, \overline{\delta}]$ so that both self-interested and reciprocal countries can maintain some cooperation in trade policies but global free trade is infeasible for either of them. The next proposition compares τ_{CR} with τ_{CS} assuming the countries are moderately de-

[28] Note that the most cooperative equilibrium is the most natural focal point of either game since (i) it is the only equilibrium of the desired class that is not Pareto dominated; and (ii) nothing precludes preplay communication between the countries.

manding from their trading partners regarding their commercial policy (i.e., assuming the fair tariff is not too low).

Proposition 3 *Let $\tau_f \geq \tau_{CS}$. Then the most cooperative equilibrium tariff of $\Gamma^s_\infty(SW)$ is higher than the one of $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$, i.e., $\tau_{CS} > \tau_{CR}$.*

The intuition underlying Proposition 3 is the same as the one behind Proposition 2.

Finally, we compare τ_{CR} with τ_{CS} assuming now that the countries are highly demanding from their trading partners with respect to their import policy (i.e., assuming that only very liberal import policies are considered fair).

Proposition 4 *Let $\tau_f < \tau_{CS}$. Then the effect of fairness and reciprocity on the most cooperative tariff equilibrium of the dynamic game is ambiguous.*

To gain some insight for Proposition 4, recall that for any cooperative tariff τ_C higher than the fair tariff τ_f, the countries attach a negative weight to their partner's self-interested welfare, i.e., they are in a negative-reciprocity state. Two observations can then be readily made for any such $\tau_C > \tau_f$. On the one hand, for any country J, the value of cooperation at τ_C is higher in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ than in $\Gamma^S_\infty(SW)$ since the punitive Nash tariff is higher in the former game. Of course, infinite Nash reversion would be costly also for country $-J$, which acts to lower the cost of the punishment phase for country J in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$, but this effect is relatively weak for a sufficiently small γ. On the other hand, country J has a stronger incentive to deviate from τ_C in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ than in $\Gamma^S_\infty(SW)$ since in the former game (i) the defect tariff is higher, because the countries are in a negative-reciprocity state;

and (ii) defection would hurt country $-J$, raising the gains from cheating for country J. Hence, it is ambiguous whether reciprocal or self-interested countries can more easily sustain any given cooperative tariff above the fair tariff. As a result, when only very liberal trade policies are considered fair, the overall effect of reciprocity on multilateral tariff cooperation could be negative. This is more clearly illustrated in the next section within the context of a simplified model.[29]

At a more general level, Propositions 3 and 4 demonstrate that if, for whatever reason, countries become more demanding from their trading partners with respect to their import policy (i.e., if the fair tariff decreases), a given cooperative equilibrium that could have been otherwise supported, might no longer be feasible. This then suggests that if countries enter a round of multilateral trade negotiations with elevated expectations due to economic and/or political reasons, they might fail to reach an agreement on further multilateral trade liberalization, even though such an agreement might have been attainable in the absence of these high expectations. Therefore, Propositions 3 and 4 provide a novel perspective on the occasional failures of multilateral trade negotiations, as in the ongoing Doha Round.

[29] At this point, we should note that asymmetries in fair-tariff perceptions would not affect the qualitative nature of our findings as long as the countries remained *symmetrically* demanding from each other with respect to their trade policy. In particular, under asymmetric fair-tariff perceptions, Proposition 3 would still hold as long as $\tau_f^{-J} \geq \tau_{CS}^{-J}$ for any country J, whereas Proposition 4 would be still valid as long as $\tau_f^{-J} < \tau_{CS}^{-J}$ for all J.

2.5 Simplified Model

In this section, we reproduce the results of the paper within a simple setup with linear demand curves and a specific functional form for $w^J(\tau^{-J}, \tau_f)$. This serves two goals. First, it enables us to better illustrate the insights from our model. At the same time, the results obtained here can be more easily related to the ones found in a substantial part of the literature on trade agreements that uses similar demand specifications (e.g., Bagwell and Staiger, 1999b; Freund, 2000; Ornelas, 2005; Saggi and Yildiz, 2010). To this end, let the demand for good j in country J be given by:

$$D\left(P_j^J\right) = \alpha - \beta P_j^J, \tag{2.13}$$

where $\alpha > \frac{1}{2}, \beta > 0$ are constants. Moreover, let us assume that the weight function w^J is of the following form:

$$w^J(\tau^{-J}, \tau_f) = \frac{\tau_f - \tau^{-J}}{\tau_f + \tau^{-J}} \in (-1, 1)\,. \tag{2.14}$$

We first look at the static game, and in particular at $\Gamma^S(SW)$. It turns out the best-response tariff of country J equals:

$$BR_S^J(\tau^{-J}) = \frac{1}{3\beta}, \tag{2.15}$$

meaning both countries have the same dominant strategy. Then, trivially:

$$\tau_{NS} = \frac{1}{3\beta}. \tag{2.16}$$

However, the analysis for $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ is slightly more involved. The countries no longer have a dominant strategy. Rather:

$$BR_R^J(\tau^{-J}) = \frac{(1+\gamma)\tau^{-J} + (1-\gamma)\tau_f}{\beta\left[3(\tau^{-J}+\tau_f) + \gamma(\tau^{-J}-\tau_f)\right]}. \tag{2.17}$$

It is direct to show that $BR_R^J(\tau^{-J})$ is strictly decreasing in τ_f and strictly increasing in τ^{-J}. Simple algebra then yields:

$$\tau_{NR} = \frac{1+\gamma-\beta(3-\gamma)\tau_f + \sqrt{4\beta(1-\gamma)(3+\gamma)\tau_f + (1+\gamma-\beta(3-\gamma)\tau_f)^2}}{2\beta(3+\gamma)}. \tag{2.18}$$

Two conclusions can be drawn from equation (2.18). First, if $\tau_f = \tau_{NS} = \frac{1}{3\beta}$, then τ_{NR} collapses to τ_{NS}. Second, we have that $\frac{\partial \tau_{NR}}{\partial \tau_f} < 0$, implying that if $\tau_f < \tau_{NS}$, then $\tau_{NR} > \tau_{NS}$, which is in line with Proposition 1.

Next, we turn to the dynamic game. Straightforward calculations reveal that the most cooperative equilibrium tariff for $\Gamma_\infty^S(SW)$ equals:

$$\tau_{CS} = \frac{3-5\delta}{\beta(9-3\delta)}, \tag{2.19}$$

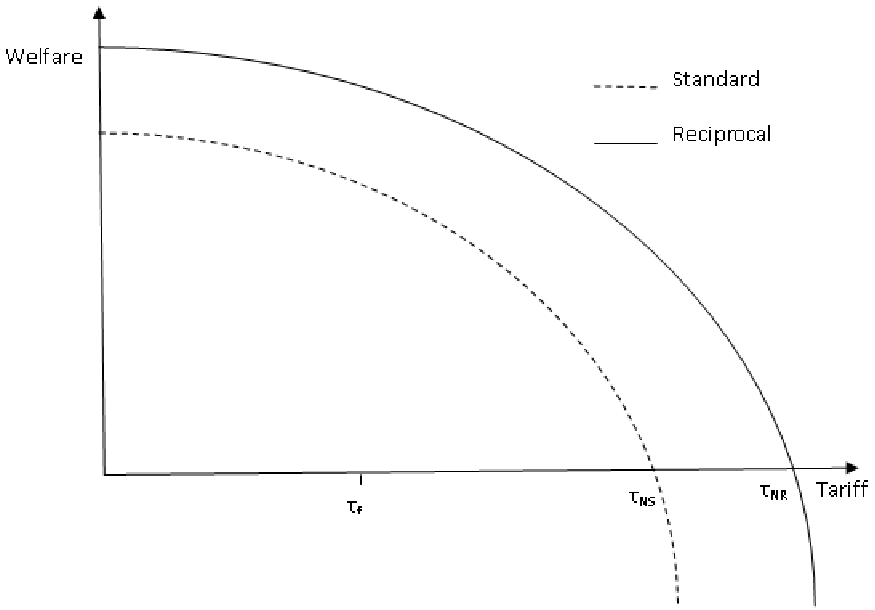
meaning that free trade could be supported by self-interested countries for $\delta \geq 3/5$.

In order to now compare τ_{CS} with the most cooperative equilibrium tariff of $\Gamma_\infty^R(RW, w, \overrightarrow{\tau}_f)$, τ_{CR}, we resort for expositional simplicity to graphical/numerical analysis, while maintaining the (nonrestrictive) assumptions that $\tau_f < \tau_{NS}$ and $\tau_C < \tau_{NS}$.[30] Let us consider first the per-period value of

[30] The analysis was carried out using Mathematica. The file is available from the authors upon request.

cooperation for country J in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ and $\Gamma^S_\infty(SW)$: ω^J_R and ω^J_S, respectively. Figure 2.1 depicts the relation between the two: $\omega^J_R > \omega^J_S$ for any $\tau_C \in (0, \tau_{NS})$. Intuitively, two forces are at work here. First, the punitive Nash tariff is higher in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$ than in $\Gamma^S_\infty(SW)$, i.e., $\tau_{NR} > \tau_{NS}$. Second, infinite Nash reversion would also be costly for J's partner. This could raise or lower the cost of the punishment phase for country J itself in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$, depending on whether the countries are in a positive- or a negative-reciprocity state, i.e., depending on whether τ_C is below or above τ_f. In any case, for sufficiently low γ, this effect is relatively weak.

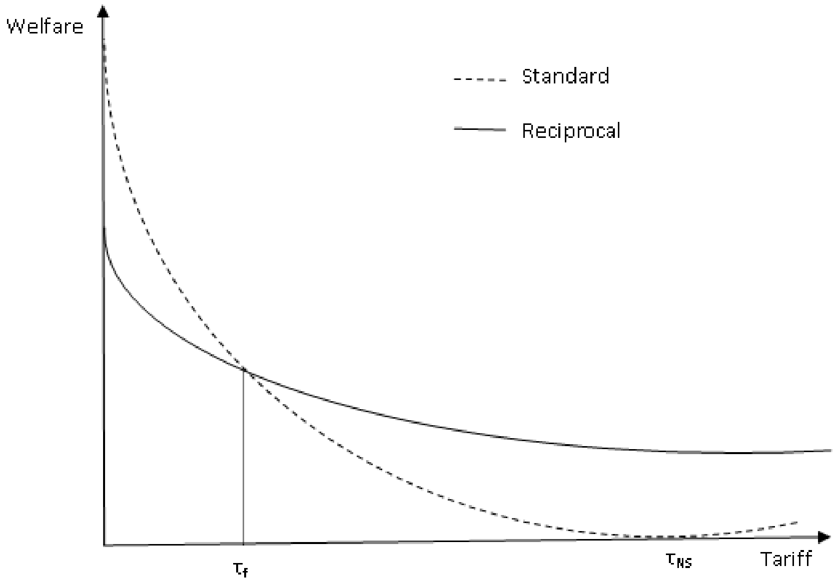
Figure 2.1: Incentive to Cooperate



We next examine the static incentive country J has to cheat in $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$

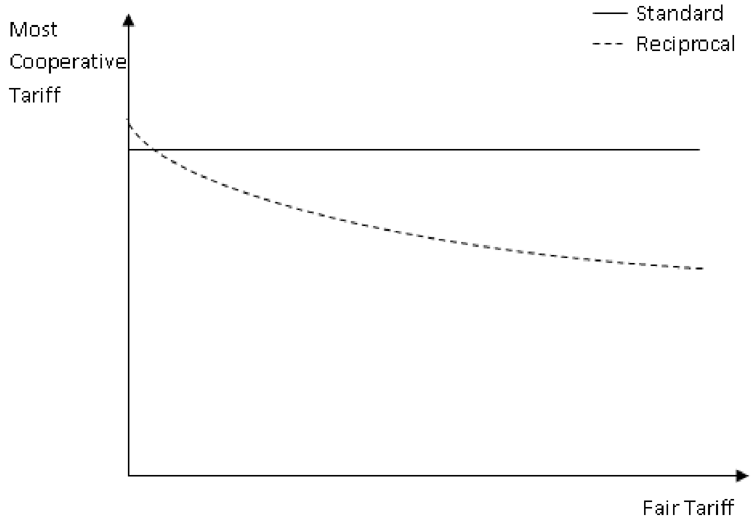
and $\Gamma_{\infty}^{S}(SW)$: Ω_R^J and Ω_S^J, correspondingly. As Figure 2.2 reveals, the former is weaker if and only if $\tau_C \in (0, \tau_f)$. The intuition is straightforward. For any given τ_C below τ_f, the countries are in a positive-reciprocity state. This has a dampening effect on the defect tariff. Moreover, defection would be costly for J's partner, which acts to mitigate the potential onetime gains from cheating for J in $\Gamma_{\infty}^{R}(RW, w, \overrightarrow{\tau}_f)$. As a result, $\Omega_R^J < \Omega_S^J$ for any such τ_C. Of course, the reverse is true for cooperative tariffs above the fair tariff, i.e., $\Omega_R^J > \Omega_S^J$ for $\tau_C > \tau_f$.

Figure 2.2: Incentive to Deviate



Therefore, for $\tau_f \in [\tau_{CS}, \tau_{NS})$, reciprocal countries have a stronger incentive to cooperate and a weaker incentive to defect than self-interested countries around τ_{CS}, implying that the former can support more liberal trade policies than the latter, or $\tau_{CR} < \tau_{CS}$. However, for $\tau_f < \tau_{CS}$, reciprocal countries have around τ_{CS} both a stronger incentive to cheat and a stronger incentive to cooperate than self-interested ones. In other words, there are two offsetting forces at play for low fair-tariff perceptions, making the comparison between τ_{CS} and τ_{CR} less clear-cut. Our simulations do confirm that for very low fair tariffs, τ_{CR} does indeed exceed τ_{CS}, as we depict in Figure 2.3. Actually, it is interesting to note that τ_{CR} is more likely to exceed τ_{CS} when δ is relatively low, i.e., when the countries are relatively impatient. This is due to the fact that a lower δ weakens the relative significance of the stronger-incentive-to-cooperate force, while it leaves the stronger-incentive-to-cheat force unaffected. To summarize, when countries are highly demanding from their trading partners regarding their commercial policy, reciprocity could have a detrimental effect on multilateral tariff cooperation, and this is more likely to occur if countries are relatively impatient.

Figure 2.3: Most Cooperative Tariff



2.6 Endogenizing Fairness

We have hitherto assumed that perceptions of fairness are exogenous and constant between periods. This is consistent with the experimental work of Fehr and Falk (1999) who in a wage-setting context find virtually no change in either behavior or perceptions of fairness over time. However, one could argue that countries' perceptions of a fair tariff might adjust during the course of the game. As Kahneman et al. (1986, pp.730-1) write: "Psychological studies of adaptation suggest that any stable state of affairs tends to become accepted eventually, at least in the sense that alternatives to it no longer readily come to mind. Terms of exchange that are initially seen as unfair may in time acquire the status of a reference transaction...they [people] adapt their views of fairness to the norms of actual behavior."[31] In this section, we extend our analysis by allowing for endogenous formation of fair-tariff perceptions, and investigate whether our main predictions so far still hold.

To this end, we adapt to our discrete-time framework an equation widely used in the habit formation literature (e.g., Ryder and Heal, 1973; Carroll et al., 2000; Fuhrer, 2000), assuming current tariffs affect future fair-tariff perceptions as follows:

$$\tau_f^{-J,t} = \alpha\tau^{-J,t-1} + (1-\alpha)\,\tau_f^{-J,t-1}, \text{ for any } J, \tag{2.20}$$

where $\alpha \in (0,1)$. Equation (2.20) indicates that country J's fair-tariff perception today is a linear combination of its last period's fair-tariff perception and of the tariff it actually faced in that period. In other words, if country $-J$ behaves kindly today by choosing an import tariff below the one country J deems fair, then country J will be more demanding next period (i.e., τ_f^{-J} decreases), which acts to lower its fairness payoff for any given τ^{-J}. On the other hand, if country $-J$ imposes an unfairly high tariff today, then country J will be less demanding next period (i.e., τ_f^{-J} increases), which acts to raise its fairness payoff given a τ^{-J}.[32] It follows that as the game unfolds, country J's fair-tariff perception converges to the tariff policy of its trading partner.

[31] See Franciosi et al. (1995) for experimental support of these ideas in a price-setting context.

[32] It is only reasonable to assume that as the multilateral trading environment becomes more liberal, countries become more demanding with respect to trade policy. For example, it is logical to expect that the tariffs deemed fair nowadays are substantially lower as compared with the ones in the late 1940s when the actual tariffs in place were significantly higher than the current ones.

More formally:

$$\left|\tau_f^{-J,t} - \tau^{-J,t}\right| \underset{t\to\infty}{\longrightarrow} 0. \tag{2.21}$$

We are now prepared to examine whether our main conclusions heretofore are affected in any fundamental way by (2.20). Let us make the assumption that α is not too big (i.e., countries do not adjust their reference levels too fast). Clearly, endogenizing fairness has no impact on the static game. At the same time, in the dynamic game, the major difference equation (2.20) introduces is that both τ_{NR} and τ_{CR} vary over time. However, given an α and an initial fair-tariff perception, we can readily derive the future fair-tariff perceptions, and hence τ_{NR}^t and τ_{CR}^t for all t.

It is direct to show that if the fair tariff initially exceeds the most cooperative equilibrium tariff of the self-interested game, then τ_{CR} remains below τ_{CS} along the equilibrium path, which is along the lines of Proposition 3. Intuitively, under this scenario, the countries start with a τ_{CR} below τ_{CS}. As the game progresses, τ_f converges to τ_{CR}, and thus, τ_{CR} converges to τ_{CS} (since w^J converges to zero). But for a sufficiently low α, τ_f never falls below τ_{CS}, implying τ_{CR} never exceeds τ_{CS} (by Proposition 3).

Moreover, a result analogous to Proposition 4 is obtained: If the fair tariff is initially smaller than τ_{CS}, then the effect of fairness and reciprocity on multilateral tariff cooperation is ambiguous, since the countries might start with a τ_{CR} either below or above τ_{CS}. Eventually though, under this scenario as well, the reciprocal game converges to the self-interested one (in infinite time). In summary, allowing for endogenously formed fair-tariff perceptions does not affect the qualitative nature of our findings (as long as α is not too high).

2.7 Conclusion

This paper has explored the impact of fairness and reciprocity on multilateral tariff cooperation. In particular, we examined whether reciprocal countries can sustain a greater degree of cooperation than self-interested ones in the context of self-enforcing multilateral tariff agreements. This is an important question given that governments and consumers seem to exhibit reciprocal preferences towards commercial policy. In our setting, a reciprocal country is willing to reward its trading partners by imposing lower tariffs if it expects them to behave kindly by setting their tariffs below the one it perceives as fair. However, when it expects its partners to behave unkindly by choosing tariffs in excess of the one it deems fair, it wishes to punish them and is willing to sacrifice some of its own self-interested welfare in order to do so.

We have demonstrated that as compared with self-interested countries, reciprocal ones that are moderately demanding from their trading partners regarding their commercial policy (i.e., when the commercial policies deemed fair are not too liberal) can support lower cooperative tariffs and therefore achieve higher welfare in an infinitely-repeated tariff game. Nevertheless, when countries are highly demanding from their partners with respect to their import policy (i.e., when only very liberal import policies are considered fair), reciprocity could have a detrimental effect on multilateral tariff cooperation. Our findings therefore provide a novel perspective on the successes and the occasional failures of multilateral trade negotiations, and suggest a plausible explanation for the problems plaguing the Doha Round since its initiation in 2001. In particular, countries might have entered the Doha negotiations with too high expectations (i.e., with very low far-tariff perceptions), hindering the efforts for further multilateral trade liberalization. Finally, we have argued that our results are robust to allowing for fair-tariff perceptions that are endogenously determined during the course of the game.

In concluding, a couple of remarks are in order. First, our framework here could be readily applied to other types of agreements among countries that are constrained to be self-enforcing, such as international environmental agreements. For instance, it would be interesting to investigate whether reciprocal countries can sustain a greater degree of cooperation in abatement standards than self-interested ones. Second, another promising avenue for future research would be to examine how regional trade agreements affect countries' fair-tariff perceptions and thus their ability to multilaterally cooperate. Given the unprecedented proliferation of such arrangements in recent years, this is a particularly important question.

2.8 Acknowledgements

We would like to thank Luis Santos-Pinto, and participants at CRETE 2009, ETSG 2009, and ASSET 2009 for very helpful comments and suggestions. A previous version of the paper was written while İriş was visiting UC San Diego, and Tabakis was an International Faculty Fellow at the MIT Sloan School of Management. Tabakis gratefully acknowledges financial support from the Fundação para a Ciência e a Tecnologia in Portugal. Any errors are ours.

Appendix 2.A Proofs

Proof of Lemma 1: We first consider $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$. Given that the number of countries is finite and that for any country J (i) Θ^J is a compact interval in $\mathcal{R}^+$; (ii) RW^J is twice continuously differentiable on Θ^J; and (iii) $\frac{\partial RW^J}{\partial \tau^J \partial \tau^{-J}} \geq 0$, we know from Theorem 4 in Milgrom and Roberts (1990) that $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ is a (smooth) supermodular game. It then follows from Theorem 5 in Milgrom and Roberts (1990) that (i) there exist largest and smallest serially undominated strategies for each country J, $\overline{\tau}^J$ and $\underline{\tau}^J$; and (ii) the strategy profiles $\overrightarrow{\overline{\tau}} \equiv \left(\overline{\tau}^J, \overline{\tau}^{-J}\right)$ and $\overrightarrow{\underline{\tau}} \equiv \left(\underline{\tau}^J, \underline{\tau}^{-J}\right)$ are pure Nash equilibrium profiles. Finally, given the overall symmetry of our model, we have that $\overline{\tau}^J_{NR} = \overline{\tau}^{-J}_{NR} \equiv \overline{\tau}_{NR}$ and $\underline{\tau}^J_{NR} = \underline{\tau}^{-J}_{NR} \equiv \underline{\tau}_{NR}$.

The second part of the lemma is straightforward once it is recalled that $\Gamma^S(SW)$ can be obtained from $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ by setting $\gamma = 0$, meaning that $\Gamma^S(SW)$ is also a (smooth) supermodular game. *Q.E.D.*

Proof of Lemma 2: Given that (i) $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$ is a supermodular game; and (ii) $\frac{\partial RW^J}{\partial \tau^J \partial \tau_f} \leq 0$ for any J, the lemma follows immediately from Theorem 6 in Milgrom and Roberts (1990). *Q.E.D.*

Proof of Proposition 1: If $\tau_f = \tau_{NS}$, then trivially $\tau_{NR} = \tau_{NS} \equiv \tau_N$, and for any J, $RW^J(\overrightarrow{\tau}_N, \tau_f) = SW^J(\overrightarrow{\tau}_N)$ since $w^J(\tau_N, \tau_f) = 0$ by (2.6). On the other hand, if $\tau_f < \tau_{NS}$, then $\tau_{NR} \geq \tau_{NS}$ by Lemma 2. These two inequalities imply $\tau_{NR} > \tau_f$, and thus for any J, $w^J(\tau_{NR}, \tau_f) < 0$ from (2.6). Moreover, for $\tau_{NR} \geq \tau_{NS}$, $SW^J(\overrightarrow{\tau}_{NR}) \leq SW^J(\overrightarrow{\tau}_{NS})$ for all J. But then it follows that for all J, $RW^J(\overrightarrow{\tau}_{NR}, \tau_f) < SW^J(\overrightarrow{\tau}_{NS})$. *Q.E.D.*

Proof of Lemma 3: If $\tau_C \leq \tau_f$, we have from (2.6) that for all J, $w^J(\tau_C, \tau_f) \geq 0$, implying:

$$RW^J(\overrightarrow{\tau}_C, \tau_f) \geq SW^J(\overrightarrow{\tau}_C). \tag{2.22}$$

In addition, we know that:

$$SW^J(\overrightarrow{\tau}_C) > SW^J(\overrightarrow{\tau}_{NS}). \tag{2.23}$$

Furthermore, we have from Proposition 1 that for any J:

$$RW^J(\overrightarrow{\tau}_{NR}, \tau_f) \leq SW^J(\overrightarrow{\tau}_{NS}). \tag{2.24}$$

From (2.22), (2.23), and (2.24), we then obtain for all J:

$$RW^J(\overrightarrow{\tau}_C, \tau_f) > RW^J(\overrightarrow{\tau}_{NR}, \tau_f),$$

which implies by Friedman (1971) that there exists a sufficiently high discount factor such that $\overrightarrow{\tau}_C$ is a subgame-perfect Nash equilibrium for $\Gamma^R_\infty(RW, w, \overrightarrow{\tau}_f)$. *Q.E.D.*

Proof of Proposition 2: We want to show that $\tau_C \leq \tau_f$ implies that $\delta^R_{\tau_C} = \frac{RW^J_D - RW^J_C}{RW^J_D - RW^J_N} < \frac{SW^J_D - SW^J_C}{SW^J_D - SW^J_N} = \delta^S_{\tau_C}$. To do so, we will prove:

(i) If $\tau_C \leq \tau_f \Rightarrow RW^J_D - RW^J_C \leq SW^J_D - SW^J_C$ for any J.

(ii) If $\tau_C \leq \tau_f \Rightarrow RW^J_D - RW^J_N > SW^J_D - SW^J_N$ for any J.

Let us start with (i). We have that for any J:

$$RW^J_C = SW^J(\overrightarrow{\tau}_C) + \gamma w^J(\tau_C, \tau_f) SW^{-J}(\overrightarrow{\tau}_C) \text{ and}$$

$$RW^J_D = SW^J(BR^J_R(\tau_C), \tau_C) + \gamma w^J(\tau_C, \tau_f) SW^{-J}(BR^J_R(\tau_C), \tau_C).$$

Therefore:

$$\begin{aligned} RW^J_D - RW^J_C &= SW^J(BR^J_R(\tau_C), \tau_C) - SW^J(\overrightarrow{\tau}_C) \\ &+ \gamma w^J(\tau_C, \tau_f)\left(SW^{-J}(BR^J_R(\tau_C), \tau_C) - SW^{-J}(\overrightarrow{\tau}_C)\right) \\ &\leq SW^J(BR^J_R(\tau_C), \tau_C) - SW^J(\overrightarrow{\tau}_C) \\ &\leq SW^J(BR^J_S(\tau_C), \tau_C) - SW^J(\overrightarrow{\tau}_C) = SW^J_D - SW^J_C. \end{aligned} \tag{2.25}$$

We know from (2.6) that $w^J(\tau_C, \tau_f) \geq 0$ if $\tau_C \leq \tau_f$. Furthermore, the welfare of self-interested country $-J$ is (weakly) lower when country J deviates while it still cooperates than when both countries cooperate, i.e., $SW^{-J}(BR^J_R(\tau_C), \tau_C) - SW^{-J}(\overrightarrow{\tau}_C) \leq 0$. The first inequality then follows. The second inequality stems from the fact that $BR^J_S(\tau_C)$ is the best reply of the self-interested country J. This concludes the proof of (i).

We now turn to (ii). Let us rewrite the result we want to show:

$$\tau_C \leq \tau_f \Rightarrow \left(RW_D^J - SW_D^J\right) - \left(RW_N^J - SW_N^J\right) > 0 \text{ for any } J.$$

By Proposition 1 we know that the Nash equilibrium tariff of $\Gamma^S(SW)$ is (weakly) smaller than that of $\Gamma^R(RW, w, \overrightarrow{\tau}_f)$, i.e., $\tau_{NR} \geq \tau_{NS}$. Thus, we have that $\tau_f \leq \tau_{NS} \leq \tau_{NR}$, implying that $w^J(\tau_{NR}, \tau_f) \leq 0$ by (2.6). Therefore, the following inequality holds for any J:

$$\begin{aligned} RW_N^J &= SW^J(\overrightarrow{\tau}_{NR}) + \gamma w^J(\tau_{NR}, \tau_f) SW^{-J}(\overrightarrow{\tau}_{NR}) \\ &\leq SW^J(\overrightarrow{\tau}_{NR}) \leq SW^J(\overrightarrow{\tau}_{NS}) = SW_N^J. \end{aligned} \tag{2.26}$$

Next we will show that $RW_D^J - SW_D^J \geq 0$ for any J. Remember that γ is assumed to be sufficiently small. Taking a first-order Taylor series expansion of $RW^J(BR_R^J(\tau_C), \tau_C, \tau_f)$ around $\gamma = 0$, we obtain:

$$\begin{gathered} RW^J(BR_R^J(\tau_C), \tau_C, \tau_f) \approx SW^J(BR_S^J(\tau_C), \tau_C) + \\ + \gamma w^J(\tau_C, \tau_f) SW^{-J}(BR_S^J(\tau_C), \tau_C) \Leftrightarrow \\ \Leftrightarrow RW^J(BR_R^J(\tau_C), \tau_C, \tau_f) - SW^J(BR_S^J(\tau_C), \tau_C) \approx \\ \approx \gamma w^J(\tau_C, \tau_f) SW^{-J}(BR_S^J(\tau_C), \tau_C) \geq 0 \Leftrightarrow \\ \Leftrightarrow RW_D^J - SW_D^J \geq 0. \end{gathered} \tag{2.27}$$

The inequality holds due to $w^J(\tau_C, \tau_f) \geq 0$. By assumption, we have that $\tau_C < \tau_{NS}$, and thus τ_f cannot be equal to both τ_C and τ_{NS} at the same time. Hence, at least one of the inequalities in (2.26) and (2.27) must be strict. This concludes the proof of part (ii). Therefore, by (i) and (ii), we finally have $\delta_{\tau_C}^R < \delta_{\tau_C}^S$. *Q.E.D.*

Proof of Proposition 3: From Proposition 2, we know that for any cooperative tariff $\tau_C \leq \tau_f$, $\delta_{\tau_C}^R < \delta_{\tau_C}^S$. So, given the assumptions of Proposition 3, this is also true for the most cooperative equilibrium tariff of the repeated game with self-interested countries, τ_{CS}: $\delta_{\tau_{CS}}^R < \delta_{\tau_{CS}}^S$. Note that both self-interested and reciprocal countries can sustain τ_{CS} at the discount factor $\delta_{\tau_{CS}}^S$, but only reciprocal countries can support τ_{CS} at $\delta_{\tau_{CS}}^R$. From (2.10) and

(2.12), we have:

$$\begin{aligned} SW_D^J - SW_C^J &= \delta_{\tau_{CS}}^S \left(SW_D^J - SW_N^J\right) \text{ and} \\ RW_D^J - RW_C^J &= \delta_{\tau_{CS}}^R \left(RW_D^J - RW_N^J\right). \end{aligned}$$

Since $\delta_{\tau_{CS}}^R < \delta_{\tau_{CS}}^S$:

$$RW_D^J - RW_C^J < \delta_{\tau_{CS}}^S \left(RW_D^J - RW_N^J\right) \Leftrightarrow$$

$$\Leftrightarrow \left(1 - \delta_{\tau_{CS}}^S\right) RW^J(BR_R^J(\tau_{CS}), \tau_{CS}, \tau_f) \tag{2.28}$$

$$< RW^J(\overrightarrow{\tau}_{CS}, \tau_f) - \delta_{\tau_{CS}}^S RW^J\left(\overrightarrow{\tau}_{NR}, \tau_f\right), \tag{2.29}$$

meaning that $\Omega_R^J(\tau_{CS}) < \frac{\delta_{\tau_{CS}}^S}{1-\delta_{\tau_{CS}}^S}\omega_R^J(\tau_{CS})$, or that the incentive-compatibility condition is not binding for a reciprocal country J at the pair $(\tau_{CS}, \delta_{\tau_{CS}}^S)$.

Note here that RW_N^J does not depend on the cooperative tariff. Moreover, for any cooperative tariff τ_C lower than the most cooperative equilibrium tariff of $\Gamma_\infty^s(SW)$, τ_{CS}, the welfare for reciprocal country J under defection from τ_C is higher than the welfare under deviation from τ_{CS}:

$$RW^J(BR_R^J(\tau_C), \tau_C, \tau_f) > RW^J(BR_R^J(\tau_{CS}), \tau_{CS}, \tau_f).$$

At the same time, for such a $\tau_C < \tau_{CS}$, country J's welfare under cooperation is also higher at τ_C than at τ_{CS}:

$$RW^J(\overrightarrow{\tau}_C, \tau_f) > RW^J(\overrightarrow{\tau}_{CS}, \tau_f).$$

By the continuity of $RW^J(\bullet)$, then there exists a cooperative tariff $\widehat{\tau}_C < \tau_{CS}$ such that (**??**) still holds, or $\Omega_R^J(\widehat{\tau}_C) < \frac{\delta_{\tau_{CS}}^S}{1-\delta_{\tau_{CS}}^S}\omega_R^J(\widehat{\tau}_C)$. Since the same analysis applies to any $(\tau_{CS}, \delta_{\tau_{CS}}^S)$ pair for $\delta_{\tau_{CS}}^S \in [\underline{\delta}, \overline{\delta}]$, we have that for any $\delta \in [\underline{\delta}, \overline{\delta}]$, $\tau_{CS} > \tau_{CR}$. *Q.E.D.*

Proof of Proposition 4: Proposition 2 holds for any cooperative tariff τ_C (weakly) lower than the fair tariff τ_f. However, for any $\tau_C > \tau_f$, it is ambiguous by (2.25) and (2.27) whether $\delta_{\tau_C}^R$ or $\delta_{\tau_C}^S$ is higher, since the weight function is negative at τ_C. Hence, it is possible that the minimum discount factor required for countries with reciprocal preferences to sustain cooperation at τ_C is higher than that for self-interested countries, i.e., $\delta_{\tau_C}^R > \delta_{\tau_C}^S$.

Let us consider this case first, and focus on the most cooperative equilibrium tariff of $\Gamma_{\infty}^{s}(SW)$, τ_{CS}. Under the scenario in question, both types of countries could sustain cooperation at τ_{CS} only with a level of discount factor equal to $\delta_{\tau_{CS}}^{R}$ or above. Moreover, let us make the assumption that $\Omega_{R}^{J}(\bullet)$ is a strictly convex function whereas $\omega_{R}^{J}(\bullet)$ is a strictly concave one.[33] From (2.12) and (2.10), we have:

$$\begin{aligned} RW_{D}^{J}-RW_{C}^{J} &= \delta_{\tau_{CS}}^{R}\left(RW_{D}^{J}-RW_{N}^{J}\right) \text{ and} \\ SW_{D}^{J}-SW_{C}^{J} &= \delta_{\tau_{CS}}^{S}\left(SW_{D}^{J}-SW_{N}^{J}\right). \end{aligned}$$

Since $\delta_{\tau_{CS}}^{R} > \delta_{\tau_{CS}}^{S}$:

$$\begin{aligned} RW_{D}^{J}-RW_{C}^{J} > \delta_{\tau_{CS}}^{S}\left(RW_{D}^{J}-RW_{N}^{J}\right) \Leftrightarrow \left(1-\delta_{\tau_{CS}}^{S}\right) RW^{J}(BR_{R}^{J}(\tau_{CS}),\tau_{CS,}\tau_{f}) \\ > RW^{J}(\overrightarrow{\tau}_{CS},\tau_{f})-\delta_{\tau_{CS}}^{S} RW^{J}\left(\overrightarrow{\tau}_{NR},\tau_{f}\right), \end{aligned}$$

meaning that $\Omega_{R}^{J}(\tau_{CS}) > \frac{\delta_{\tau_{CS}}^{S}}{1-\delta_{\tau_{CS}}^{S}}\omega_{R}^{J}(\tau_{CS})$, or that the incentive-compatibility condition is violated for a reciprocal country J at the pair $(\tau_{CS},\delta_{\tau_{CS}}^{S})$.

For any cooperative tariff τ_{C} higher (lower) than the most cooperative equilibrium tariff of $\Gamma_{\infty}^{s}(SW)$, τ_{CS}, the onetime gain for reciprocal country J under defection from τ_{C} is lower (higher) than the static gain under deviation from τ_{CS}:

$$\Omega_{R}^{J}(\tau_{C}) < (>)\Omega_{R}^{J}(\tau_{CS}).$$

At the same time, for such a $\tau_{C} > (<)\tau_{CS}$, country J's per-period gain from cooperation is also lower (higher) at τ_{C} than at τ_{CS}:

$$\omega_{R}^{J}(\tau_{C}) < (>)\omega_{R}^{J}(\tau_{CS}).$$

Given the strict convexity of $\Omega_{R}^{J}(\bullet)$ and the strict concavity of $\omega_{R}^{J}(\bullet)$, it follows that the incentive-compatibility condition for reciprocal country J can only be restored at a cooperative tariff $\widehat{\tau}_{C} > \tau_{CS}$. Since the same analysis applies to any $(\tau_{CS},\delta_{\tau_{CS}}^{S})$ pair for $\delta_{\tau_{CS}}^{S} \in [\underline{\delta},\overline{\delta}]$, we have that for any $\delta \in [\underline{\delta},\overline{\delta}]$, $\tau_{CS} < \tau_{CR}$.

Nevertheless, $\delta_{\tau_{CS}}^{R} < \delta_{\tau_{CS}}^{S}$ is also possible by (**??**) and (**??**). In this case, as we showed in the proof of Proposition 3, $\tau_{CS} > \tau_{CR}$. Therefore, when $\tau_{f} < \tau_{CS}$, it is ambiguous whether τ_{CR} or τ_{CS} is higher due to the ambiguity

[33] This assumption is clearly not restrictive given the type of result we are here after.

of whether $\delta^R_{\tau_{CS}}$ or $\delta^S_{\tau_{CS}}$ is higher. *Q.E.D.*

Chapter 3

Tacit Collusion under Fairness and Reciprocity

This paper departs from the standard profit-maximizing model of firm behavior by assuming that managers are motivated in part by personal animosity–or respect–towards their rivals. A reciprocal manager responds to unkind behavior of rivals with unkind actions (negative reciprocity), while at the same time, it responds to kind behavior of rivals with kind actions (positive reciprocity). We find that if fairness payoffs are small by comparison with monetary payoffs, then collusive action profiles (prices or quantities) are easier to sustain when firms have reciprocal managers. Thus, fairness concerns among firms with reciprocal managers can have adverse welfare consequences for consumers.

Keywords: Fairness; Reciprocity; Collusion; Repeated Games.
JEL Classification Numbers: D43, D63, L13, L21.

3.1 Introduction

The assumption that individuals behave as if maximizing their material payoffs, despite its central role in economic analysis, is at odds with a large body of evidence from psychology and from experimental economics. Economic agents often pursue objectives other than actual payoff maximization. Many observed departures from material payoff maximizing behavior arise through actions that favor fairness or reciprocity.

Fairness and reciprocity have been shown to explain behavior in bargaining games and in trust games. For example, in ultimatum games offers are usually much more generous than predicted by equilibrium and low offers are often rejected. These offers are consistent with an equilibrium in which proposers make offers knowing that responders may reject allocations that appear unfair.

The impact of fairness and reciprocity on market outcomes is an active area of research. Rabin (1993) and Rotemberg (forthcoming) show that fairness concerns by the part of consumers can improve consumer welfare. For example, Rabin (1993) finds that a monopolist ought to set price lower than “the monopoly price” if consumers have concerns about fairness.

In this paper we ask whether reciprocity may help to sustain collusive behavior. For instance, if a collusive agreement is seen by the parties as a fair outcome, then if one party reneges on the agreement and undercuts the price (or boosts its output), its rivals may be offended and hence punish the deviator aggressively (even at extra cost to themselves).

To perform this analysis we rule out fairness concerns by the part of consumers with respect to firms and vice-versa. This allows us to focus on the impact of fairness concerns among firms on collusive outcomes. The assumption that firms have fairness concerns and behave reciprocally towards their rivals finds support on experimental evidence where subjects play the role of firms.

In Lehman (2001), individuals placed in the role of a manager were asked to report satisfaction with various combinations of sales figures for their own firm, as well as for a competing firm. Attention to fairness was found to be a significant factor.

Huck et al. (2001) show that a Stackelberg leader finds it hard to exploit that advantage in experimental markets. The reason is that the Stackelberg follower acts more aggressively than predicted by the subgame perfect equilibrium of these market games. In fact, followers punish the leader by supplying a higher quantity than their most profitable response to the leader's quantity. This behavior is in line with the observed negative reciprocity of responders in the ultimatum game when the proposer tries to exploit his first-mover advantage.

Armstrong and Huck (2010) argue that sometimes managers are motivated in part by personal animosity–or respect–towards a rival. Thus, firms might punish rivals who behave "unfairly" towards them. For example, firms might sometimes care when their rivals obtain an "unfair" share of industry profits, for instance by cheating on a collusive agreement.

To model reciprocity we follow Segal and Sobel (2007) and assume that players in a strategic environment have preferences not only over the outcomes but also the strategies. A player's utility is additively separable in monetary and fairness payoffs. Monetary payoffs are revenues minus costs and fairness payoffs are a weighted average of the rivals' monetary payoffs where the weights depend on how the rivals' choices are expected to differ from the fair ones. If a player expects a rival to play a kind (mean) strategy, then he places a positive (negative) weight on that rival's monetary payoff. If a player expects a rival to play a fair strategy then he places zero weight on that rivals' monetary payoff.

We start by showing that reciprocity can lead to more or less competitive outcomes under static price competition. If players think that the fair prices of the rivals are at most the smallest equilibrium prices of the rivals in the game with self-interested players, then the equilibrium attained is a positive reciprocity state. In this case reciprocity leads to a less competitive outcome since equilibrium prices are higher with reciprocators than with self-interested players. This happens because in a positive reciprocity state, reciprocators want to reward the rivals for having set prices higher than the fair ones. In contrast, if players believe that the fair prices of the rivals are at least the largest equilibrium prices of the rivals in the game with self-interested players, then the equilibrium attained is a negative reciprocity state. In this case reciprocity leads to a more competitive outcome since equilibrium prices are lower with reciprocators than with self-interested players. This happens because in a negative reciprocity state, reciprocators want to punish the rivals for having set prices lower than the fair ones.

In a standard setting, collusion can be sustained as an equilibrium by self-interested players if they interact infinitely often and are sufficiently patient. A player is said to be patient if his discount factor is sufficiently close to one. In order to determine whether collusion is or is not facilitated by reciprocity we compare the minimal discount factor that allows for collusion in the case with reciprocity and in the standard case.

The main result of the paper shows that reciprocity facilitates collusion when players think that the fair prices of the rivals are at most the collusive prices and at least the largest equilibrium prices of the static game with self-interested players. The intuition as follows.

If players think that the fair prices of the rivals are at most the collusive prices, then collusion becomes a positive reciprocity state. In this case players' monetary payoffs from collusion are the same as the ones obtained in the game with self-interested players but in addition there are fairness payoff gains since players think that their rivals are being kind. This effect makes collusion *more* attractive to reciprocal players. Additionally, if players think that the fair prices of the rivals are at least the largest equilibrium prices of the static game with self-interested players, then Nash reversion becomes a negative reciprocity state. This implies that the punishment imposed after cheating occurs is more severe when players are reciprocal. This effect also makes collusion *more* attractive to reciprocal players. However, the short-run benefit to deviating is larger with reciprocal players because it includes the payoff a player derives from being treated kindly by the rivals (the rivals are playing their collusive prices). This effect makes collusion *less* attractive to reciprocal players. The assumption that monetary payoffs are large by comparison with fairness payoffs implies that the increase in collusive payoff is of first-order whereas the increase in the short-run benefit to deviating is of second-order.

We show that our main result also holds under quantity competition.

In this case, reciprocity facilitates collusion when players think that the fair output of the rivals is at least the collusive output and at most the equilibrium output of the rivals in static game with self-interested players.

The analysis is mainly conducted assuming that players play Nash reversion punishments. In the appendix we extend the analysis to the case where players can play any credible punishment using penal codes à la Abreu (1988).

Our paper is an additional contribution to the literature on the factors that help or hinder collusion. It is now well known that concentration, barriers to entry, cross-ownership, symmetry and multi-market contracts facilitate collusion–see Feuerstein (2005). We find that reciprocity by the part of firms can facilitate collusion.

The main policy implication of our paper is that fairness concerns by firms with reciprocal managers can have adverse welfare consequences for consumers. In contrast, Rabin (1993) and Rotemberg (forthcoming) find that fairness concerns by the part of consumers can increase consumer welfare. Thus, social preferences in imperfectly competitive markets might lead to different outcomes depending on who has such preferences (producers or consumers) and what is the comparison group.

The rest of the paper proceeds as follows. Section 2 sets-up the model. Section 3 analyzes the impact that fairness and reciprocity have on incentives for collusion when action choices are strategic complements. Section 4 considers the case of strategic substitutes. Section 5 discusses the findings. Section 6 concludes the paper. Appendix A contains the proofs of all results in the main text. Appendix B states and proves results when players use optimal punishments.

3.2 Set-up

The existing theories of social preferences can be classified into three broad categories. The first one is the distributional preference approach where social preferences only depend on the distribution of material payoffs. This includes Fehr and Schmidt (1999) and Bolton and Ockenfels (2000). These models are highly tractable and capture a wide range of phenomena but fail to explain the fact that preferences depend on more than outcomes, namely, intentions also matter.

The second category consists of intention-based models and includes Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006), among others. These models assume that reciprocity depends on overall strategies and beliefs (and beliefs about beliefs) building on Geanakoplos et al. (1989) theory of psychological games. In Rabin (1993) utility is additively separable in monetary and fairness payoffs and the weight a player places on rivals' monetary payoffs depends on his perception of the rivals' intentions, which are evaluated using (i) beliefs about the rivals' strategy choices, and (ii) beliefs about the rivals' beliefs about his strategy. Dufwenberg and Kirchsteiger (2004) develop a theory of reciprocity for extensive form games where players update beliefs about intentions as the game unfolds and make a choice accordingly. Falk and Fischbacher (2006) model reciprocity in incomplete information games. Intention-based models have two major weaknesses: they use specific functional forms and are highly intractable.[1]

[1] Sobel (2005) points out some of the drawbacks of the distributional-preferences and intention-based approaches to reciprocity.

The third category explores the axiomatic foundations that generate utility functions that display social preferences. Nielson (2006) proposes a preference axiom which leads to a foundation of Fehr and Schmidt (1999) inequity aversion model. Segal and Sobel (2007) provide an axiomatic foundation for interdependent preferences that can reflect reciprocity, inequity aversion, altruism as well as spitefulness. The key innovation of their approach is that, in addition to conventional preferences over outcomes, players in a strategic environment also have preferences over strategy profiles. This allows one to study situations where a player's preference is affected by the behavior of other players.

Their representation theorem shows that the payoff function of a player with such preferences is of the form

$$V_i(\sigma_i, \sigma^*_{-i}) = u_i(\sigma_i, \sigma^*_{-i}) + \sum_{j \neq i} w_{ij}(\sigma^*) u_j(\sigma_i, \sigma^*_{-i}), \tag{3.1}$$

where σ_i is the strategy of player i, σ^* is how the game is expected to be played, u_i is the utility from outcomes of player i, u_j is the utility from outcomes of player $j \neq i$, and $w_{ij}(\sigma^*)$ is a coefficient that measures the weight player i gives to player j's utility, which is a function of the entire strategy profile. Positive values of the coefficient mean that player i is willing to sacrifice his utility from outcomes in order to increase the payoff of player j. Negative values mean that player i is willing to sacrifice his utility from outcomes in order to lower player j's payoff. Since the coefficient depends on the strategy chosen by player j, there is scope to model reciprocity.[2]

[2] The underlying preferences in (3.1) are defined over outcomes. If an outcome specifies a material payoff to each player, then it is permissible for u_i to depend on other players'

We apply Segal and Sobel's (2007) approach to a dynamic game where n players, $n > 2$, play the same stage game over an infinite horizon $t = 0, 1, 2, \ldots$. The repeated game monetary payoff of player i of choosing strategy $s_i = (a_i^1, a_i^2, ...)$ when rivals play strategies s_{-i} is given by

$$\Pi_i(s_i, s_{-i}) = \sum_{t=1}^{\infty} \delta^{t-1} \pi_i(a_i^t, a_{-i}^t), \tag{3.2}$$

where $\pi_i(a_i^t, a_{-i}^t)$ represents player i's monetary payoff at stage t, a function of player i's action at t, a_i^t, and the actions of the rivals at t, a_{-i}^t. Players discount the future at rate $\delta \in (0,1)$. To model reciprocity we assume that the weight player i places on player j's repeated game monetary payoff depends only on player j's strategy and on player i's perception of what is the fair strategy of player j, s_{ij}^f. We also assume throughout that players' preferences as well as their exogenous perceptions of the fair strategies of the rivals are common knowledge. The repeated game payoff of reciprocal player i of choosing strategy $s_i = (a_i^1, a_i^2, ...)$ when rivals play strategies s_{-i} is given by

$$U_i(s_i, s_{-i}, s_{-i}^f) = \sum_{t=1}^{\infty} \delta^{t-1} \pi_i(a_i^t, a_{-i}^t) + \alpha \sum_{j \neq i} \sum_{t=1}^{\infty} \delta^{t-1} w_{ij}(a_j^t, a_{ij}^f) \pi_j(a_i^t, a_{-i}^t) \tag{3.3}$$

where $\alpha > 0$ is a normalization. The central behavioral feature of these preferences is the assumption that players care about the intentions of the rivals. If player i expects player j to treat him kindly, then w_{ij} will be positive, and player i will wish to treat player j kindly. If player i expects

material payoffs. Thus, this approach also generalizes the inequity aversion approach.

player j to treat him badly, then w_{ij} will be negative, and player i will wish to treat player j badly. If player i expects player j to be fair, then w_{ij} will be zero, and there is no issue of reciprocity.

Denote the dynamic game with reciprocal players by $\Gamma^r_\infty(u, s)$, where $u = (u_1, ..., u_n)$ and $s = (s_1, ..., s_n)$ and the dynamic game with self-interested players by $\Gamma^s_\infty(\pi, s)$, where $\pi = (\pi_1, ..., \pi_n)$. Players are able to sustain a collusive outcome when the payoff from collusion is no less than the payoff from deviation. To understand how fairness and reciprocity influence collusion we will compare the incentive compatibility condition of self-interested players in $\Gamma^s_\infty(\pi, s)$ to that of reciprocal players in $\Gamma^r_\infty(u, s)$ assuming that these two games are identical in all respects (monetary payoffs and the number of players) with the exception of players' preferences.

To perform this analysis we consider the cases where players' actions are strategic complements (e.g., price competition with products that are imperfect substitutes) and strategic substitutes (e.g., quantity competition with products that are perfect substitutes). We also consider two alternative modes of punishments after deviations: Nash reversion and optimal punishments.

The standard approach to study collusion in infinitely repeated games assumes that players use grim trigger strategies to punish any deviation from collusion, that is, following a deviation players switch to a Nash equilibrium of the stage game forever after. Thus, when self-interested player uses grim trigger punishments in $\Gamma^s_\infty(\pi, s)$, each player i will prefer to play his collusive strategy $s^c_i = (a^c_i, a^c_i, ...)$ if the payoff from collusion, $\pi_i(a^c)/(1 - \delta)$, is no less than the payoff from defection which consists of the one period gain from deviating $\pi_i(BR_i^s(a_{-i}^c), a_{-i}^c)$ plus the discounted payoff of inducing Nash reversion forever $\delta\pi_i(a^{ns})/(1-\delta)$, that is,

$$\pi_i(BR_i^s(a_{-i}^c), a_{-i}^c) + \frac{\delta}{1-\delta}\pi_i(a^{ns}) \leq \frac{1}{1-\delta}\pi_i(a^c). \tag{3.4}$$

Solving for δ we obtain

$$\delta_{a^c}^s = \frac{\pi_i(BR_i^s(a_{-i}^c), a_{-i}^c) - \pi_i(a^c)}{\pi_i(BR_i^s(a_{-i}^c), a_{-i}^c) - \pi_i(a^{ns})} \leq \delta. \tag{3.5}$$

The collusion strategy profile s^c can be sustained by self-interested players who are patient enough such that $\delta_{a^c}^s \leq \delta$ where $\delta_{a^c}^s$ is the critical discount factor above which s^c can be sustained by self-interested players.

The same reasoning applies when players have reciprocal preferences. A reciprocal player i plays the collusive strategy s_i^c in $\Gamma_\infty^r(u, x)$ using a grim trigger strategy as long as the following condition holds

$$u_i(BR_i^r(a_{-i}^c), a_{-i}^c, a_{-i}^f) + \frac{\delta}{1-\delta}u_i(a^{nr}, a_{-i}^f) \leq \frac{1}{1-\delta}u_i(a^c, a_{-i}^f), \tag{3.6}$$

where u_i denotes the stage game payoff of a reciprocal player, a function of the actions played and perceptions of the fair actions of the rivals. Solving for δ we obtain

$$\delta_{a^c}^r = \frac{u_i(BR_i^r(a_{-i}^c), a_{-i}^c, a_{-i}^f) - u_i(a^c, a_{-i}^f)}{u_i(BR_i^r(a_{-i}^c), a_{-i}^c, a_{-i}^f) - u_i(a^{nr}, a_{-i}^f)} \leq \delta. \tag{3.7}$$

When players have reciprocal preferences it follows that the collusive strategy profile s^c can be sustained if players are patient enough such that $\delta_{a^c}^r \leq \delta$

where $\delta^r_{a^c}$ is the critical discount factor above which s^c can be sustained by reciprocal players.

We will use (3.5) and (3.7) to characterize the impact that fairness and reciprocity have on collusion when players use grim trigger strategies. To perform this analysis we compare the critical discount factor above which the collusive strategy profile can be sustained when players are self-interested to the critical discount factor when players are reciprocal. We say that fairness and reciprocity facilitate collusion when the collusive strategy profile can be sustained at a lower critical discount factor when players are reciprocal than when they are self-interested. If the opposite happens we say that fairness and reciprocity make collusion harder.

3.3 Strategic Complements

We now specialize the model by assuming that players' actions are strategic complements. This assumption means that a player's incremental returns from increasing his own action are increasing in the rivals' actions. The canonical market game where players' actions are strategic complements is price competition with imperfect substitutes. We use this game to study the impact of fairness and reciprocity on collusion when players' actions are strategic complements.

In each stage player i chooses price, p_i, and his payoff in that stage is

$$u_i(p_i, p_{-i}, p^f_{-i}) = \pi_i(p_i, p_{-i}) + \alpha \sum_{j \neq i} w_{ij}(p_j, p^f_{ij})\pi_j(p_i, p_{-i}), \qquad (3.8)$$

where $\pi_i(p_i, p_{-i})$ is the monetary payoff and $\alpha \sum_{j \neq i} w_{ij}(p_j, p_{ij}^f)\pi_j(p_i, p_{-i})$ the fairness payoff. The monetary payoff is the difference between revenue and cost, that is,

$$\begin{aligned} \pi_i(p_i, p_{-i}) &= R_i(p_i, p_{-i}) - C_i(D_i(p_i, p_{-i})) \\ &= p_i D_i(p_i, p_{-i}) - C_i(D_i(p_i, p_{-i})), \end{aligned} \tag{3.9}$$

where $R_i(p_i, p_{-i})$ is revenue, $C_i(D_i(\cdot))$ is the cost of production, and $D_i(p_i, p_{-i})$ is the demand faced by player i. We assume that $D_i(\cdot)$ is decreasing with p_i, increasing with p_{-i}, and $C_i(\cdot)$ is increasing with $D_i(\cdot)$. Furthermore, we assume that

$$w_{ij}(p_j, p_{ij}^f) \begin{cases} > 0 \text{ if } p_j > p_{ij}^f \\ = 0 \text{ if } p_j = p^f \\ < 0 \text{ otherwise} \end{cases} . \tag{3.10}$$

The assumptions on $w_{ij}(p_j, p_{ij}^f)$ capture the fact that a reciprocal player cares about the intentions of the rivals. The first condition expresses positive or constructive reciprocity. If a player expects one of her rivals to charge a price higher than the fair price, then she puts a positive weight on that rival's profit and she is willing to sacrifice some of her profit to increase that rival's profit. The second condition says that if a player expects one of her rivals to choose the fair price, then she places no weight on that rival's profit. The third condition expresses negative or destructive reciprocity. If player a expects one of her rivals to undercut her perception of fair price, then she puts a negative weight on that rival's profit and she is willing to sacrifice some of her profit to reduce that rival's profit.

Let

$$A_i(p_i, p_{-i}, p_{-i}^f) = \arg\max_{p_i \in P_i} \pi_i(p_i, p_{-i}) + \alpha \sum_{j \neq i} w_{ij}(p_j, p_{ij}^f)\pi_j(p_i, p_{-i}), \quad (3.11)$$

denote the set of maximizers of player i's stage game problem as a function of p_i, p_{-i} and p_{-i}^f. For finite quantities, the players will never choose an infinite price. Hence, the players' price choice set is compact set in $\mathcal{R}$. We assume that u_i is order upper semi-continuous in p_i. The choice set being compact with this assumption guarantees that the set of maximizers $A_i(p_i, p_{-i}, p_{-i}^f)$ is nonempty.

We also assume that u_i has increasing differences in (p_i, p_{-i}), that is, for any fixed p_{-i}^f, $u_i(p_i, p'_{-i}, p_{-i}^f) - u_i(p_i, p''_{-i}, p_{-i}^f)$ is increasing in p_i for all $p'_{-i} \geq p''_{-i}$. This assumption implies that fairness payoffs are small by comparison with monetary payoffs which guarantees that prices are strategic complements.[3] Lemma 0, stated and proved formally in Appendix A, shows that these assumptions imply that $\Gamma^r(u,p)$ is a supermodular game. By Milgrom and Roberts (1990) we know that if $\Gamma^r(u,p)$ is a supermodular game, then there exist largest and smallest serially undominated strategies for each

[3] If the payoff function is differentiable, then u_i having increasing differences in (p_i, p_{-i}), is equivalent to the assumption that the cross partial derivatives of u_i with respect to p_i and p_j for any player j, is non-negative, that is,

$$\frac{\partial^2 u_i}{\partial^2 p_i p_j} = \underbrace{\frac{\partial^2 \pi_i}{\partial^2 p_i p_j}}_{\geq 0} + \alpha w_{ij}(p_j, p_{ij}^f) \underbrace{\frac{\partial^2 \pi_j}{\partial^2 p_i p_j}}_{\geq 0} + \alpha \frac{\partial w_{ij}(p_j, p_{ij}^f)}{\partial p_j} \frac{\partial \pi_j}{\partial p_i} \geq 0.$$

Note that if a player cares only about monetary payoffs and if the payoff function is differentiable, then the increasing differences assumption boils down to $\frac{\partial^2 \pi_i}{\partial^2 p_i p_j} > 0$. In the game with reciprocal players and differentiable payoff functions, the assumption will be satisfied if $\frac{\partial^2 \pi_i}{\partial^2 p_i p_j} > 0$ and α is sufficiently small.

player, $\overline{p}_i$ and $\underline{p_i}$. Moreover, the strategy profiles $\underline{p}$ and $\overline{p}$ are pure-strategy Nash equilibrium profiles. Thus, the existence of a Nash equilibrium of the stage game is guaranteed. Let us further assume that u_i has decreasing differences in (p_i, p_{-i}^f). The following result shows how players' perceptions of the fair prices of the rivals influence the extremal equilibrium prices of this game.

Lemma 1: *If* $\Gamma^r(u,p)$ *his a supermodular game and* u_i *has decreasing differences in* (p_i, p_{-i}^f) *then, the smallest and the largest pure-strategy Nash equilibria of* $\Gamma^r(u,p)$*, i.e.,* $\underline{p}^{nr}$ *and* $\overline{p}^{nr}$*, are nonincreasing functions of* $p^f = (p_{-1}^f, ..., p_{-n}^f)$.

Lemma 1 is a comparative statics result that characterizes the impact that players' perceptions of the fair prices of their rivals have on the Nash equilibrium prices of the stage game. This result says that the higher are players' perceptions of what the fair prices of the rivals should be, the lower will the equilibrium prices be. This happens because an increase in p_{-i}^f shifts the best reply of a reciprocal player i towards origin. In other words, the higher player i perceives the fair price for the other players to be, the more he would like to set a smaller price for any price of the other players. The critical assumption that drives this result is that u_i has decreasing differences in (p_i, p_{-i}^f), that is, the marginal returns from increasing prices are decreasing with a player's perception of the fair prices of the rivals.[4]

Next we show how preferences for fairness and reciprocity change the outcome of static price competition. To do that we compare the Nash equilibria

[4]If u_i is differentiable this assumption is equivalent to $\frac{\partial w_{ij}(p_j, p_{ij}^f)}{\partial p_{ij}^f} \frac{\partial \pi_j}{\partial p_i} < 0$ for all j.

of the stage game with self-interested players to that of the stage game with reciprocal players. The findings are summarized in Proposition 1.

Proposition 1:

(i) If reciprocal players believe that the fair prices of the rivals are at least the largest equilibrium prices of the rivals in $\Gamma^s(\pi, p)$, that is, $p^f_{-i} \geq \overline{p}^{ns}_{-i}$, for all i, then (a) $\overline{p}^{nr} \leq \overline{p}^{ns}$, and $u_i(\overline{p}^{nr}, p^f_{-i}) \leq \pi_i(\overline{p}^{ns})$, and (b) $\underline{p}^{nr} \leq \underline{p}^{ns}$ and $u_i(\underline{p}^{nr}, p^f_{-i}) \leq \pi_i(\underline{p}^{ns})$.

(ii) If reciprocal players think that the fair prices of the rivals are at most the smallest equilibrium prices of the rivals in $\Gamma^s(\pi, p)$, that is, $p^f_{-i} \leq \underline{p}^{ns}_{-i}$, for all i, then (c) $\overline{p}^{nr} \geq \overline{p}^{ns}$, and $u_i(\overline{p}^{nr}, p^f_{-i}) \geq \pi_i(\overline{p}^{ns})$, and (d) $\underline{p}^{nr} \geq \underline{p}^{ns}$ and $u_i(\underline{p}^{nr}, p^f_{-i}) \geq \pi_i(\underline{p}^{ns})$.

This result tells us how fairness and reciprocity change the nature of static price competition. Part (i) tells us that if reciprocal players believe that the fair prices of the rivals are at least the largest equilibrium prices of the rivals in the game with self-interested players, then prices set by reciprocators will be lower than those set by self-interested players. In this case, fairness and reciprocity lead to a more competitive outcome. In contrast, part (ii) tells us that if reciprocal players think that the fair prices of the rivals are at most the smallest equilibrium prices of the rivals in the game with self-interested players, then prices set by reciprocators will be higher than those set by self-interested players. In this case, fairness and reciprocity lead to a less competitive outcome.

The intuition behind Proposition 1 is as follows. When reciprocal players believe that the fair prices of the rivals are at least the largest equilibrium prices of the rivals in the game with self-interested players, the smallest and the largest Nash equilibria of the game with reciprocal players are negative reciprocity states: reciprocal players expect their rivals to set unfair prices. This implies that reciprocal players wish to punish their rivals. They do it by setting a price lower than the price a self-interested player would set. The lower equilibrium prices reduce players' monetary payoffs and in addition lead to payoff losses due to the unkind behavior of the rivals. In contrast, when reciprocal players think that the fair prices of the rivals are at most the smallest equilibrium prices of the rivals in the game with self-interested players, the smallest and the largest Nash equilibria of the game with reciprocal players are positive reciprocity states: reciprocal players expect their rivals to set kind prices. This implies that reciprocal players wish to reward their rivals. They do it by setting a higher price than the price a self-interested player would set. The higher equilibrium prices increase players' monetary payoffs and in addition lead to payoff gains due to the kind behavior of the rivals.

We now turn our attention to how fairness and reciprocity change the nature of dynamic price competition. The repeated game payoff of strategy $p_i = (p_i^1, p_i^2, ...)$ when rivals play strategies p_{-i} is given by

$$U_i(p_i, p_{-i}, p_{-i}^f) = \sum_{t=1}^{\infty} \delta^{t-1} \pi_i(p_i^t, p_{-i}^t) + \alpha \sum_{j \neq i} \sum_{t=1}^{\infty} \delta^{t-1} w_{ij}(p_j^t, p_{ij}^f) \pi_j(p_i^t, p_{-i}^t) \tag{3.12}$$

When players use stationary strategies the repeated game payoff becomes

$$\begin{aligned} U_i(p_i, p_{-i}, p^f_{-i}) &= \sum_{t=1}^{\infty} \delta^{t-1} \pi_i(p_i, p_{-i}) + \alpha \sum_{j \neq i} w_{ij}(p_j, p^f_{ij}) \sum_{t=1}^{\infty} \delta^{t-1} \pi_j(p_i, p_{-i}) \\ &= \frac{1}{1-\delta} \left[\pi_i(p_i, p_{-i}) + \alpha \sum_{j \neq i} w_{ij}(p_j, p^f_{ij}) \pi_j(p_i, p_{-i}) \right] \\ &= \frac{1}{1-\delta} u_i(p, p_{-i}, p^f_{-i}). \end{aligned} \tag{3.13}$$

For the dynamic game with self-interested players, $\Gamma^s_\infty(\pi, p)$, we know from Friedman (1971) that for a sufficiently high discount factor, there is a subgame-perfect Nash equilibrium of $\Gamma^s_\infty(\pi, p)$ at a collusive price p^c with payoff $\pi(p^c)$ where $\pi(p^c)$ is any payoff which gives every player strictly more than the payoff of the largest Nash equilibrium of $\Gamma^s(\pi, p)$, that is, $\pi_i(p^c) > \pi_i(\overline{p}^{ns})$, for all i. Lemma 2 applies this result to the dynamic game with reciprocal players, $\Gamma^r_\infty(u, p)$.

Lemma 2: *If $p^f_{ij} \in [\overline{p}^{ns}_j, p^c_j]$ for all i and $j \neq i$, then there is a sufficiently high discount factor such that there exists a subgame-perfect Nash equilibrium of $\Gamma^r_\infty(u, p)$ at p^c.*

This result states that given the fair prices profile, p^f, for any p^c such that the players' payoffs at the collusive prices are higher than their payoffs at the largest Nash equilibrium of the stage game, collusion can be sustained by reciprocal players at the strategy profile p^c. From now on we assume that Nash punishments in $\Gamma^r_\infty(u, p)$ and in $\Gamma^s_\infty(\pi, p)$ are either at the smallest or largest pure strategy Nash equilibria of $\Gamma^r(u, p)$ and $\Gamma^s(\pi, p)$, respectively.

Our next result shows that reciprocity facilitates collusion when there is price competition, marginal costs are constant, and players think that fair prices of the rivals are at least the largest Nash prices of the stage game with self-interested players and at most the collusive prices.

Proposition 2: *Let* $\pi_i(p_i, p_{-i}) = (p_i - c_i)D_i(p_i, p_{-i})$, *and* $p^f_{ij} \in [\overline{p}^{ns}_j, p^c_j]$ *for all* i *and* $j \neq i$. *Then the critical (minimum) discount factor needed to sustain collusion at* p^c *is lower in* $\Gamma^r_\infty(u, p)$ *than in* $\Gamma^s_\infty(\pi, p)$, *that is,* $\delta^r_{p^c} < \delta^s_{p^c}$.

The intuition for this result is as follows. If players think that the fair prices of the rivals are less than the collusive prices, then collusion becomes a positive reciprocity state. In this case players' monetary payoffs from collusion are the same as the ones obtained in the game with self-interested players but in addition there are fairness payoff gains since players think that their rivals are being kind. This effect makes collusion more attractive when players are reciprocal than when they are self-interested.

Additionally, if players think that the fair prices of the rivals are greater than the largest Nash equilibrium prices of the stage game with self-interested players, then Nash reversion becomes a negative reciprocity state. This implies that the punishment imposed after cheating occurs is more severe when players are reciprocal than when they are self-interested. This happens because monetary payoffs are lower than the payoffs of self-interested players and in addition there are fairness payoff loses since players think that the rivals are being unkind. This effect also makes collusion more attractive when players are reciprocal than when they are self-interested.

Clearly, these two effects make collusion *more* attractive to reciprocal players than to self-interested ones. However, the unilateral single period deviation payoff is higher with reciprocal players than with self-interested ones. This happens because the unilateral single period deviation payoff of a reciprocal player also includes the benefit that player derives from being treated kindly by the rivals (the rivals are playing their collusive prices). This effect of fairness and reciprocity makes collusion *less* attractive to reciprocal players than to self-interested ones. The assumption that monetary payoffs are large by comparison with fairness payoffs implies that the increase in collusive payoff is of first-order whereas the increase in the unilateral single period deviation payoff is of second-order.

3.4 Strategic Substitutes

We now show that fairness and reciprocity facilitate collusion in dynamic quantity-setting games with grim trigger punishments. Thus, when players are reciprocal, collusive action profiles are easier to sustain not only when players' actions are strategic complements but also when they are strategic substitutes.

When players' actions are strategic substitutes a player's incremental returns from increasing his own action are decreasing in the rivals' actions. The canonical market game where players' actions are strategic substitutes is quantity competition with products that are perfect substitutes. We use this game to study the impact of fairness and reciprocity on collusion when players' actions are strategic substitutes.

Assume that in each period player i chooses quantity, q_i, and his payoff in that period is given by

$$u_i(q_i, Q_{-i}) = \pi_i(q_i, Q_{-i}) + \alpha w_i(Q_{-i}, Q^f_{-i}) \sum_{j \neq i} \pi_j(q_i, Q_{-i}), \tag{3.14}$$

where $\pi_i(q_i, Q_{-i})$ is the monetary payoff and $\alpha w_i(Q_{-i}, Q^f_{-i}) \sum_{j \neq i} \pi_j(q_i, Q_{-i})$ is the fairness payoff, with $\alpha > 0$. Player i's monetary payoff, $\pi_i(q_i, q_{-i})$, is the difference between revenue and cost, that is,

$$\begin{aligned} \pi_i(q_i, Q_{-i}) &= R_i\left(q_i, Q_{-i}\right) - C_i(q_i) \\ &= P(Q)q_i - C_i(q_i), \end{aligned} \tag{3.15}$$

where $R_i(q_i, Q_{-i})$ is revenue, $C_i(q_i)$ is the cost of production, and $P(Q)$ is the inverse market demand with $Q = \sum q_i$. We assume that $P(Q)$ is strictly positive on some bounded interval $(0, \bar{Q})$ with $P(Q) = 0$ for $Q \geq \bar{Q}$. We also assume that $P(Q)$ is twice continuously differentiable with $P'(Q) < 0$ (in the interval for which $P(Q) > 0$). Players' costs of production are assumed to be twice continuously differentiable with $C_i'(q_i) \geq 0$. It is also assumed that the decreasing marginal revenue property holds, that is, $P'(Q) + P''\left(Q\right) q_i \leq 0$, and $P'(Q) - C_i''(q_i) < 0$. Furthermore, we assume that the weight that player i places on the rivals' aggregate monetary payoffs depends on player i's perception of the fair aggregate output of the rivals, Q^f_{-i}, and on the actual aggregate output of the rivals such that

$$w_i(Q_{-i}, Q^f_{-i}) \begin{cases} > 0 \text{ if } Q_{-i} < Q^f_{-i} \\ = 0 \text{ if } Q_{-i} = Q^f_{-i} \\ < 0 \text{ otherwise} \end{cases}, \tag{3.16}$$

where $w_i(Q_{-i}, Q^F_{-i})$ is assumed to be differentiable in both arguments with $\partial w_i/\partial Q_{-i} < 0$ and $\partial w_i/\partial Q^F_{-i} > 0$. The first condition in (3.16) expresses positive reciprocity. If a player expects her rivals to produce less than her perception of fair output, then she is willing to sacrifice some of her profit to increase the rivals' profits. The third condition in (3.16) expresses negative reciprocity. If a player expects her rivals to produce more than her perception of fair output, then she is willing to sacrifice some of her profit to reduce the rivals' profits.

Finally, we assume that monetary payoffs are large by comparison with fairness payoffs otherwise best replies of reciprocal players in a static Cournot oligopoly might no longer have a negative slope across all quantities.

Proposition 3: *If $\Gamma^r(u,p)$ and $\Gamma^s(\pi,p)$ satisfy the conditions stated and $Q^f_{-i} \in [Q^c_{-i}, Q^{ns}_{-i}]$ for all i, then the critical (minimum) discount factor needed to sustain collusion at q^c is lower in $\Gamma^r_\infty(u,q)$ than in $\Gamma^s_\infty(\pi,q)$, that is, $\delta^r_{q^c} < \delta^s_{q^c}$.*

Proposition 3 shows that fairness and reciprocity also facilitate collusion when players' choices are strategic substitutes. It says that if players think that the fair output of the rivals is at least the collusive output and at most the equilibrium output of the rivals in static game with self-interested players, then it is easier to sustain collusion when players are reciprocal than when they are self-interested.

The intuition is similar to that of Proposition 2. If reciprocal players think that the fair output of their rivals is greater than their collusive output, then playing the collusive output is more attractive in the dynamic quantity-setting game with reciprocal players than in the game with self-interested players. This happens because the collusive monetary payoffs are the same as the ones obtained in the game with self-interested players but in addition there are payoff gains from positive reciprocity since reciprocal players think that their rivals are being kind.

Additionally, if reciprocal players perceive that the fair output of their rivals is smaller than the equilibrium output of the rivals in static game with self-interested players, then the punishment imposed after cheating occurs becomes more severe with reciprocal players than with self-interested players. This happens because, the Nash equilibrium of the stage game with reciprocal players becomes a negative reciprocity state. This is bad for players since it reduces monetary payoffs (by comparison with the monetary payoffs of self-interested players) and leads to payoff loses from negative reciprocity since reciprocal players think that the rivals are being mean.

In contrast, the single period deviation payoff in the game with reciprocal players is larger than the single period deviation payoff in the game with self-interested players. This happens because the unilateral single period deviation payoff of a reciprocal player also includes the benefit that player derives from being treated kindly by the rivals (the rivals are playing their collusive outputs). However, this effect is of second-order since monetary payoffs are large by comparison with fairness payoffs.

3.5 Discussion

Our main results hold provided certain conditions are met. For example, we rule out fairness concerns by the part of consumers. This assumption was made on methodological grounds, to better isolate the effect of fairness and reciprocity among firms on collusive outcomes.

We also rule out that firms have fairness considerations with respect to consumers. Contrary to this assumption, Engel (2007) reports that when subjects know that they are playing against human buyers (instead of simulated demand), collusion rates decrease substantially. This might undermine the effects predicted by the model.

The assumption that Nash punishments are either at the smallest or largest pure strategy Nash equilibria is essentially a technical condition. This condition is necessary when the stage game has multiple equilibria since in a supermodular game we can state unambiguous comparative static results for the largest and the smallest Nash equilibria but not for other Nash equilibria.

So far the paper has indicated that fairness and reciprocity facilitate collusion when players use Nash reversion to punish deviations. However, Abreu's (1988) theory of optimal punishments can be an alternative framework of analysis.

The existence of penal code punishments gives necessary and sufficient conditions for an outcome to be a subgame perfect equilibrium. On the contrary, Nash reversion punishments give only sufficient conditions. This is a problem since sufficient conditions do not prevent the existence of a harsher punishment in the self-interested case, which is not a Nash reversion punishment, such that the target payoff is a subgame perfect equilibrium for a smaller discount factor in the self-interested case than in the reciprocity case.

We have chosen to conduct the main analysis under Nash reversion punishments because simple strategies are more appealing since it is not very realistic that economic agents play complex strategies. Nevertheless, we show in Appendix B that our findings also extends to the optimal punishments framework (we only analyze the dynamic price-setting market game since the quantity-setting case is similar).

The intuition behind this result is as follows. First, the benefit of deviating today (the unilateral single period deviation payoff minus the collusive payoff) when players use optimal punishments is the same as when they use grim trigger punishments. We already know from Proposition 2 that if monetary payoffs are large by comparison with fairness payoffs, then the increase in the collusive payoff due to fairness considerations is of first-order whereas the increase in the unilateral single period deviation payoff is of second-order. Thus, the benefit of deviating is smaller for reciprocators than for self-interested players no matter if players use optimal punishments or grim trigger punishments.

Second, if reciprocal players think that the fair prices are smaller than the collusive prices, then the prices set on the initial path are perceived as kind behavior by the other players and lead to positive fairness payoffs. Therefore, when the prices of the initial path are set, the payoffs for reciprocal players are higher than those for self-interested players.

Third, it is well known that punishments are more severe when players use optimal punishments than when they use Nash reversion strategies. If reciprocal players think that the fair prices of the rivals are greater than the largest Nash prices of the stage game with self-interested players, then seeing the rivals setting punishment prices lower than Nash prices will be perceived as nastier behavior than seeing the rivals setting Nash prices. Therefore, reciprocal players will set lower prices than self-interested players during the punishment phase under optimal punishments.

The second and the third effects imply that the cost of deviating (the collusive payoff minus the payoff of entering a punishment stage) is larger for reciprocal players than for self-interested players when players use optimal punishments.

3.6 Conclusion

This paper contributes to the literature on how fairness and reciprocity might affect market outcomes. Most of this literature has focused on the impact of fairness concerns by consumers on welfare. Here we take a complementary approach and focus instead on the role of fairness concerns among firms on collusive behavior.

Our main departure from the standard model of firm behavior is the assumption that managers in firms are motivated in part by personal animosity–or respect–towards a rival. Hence, firms might punish rivals who behave "unfairly" towards them and reward rivals who behave "fairly."

We find that fairness and reciprocity among firms can facilitate collusive behavior. We show that this result is valid not only when players' choices are strategic complements but also when they are strategic substitutes. The result also holds no matter if players use grim trigger punishments or optimal punishments. Thus, fairness concerns among producers with reciprocal preferences who interact repeatedly can have adverse welfare consequences for consumers.

3.7 Acknowledgements

We gratefully acknowledge comments from Matthew Rabin, Joel Sobel, and seminar audiences at University of Copenhagen, University Cergy-Pontoise, Free University of Amsterdam, University of Geneva, University of Lausanne, and New University of Lisbon.

Appendix 3.A Proofs

Lemma 0: *If* (i) P_i *is a compact interval in* $\mathcal{R}$, (ii) $u_i(p_i, p_{-i}, p^f_{-i})$ *is order upper semi-continuous in* p_i *for fixed* p_{-i} *and order continuous in* p_{-i} *for a fixed* p_i*, and* $u_i(p_i, p_{-i}, p^f_{-i})$ *has a finite upper bound,* (iii) u_i *is supermodular in* p_i *for fixed* p_{-i}*, and* (iv) $u_i(p_i, p_{-i}, p^f_{-i})$ *has increasing differences in* (p_i, p_{-i})*, then* $\Gamma^r(u,p)$ *is a supermodular game.*

Proof of Lemma 0: According to Milgrom and Roberts (1990), a game $\Gamma(u,x)$ is supermodular if (i) the choice set is a compact interval in $\mathcal{R}$, (ii) u_i is order upper semi-continuous in x_i for x_{-i} and order continuous in x_{-i} for a fixed x_i, and it has a finite upper bound, (iii) u_i is supermodular in x_i for fixed x_{-i}, and (iv) u_i has increasing differences in (x_i, x_{-i}).
The price stage game with reciprocal players $\Gamma^r(u,p)$ satisfies condition (i) since it is never optimal for players to choose an infinite price for any finite quantity. We have assumed that u_i also satisfies all the requirements of condition (ii). Condition (iii) is satisfied since the choice variables of players are scalars. Condition (iv) is satisfied if for any two aggregate actions of the others p'_{-i}, p''_{-i} with $p'_{-i} \geq p''_{-i}$ (product order) the difference $u_i(p_i, p'_{-i}, P^f_i) - u_i(p_i, p''_{-i}, P^f_i)$ is increasing (or non-decreasing) in p_i, which is assumed as well. Therefore $\Gamma^r(u,p)$ is supermodular game. *Q.E.D.*

Proof of Lemma 1: It is an application of Theorem 6 in Milgrom and Roberts (1990) with a slight difference. In their setting, the smallest and largest pure strategy equilibria of the game depends on a scalar, but in our model it depends on a vector. Nevertheless, the proof is immediate since we propose the smallest and largest equilibria is nonincreasing with the fair price perception for any player j, which is a scalar. As a result, if the vector increases in every component, then the smallest and largest equilibria do not increase. *Q.E.D.*

Proof of Proposition 1: The stage game $\Gamma^s(\pi,p)$ is obtained from the stage game $\Gamma^r(u,p)$ by setting $\alpha = 0$. Thus, if $\Gamma^r(u,p)$ is a supermodular game so is $\Gamma^s(\pi,p)$. This means that $\Gamma^s(\pi,p)$ also has a smallest and a largest Nash equilibria in pure-strategies. Denote these two equilibria by $\underline{p}^{ns}$ and $\overline{p}^{ns}$, respectively. Observe that if $p^f_{-i} = \overline{p}^{ns}_{-i}$ then $\overline{p}^{nr} = \overline{p}^{ns} = \overline{p}^n$ and $u_i(\overline{p}^n, p^f_{-i}) = \pi_i(\overline{p}^n)$ since $w_{ij}(\overline{p}^n_j, p^f_{ij}) = 0$ for all j. (i) If $\overline{p}^{ns}_{-i} < p^f_{-i}$, then $\overline{p}^{nr} \leq \overline{p}^{ns}$ by

Lemma 1. These two inequalities imply $\overline{p}^{nr}_{-i} < p^f_{-i}$ which together with (3.10) imply $w_{ij}(\overline{p}^{nr}_j, p^f_{ij}) < 0$ for all j. But then it follows that $u_i(\overline{p}^{nr}, p^f_{-i}) < \pi_i(\overline{p}^{ns})$ by the fact that $w_{ij}(\overline{p}^{nr}_j, p^f_{ij}) < 0$ for all j and $\pi_i(\overline{p}^{nr}) \leq \pi_i(\overline{p}^{ns})$ for all i. Similarly, (ii) if $p^f_{-i} < \underline{p}^{ns}_{-i}$, then $\overline{p}^{nr} \geq \overline{p}^{ns}$ by Lemma 1. These two inequalities imply $p^f_{-i} < \underline{p}^{nr}_{-i}$ which together with (3.10) imply $w_{ij}(\overline{p}^{nr}_j, p^f_{ij}) > 0$ for all j. But then it follows that $u_i(\overline{p}^{nr}, p^f_{-i}) > \pi_i(\overline{p}^{ns})$ by the fact that $w_{ij}(\overline{p}^{nr}_j, p^f_{ij}) > 0$ for all j and $\pi_i(\overline{p}^{nr}) \geq \pi_i(\overline{p}^{ns})$ for all i. *Q.E.D.*

Proof of Lemma 2: If $p^f_{ij} \in [\overline{p}^{ns}_j, p^c_j]$ for all i and $j \neq i$, then by (3.10) $w_{ij}(p^c_j, p^f_{ij}) \geq 0$ and $w_{ij}(\overline{p}^{nr}_j, p^f_{ij}) \leq 0$, for all i and $j \neq i$. This in turn implies that

$$u_i(p^c, p^f_{-i}) \geq \pi_i(p^c). \tag{3.17}$$

We also know that

$$\pi_i(p^c) > \pi_i(\overline{p}^{ns}) > \pi_i(\underline{p}^{ns}). \tag{3.18}$$

If $p^f_{ij} \geq \overline{p}^{ns}_j$ for all i and $j \neq i$, then we know from Proposition 1 that

$$u_i(\overline{p}^{nr}, p^f_{-i}) \leq \pi_i(\overline{p}^{ns}), \text{ and } u_i(\underline{p}^{nr}, p^f_{-i}) \leq \pi_i(\underline{p}^{ns}) \tag{3.19}$$

for all i. From (3.17), (3.18) and (3.19) we obtain

$$u_i(p^c, p^f_{-i}) > u_i(\overline{p}^{nr}, p^f_{-i}) \text{ and } u_i(p^c, p^f_{-i}) > u_i(\underline{p}^{nr}, p^f_{-i})$$

for all i, which by Friedman (1971) implies that there exists a sufficiently high discount factor such that p^c is a subgame-perfect Nash equilibrium of $\Gamma^r(u, p)$. *Q.E.D.*

Proof of Proposition 2: By Friedman (1971) and Lemma 2, the assumptions made imply that p^c is a subgame-perfect Nash equilibrium of $\Gamma^s(\pi, p)$ and of $\Gamma^r(u, p)$. We want to show that the critical discount factor at which p^c can be sustained using grim trigger punishments in $\Gamma^r_\infty(u, p)$ is lower than the critical discount factor at which p^c can be sustained using grim trigger punishments in $\Gamma^s_\infty(\pi, p)$, that is, $\delta^r_{p^c} < \delta^s_{p^c}$. From (3.5) and (3.7) sufficient conditions for $\delta^r_{p^c} < \delta^s_{p^c}$ are

$$u_i(BR^r_i(p^c_{-i}), p^c_{-i}, p^f_{-i}) - u_i(p^c, p^f_{-i}) \leq \pi_i(BR^s_i(p^c_{-i}), p^c_{-i}) - \pi_i(p^c), \tag{3.20}$$

and

$$u_i(BR_i^r(p_{-i}^c), p_{-i}^c, p_{-i}^f) - u_i(p^{nr}, p_{-i}^f) \geq \pi_i(BR_i^s(p_{-i}^c), p_{-i}^c) - \pi_i(p^{ns}), \quad (3.21)$$

and at least one inequality holds strictly.
We start by showing that $\pi_i(p_i, p_{-i}) = (p_i - c_i)D_i(p_i, p_{-i})$ and $p_{ij}^f \leq p_j^c$ for all $j \neq i$ imply that (3.20) is satisfied as a strict inequality. We have that

$$\begin{aligned} u_i(BR_i^r(p_{-i}^c), p_{-i}^c, p_{-i}^f) - u_i(p^c, p_{-i}^f) &= \pi_i(BR_i^r(p_{-i}^c), p_{-i}^c) - \pi_i(p^c) \\ + \alpha \sum_{j \neq i} w_{ij}(p_j^c, p_{ij}^f)(p_j^c - c_j)&[D_j(BR_i^r(p_{-i}^c), p_{-i}^c) - D_j(p^c)] \\ \leq \pi_i(BR_i^r(p_{-i}^c), p_{-i}^c) - \pi_i(p^c) &< \pi_i(BR_i^s(p_{-i}^c), p_{-i}^c) - \pi_i(p^c) \end{aligned}$$

The equality is obtained from (3.8) and from the assumption $\pi_i(p_i, p_{-i}) = (p_i - c_i)D_i(p_i, p_{-i})$. The weak inequality comes from the assumption that $p_{ij}^f \leq p_j^c$ which implies $w_{ij}(p_j^c, p_{ij}^f) \geq 0$, and the assumption that D_j is increasing with p_i which together with $p_i^{dr} < p_i^c$ imply $D_j(BR_i^r(p_{-i}^c), p_{-i}^c) - D_j(p^c) < 0$. The strict inequality comes from the fact that $BR_i^s(p_{-i}^c)$ is the best-reply to p_{-i}^c by a self-interested player.
We now show that if $\overline{p}_j^{ns} \leq p_{ij}^f$ for all $j \neq i$ and α is sufficiently small, then (3.21) is satisfied. Rewrite (3.21) as

$$[u_i(BR_i^r(p_{-i}^c), p_{-i}^c, p_{-i}^f) - \pi_i(BR_i^s(p_{-i}^c), p_{-i}^c)] + [\pi_i(p^{ns}) - u_i(p^{nr}, p_{-i}^f)] \geq 0.$$

From Proposition 1 we have that

$$\pi_i(p^{ns}) \geq u_i(p^{nr}, p_{-i}^f).$$

If $\overline{p}_j^{ns} \leq p_{ij}^f$ for all $j \neq i$, then $w_{ij}(p_j, p_{ij}^f) \geq 0$ for all $j \neq i$. Taking a first-order Taylor series expansion of $u_i(BR_i^r(p_{-i}^c), p_{-i}^c, p_{-i}^f)$ around $\alpha = 0$ we obtain

$$\begin{aligned} &u_i(BR_i^r(p_{-i}^c), p_{-i}^c, p_{-i}^f) \approx \\ &\qquad \pi_i(BR_i^s(p_{-i}^c), p_{-i}^c)) + \alpha[\sum_{j \neq i} w_{ij}(p_j, p_{ij}^f)\pi_j(BR_i^s(p_{-i}^c), p_{-i}^c)], \end{aligned}$$

which is equivalent to

$$u_i(BR_i^r(p_{-i}^c), p_{-i}^c, p_{-i}^f) - \pi_i(BR_i^s(p_{-i}^c), p_{-i}^c) \approx \alpha[\sum_{j \neq i} w_{ij}(p_j, p_{ij}^f)\pi_j(BR_i^s(p_{-i}^c), p_{-i}^c)] \geq 0.$$

Thus $\delta_{p^c}^r < \delta_{p^c}^s$. *Q.E.D.*

Proof of Proposition 3: We need to show that $Q_{-i}^f \in [Q_{-i}^c, Q_{-i}^{ns}]$ for all i, implies $\delta_{q^c}^r < \delta_{q^c}^s$, where $\delta_{q^c}^r$ is the critical discount factor above which q^c can be sustained in $\Gamma_\infty^r(u,q)$ and $\delta_{q^c}^s$ is the critical discount factor above which q^c can be sustained in $\Gamma_\infty^s(\pi,q)$. From (3.5) and (3.7) sufficient conditions are that

$$u_i(BR_i^r(Q_{-i}^c), Q_{-i}^c) - u_i(q^c) \leq \pi_i(BR_i^s(Q_{-i}^c), Q_{-i}^c) - \pi_i(q^c) \tag{3.22}$$

and

$$u_i(BR_i^r(Q_{-i}^c), Q_{-i}^{cs}) - u_i(q^{nr}) \geq \pi_i(BR_i^s(Q_{-i}^c), Q_{-i}^{cs}) - \pi_i(q^{ns}). \tag{3.23}$$

(i) We start by showing that $Q_{-i}^f \in [Q_{-i}^c, Q_{-i}^{ns}]$ implies (3.22) is satisfied as a strict inequality. We have that

$$\begin{aligned} u_i(BR_i^r(Q_{-i}^c), Q_{-i}^c) - u_i(q^c) &= \pi_i(BR_i^r(Q_{-i}^c), Q_{-i}^c) - \pi_i(q^c) \\ &+ \alpha w_i(Q_{-i}^c, Q_{-i}^f)\left[P(BR_i^r(Q_{-i}^c) + Q_{-i}^c) - P(Q^c)\right] Q_{-i}^c \\ &\leq \pi_i(BR_i^r(Q_{-i}^c), Q_{-i}^c) - \pi_i(q^c) < \pi_i(BR_i^s(Q_{-i}^c), Q_{-i}^c) - \pi_i(q^c) \end{aligned}$$

The strict inequality follows from the fact that $BR_i^s(Q_{-i}^c)$ is the best reply to Q_{-i}^c for self-interested players. If $Q_{-i}^c \leq Q_{-i}^f$ then $w_i(Q_{-i}^c, Q_{-i}^f) \geq 0$. Furthermore, $Q_{-i}^f \leq Q_{-i}^{ns}$ implies $BR_i^r(Q_{-i}^c) > q_i^c$ which in turn implies $P(BR_i^r(Q_{-i}^c) + Q_{-i}^c) < P(Q^c)$, since $P'(\cdot) < 0$.

(ii) We now show that $Q_{-i}^f \in [Q_{-i}^c, Q_{-i}^{ns}]$ implies that (3.23) is satisfied. Rewrite (3.23) as

$$[u_i(BR_i^r(Q_{-i}^c), Q_{-i}^c) - \pi_i(BR_i^s(Q_{-i}^c), Q_{-i}^c)] + [\pi_i(q^{ns}) - u_i(q^{nr})] \geq 0.$$

We have that

$$u_i(q^{nr}) = \pi_i(q^{nr}) + \alpha w_i(Q^{nr}_{-i}, Q^f_{-i}) \sum_{j \neq i} \pi_j(q^{nr}) \leq \pi_i(q^{ns}).$$

The inequality follows from $w_i(Q^{nr}_{-i}, Q^f_{-i}) \leq 0$ and the fact that $Q^f_{-i} \leq Q^{ns}_{-i}$ implies $q^{ns}_i \leq q^{nr}_i$ and $\pi_i(q^{nr}) \leq \pi_i(q^{ns})$, for all i. Taking a first-order Taylor series expansion of $u_i(BR^r_i(Q^c_{-i}), Q^{cs}_{-i})$ around $\alpha = 0$ we have

$$u_i(BR^r_i(Q^c_{-i}), Q^c_{-i}) \approx \pi_i(BR^s_i(Q^c_{-i}), Q^c_{-i}) + \alpha[w_i(Q^c_{-i}, Q^f_{-i}) \sum_{j \neq i} \pi_j(BR^s_i(Q^c_{-i}), Q^c_{-i})].$$

which is equivalent to

$$u_i(BR^r_i(Q^c_{-i}), Q^c_{-i}) - \pi_i(BR^s_i(Q^c_{-i}), Q^c_{-i}) \approx \alpha[w_i(Q^c_{-i}, Q^f_{-i}) \sum_{j \neq i} \pi_j(BR^s_i(Q^c_{-i}), Q^c_{-i})] \geq 0$$

since $Q^c_{-i} \leq Q^f_{-i}$ implies that $w_i(Q^c_{-i}, Q^f_{-i}) \geq 0$. Thus, $Q^f_{-i} \in [Q^c_{-i}, Q^{ns}_{-i}]$ for all i, implies $\delta^r_{q^c} < \delta^s_{q^c}$. *Q.E.D.*

Appendix 3.B Optimal Punishments

Abreu (1988) introduces a rule which consists of an initial path (that is an infinite stream of one period action profiles) and punishments (that are also infinite streams for any deviation from the initial path or from a prescribed punishment). He introduces the notion of *simple* strategy profile in which a specific punishment takes place after any deviation for each particular player. Thus, the simple strategy profiles have a description of $(n+1)$ paths for an n-player game. On the other hand, an arbitrary strategy profile may consist of infinite amount of punishments and depends on complex history-dependent formulas.

We begin by introducing additional notations and definitions, after we show an optimal simple penal code exists. Finally, we state conditions under which it is easier to sustain collusion with reciprocal players than with self-interested ones under optimal punishments.

A pure strategy of player i is denoted σ_i. Each σ_i is a sequence of functions, $\sigma_i(1), \sigma_i(2), ..., \sigma_i(t), ...$, one for each t. The function for all periods t determines player i's action at t as a function of the actions of all players in previous periods. Formally, at $t = 1, \sigma_i(1) \in P_i$ and for $t = 2, 3, ...$, $\sigma_i(t) : P^{t-1} \to P_i$. Player i's strategy set is denoted Σ_i, and the set of strategy profiles is denoted $\Sigma \equiv \Sigma_1 \times \Sigma_2 \times ... \times \Sigma_n$.

A path (or punishment), $\widetilde{P}$, is a stream of action profiles $\{p(t)\}_{t=1}^{\infty}$ and let $\Omega \equiv P^{\infty}$ be the set of punishments. Any strategy profile $\sigma \in \Sigma$ generates a path denoted $\widetilde{P}(\sigma) = \{p(\sigma)(t)\}_{t=1}^{\infty}$, and it is defined as follows:

$$p(\sigma)(1) = \sigma(1) \text{ and}$$
$$p(\sigma)(t) = \sigma(t)(p(\sigma)(1), ..., (p(\sigma)(t)).$$

Player i's payoff from path $\widetilde{P} \in \Omega$ is given by $v_i^x : \Omega \to \mathcal{R}$ for $x = \{r, s\}$ such that

$$v_i^x(\widetilde{P}) = \begin{cases} \sum_{t=1}^{\infty} \delta^t u_i(p(t)) \text{ if } x = r \\ \sum_{t=1}^{\infty} \delta^t \pi_i(p(t)) \text{ if } x = s \end{cases} \tag{3.24}$$

where u_i is given by (3.8) and (3.10). Player i's payoff function is given by $\widetilde{v}_i^x : \Sigma \to \mathcal{R}$ such that $\widetilde{v}_i^x(\sigma) = v_i(\widetilde{P}(\sigma))$.

Abreu (1988) introduces the simple strategy profile, which is defined by $(n+1)$-vector of paths $(\widetilde{P}^0, \widetilde{P}^1, ..., \widetilde{P}^n)$ and a rule. The initial path is $\widetilde{P}^0$, and for each player $i \in \{1, ..., n\}$, $\widetilde{P}^i$ is the punishment for player i. Any unilateral deviation of player i from the ongoing path is responded by imposing $\widetilde{P}^i$. If more than one player deviate, the ongoing path continues to be followed and deviators will not be punished. Formally:

Let $\widetilde{P}^i \in \Omega$, $i = 0, 1, ..., n$. The *simple strategy profile* $\sigma(\widetilde{P}^0, \widetilde{P}^1, ..., \widetilde{P}^n)$ specifies: (i) play $\widetilde{P}^0$ until some player deviates unilaterally from $\widetilde{P}^0$; (ii) for any $j \in \{1, ..., n\}$, play $\widetilde{P}^j$ if the jth player deviates unilaterally from $\widetilde{P}^i$, $i = 0, 1, ..., n$, where $\widetilde{P}^i$ is an ongoing previously specified path; continue with $\widetilde{P}^i$ if no deviations occur or if two or more players deviate simultaneously.

A simple strategy $\sigma(\widetilde{P}^0, \widetilde{P}^1, ..., \widetilde{P}^n)$ profile is *perfect* if and only if no one-shot deviation by any player $j \in \{1, ..., n\}$ from $\widetilde{P}^i$, $i = 0, 1, ..., n$, yields player j a higher payoff, when all players conform with $\widetilde{P}^j$ after the deviation.[5] Let Σ^p denote the set of perfect equilibrium strategy profiles of $\Gamma_\infty(\delta)$. The perfect equilibrium paths $\Omega^p = \{\widetilde{P}(\sigma)|\sigma \in \Sigma^p\}$, and payoffs $V = \{v(\widetilde{P})|\widetilde{P} \in \Omega^p\}$.

We introduce three more definitions from Abreu (1988) before stating the existence result. An *optimal penal code* is an n-vector of the strategy profiles $\{\underline{\sigma}^1, ..., \underline{\sigma}^n\}$ such that for all i,

$$\underline{\sigma}^i \in \Sigma^p \text{ and } \widetilde{v}_i(\underline{\sigma}^i) = \min\{\widetilde{v}_i(\sigma)|\sigma \in \Sigma^p\}.$$

Let $\sigma^i(\widetilde{P}^1, ..., \widetilde{P}^n) = \sigma(\widetilde{P}^i, \widetilde{P}^1, ..., \widetilde{P}^n)$. The *simple penal code* $(\widetilde{P}^1, ..., \widetilde{P}^n)$ is the n-vector of the strategy profiles $\sigma^1(\widetilde{P}^1, ..., \widetilde{P}^n), ..., \sigma^n(\widetilde{P}^1, ..., \widetilde{P}^n)$. Finally, a simple penal code $(\widetilde{P}^1, ..., \widetilde{P}^n)$ is an *optimal simple penal code* if it is an optimal penal code.

Lemma 3: *If Σ^p is non-empty, P is a compact topological space and given p^f, $u : P \times p^f \to R^n$ is continuous, then an optimal simple penal code exists.*
Proof of Lemma 3: The lemma follows from Abreu (1988) under the assumptions of $u(.)$. *Q.E.D.*

Similarly, an optimal simple penal code exists for a continuous payoff function $\pi : P \to \mathcal{R}^n$. Let present discounted value of player i's payoffs from the period $t+1$ to ∞ along the path $\widetilde{P}$ be

$$v_i^x(\widetilde{P}; t+1) = \begin{cases} \sum_{k=1}^{\infty} \delta^k u_i(p(t+k)) \text{ if } x = r \\ \sum_{k=1}^{\infty} \delta^k \pi_i(p(t+k)) \text{ if } x = s \end{cases} \tag{3.25}$$

[5] This relation holds if the set of payoffs of the stage game is bounded (i.e. $\{u(p)|p \in P\}$ is bounded).

and player i's payoff under her optimal penal code, $\underline{v}_i^x = \widetilde{v}_i^x(\underline{\sigma}^i)$. The following result indicates the use of optimal penal code to characterize the set of perfect equilibrium paths.

Lemma 4: If an optimal penal code exists, then $\widetilde{P}^0 \in \Omega^p$ if and only if

$$\begin{aligned} u_i(p_i^{dr}, p_{-i}^0(t)) - u_i(p^0) &\leq v_i^r(\widetilde{P}^0; t+1) - \underline{v}_i^r && (3.26) \\ \pi_i(p_i^{ds}, p_{-i}^0(t)) - \pi_i(p^0) &\leq v_i^s(\widetilde{P}^0; t+1) - \underline{v}_i^s && (3.27) \end{aligned}$$

Proof of Lemma 4: The lemma follows from Abreu (1988). *Q.E.D.*

The left-hand-side of inequalities (3.26) and (3.27) are the benefit of deviating today for reciprocators and self interested players, respectively. The right-hand-side is the cost of deviating. Observe that the prices in each period of the initial path can be considered as any collusive prices.

Since the existence of optimal simple penal code is guaranteed under the given assumptions, our final result shows that fairness and reciprocity facilitate collusion when players use optimal simple penal codes.

Proposition 5: *Assume (i) u_i has decreasing differences in (p_i, p_{-i}^f), for all i, (ii) $\pi_i(p_i, p_{-i}) = (p_i - c_i)D_i(p_i, p_{-i})$, and (iii) $p_{ij}^f \in [\overline{p}_j^{ns}, p_j^c]$ for all i and $j \neq i$. Let p^0 satisfy $\pi_i(p^o) > \pi_i(\overline{p}^{ns})$ for all i. If an optimal simple penal code exist, then the critical (minimum) discount level to sustain collusion at $\widetilde{P}^0$is lower in the game with reciprocal players $\Gamma_\infty^r(n, u, p, P^f)$ than in the game with self-interested players $\Gamma_\infty^s(n, \pi, p)$, that is $\delta_{p^0}^r < \delta_{p^0}^s$.*

Proof of Proposition 5: The minimum critical discount factor will be obtained if the inequality (3.26) and (3.27) hold with equality respectively for reciprocators and self-interested players, otherwise the discount factor can be decreased by a small amount without violating the inequality. In Proposition 3, we proved the LHS of the equations being smaller for reciprocators, hence a smaller discount level is possible for the reciprocators. In addition, the following condition is immediate

$$v_i^r(\widetilde{P}^0; t+1) \geq v_i^s(\widetilde{P}^0; t+1) \qquad (3.28)$$

considering the initial path where each player i sets at least the collusive price p_i^c for each stage, until one deviates. Hence for any fair price perception $p_{ij}^f \in [\overline{p}_j^{ns}, p_j^c]$ for all i and $j \neq i$, the prices set at the initial path will be perceived as kind behavior, thus the condition holds. Note that, if the prices set at the initial path are equal to collusive prices p_i^c for each player i and $p_{ij}^f = p_j^c$ for all i and $i \neq i$, then the condition holds with equality. Finally, to complete the proof we need to compare the payoff of any player i in the optimal penal code $\underline{v}_i^x$. In the optimal penal code, the players punish the deviated player i via playing a pure strategy profile $\underline{\sigma}^i \in \Sigma^p$, which gives the lowest possible payoff to player i. Let ${}^{nx}\underline{\sigma}$ denote the strategy profile where in each stage players set Nash prices. Since ${}^{nx}\underline{\sigma} \in \Sigma^p$, in each stage the optimal penal code for player i, $\underline{v}_i^x$, is at least as severe as ${}^{nx}\underline{\sigma}$, which means that the optimal punishment of player j in the reciprocity case, $\underline{p}_j$, satisfies $\underline{p}_j \leq p_j^{ns}$. Note that, if the prices set in the penal code are such that $\underline{p}_j = p_j^{ns} = p_{ji}^f$ for all j and $i \neq j$, then the payoffs from the penal code are equal for reciprocal and self-interested players, that is, $\underline{v}_i^r = \underline{v}_i^s$. Otherwise, the reciprocal players perceive the unkind behavior of their rivals and negative reciprocity implies the payoff under the optimal penal code is harsher for reciprocal players than self-interested players, that is $\underline{v}_i^r < \underline{v}_i^s$. Hence $\delta_{p^0}^r < \delta_{p^0}^s$. *Q.E.D.*

Chapter 4

Experimental Cournot Oligopoly and Inequity Aversion

This paper explores the role of inequity aversion as an explanation for observed behavior in experimental Cournot oligopoly. We show that inequity aversion can change the nature of the strategic interaction: quantities are strategic substitutes for sufficiently asymmetric output levels but strategic complements otherwise. We find that inequity aversion can explain why: (i) some experiments result in higher than Cournot-Nash production levels while others result in lower, (ii) collusion often occurs with only two players whereas with three or more players market outcomes are very close to Cournot-Nash, and (iii) players often achieve equal profits in asymmetric Cournot oligopoly.

Keywords: Inequity Aversion; Cournot Oligopoly; Experiments.
JEL Classification Numbers: D43, D63, L13, L21.

4.1 Introduction

Although quantity-setting oligopoly is one of the "workhorse models" of industrial organization, experimentally there is much ambiguity about its outcome. A recent survey by Georgantzis (2006) indicates that many experimental Cournot oligopoly games reject the hypothesis that the outcome is in line with the Cournot-Nash equilibrium of the corresponding one-shot game. Interestingly, however, outcomes on both sides of the Cournot-Nash outcome are found: some experiments result in higher than Cournot-Nash production levels while others result in lower production levels (Holt, 1995).[1]

Why does the theory perform poorly in the experiments? One possibility is that players are averse to inequality in earnings, that is, they are concerned about their own material payoff but also about the consequences of their acts on payoff distributions.

Inequity aversion has been shown to explain a broad range of data for many different games. The clearest evidence for these type of preferences comes from bargaining and trust games. For example, in ultimatum games offers are usually much more generous than predicted by equilibrium, and low offers are often rejected. According to the inequity aversion explanation, these offers are consistent with an equilibrium in which players make offers knowing that other players may reject allocations that appear unfair.[2]

[1]A rather general finding in finitely-repeated symmetric experimental Cournot oligopoly is that, while some learning occurs during the session, in many sessions total output is not significantly different from the collusive prediction, while in other sessions, total output oscilates between the collusive and the Cournot outcome. Additionally, in finitely-repeated asymmetric experimental Cournot oligopoly, subjects' strategies fail to converge towards the Nash equilibrium prediction. See Rassenti et al. (2000) and Huck et al. (2000, 2001).

[2]Camerer (2003) and Sobel (2005) provide excellent reviews of this literature.

In this paper we study formally the role of inequity aversion on Cournot competition. We assume that a player cares about her own monetary payoff and, in addition, would like to reduce the difference between her payoff and those of her rivals. More specifically, a inequity averse player dislikes advantageous inequity: she feels compassion towards her rivals when the average material payoff of her rivals is smaller than her own material payoff. Additionally, an inequity averse player also dislikes disadvantageous inequity: she feels envy towards her rivals when the average material payoff of her rivals is greater than her own material payoff.

We find that inequity aversion can change the nature of the strategic interaction: quantities are strategic substitutes when players choose asymmetric output levels but strategic complements when they choose similar output levels. This can give rise to a continuum of equilibria. We show that the set of Nash equilibria of Cournot competition with inequity averse players changes monotonically with compassion and envy. If players' degree of envy increases, then the largest Nash equilibria of the Cournot game moves closer to the Walrasian outcome. In contrast, if players' degree of compassion increases, then the smallest Nash equilibria of the Cournot game moves closer to the collusive outcome. However, as the number of players grows the impact of inequity aversion vanishes. This happens because it takes only one self-interested player to destroy the continuum of equilibria generated by inequity aversion.

We find that relatively low levels of inequity aversion generate less asymmetries in profits than those predicted when self-interested players play asymmetric Cournot oligopolies. We also show that relatively high levels of inequity aversion can explain why often players attain equal profits in asymmetric experimental Cournot oligopolies. The intuition for this result is straightforward. For relatively high levels of inequity aversion, attaining asymmetric profits imposes inequity costs that are too high in relation to the material benefits.

This paper is related to a recent strand of literature in economics that studies the consequences of relaxing the assumption of pure self interest. Rabin (1993) is the first using fairness considerations in game theory. Sappington and Desiraju (2007) study inequity aversion in adverse selection contexts. Biel (2008) studies how the optimal incentive contract in team production is affected when workers are averse to inequity. Santos-Pinto (2008) shows that inequity aversion is able to organize several experimental regularities of endogenous timing games. Englmaier and Wambach (2010) study optimal contracts when the agent suffers from being better off or worse off than the principal.

The paper proceeds as follows. Section 2 sets-up the model. Section 3 characterizes equilibria of Cournot oligopoly with symmetric costs and inequity averse players. Section 4 considers Cournot oligopoly with asymmetric costs. Section 5 concludes the paper. All proofs are in the Appendix.

4.2 The Model

Many experiments indicate that individuals are motivated not only by material self-interest, but also by the distribution of payoffs. We incorporate this possibility in the Cournot oligopoly game by assuming that players are averse to inequality in profits.

There are two main theories of inequity aversion: Fehr and Schmidt's (1999) and Bolton and Ockenfels (2000). According to Fehr and Schmidt a player cares about his own payoff and dislikes absolute payoff differences between his own payoff and the payoff of any other player.[3]

According to Bolton and Ockenfels's (2000) an inequity averse player is concerned with both his own payoff and his relative share of the total group payoff. So, a player would be equally happy if all players received the same payoff or if some were rich and some were poor as long as he received the average payoff, while according to Fehr and Schmidt (1999) he would clearly prefer that all players get the same.[4]

We follow Fehr and Schmidt's (1999) approach to model inequity aversion. Consider a Cournot oligopoly with n players where the profit of player i is the difference between revenue and cost, that is,

$$\pi_i(q_i, Q_{-i}) = R_i\left(q_i, Q_{-i}\right) - C_i(q_i) = P(Q)q_i - C_i(q_i), \tag{4.1}$$

where $R_i(q_i, Q_{-i})$ is revenue, $C_i(q_i)$ is the cost of production, and $P(Q)$ is the inverse market demand with $Q = \sum q_i$. We assume that $P(Q)$ is strictly positive on some bounded interval $(0, \bar{Q})$ with $P(Q) = 0$ for $Q \geq \bar{Q}$. We also assume that $P(Q)$ is twice continuously differentiable with $P'(Q) < 0$ (in the interval for which $P(Q) > 0$). Players costs of production are assumed to

[3] Neilson (2000) provides an axiomatic foundation for Fehr and Schmidt (1999).

[4] Bolton and Ockenfels's (2000) payoff function is $U_i(\pi) = v(\pi_i, \pi_i / \sum_{j=1}^{n} \pi_j)$, where v is assumed to be globally non-decreasing and concave in the first argument, to be strictly concave in the second argument (relative payoff), and to satisfy $v(\pi_i, 1/n) = 0$ for all π_i. This type of inequity aversion has no impact on equilibrium outcomes in symmetric Cournot games. This result is driven by the assumption that $v(\pi_i, 1/n) = 0$ for all π_i.

be twice continuously differentiable with $C_i'(q_i) \geq 0$. It is also assumed that the decreasing marginal revenue property holds, that is, $P'(Q) + P''(Q)\, q_i < 0$ (this implies that quantities are strategic substitutes). Furthermore, we assume that $P'(Q) - C_i''(q_i) \leq 0$ (this implies the profit function is strictly concave).

According to Fehr and Schmidt (1999), the payoff function of player i is

$$U_i(\pi_i, \pi_{-i}) = \pi_i - \left[\frac{\alpha_i}{n-1} \sum_{j \neq i} \max\left(\pi_j - \pi_i, 0\right) + \frac{\beta_i}{n-1} \max \sum_{j \neq i} \left(\pi_i - \pi_j, 0\right) \right]. \tag{4.2}$$

The terms in the square bracket are the payoff effects of compassion β_i and envy α_i. We see that if player i's profits are greater than the average profits of its rivals then player i feels compassion towards its rivals. However, if player i's profits are smaller than the average profits of its rivals then player i feels envious of his rivals.[5] This model of inequity aversion has piecewise linear indifference curves over a player's own profits and its rivals' profits.

Player i's inequity aversion towards its rivals is characterized by the pair of parameters (α_i, β_i), $i = 1, 2, \ldots, n$.[6] player i exhibits strict inequity aversion when both α_i and β_i are strictly greater than zero. player i only cares about maximizing profits when $\alpha_i = \beta_i = 0$. In all other cases player is (weakly) averse to inequity. We assume that α_i and β_i, $i = 1, \ldots, n$, are common knowledge. Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$.

Fehr and Schmidt assume that the dislike of disadvantageous inequity

[5] When there are only two players in the market, player i's payoff function becomes $U_i(\pi_i, \pi_j) = \pi_i - [\alpha_i \max(\pi_j - \pi_i, 0) + \beta_i \max(\pi_i - \pi_j, 0)]$, $i \neq j = 1, 2$.

[6] Alternatively, we could have assumed that player i has different feelings of compassion and envy towards each rival. To simplify the analysis, we assume that player i feels the same degree of compassion and envy towards all rivals.

is stronger than that of advantageous inequity, i.e. $\alpha_i > \beta_i$ and that β_i is smaller than 1. We make no assumptions about the relation between α_i and β_i but we assume, like Fehr and Schmidt, that β_i is smaller than 1.

4.3 Equilibria with Symmetric Costs

In this section we characterize the impact of inequity aversion on Cournot oligopoly with symmetric costs. Our first result characterizes the best reply of an inequity averse player.

Proposition 1: *The best reply of player i in a Cournot oligopoly with symmetric costs and inequity averse players is*

$$r_i(Q_{-i}) = \begin{cases} s_i(Q_{-i}), \ \textit{if} & 0 \leq Q_{-i} \leq (n-1)q(\beta_i) \\ \frac{Q_{-i}}{n-1}, \ \textit{if} & (n-1)q\left(\beta_i\right) \leq Q_{-i} \leq (n-1)q(\alpha_i) \\ t_i(Q_{-i}), \ \textit{if} & (n-1)q(\alpha_i) \leq Q_{-i} \end{cases}, \quad (4.3)$$

where

$$s_i(Q_{-i}) = \arg\max_{q_i} \left[(1-\beta_i)\, \pi_i(q_i, Q_{-i}) + \frac{\beta_i}{n-1} \sum_{j \neq i} \pi_j(q_i, Q_{-i}) \right], \quad (4.4)$$

$$t_i(Q_{-i}) = \arg\max_{q_i} \left[(1+\alpha_i)\, \pi_i(q_i, Q_{-i}) - \frac{\alpha_i}{n-1} \sum_{j \neq i} \pi_j(q_i, Q_{-i}) \right], \quad (4.5)$$

$q(\beta_i)$ is the solution to $(1-\beta_i)\,[P(nq) - C'(q)] + P'(nq)q = 0$, and $q(\alpha_i)$ is the solution to $(1+\alpha_i)\,[P(nq) - C'(q)] + P'(nq)q = 0$.

The best reply has three different segments. When the rivals produce low output levels the best reply has a negative slope and consists of a smaller output than the output of a self-interested player due to compassion. However, when the rivals produce intermediate output levels the best reply has a positive slope and consists in producing the average output level of the rivals. Finally, when the rivals produce high output levels the best reply has a negative slope and consists of a larger output level than the output a self-interested player due to envy.

We see that the best reply of an inequity averse player is continuous like the best reply of self-interested player. However, the best reply of an inequity averse player is non-monotonic whereas a self-interested player has a monotonic best reply. Thus, under inequity aversion quantities are strategic substitutes over low and high output levels of the rivals but strategic complements over intermediate output levels of the rivals.

Proposition 2: *The set of Nash equilibria of a Cournot oligopoly with symmetric costs and inequity averse players is*

$$N^{IA} = \{(q_1, \ldots, q_n) : q_i = q_j,\ \forall i \neq j,\ \textit{and}\ q(\beta) \leq q_i \leq q(\alpha),\ i = 1, \ldots, n\},$$

where $q(\beta) = \max\left[q\left(\beta_1\right), \ldots, q(\beta_n)\right]$, *and* $q(\alpha) = \min\left[q(\alpha_1), \ldots, q(\alpha_n)\right]$.

Proposition 2 tells us that if all players are averse to inequity, then there is a continuum of equilibria in a Cournot oligopoly with symmetric costs. The smallest Nash equilibrium is determined by the preferences of the player with the highest level of compassion and the largest Nash equilibrium is determined by the preferences of the player with the lowest level of envy. Proposition 2 also tells that the market output with inequity averse players may be higher or lower than the market output with self-interested players.

This depends on players' degree of envy and compassion.

Proposition 3: *The smallest Nash equilibrium of a Cournot oligopoly with symmetric costs and inequity averse players is a nonincreasing function of β. The largest Nash equilibrium is a nondecreasing function of α.*

This result characterizes the impact of compassion and envy on the set of Nash equilibria of the Cournot oligopoly with inequity averse players. It tells us that an increase in compassion reduces the market output produced in the smallest Nash equilibria with inequity averse players. This result is quite intuitive. In fact, Fehr and Schmidt's (1999) payoff function implies that if player i has a higher monetary payoff than the average payoff of his opponents and $\beta_i = 1/2$, then player i is just as willing to keep one dollar to himself as to give it to his rivals. If all players have similar preferences, then they act as if they are maximizing the joint profit, $\sum \pi_i$. So, if $\beta_i = 1/2$, for all i, then compassion leads to the best collusive outcome.[7] In contrast, an increase in envy raises the market output produced in the largest Nash equilibria with inequity averse players.

Propositions 2 and 3 show that inequity aversion can change qualitatively the predictions of a symmetric Cournot oligopoly. In the benchmark game with self-interested players ($\alpha_i = \beta_i = 0$, for all i) there is a unique Nash

[7]Experimental evidence shows that the amount of collusion observed in repeated Cournot oligopoly depends on communication, playing with the same rival(s) and the size of the market. Daughety and Forsythe (1987a,1987b) report that face to face non-binding groups discussions increase price in repeated Cournot games in which the quantity decision are made afterwards. Similarly, Isaac et al. (1984) report that posted-offer prices are increased when sellers are given chance to meet face to face prior to each period. Holt (1985) finds that collusion occurs only when the same subjects are matched in fixed groups for the entire experiment. With random matching the Cournot-Nash solution is a good prediction.

equilibrium. Proposition 2 shows that inequity aversion can give rise to a multiplicity of symmetric equilibria. Proposition 3 shows that compassion can generate collusive outcomes in Cournot markets whereas envy can generate perfectly competitive outcomes. The existence of a multiplicity of equilibria might be the reason why in many experiments with Cournot duopolies outcomes fall on both sides of the Cournot prediction and may range from perfectly collusive to relatively competitive (Holt, 1995).

Huck et al. (2004) review the literature on the role of the number of players on the outcome of Cournot oligopolies with symmetric costs. They find that (pp. 440) "(...) collusion sometimes occurs in duopolies and is very rare in markets with more than two firms. On average, total outputs in markets with more than two firms slightly exceed the Cournot prediction." They also test for number effects in oligopoly in a unified economic frame and find that (pp. 443) "collusion sometimes occurs with two firms. For three-firm oligopolies Nash equilibrium seems to be a good predictor. Markets with four or more firms are never collusive and typically settle around the Cournot outcome while some of them are very competitive with outputs close to the Walrasian outcome." Our next result shows that these findings are consistent with our model.

Proposition 4: *Assume that* α_i *and* β_i, *for all* i, *are drawn from a uniform distribution with support on* $[0,1]$. *As the number of players increases the set of Nash equilibria of a Cournot oligopoly with symmetric costs and inequity averse players converges to the unique Nash equilibrium of a Cournot oligopoly with symmetric costs and self-interested players.*

This result shows that increasing the number of players reduces the impact of inequity aversion on the set of Nash equilibria of a Cournot oligopoly with symmetric costs. This happens because when there are n players, the smallest Nash equilibrium of the game is determined by the preferences of the player with the lowest degree of compassion. Similarly, the largest Nash equilibrium of the game is determined by the player with the lowest degree of envy. If the levels of compassion and envy of each player are drawn from a uniform distribution with support on $[0,1]$, then an increase in the number of players makes it more likely that the lowest level of compassion as well as the lowest level of envy are both very close to zero. Thus, as the number of players increases the smallest and the largest Nash equilibria of a Cournot oligopoly with symmetric costs and inequity averse players converge to the Nash equilibrium of a Cournot oligopoly with symmetric costs and self-interested players.

4.4 Equilibria with Asymmetric Costs

In this section we analyze the impact of inequity aversion on Cournot oligopoly with asymmetric costs. To simplify the analysis let $n = 2$. Furthermore, suppose that there are no fixed costs and that player 1 has a lower marginal cost than player 2, that is, $C_1'(q) < C_2'(q)$ for all q.

Players will attain equal profits when $\pi_1(q_1, q_2) = \pi_2(q_1, q_2)$ or

$$P(q_1 + q_2)(q_2 - q_1) = C_2(q_2) - C_1(q_1). \tag{4.6}$$

Denote the solution of (4.6) with respect to q_i as $q_i = e_i(q_j)$. The slope of the equal profit curve is

$$\frac{dq_2}{dq_1} = \frac{MR_1 - C_1' - P'q_2}{MR_2 - C_2' - P'q_1}. \tag{4.7}$$

We see from (4.7) that if (q_1, q_2) is a point in the equal profit curve that satisfies $MR_1 - C_1' - P'q_2 > 0$ and $MR_2 - C_2' - P'q_1 > 0$, then the slope of the equal profit curve at that point is well-defined and positive. We assume from now on that the cost asymmetry is not too high such that the equal profit curve has a positive slope at all points (q_1, q_2) with $q_2 > q_1$.

Proposition 5: *The best reply of player i in a Cournot duopoly with asymmetric costs and inequity averse players is*

$$r_i(q_j) = \begin{cases} s_i(q_j), \text{ if } & 0 \leq q_j \leq e_j(q_i(\beta_i)) \\ e_i(q_j), \text{ if } & e_j(q_i(\beta_i)) \leq q_j \leq e_j(q_i(\alpha_i)) \\ t_i(q_j), \text{ if } & e_j(q_i(\alpha_i)) \leq q_j \end{cases}, \tag{4.8}$$

where

$$s_i(q_j) = \arg\max_{q_i} \left[(1-\beta_i)\,\pi_i(q_i,q_j) + \beta_i \pi_j(q_i,q_j)\right], \tag{4.9}$$

$$t_i(q_j) = \arg\max_{q_i} \left[(1+\alpha_i)\,\pi_i(q_i,q_j) - \alpha_i \pi_j(q_i,q_j)\right], \tag{4.10}$$

$q_i(\beta_i)$ is the solution to

$$(1-\beta_i)\left[P'(q_i + e_j(q_i))q_i + P(q_i + e_j(q_i)) - C_i'(q_i)\right] + \beta_i P'(q_i + e_j(q_i))e_j(q_i) = 0,$$

and $q_i(\alpha_i)$ *is the solution to*

$$(1+\alpha_i)\left[P'(q_i+e_j(q_i))q_i + P(q_i+e_j(q_i)) - C_i'(q_i)\right] \\ - \alpha_i P'(q_i+e_j(q_i))e_j(q_i) = 0.$$

We see from (4.3) and (4.8) that the best reply of an inequity averse player in a Cournot duopoly with asymmetric costs is qualitatively similar to the best reply of an inequity averse player in a Cournot duopoly with symmetric costs. Quantities are strategic substitutes for low and high output levels of the rival but strategic complements for intermediate output levels of the rival. The only difference is that for intermediate output levels inequity averse players wish to equalize profits. Since costs are asymmetric it is not possible to equalize profits by producing the same output level of as the rival. Thus, players will choose different output levels to equalize profits.

Proposition 6: *Consider a Cournot duopoly where player 1 has lower marginal cost than player 2 and players are inequity averse. If* β_1 *and* α_2 *are sufficiently small, that is,* $q_1(\beta_1) \geq e_1(q_2(\alpha_2))$*, then this game has a unique Nash equilibrium* (q_1^{IA}, q_2^{IA})*, which is the solution to* $q_1 = s_1(q_2)$ *and* $q_2 = t_2(q_1)$*. In this equilibrium: (i) player 1 feels compassion of player 2, (ii) player 2 feels envy of player 1, and (iii)* $q_2^S < q_2^{IA} < q_1^{IA} < q_1^S$*, where* (q_1^S, q_2^S) *is the Nash equilibrium of the game with self-interested players.*

This result says that if the low cost player has a small dislike of advantageous inequity and the high cost player has a small dislike of disadvantageous inequity, then the Cournot duopoly with asymmetric costs and inequity averse players has an asymmetric Nash equilibrium where the low cost player attains a higher profit than the high cost player. Furthermore, the inequity averse low cost player chooses a lower output level than a low cost self-interested player and the inequity averse high cost player chooses a higher output than a high cost self-interested player. The intuition behind this result is as follows.

A low cost player with a small dislike of advantageous inequity chooses a lower output than a low cost self-interested player because she knows that in equilibrium she will attain higher profits than her rival and this induces compassion towards the rival. A high cost player with a small dislike of disadvantageous inequity chooses a higher output than a high cost self-interested player because he knows that in equilibrium he will attain lower profits than his rival and this induces envy towards the rival.

Proposition 7: *Consider a Cournot duopoly where player 1 has lower marginal cost than player 2 and players are inequity averse. If β_1 and α_2 are sufficiently large, that is, $q_1(\beta_1) < e_1(q_2(\alpha_2))$, then the set of Nash equilibria of this game is*

$$N^{IA} = \{(q_1, q_2) : \pi_1(q_1, q_2) = \pi_2(q_1, q_2), \text{ and } q(\beta) \leq q_1 \leq q(\alpha)\}, \quad (4.11)$$

where $q(\beta) = \max[q_1(\beta_1), e_1(q_2(\beta_2))]$, *and* $q(\alpha) = \min[q_1(\alpha_1), e_1(q_2(\alpha_2))]$.

Proposition 7 tells us that if the low cost player has a high dislike of advantageous inequity and the high cost player has a high dislike of disadvantageous inequity, then the Cournot duopoly with asymmetric costs and inequity averse players has a continuum of asymmetric Nash equilibria where players attain equal profits.

Propositions 6 and 7 are consistent with experimental evidence on Cournot oligopolies with asymmetric costs. Keser (1993) studies two stage duopoly games with asymmetric costs and demand inertia. She finds that the high cost player has higher profits and the low cost player lower profits than the self-interested subgame perfect equilibrium profits. Selten et al. (1997) study a 20-period repeated Cournot duopoly with asymmetric costs and find that players often try to achieve equal profits.

4.5 Conclusion

This paper studies the impact of inequity aversion on Cournot competition. We find that inequity aversion can change the nature of the strategic interaction: quantities are strategic substitutes when players choose asymmetric output levels but strategic complements when they choose similar output levels. We show that inequity aversion is able to organize at least three behavioral regularities in experimental Cournot oligopoly.

Appendix 4.A Proofs

Proof of Proposition 1: The quantity $q(\alpha_i)$ is the interception of $t_i(Q_{-i})$ with the 45 degree line. From the definition of $t_i(Q_{-i})$ we have

$$(1+\alpha_i)\left[P'(Q)q_i + P(Q) - C'(q_i)\right] - \frac{\alpha_i}{n-1}\sum_{j\neq i} P'(Q)q_j = 0.$$

In a symmetric equilibrium we have $q_1 = ... = q_n = q$. So,

$$(1+\alpha_i)\left[P'(nq)q + P(nq) - C'(q)\right] - \alpha_i P'(nq)q = 0.$$

or

$$(1+\alpha_i)\left[P(nq) - C'(q)\right] + P'(nq)q = 0.$$

Similarly, the quantity $q(\beta_i)$ is the interception of $s_i(Q_{-i})$ with the 45 degree line. From the definition of $s_i(Q_{-i})$ we have

$$(1-\beta_i)\left[P'(Q)q_i + P(Q) - C'(q_i)\right] + \frac{\beta_i}{n-1}\sum_{j\neq i} P'(Q)q_j = 0.$$

In a symmetric equilibrium we have $q_1 = ... = q_n = q$. So,

$$(1-\beta_i)\left[P'(nq)q + P(nq) - C'(q)\right] + \beta_i P'(nq)q = 0.$$

or

$$(1-\beta_i)\left[P(nq) - C'(q)\right] + P'(nq)q = 0.$$

We now show that $q(\alpha_i)$ is an increasing function of α_i and $q(\beta_i)$ a decreasing function of β_i for $i = 1, \ldots, n$. Let

$$\begin{aligned} h(q,\alpha_i) &= (1+\alpha_i)\left[P(nq) - C'(q)\right] + P'(nq)q = 0, \\ g(q,\beta_i) &= (1-\beta_i)\left[P(nq) - C'(q)\right] + P'(nq)q = 0, \end{aligned}$$

which imply

$$\begin{aligned}
\frac{\partial q}{\partial \alpha_i} &= -\frac{\partial h/\partial \alpha_i}{\partial h/\partial q} = -\frac{P(Q)-C'(q)}{(1+n(1+\alpha_i))\,P'(Q)+nP''(Q)q-C_i''(q)} > 0, \\
\frac{\partial q}{\partial \beta_i} &= -\frac{\partial g/\partial \beta_i}{\partial g/\partial q} = -\frac{-\left[P(Q)-C'(q)\right]}{(1+n(1-\beta_i))\,P'(Q)+nP''(Q)q-C_i''(q)} < 0,
\end{aligned}$$

since $P'(Q) < 0$, $P'(Q) \leq 0$, and $C''(q_i) \geq 0$.

We will now show that $q_i = \frac{1}{n-1}\sum_{j\neq i} q_j$ is a best response for player i when the rivals produce

$$q_i^N \leq \bar{q}_j \leq q(\alpha_i), \tag{4.12}$$

where $\bar{q}_j = \frac{1}{n-1}\sum_{j\neq i} q_j$. To do that we will show that player i can not gain from deviating from $q_i = \bar{q}_j$ when (4.12) holds. Suppose, that (4.12) holds and that player i produces $q_i = \bar{q}_j + \varepsilon$, with $\varepsilon > 0$. In this case player i's payoff is

$$U_i = (1-\beta_i)\left[P\left(Q\right)q_i - C(q_i)\right] + \frac{\beta_i}{n-1}\sum_{j\neq i}\left[P\left(Q\right)q_j - C(q_j)\right]$$

and the change in player i's payoff from producing $q_i = \bar{q}_j + \varepsilon$, $\varepsilon > 0$, instead of $\bar{q}_j$ is approximately equal to

$$\begin{aligned}
dU_i &\approx \left.(1-\beta_i)\left[P'\left(Q\right)q_i + P(Q) - C'(q_i)\right] + \frac{\beta_i}{n-1}\sum_{j\neq i} P'\left(Q\right)q_j\right|_{q_i=\bar{q}_j} (\varepsilon) \\
&= \left[\left(P'\left(n\bar{q}_j\right)\bar{q}_j + P(n\bar{q}_j) - C'(\bar{q}_j)\right) - \beta_i\left(P(n\bar{q}_j) - C'(\bar{q}_j)\right)\right]\varepsilon.
\end{aligned}$$

The square brackets are negative since $q_i = \bar{q}_j > \arg\max\left[P\left(Q\right)q_i - C(q_i)\right]$ and $P(n\bar{q}_j) - C'(\bar{q}_j) > 0$. So, when (4.12) holds, player i can not gain by producing more than $\bar{q}_j$. Now, suppose that (4.12) holds and that player i produces $q_i = \bar{q}_j + \varepsilon$, with $\varepsilon < 0$. In this case player i's payoff is

$$U_i = (1+\alpha_i)\left[P\left(Q\right)q_i - C(q_i)\right] - \frac{\alpha_i}{n-1}\sum_{j\neq i}\left[P\left(Q\right)q_j - C(q_j)\right],$$

and the change in player i's payoff from producing $q_i = \bar{q}_j + \varepsilon$, $\varepsilon < 0$, instead of $\bar{q}_j$ is approximately equal to

$$\begin{aligned} dU_i &\approx \left.(1+\alpha_i)\left[P'(Q)\,q_i + P(Q) - C'(q_i)\right] - \frac{\alpha_i}{n-1}\sum_{j\neq i} P'(Q)\,q_j\right|_{q_i=\bar{q}_j} (\varepsilon) \\ &= \left[(1+\alpha_i)\left[P(n\bar{q}_j) - C'(\bar{q}_j)\right] + P'(n\bar{q}_j)\bar{q}_j\right]\varepsilon = \left.h(q,\alpha_i)\right|_{q=\bar{q}_j}(\varepsilon)\,. \end{aligned}$$

Since $\varepsilon < 0$, we have that $sign\ dU_i = -sign\ \left.h(q,\alpha_i)\right|_{q=\bar{q}_j}$. If $\bar{q}_j = q(\alpha_i)$ we have that $sign\ dU_i = 0$. If $q_i^N \leq \bar{q}_j < q(\alpha_i)$, the fact $h(q,\alpha_i)$ is a decreasing function of q implies that $\left.h(q,\alpha_i)\right|_{q=\bar{q}_j} > 0$, which in turn implies that $sign$ $dU_i < 0$. So, when (4.12) holds, player i can not gain by producing less than $\bar{q}_j$. From this result is follows immediately that if player i's rivals produce $q(\alpha_i) < \frac{1}{n-1}\sum_{j\neq i} q_j$, then the best response of player i is given by $t_i(q_{-i})$. We will now show that $q_i = \frac{1}{n-1}\sum_{j\neq i} q_j$ is a best response for player i when the rivals produce

$$q(\beta_i) \leq \bar{q}_j \leq q_i^N, \tag{4.13}$$

To do that we will show that player i can not gain from deviating from $q_i = \bar{q}_j$ when (4.13) holds. Suppose, that (4.13) holds and that player i produces $q_i = \bar{q}_j + \varepsilon$, with $\varepsilon < 0$. In this case player i's payoff is given by

$$U_i = (1+\alpha_i)\left[P(Q)\,q_i - C(q_i)\right] - \frac{\alpha_i}{n-1}\sum_{j\neq i}\left[P(Q)\,q_j - C(q_j)\right],$$

and the change in player i's payoff from producing $q_i = \bar{q}_j + \varepsilon$, $\varepsilon < 0$, instead of $\bar{q}_j$ is approximately equal to

$$\begin{aligned} dU_i &\approx \left.(1+\alpha_i)\left[P'(Q)\,q_i + P(Q) - C'(q_i)\right] - \frac{\alpha_i}{n-1}\sum_{j\neq i} P'(Q)\,q_j\right|_{q_i=\bar{q}_j} (\varepsilon) \\ &= \left[(1+\alpha_i)\left[P'(n\bar{q}_j)\bar{q}_j + P(n\bar{q}_j) - C'(\bar{q}_j)\right] - \alpha_i P'(n\bar{q}_j)\bar{q}_j\right]\varepsilon. \end{aligned}$$

The square brackets are positive since $q_i = \bar{q}_j < \arg\max\left[P(Q)\,q_i - C(q_i)\right]$ and $P'(n\bar{q}_j) < 0$. So, when (4.13) holds, player i can not gain by producing less than $\bar{q}_j$. Now, suppose that (4.13) holds and that player i produces $q_i = \bar{q}_j + \varepsilon$, with $\varepsilon > 0$. In this case player i's payoff is given by

$$U_i = (1-\beta_i)\left[P(Q)\,q_i - C(q_i)\right] + \frac{\beta_i}{n-1}\sum_{j\neq i}\left[P(Q)\,q_j - C(q_j)\right]$$

and the change in player i's payoff from producing $q_i = \bar{q}_j + \varepsilon$, $\varepsilon > 0$, instead of $\bar{q}_j$ is approximately equal to

$$\begin{aligned} dU_i &\approx (1-\beta_i)\left[P'(Q)\,q_i + P(Q) - C'(q_i)\right] + \frac{\beta_i}{n-1}\sum_{j\neq i} P'(Q)\,q_j \Bigg|_{q_i=\bar{q}_j} (\varepsilon) \\ &= \left[(1-\beta_i)\left[P(n\bar{q}_j) - C'(\bar{q}_j)\right] + P'(n\bar{q}_j)\,\bar{q}_j\right]\varepsilon = g(q,\beta_i)|_{q=\bar{q}_j}(\varepsilon). \end{aligned}$$

Since $\varepsilon > 0$, we have that $sign\ dU_i = sign\ \ g(q,\beta_i)|_{q=\bar{q}_j}$. If $\bar{q}_j = q(\beta_i)$ we have that $sign\ dU_i = 0$. If $q(\beta_i) < \bar{q}_j \leq q_i^N$, the fact $g(q,\beta_i)$ is a decreasing function of q implies that $g(q,\beta_i)|_{q=\bar{q}_j} < 0$, which in turn implies that $sign$ $dU_i < 0$. So, when (4.13) holds, player i can not gain by producing more than $\bar{q}_j$. From this result is follows immediately that if player i's rivals produce $0 \leq \frac{1}{n-1}\sum_{j\neq i} q_j < q(\beta_i)$, then the best response of player i is given by $s_i(q_{-i})$. *Q.E.D.*

Proof of Proposition 2: The proof proceeds in two steps. First we show that the set of equilibria is non-empty. Second, we show that if all players are strictly averse to inequality, then there is a continnum of equilibria and we characterize the largest and the smallest one.
We now show that $q_i = q_i^N$ is the best reply to $q_{-i}^N = \left(q_1^N, \ldots, q_{i-1}^N, q_{i+1}^N, \ldots q_n^N\right)$ in the Cournot oligopoly with symmetric costs and inequity averse players. The welfare of player 1 under outcome q^N is $\pi_1(q^N) = \left[P\left(nq_i^N\right) - C_i(q_i^N)\right] q_i^N$, where $q_i^N = \arg_{q_1}\max\left[P\left(q_i + \sum_{j\neq i} q_j^N\right) - C_i(q_i)\right] q_i$.
If player i produces $q_i^N + \varepsilon$, with $\varepsilon > 0$, and all other players produce q_{-i}^N, then the change in player i's profit is approximately equal to

$$\begin{aligned} d\pi_i &\approx \varepsilon\,\partial\pi_i/\partial q_i|_{q_i=q_i^N} + \frac{1}{2}\varepsilon^2\,\partial^2\pi_i/\partial q_i^2|_{q_i=q_i^N} \\ &= \frac{1}{2}\varepsilon^2\left[2P'(Q^N) + P''(Q^N)q_i^N - C''(q_i^N)\right]. \end{aligned} \tag{4.14}$$

The assumption that $P' < 0$, $P'' \leq 0$, and $C'' \geq 0$ imply that $d\pi_i < 0$. The change in the profit of one of player i's rivals, say j, is approximately equal to

$$\begin{aligned} d\pi_j &\approx \varepsilon\, \partial\pi_j/\partial q_i\big|_{q_i=q_i^N} + \frac{1}{2}\varepsilon^2\, \partial^2\pi_j/\partial q_i^2\big|_{q_i=q_i^N} \\ &= \varepsilon P'(Q^N)q_j^N + \frac{1}{2}\varepsilon^2 P''(Q^N)q_j^N. \end{aligned}$$

Note that the change in the average profit of player i's rivals is the same as the change in the profit of a single rival since

$$\begin{aligned} \frac{1}{n-1}\sum_{j\neq i} d\pi_j &\approx \frac{1}{n-1}\varepsilon P'(Q^N)\sum_{j\neq i} q_j^N + \frac{1}{2}\varepsilon^2 P''(Q^N)\sum_{j\neq i} q_j^N \\ &= \varepsilon P'(Q^N)q_j^N + \frac{1}{2}\varepsilon^2 P''(Q^N)q_j^N. \end{aligned} \tag{4.15}$$

The assumption that $P' < 0$ and $P'' \leq 0$ imply that $\frac{1}{n-1}\sum_{j\neq i} d\pi_j < 0$. We see from (4.14) and (4.15) that if player i produces $q_i^N + \varepsilon$, with $\varepsilon > 0$, and all other players produce q_{-i}^N, then there is a first order decrease in profits of player i and a second order decrease in the average profit of player i's rivals. Thus, if player i produces $q_i^N + \varepsilon$, with $\varepsilon > 0$, it suffers a loss in profits and also a loss from an increase in inequity aversion given that the average profit of the rivals becomes smaller than player i's profit. If that is the case, then player i can not gain by producing $q_i^N + \varepsilon$, with $\varepsilon > 0$, instead of producing q_i^N.

If player i produces $q_i^N + \varepsilon$, with $\varepsilon < 0$, and all other players produce q_{-i}^N, then the change in player i's profit is given by (4.14) and we have that $d\pi_i < 0$. The change in the average profit of player i's rivals is given by (4.15) and we have that $\frac{1}{n-1}\sum_{j\neq i} d\pi_j > 0$ since $\varepsilon < 0$ and the first term is of first order while the second term is of second order. Thus, if player i produces $q_i^N + \varepsilon$, with $\varepsilon < 0$, it suffers a loss in profits and also a loss from an increase in inequity aversion given that the average profit of the rivals becomes greater than player i's profit. If that is the case, then player i can not gain by producing $q_i^N + \varepsilon$, with $\varepsilon < 0$, instead of producing q_i^N. This proves that $q_i = q_i^N$ is the best reply to $q_{-i}^N = \left(q_1^N, \ldots, q_{i-1}^N, q_{i+1}^N, \ldots q_n^N\right)$ and so q^N is a Nash equilibrium of a Cournot oligopoly with inequity averse players.

We now know that the set N^{IA} is non-empty. We still need to show that if all players are strictly averse to inequity, then $q(\beta) < q(\alpha)$, that is, N^{IA} is an interval. We know that $q(\alpha_i)$ is an increasing function of α_i and that $q(\beta_i)$ is a decreasing function of β_i for $i = 1, \ldots, n$. Note that if at least one player does not feel inequity aversion then $q(\beta) = q(\alpha)$, and N^{IA} is a singleton. To see this suppose that player i is not inequity averse, that is, $\alpha_i = \beta_i = 0$. If that is the case, then $h(q, \alpha_i) = 0$ and $g(q, \beta_i) = 0$ imply that $q(0) = q^N$. If $q(\alpha_i)$ is an increasing function of α_i and $q(0) = q^N$, then $q(\alpha) = q^N$. Similarly, if $q(\beta_i)$ is a decreasing function of β_i and $q(0) = q^N$, then $q(\beta) = q^N$. So, if at least one player feels aversion to inequity we have that $q(\beta) = q(\alpha) = q^N = N^{IA}$. We will now show that if all players are strictly averse to inequity, then $q(\beta) < q(\alpha)$, that is, N^{IA} is an interval. If all players are strictly averse to inequity, $q(\alpha_i)$ is an increasing function of α_i and $q(0) = q^N$, then $q(\alpha) > q^N = q(0)$. Also, if all players are strictly inequity averse, $q(\alpha_i)$ is an decreasing function of β_i and $q(0) = q^N$, then $q(\beta) < q^N = q(0)$. This shows that $q(\beta) < q(\alpha)$ when all players are strictly inequity averse, that is the set N^{IA} is an interval. All outcomes in the set N^{IA} are equilibria of the symmetric Cournot game with inequity aversion since for any profile of quantities, q_{-i}, the quantity q_i belongs to the best response of player i, $i = 1, \ldots n$. *Q.E.D.*

Proof of Proposition 3: The quantity produced by each player in the largest Nash equilibria of N^{IA} is given by $q(\alpha) = \min\left[q(\alpha_1), \ldots, q(\alpha_n)\right].$ The largest Nash equilibria of N^{IA} is nondecreasing in α since $\min\left[q(\alpha_1), \ldots, q(\alpha_n)\right]$ is nondecreasing in α. Similarly, the quantity produced by each player in the smallest Nash equilibria of N^{IA} is given by $q(\beta) = \max\left[q(\beta_1), \ldots, q(\beta_n)\right].$ The smallest Nash equilibria of N^{IA} is nonincreasing in β since $\max\left[q(\beta_1), \ldots, q(\beta_n)\right]$ is nonincreasing in β. *Q.E.D.*

Proof of Proposition 4: When all players are strictly averse to inequity we have $q(\beta) < q^N < q(\alpha)$. Since α_i is drawn from a uniform distribution with support on $[0, 1]$, the larger is n the most likely it becomes that $\min(\alpha_1, \ldots \alpha_n)$ is closer to zero, that is, $N(\alpha)$ is closer to q^N. Similarly, since β_i is drawn from a uniform distribution with support on $[0, 1]$, the larger is n the most likely it becomes that $\min(\beta_1, \ldots, \beta_n)$ is closer to zero, that is, that $N(\beta)$ is closer to q^N. *Q.E.D.*

Proof of Proposition 5: The quantity $q_i(\alpha_i)$ is the interception of $t_i(q_j)$ with the equal profit curve. From the definition of $t_i(q_j)$ we have

$$(1 + \alpha_i)\left[P'(Q)q_i + P(Q) - C_i'(q_i)\right] - \alpha_i P'(Q)q_j = 0.$$

In the equal profit curve we have $q_j = e_j(q_i)$. So,

$$(1+\alpha_i)\left[P'(q_i+e_j(q_i))q_i + P(q_i+e_j(q_i)) - C_i'(q_i)\right] - \alpha_i P'(q_i+e_j(q_i))e_j(q_i) = 0.$$

Similarly, the quantity $q_i(\beta_i)$ is the interception of $s_i(q_j)$ with the equal profit curve. From the definition of $s_i(q_j)$ we have

$$(1-\beta_i)\left[P'(Q)q_i + P(Q) - C'(q_i)\right] + \beta_i P'(Q)q_j = 0.$$

In the equal profit curve we have $q_j = e_j(q_i)$. So,

$$(1-\beta_i)\left[P'(q_i+e_j(q_i))q_i + P(q_i+e_j(q_i)) - C_i'(q_i)\right] + \beta_i P'(q_i+e_j(q_i))e_j(q_i) = 0.$$

The rest of the proof is similar to that of Proposition 1. *Q.E.D.*

Proof of Proposition 6: It follows from $e_1(q_2(\alpha_2)) \leq q_1(\beta_1)$ that the best replies of players 1 and 2 only intersect when $q_1 \geq e_1(q_2(\alpha_2))$ and $q_2 \leq e_2(q_1(\beta_1))$. This together with (4.8) implies that the Nash equilibrium is the solution to $q_1 = s_1(q_2)$ and $q_2 = t_2(q_1)$. Hence, player 1 fells compassion of player 2 and player 2 feels envy of player 1. It follows from the definitions of $s_1(q_2)$ and $t_2(q_1)$ that $q_2^S < q_2^{IA} < q_1^{IA} < q_1^S$. *Q.E.D.*

Proof of Proposition 7: It follows from $q_1(\beta_1) < e_1(q_2(\alpha_2))$, that the best replies of players 1 and 2 intersect when $e_2(q_1(\beta_1)) \leq q_2 \leq e_2(q_1(\alpha_1))$ and $e_1(q_2(\beta_2)) \leq q_1 \leq e_1(q_2(\alpha_2))$. This together with (4.8) implies that the set of Nash equilibria is given by (4.11). *Q.E.D.*

.

Bibliography

Abreu, D., 1988. On the Theory of Infinitely Repeated Games with Discounting. Econometrica 56, 383–396.

Aidt, T.S., 1998. Political internalization of economic externalities and environmental policy. Journal of Public Economics 69, 1–16.

Andreoni, J., 1988. Why free ride? Strategies and learning in public good experiments. Journal of Public Economics 37, 291–304.

Armstrong, M., Huck, S., 2010. Behavioral Economics as Applied to Firms: A Primer. Competition Policy International.

Bagwell, K., Staiger, R.W., 1999a. An economic theory of GATT. American Economic Review 89, 215-248.

Bagwell, K., Staiger, R.W., 1999b. Regionalism and multilateral tariff cooperation, in: Piggott, J., Woodland, A. (Eds.), International Trade Policy and the Pacific Rim. Macmillan, London, pp. 157-185.

Bagwell, K., Staiger, R.W., 2002. The Economics of the World Trading System. MIT Press, Cambridge.

Baldwin, R., 1987. Politically realistic objective functions and trade policy. Economic Letters 24, 287-290.

Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. Games and Economic Behavior 10, 122–142.

Barrett, S., 1994. Self-enforcing international environmental agreements. Oxford Economic Papers 46, 878–894.

Barrett, S., 2005. The theory of international environmental agreements, in: Maler, K.G., Vincent, J.R. (Eds.), Handbook of Environmental Economics, Volume 3. Elsevier, Amsterdam, pp. 1457–1516.

Biel, P., 2008. Inequity Aversion and Team Incentives. Scandinavian Journal of Economics 110, 297-320.

Bolton, G., Ockenfels, A., 2000. ERC: A Theory of Equity, Reciprocity, and Competition. American Economic Review, 90, 166–193.

Camerer, C., 2003. Behavioral Game Theory: Experiments on Strategic Interaction. Princeton University Press.

Carroll, C.D., Overland, J., Weil, D.N., 2000. Saving and growth with habit formation. American Economic Review 90, 341–355.

Croson, R.T.A., 2007. Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. Economic Inquiry 45, 199–216.

Dam, K., 1970. The GATT: Law and International Economic Organization. University of Chicago Press, Chicago.

Daughety, A. F., Forsythe, R., 1987a. Industrywide Regulation and Formation of Reputations: A laboratory Analysis, in E. Bailey (ed), Public Regulation: New Perspectives on Institutions and Policies. MIT Press. Cambridge, Mass., pp. 347–398.

Daughety, A. F., Forsythe, R., 1987b. Regulatory-induced Industrial Organization. Journal of Law and Economics 3, 397–434.

Dixit, A., 1987. Strategic aspects of trade policy, in: Truman, B. (Ed.), Advances in Economic Theory: Fifth World Congress. Cambridge University Press, Cambridge, pp. 329–362.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2009. Homo reciprocans: Survey evidence on behavioural outcomes. Economic Journal 119, 592–612.

Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games and Economic Behavior 47, 268-298.

Engel, C., 2007. How Much Collusion? A Meta-Analysis of Oligopoly Experiments. Journal of Competition Law and Economics 3, 491–549.

Englmaier, F. Wambach, A., 2010. Contracts and Inequity Aversion. Games and Economic Behavior 69, 312–328.

Eriksson, C., Persson, J., 2003. Economic growth, inequality, democratization, and the environment. Environmental and Resource Economics 25, 1–16.

Falk, A., Fischbacher, U., 2006. A Theory of Reciprocity. Games and Economic Behavior 54, 293–315.

Fehr, E., Falk, A., 1999. Wage rigidities in a competitive incomplete contract market. Journal of Political Economy 107, 106–134.

Fehr, E., Gächter, S., Kirchsteiger, G., 1997. Reciprocity as a contract enforcement device: Experimental evidence. Econometrica 65, 833–860.

Fehr, E., Kirchsteiger, G., Riedl, A., 1998. Gift exchange and reciprocity in competitive experimental markets. European Economic Review 42, 1–34.

Fehr, E., Schmidt, K., 1999. A Theory of Fairness, Competition, and Cooperation. Quarterly Journal of Economics 114, 817–868.

Ferrera, I., Missios, P., Yildiz, H.M., 2009. Trading rules and the environment: Does equal treatment lead to a cleaner world? Journal of Environmental Economics and Management 58, 206–225.

Feuerstein, S., 2005. Collusion in Industrial Economics–A Survey. Journal of Industry, Competition and Trade 5, 163–198.

Franciosi, R., Kujal, P., Michelitsch, R., Smith, V., Deng, G., 1995. Fairness: Effect on temporary and equilibrium prices in posted-offer markets. Economic Journal 105, 938–950.

Fredriksson, P.G., 1997. The political economy of pollution taxes in a small open economy. Journal of Environmental Economics and Management 33, 44–58.

Freund, C., 2000. Multilateralism and the endogenous formation of preferential trade agreements. Journal of International Economics 52, 359-376.

Freund, C., 2003. Reciprocity in free trade agreements. Policy Research Working Paper #3061, World Bank.

Friedman, J.W., 1971. A non-cooperative equilibrium for supergames. Review of Economic Studies 38, 1-12.

Fuhrer, J.C., 2000. Habit formation in consumption and its implications for monetary-policy models. American Economic Review 90, 367–390.

Gawande, K., Bandyopadhyay, U., 2000. Is protection for sale? Evidence on the Grossman-Helpman theory of endogenous protection. Review of Economics and Statistics 82, 139-152.

Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological Games and Sequential Rationality. Games and Economic Behavior 1, 60–79.

Georgantzis, N., 2006. Testing Oligopoly Theory in the Lab. Anales de Estudios Economicos y Empresariales 16, 37-74.

Goldberg, P.K., Maggi, G., 1999. Protection for sale: An empirical investigation. American Economic Review 89, 1135-1155.

Grossman, G.M., Helpman, E., 1994. Protection for sale. American Economic Review 84, 833–850.

Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 3, 367–388.

Helson, H., 1964. Adaptation Level Theory: An Experimental and Systematic Approach to Behavior. Harper & Row, New York.

Holt, C. A., 1985. An Experimental Test of the Consistent Conjectures Hypothesis. American Economic Review 75, 314–325.

Holt, C . A., 1995. Industrial Organization: A Survey of Laboratory Research, in Kagel, J. H., Roth, A. E. (eds.) The Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 349–443.

Huck, S., Muller, W., Normann, H.-T., 2001. Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets. Economic Journal 111, 749–765.

Huck, S., Normann, H.-T., Oechssler, J., 2004. Two are Few and Four are Many: Number Effects in Experimental Oligopolies. Journal of Economic Behavior and Organization 53, 435–446.

Isaac, R. M., Ramey, V., Williams, A. W., 1984. The Effects of Market Organization on Conspiracies in Restraint of Trade. Journal of Economic Behavior and Organization 5, 191–222.

Kahneman, D., Knetsch, J.L., Thaler, R., 1986. Fairness as a constraint on profit seeking: Entitlements in the market. American Economic Review 76, 728–741.

Keser, C., 1993. "Some Results of Experimental Duopoly Markets with Demand Inertia. Journal of Industrial Economics 41, 133–151.

Krishna, P., Mitra, D., 2005. Reciprocated unilateralism in trade policy. Journal of International Economics 65, 461-487.

Lange, A., Vogt, C., 2003. Cooperation in international environmental negotiations due to a preference for equity. Journal of Public Economics 87, 2049–2067.

Lehmann, D. 2001. The Impact of Altruism and Envy on Competitive Behavior and Satisfaction. International Journal of Research in Marketing 18, 5–17.

Mayer, W., 1984. Endogenous tariff formation. American Economic Review 74, 970-985.

McAusland, C., 2003. Voting for pollution policy: The importance of income inequality and openness to trade. Journal of International Economics 61, 425–451.

Milgrom, P., Roberts, J., 1990. Rationalizability, learning, and equilibrium in games with strategic complementarities. Econometrica 58, 1255–1277.

Neilson, W., 2006. Axiomatic Reference-Dependence in Behavior Toward Others and Toward Risk. Economic Theory 28, 681–692.

Ornelas, E., 2005. Trade creating free trade areas and the undermining of multilateralism. European Economic Review 49, 1717-1735.

Palfrey, T.R., Prisbrey, J.E., 1997. Anomalous behavior in public goods experiments: How much and why? American Economic Review 87, 829–846.

Rabin, M., 1993. Incorporating Fairness into Game Theory and Economics. American Economic Review 83, 1281–1302.

Rassenti, S., Reynolds, S., Smith, V., Szidarovsky, F., 2000. Adaptation and Convergence of Behavior in Repeated Experimental Cournot Games. Journal of Economic Behavior and Organization 41, 117–146.

Rotemberg, J., forthcoming. Fair Pricing. Journal of the European Economic Association.

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study. American Economic Review 81, 1068–1095.

Ryder, H.E. Jr., Heal, G.M., 1973. Optimal growth with intertemporally dependent preferences. Review of Economic Studies 40, 1–31.

Saggi, K., Yildiz, H.M., 2010. Bilateralism, multilateralism, and the quest for global free trade. Journal of International Economics 81, 26-37.

Santos-Pinto, L., 2008. Making Sense of the Experimental Evidence on Endogenous Timing in Duopoly Markets. Journal of Economic Behavior and Organization 68, 657–666.

Sappington, D., Desiraju, R., 2007. Equity and Adverse Selection. The Journal of Economics and Management Strategy 16, 285–318.

Segal, U., Sobel, J., 2007. Tit for tat: Foundations of preferences for reciprocity in strategic settings. Journal of Economic Theory 136, 197–216.

Selten, R., Mitzkewitz, M., Uhlich, G. R., 1997. Duopoly Strategies Programmed by Experienced Players. Econometrica 65, 517–555.

Sobel, J., 2005. Interdependent Preferences and Reciprocity. Journal of Economic Literature 43, 392–436.

Tversky, A., Kahneman, D., 1991. Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics 106, 1039–1061.

Vives, X., 1999. Oligopoly Pricing: Old Ideas and New Tools. MIT Press, Cambridge.

Wagner, U.J., 2001. The design of stable international environmental agreements: Economic theory and political economy. Journal of Economic Surveys 15, 377–411.