# Essays in International Trade Policy and Social Choice under Ambiguity

A dissertation submitted
to the Faculty of Business Administration and Economics
of Bielefeld University
in partial fulfillment of the requirements
for the doctoral degree in Economics

**Lasha Chochua**

First Supervisor: Prof. Gerald Willmann

Second Supervisor: Prof. Hale Utar

October, 2018

Abstract. In the first chapter of the thesis, We discuss a general context and provide an overview of the dissertation. Furthermore, we outline the contributions.

In the second chapter, based on the Protection for Sale approach of Grossman and Helpman (1994), we develop a theoretical model in which exogenously organised groups provide political contributions to influence trade policy. The incumbent government cares about contributions, yet at the same time, takes into account the reactions of voters. We formally consider voting decisions of citizens, assuming they have heterogeneous ignorance thresholds, and explicitly derive the objective function of the policy-maker. We find that the resulting equilibrium structure of protection differs from the standard case. Free trade obtains only if no group lobbies and ignorance levels are the same. On average, more ignorant groups will have lower (if any) protection from the policy-maker and, also, groups represented by lobbies will not always be supported by the incumbent government.

In chapter three, we study and compare the stability of trade policy arrangements in two different regulatory scenarios, one with and one without Preferential Trade Agreements (PTAs), i.e. current vs modified WTO rules. Unlike the existing literature, our work considers an extensive choice set of trade constellations, containing both available PTAs, Customs Unions (CUs) and Free Trade Agreements (FTAs), as well as Multilateral Trade Agreements (MTAs), while assuming unlimited farsightedness of the negotiating parties. With symmetric countries and under both the current and the modified WTO rules, the Global Free Trade (GFT) regime emerges as the unique stable outcome. In the case of asymmetry, the results are driven by the relative size of the countries. If the world is in the vicinity of symmetry and two out of three countries are close to identical while relatively smaller than the other one, the area where the GFT regime is stable increases when prohibiting PTAs. However, when two similar countries are relatively larger, the availability of PTAs is conducive to the stability of the GFT regime. Finally, if the world is further away from symmetry, full trade liberalisation is not attainable at all, and an area where the Most-Favoured-Nation (MFN) regime is stable appears in the scenario without PTAs. Thus, the direction of the effect of PTAs on trade liberalisation depends on the degree of asymmetry among countries.

In the last chapter, we deal with normative social decision making under ambiguity. Imagine an individual facing a decision problem affecting society as a whole and suppose that the choice features uncertainty in the form of belief systems for each member of the group. By combining the concept of an impartial observer with that of second-order beliefs, this chapter derives a representation of the preferences of an individual in such a situation using an axiomatic approach. It ultimately extends (a generalised version of) Harsanyi's Impartial Observer Theorem (1953) to include uncertainty by building on the work of Grant et al. (2010) and Seo (2009).

In memory of my father.
To my mother and my sister.

# Contents

# List of Figures

# List of Tables

## Acknowledgement

This thesis is not an outcome of an individual journey. Many people around me contributed to it in various ways, and all of them are important to me. They made my time as a PhD student very comfortable and smooth.

Firstly, I would like to express my sincere gratitude to my advisor Prof. Gerald Willmann for the continuous support of my PhD study and related research, for his patience, motivation, knowledge and friendly attitude. His guidance helped me in all the time of research and writing of this thesis. But what is the most crucial for me, I have always felt his trust, which was motivating me the most. My sincere thanks also go to my second advisor Prof. Hale Utar for her discussions, guidance and assistance.

Besides my advisors, I would like to thank Prof. Herbert Dawid for his insightful comments, which have always challenged my co-author and me. Also, I am very thankful to Prof. Frank Riedel for reading our paper on the impartial observer and giving constructive comments. My thanks also go to Prof. Christoph Kuzmics for his inspiring approach to science in general and his discussions on the second chapter of this thesis.

My time as a PhD student would have been far less exciting without my co-authors: Stefan Berens and Giorgi Papava. Long discussions, debates, "collective" thinking, and what's the most important to me their friendship were the most significant ingredients while working on the thesis. I am also very thankful to all of the PhD students I have interacted here at Bielefeld University.

I was lucky to work in Free University at Tbilisi before coming to Bielefeld and then conducting on-line lectures from Bielefeld at 6:00 a.m. for the first three years of my stay here. My interactions with Mr Kakha Benduqidze and my students there are a precious experience. Also, I am very thankful to PMCG and Mr Aleksi Aleksishvili for giving me the possibility to apply my knowledge to some practical economic issues in Georgia.

There are so many people whose support and friendship were critical to me. Lika Nadiradze, Salome Gvetadze, Levan Chachava, Mariam Lortkipanidze, Natia Khantadze, Maka Chitanava, Tamar Jugheli, Salome Baslandze, Lasha Meskhia, Zaza Chanturia, Lasha Bokuchava, Quji Bichia, and many others. Thank you all!

Last but not the least, I would like to thank my family: my mother, my sister and her family: Maro, Gio and Levan!

CHAPTER 1

# Introduction

Broadly speaking this thesis is a work in applied economic theory. We extensively exploit the theoretical tools to examine economic problems of high practical significance. In chapters 2, we develop a theoretical model in the context of international trade to understand the endogenous trade policy formation process. The focus is on the domestic parties involved. Chapter 3 deals with the issue of preferential vs multilateral liberalisation under the rules of the World Trade Organisation, implying the change of our focus from a local to a global perspective. In both chapters, the strategic interactions among the decision-makers play a crucial role in determining the trade policy. In chapter 4, we abstract from strategic interactions and contribute to the normative theory of social decision-making under ambiguity.

## 1.1. Context and overview

The famous economic historian Douglas Irwin[1] starts his book on the history of the U.S. trade policy with the words of Madison from *Federalist 10*. There Madison says "A landed interest, a manufacturing interest, a mercantile interest, a moneyed interest, with many lesser interests, grow up of necessity in civilised nations, and divide them into different classes, actuated by different sentiments and views. The regulation of these various and interfering interests forms the principal task of modern legislation, and involves the spirit of party and faction in the necessary and ordinary operations of the government... Shall domestic manufactures be encouraged, and in what degree, by restrictions on foreign manufactures? are questions which would be differently decided by the landed and the manufacturing classes, and probably by neither with a sole regard to justice and the public good...It is in vain to say that enlightened statesmen will be able to adjust these clashing interests, and render them all subservient to the public good."[2]

Madison was correct pointing to the importance of clashing interests in determining the policy outcomes, but do governments, in general, adjust various benefits and "render them all subservient to the public good"? Special interest groups, voters, policy-makers, all of them have their interests. How are these interests aggregated when a government decides the policy outcome? This is the central question that we aim at answering in the context of international trade policy in chapter 2.

[1] see Irwin (2017)

[2] The Federalist Papers : No. 10. `http://avalon.law.yale.edu/18th_century/fed10.asp`

McLaren (2016), reviewing the literature on commercial policy, writes: ”In a democracy, politicians campaign to win an election and thereby gain power; once they have gained power, they bargain with each other while being influenced by lobbyists, and the result is a realised policy. A full model would involve all three of the mechanisms, (i) electoral competition, (ii) legislative bargaining, and (iii) lobbying; but in practice, these three have tended to be studied separately.”

In chapter 2, we build a model where lobbying and electoral competition are integrated and, additionally, we consider voters' heterogeneous ignorance thresholds, as their reactions to trade policy choices differ in practice (See Ponzeto (2011) and Wegenast (2010)). How the voters respond to policy choices is an important determinant neglected in political economy models of endogenous trade policy. If citizens are ignorant, in other words, if they cannot map policy decisions to their welfare, then it is costless for policy-makers to design and implement a policy that is favourable only for a specific interest group. But if the voters can map the decisions of politicians on the changes to their welfare, then the policy-makers do not possess such freedom in policy choices. Voters' potential reactions constrain them. In the end, when a policy-maker chooses a trade policy, his/her motivation to win the election forces a policy-maker to base the decision on a clear trade-off between the interests of special interest groups and those of voters.

In our model, organised groups provide political contributions to the incumbent government to influence its decision on trade policy. The incumbent is exogenously given, and it faces the utilitarian challenger in the upcoming election. Therefore, the policy-maker cares about donations, but at the same time, takes the possible reactions of voters into consideration. Contrary to the tradition in the literature, we formally define the voting decision of citizens, assuming they are heterogeneously ignorant and explicitly derive the objective function of a policy-maker. Then we study the structure of the protection that arises in the political equilibrium of the model.

There are two stages in the model: in the first stage, lobbyists, representing organised groups, non-cooperatively and simultaneously decide on contribution functions contingent on trade policy choices of the incumbent government. In the second stage, the incumbent government observes the contribution functions offered by lobbyists, takes the possible responses from the voters into consideration, sets a trade policy and collects from each lobby the contribution associated with its policy. At the end an election takes place. The citizens cannot abstain from voting.

At large, in the political equilibrium of a trade policy game that we consider, the free trade regime, even though it is a social welfare maximising policy in a small, competitive economy, is less likely as it is attainable only in a particular case. So, in our model politicians do adjust the clashing interests, but do not necessarily "render them subservient to the public good."

In addition to standard import/export taxes/subsidies, trade agreements are another commonly used instruments in the hands of the governments to conduct trade policy. In 1860 the first trade agreement, the Anglo-French commercial treaty was inked. This treaty was significant in many respect, but in connection to the second chapter of the thesis, one crucial political economy reason stands out. The agreement allowed the French Emperor "to circumvent domestic protectionist interests."[3] The clashing domestic interests obviously might be one of the reasons why countries sign trade agreements and, in general, understanding the motives for trade agreements is a captivating undertaking and subject of the vast literature, but in chapter 3 we explore different aspects of trade agreements.

The Anglo-French commercial treaty of 1860 unquestionably was the turning point. A purely bilateral arrangement between Britain and France sparked the interest of other European countries to conclude similar agreements with their trading partners all over Europe. Within 15 years an additional 56 trade treaties went into effect and "by 1875, virtually all of Europe was party to a low-tariff zone by dint of a web of agreements that included the linchpin MFN (most-favored-nation) clause."[4] Though, World War I ended the bilateral trade regime of the 19th century. Tariff and non-tariff measures became the conventional instruments of trade policies of all major countries, and despite a few attempts, nothing significant has changed until the end of World War II. Only in 1944, the Bretton Woods conference offered new contours of the postwar economic order. The General Agreement on Tariffs and Trade (GATT) of 1947 was one of the mainsprings of this new economic worldview.

Following the GATT, later the World Trade Organisation (WTO), an increasing number of signatory countries liberalised their trade policies primarily via two channels: bilateral (discriminatory) and multilateral (non-discriminatory) negotiations.[5] To the present day, there have been eight rounds of multilateral trade negotiations[6] with the current ninth one, the Doha Round, still ongoing. At the same time, parallel to the arrangements observed on the multilateral level, the world has seen an ever-increasing number of Preferential Trade Agreements (PTAs) mainly in the wake of bilateral negotiations. Currently, about forty percent of all countries/territories are a member of more than five PTAs while about a quarter participates in more than

[3]see Irwin (1993)

[4]see Grossmann (2016)

[5]According to Article I of the GATT (MFN principle): Any concession granted to one member needs to be extended to all other members of the WTO. Contrary to the core MFN principle of GATT Article XXIV explicitly allows countries to form PTAs, specifically Customs Unions (CUs) and Free Trade Agreements (FTAs). Under PTAs countries do not need to extend the concessions granted within the arrangement to other countries.

[6]Under such negotiations agreements reached among the set of countries are called Multilateral Trade Agreements (MTAs).

ten. In chapter 3, it is our intent to examine whether PTAs act as 'building blocks' or 'stumbling blocks' on the path towards global free trade.

The post World War II experience shows that countries have extensively engaged in both multilateral and preferential trade liberalisation simultaneously, but in contrast with this historical background, the existing theoretical literature on the effects of PTAs on the global trading system usually considers only a limited selection of possible trade policy arrangements. Therefore, in a large extent, the literature neglects the full story of strategic interactions among the policy-makers when they decide on the mode of liberalisation. Besides, in many cases, the game-theoretic solution concepts used in the literature assume shortsighted or limited farsighted decision-makers. In chapter 3 we analyse the impact of PTAs on the global trading system under the choice set of trade policy arrangements as extensive as in reality and with the assumption of farsighted policy-makers.

To our mind, the farsightedness of policy-makers is an essential ingredient of the reality in addition to the extensive set of trade policy constellations. The literature suggests that on average it takes 3-4 years between the start of negotiations and the implementation of PTAs (see Mölders (2015) and Freund and McDaniel (2016)). Such trade negotiations are complicated processes with a significant effect on the countries' economies, and they are accompanied by elaborate studies about the feasibility and possible scenario analysis. Therefore, it is natural to assume that the policy-makers take into consideration the long-run effects of the full strategic interactions among involved parties.

To account for the farsightedness of policy-makers, we employ the solution concept of the largest consistent sets (LCS) proposed by Chwe (1994). The solution concept captures the foresight based on the indirect dominance relation, according to which decision-makers care for the final outcomes that their actions may lead to. The sequence of moves is determined by the "effectiveness relations" which shows what all possible coalitions can do. No pre-determined restrictions are placed on the effectiveness relations and their nature is determined by the problem to be analysed. Furthermore, the LCS allows unrestrained negotiations and imposes little structure on the negotiation processes. The membership of a coalition is not binding as well. In the end, according to the definition of the LCS, a coalition rejects or deviates from the outcome only if its deviation leads only to alternatives that benefit its members. To our mind, the properties of the LCS are well-suited for the analysis of the problem of interest and it is superior to other solution concepts used in the literature as well. For example, with respect to 'Coalition-Proof Nash Equilibria' Bernheim, Peleg and Whinston (1987) note that their notion of self-enforceability is too restrictive in one crucial aspect as "[it] rules out the possibility that some member of the deviating coalition might form a pact to deviate further with someone not included in this coalition." Importantly, this limitation does not affect the concept of LCS. Furthermore, the notion of the 'Stable Sets' developed by Neumann and Morgestern (1944) is based on the concept of the direct dominance; thus, it is shortsighted. The same applies to the solution concept of 'Strong Nash Equilibria' developed by Aumann (1959). Another alternative solution concept developed by Ray and Vohra (2015) is limited as well as their definition of the effectiveness relations is too restrictive for the context we study.

In chapter 3, we utilise a three-country, general equilibrium trade model with endowments and perfect competition via exports. The model is taken from Saggi and Yildiz (2010), which itself is a modification of the model in Bagwell and Staiger (1997). The underlying trade model is an endogenous device based on which countries determine their optimal trade policies (tariffs) and rank all possible trade arrangements that might arise. The preference rankings and "effectiveness relations" then determine the indirect dominance relations, which in combination with preference rankings, at the end, establish the stable set of the trade policy arrangements. The trade policy constellations in our model is of one of four types: MFN,[7] CU, FTA, and MTA. So, we consider all relevant possibilities under the WTO.

The answer to the question whether PTAs are 'building blocks' or 'stumbling blocks' on the path towards global free trade is not as straightforward as one would like it to be. In the end, the results presented in chapter 3 are mixed and depend on the size distribution of the countries. Under symmetry, GFT is the unique stable trade constellation in both regulatory scenarios, with and without PTAs. But as soon as one moves away from symmetry, GFT might not be reached at all. In between, the effect of switching off the availability of PTAs depends on the exact asymmetry. In case two countries are relatively smaller, prohibiting PTAs increases the area of stability of the GFT regimes. When two countries are relatively larger, it reduces the area. Once the world is further away from symmetry, abolishing the exception for PTAs might result in the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime. Therefore, under such circumstances, PTAs act as a mechanism that prevents the MFN regime.

With chapter 3 we finish analysing the elements of international trade policy. In the final section of the thesis, we build a model to deal with a rational theory of societal judgements under ambiguity.

In the mid of the previous century, Harsanyi (1953, 1955, 1977) developed a constructive approach to put the ethical theory in the modern Bayesian theory of rational behaviour. He used to divide the general theory of rational behaviour into individual decision theory, ethics and game theory. As by that time there were

[7]MFN trade arrangement corresponds to the situation, when countries choose individually rational, non-cooperative Nash policies under the most-favoured-nation principle.

strong axiomatic foundations for individual decision making and game theory, his desire to put ethics in the same set-up seemed a relevant aspiration. By combining Adam Smith's hypothetical construct of the impartial spectator with Kant's principle of universality and utilitarian tradition, indeed Harsanyi successfully managed to achieve his goal. In a nutshell, Harsanyi argued that if in a society individuals are facing risky prospects over social outcomes and a hypothetical lottery over identities and if each imagines himself/herself as an impartial observer, then they should be (weighted) utilitarians.

Individuals, as members of society, continually face choices among moral rules, institutional arrangements, government policies or patterns of wealth distribution; therefore, they are repeatedly involved in value judgements about which social alternative to choose. The areas of application for Harsanyi's theory is vast. However, Harsanyi's Impartial Observer Theorem only considers scenarios where each of the involved individuals faces objective risk. It is a theory analysing societal judgements when each member of society knows objective probabilities over a set of social outcomes. But what if the members of a society do not have an objective probability distribution over the possible social outcomes. In chapter 4 our goal is to extend Harsanyi's Impartial Observer Theorem including Knightian Uncertainty in the model by introducing individual belief systems about the likelihood of the social outcomes (which the impartial observer necessarily takes into consideration).

Our framework is based on the generalised version of Harsanyi (1953) by Grant et al. (2010), which accommodates common criticism of Harsanyi's approach, specifically the issue of fairness and attitude towards mixing. The introduction of individual belief systems to our framework follows Seo (2009).

During working on the chapter, we had to decide which axiomatic approach to use for modelling preferences. Existing models in this stream of literature essentially differ in the choice of the domain of preference. Seo (2009) takes the domain of Anscombe and Aumann (1963) and a similar axiomatic foundation. In Seo (2009) the preferences are defined over the set of act lotteries (an act is a measurable function from the state space into the set of outcomes). Another famous alternative approach of Klibanoff, Marinacci and Mukerji (2005) by contrast requires two domains with preferences. There an individual has a preference over the acts on a state space $S$ and preference over another state space $\Delta S$, the set of all probability measures over the state space. The elements of $\Delta S$ are called second-order acts. Similarly to Klibanoff, Marinacci and Mukerji (2005), Nau (2006) and Ergin and Gul (2009) also assume state space bigger than Seo. To our mind, the domain selection of Seo allows us to introduce uncertainty to the (generalised) framework without any additional modifications. Moreover, the informational requirement for the hypothetical construct of the impartial spectator is minimal under the approach of Seo.

In the end, Our main result shows that the impartial observer's preferences admit a representation in the form of a weighted average of the individual (transformed) second-order subjective expected utilities.

Apart from the axiomatic derivation of the impartial observer's utility representation under ambiguity, in chapter 4 we provide two illustrative examples of our approach. In the first example, we consider the famous case in moral philosophy known as the moral dilemma of the 'Afghan Goatherds'. While in the second example, we examine a simple exchange economy with two goods and two individuals, with each of them receiving endowments and compare two alternative re-distributions rules, namely the Walrasian auctioneer and the Egalitarian rule when ambiguity is presented. The second example is primarily from the realm of constitutional design. Both examples demonstrate the importance of individual belief systems and thus justifies the introduction of uncertainty to the framework.

## 1.2. contributions

Chapter 2 of the thesis is joint work with George Papava. He was a PhD student at Chicago University when we actively worked on the project. Chapter 3 and chapter 4 of the thesis are joint works with Stefan Berens, who is a PhD student at Bielefeld University.

All projects in this thesis are results of many hours and days of discussions and debates, which were the most interesting and exciting part of my time while being a PhD student here. We have many times changed the focus of the projects and, in the end, jointly developed them. In chapter 3 and 4, Stefan Berens' expertise in programming and his mathematical background were of great benefit.

All work was carried out jointly, and all authors contributed equally to the relevant chapters of the thesis. Besides, We have greatly benefited from participating in different conferences, workshops and research meetings. Comments from Professors at Bielefeld University largely contributed to the improvement of the chapters in this thesis as well.

Except where otherwise indicated, this thesis is my and my co-authors' original work.

CHAPTER 2

# Endogenous Trade Policy in the presence of Lobbying and Heterogeneously Ignorant Voters

## 2.1. Introduction

The idea that under democracies trade policy outcomes are the consequence of the interactions between organised groups and policy-makers is not new among economists. Many policy outcomes have been studied and explained by analysing the interest groups' activities behind the processes. According to Anderson and Tollison (1985), the repeal of the import duties on corn by the House of Commons in 1846 was largely due to Anti-corn Law League's activities backed by a cohesive and well-defined organised group, the British cotton textile industry.

More recently, after eliminating the restrictions on the purchase of sugar from Mexico under the North American Free Trade Agreement in 2008 the USA government, from the year 2014, has experienced well-organised influence from the U.S. sugar producers against the decision, accusing the Mexican producers of price dumping. As a result, within less than a year, the United States and Mexico reached an agreement to avoid anti-dumping and countervailing duties on U.S. sugar imports from Mexico and the deal reintroduced the limitations on trade. More precisely, the U.S. government brought into the minimum prices for raw and refined sugar imports as well as volume and timing restrictions. It should be noted that sugar and sugarcane farms account only for 1.3 per cent of the value of total farm and livestock production in the U.S.A, while they provide 33 per cent of crop industries' total campaign donations and 40 per cent of crop industries' total lobbying expenditures.[1]

The special interest groups' activities are one side of the story. The policy-makers care about the political contributions, yet at the same time, they acknowledge that the voters' potential reactions to the policy choices are important enough to be taken into consideration. Indeed, Wegenast's (2010) empirical study does find that informed electorate reduces the total amount of campaign contributions, possibly as a result of the reduced freedom of policy-makers to design policies favourable to some interest groups. Furthermore, Ponzeto (2011) shows that more news coverage of trade policy in a given industry increases the demand for trade liberalisation. He demonstrates that non-tariff barriers are used less in the industries which have more media coverage. Bhagwati (1985) coined the term "Dracula Effect" in his book "Protectionism"

[1] see Bryan (2014).

to describe the role of information in a trade liberalisation process: "the mere act of recognising [protectionism] will help trigger a more corrective response. In the matters, we can count on assistance from...the Dracula Effect: exposing evil to sunlight helps destroy it."

The organised groups offer political contributions to influence policy outcomes. The policy-makers are interested in such contributions as they have to finance their election campaigns. If citizens are ignorant, in other words, if they cannot map policy decisions to their welfare, then it is costless for a policy-maker to design and implement a policy that is favourable only for a specific organised group and influence citizens' voting decisions through campaign expenditures. But if the voters can map the decisions of the politicians to the changes in their welfare, then the policy-makers do not have such freedom in policy choices. Voters' potential reactions constrain them.

The lobbying activities have occupied the stage of research for many years. There are vast empirical and theoretical literature analysing the role of special interest groups in policy determination process in general,[2] but little or improper attention has been dedicated to the importance of voters' ignorance in the political economy models of trade policy formation.[3]

In this chapter, we build a theoretical model in which organised groups provide political contributions to influence the incumbent government's decisions on trade policy. The incumbent government is exogenously given, and it faces the utilitarian challenger in the upcoming election. Therefore, the policy-maker cares about donations, but at the same time, takes into consideration the possible reactions of voters. In the model we formally define the voting decisions of citizens, assuming they are heterogeneously ignorant and explicitly derive the objective function of a policy-maker. Then we study the structure of the protection that arises in the political equilibrium of the model.

We get that free trade prevails only if lobbyists represent none of the groups owning specific factors and, at the same time, all groups share the same ignorance density. Moreover, if all groups are unorganised, our model predicts that groups with average ignorance level lower than overall average ignorance will receive positive government protection and vice versa. Contrary to the standard result of Grossman and Helpman (1994), organised groups are not always supported in the political equilibrium of our model and, depending on the parameters' values, the policy-maker

[2]For example, see Goldberg and Maggi (1999), Gawande and Bandyopadhyay (2000), Mitra et al. (2002), Tovar (2009), Bombardini and Trebi (2012) and Imai et al. (2013).

[3]To our knowledge, Downs (1957) was one of the first who acknowledged the importance of voters' ignorance in the political process. In his seminal book, Economic Theory of Democracy, he writes: "Many citizens who vote and consider voting important are nevertheless not well-informed on the issues involved in the election." (p.298)

might support unorganised groups as well. On average, more ignorant groups will have lower (if any) protection from the government.

The chapter is organised as follows. First, we briefly discuss the relevant literature and outline the major approaches modelling the political process in determining a trade policy. Then we present the general set-up for the model and elaborate on the technical details. Next, we explicitly model the decisions of the agents involved, define equilibrium and analyse the results. In the end, we present a summary and concluding remarks.

## 2.2. Related Literature

Over last 50 years several approaches have been developed to explain the endogenous formation of trade policy.[4] Here we briefly discuss the seminal papers that mostly influenced the directions of theorising in the literature.

Based on the works of Stigler (1971) and Peltzman (1976), Hillman (1982) developed a model where the choice of tariffs could have been viewed as the solution to the government's trade-off between the political support from industry and the dissatisfaction of consumers. In the model, rising domestic prices through trade taxes benefit the industry profits and provide a basis for more political support from that industry, but at the same time, the higher prices, caused by the free trade distortions, deteriorate consumers' welfare. According to Hillman (1982), this trade-off between opposing interests determines the tariff structure, as government maximises aggregate support. The problem with the approach is the fact that a reduced-form function represents the aggregate support and the model lacks the micro foundations.

For explaining trade policy outcomes, Mayer (1984) developed a majority voting model. According to his approach, the ownership structure of the economy and a median voter determine tariff rates. The higher the median voter's share of ownership of the sector-specific input and the larger the sector is in terms of output, the higher the tariff rate will be. But as Helpman (1995) notes the main shortcoming of the model is the fact that if we consider a highly concentrated distribution of ownership of the specific factors, which is not unrealistic, then we should not observe tariffs in such sectors.

Magee, Brock and Young (1989) developed an electoral competition model for understanding the endogenous formation of trade policy. They construct the model where two candidates and two lobbies interact. One candidate is assumed to be pro-trade, while the other one is pro-protection. Lobbies give contributions to increase the probability of winning for their candidate. A two-stage game is considered. In the first stage, candidates commit to their trade policies, and in the second stage,

[4]For detailed review of the literature see Helpman (1995), Rodrik (1995), Winden (2004) and McLaren (2016)

lobbies decide on contributions. The Nash equilibrium of the game identifies the rates of protection. Magee, Brock and Young's model is not explicit about the political process, who the voters are and how they vote. Besides, the restriction on candidates' platforms is artificial. For these reasons their approach was vigorously criticised (for example, see Austen-Smith (1991) and Mayer and Li (1994)).

Magee, Brock and Young (1989) were the first who explicitly consider the role of political contributions, albeit with strong restriction, as in their model contributions influenced only the election outcomes and not the choice of trade policy. Grossman and Helpman (1994) have developed a theory where the influence motive of campaign contributions plays the central role. According to this approach, interest groups offer politicians campaign contributions contingent on policy choices. Subsequently, the politicians choose the policy to implement knowing how the policy choice affects the decisions of organised groups. However, contributions are not the only factor that politicians take into consideration; they also care for the well-being of the general public. The political objective function is a weighted sum of total political contributions and aggregate welfare, where the weights are exogenously given. Based on the work of Bernheim and Whinston (1986), the authors analyse a two-stage game and determine the equilibrium outcome. Like Magee, Brock and Young (1989), Grossman and Helpman (1994) do not provide an explicit picture of the political process. Moreover, in the most of the models discussed above, full information assumption is made.[5]

Somewhat related to our work is the paper of Ponzeto (2011). He models tariff formation as the outcome of an electoral competition, where each agent endogenously acquires information about his sector of employment. In equilibrium trade policy for an industry is less protective when there is more public information about it. Like Ponzeto (2011), we also consider the informational aspect of trade policy in voting decisions of citizens, though we treat it as exogenously given. Moreover, lobbying activities play the significant role in our work contrary to Ponzeto (2011), where it is absent.

In our work, we retain the basic structure of Grossman and Helpman's (1994) model and their notion of equilibrium, but we explicitly model the political process, assuming that voters are heterogeneously ignorant about the effects of policy choices. Based on the voting decision of citizens we derive the governmental objective function when voting is probabilistic. Adding these new features allows us to formulate an explicit micro-founded model of endogenous trade policy determination.

[5]Several papers consider the role of information in the policy determination process, but the approaches are dichotomous in the sense that the authors only consider informed and uninformed voters (for example, see Baron (1994), Grossmann and Helpman (1996) and Bombardini and Trebi (2011)).

## 2.3. Model

**2.3.1. General Framework.** We consider a small open economy. The economy is populated by individuals who are assumed to have identical preferences, and each faces the following utility maximisation problem:

$$U(x) = x_0 + \sum_{m=1}^{M} u_m(x_m) \quad \text{s.t.} \quad x_0 + \sum_{m=1}^{M} p_m x_m = E \tag{2.1}$$

where $x_0$ is the consumption of a numeraire good produced by labour alone. Production technology for a numeraire good is constant returns to scale, and input-output coefficient equals to one. Furthermore, both the world and the domestic price of a good 0 is assumed to be equal to one. The aggregate supply of labour is large enough to ensure a positive supply of good 0. Note that given assumptions imply the equilibrium wage rate to be one as well. $M$ non-numeraire goods are produced and $p_m$ is the domestic price of good $m$, where $m \in M$. Since a small open economy is considered, for each good $m$ there exists an exogenously given world price, which we denote by $p_m^w$. The production technology for all non-numeraire goods is constant returns to scale and, contrary to numeraire good, manufacturing of each good $m$ requires labour, $L_m$ and a sector-specific input, $K_m$. The size of the total population is $N$, and every individual $i = (1, 2, ..., N)$ owns at most one specific factor of production; hence, each belongs to only one group of people owning exactly one type of specific factor. The number of people in the groups can be various, which is denoted by $N_m$ and $N = \sum_{m=1}^{M} N_m$. $E$ is the total income of an individual. The sub-utility function in equation (2.1) is differentiable, increasing and strictly concave.

The first order conditions of an individual's utility maximisation problem imply that:

$$x_m = d_m(p_m) = u_m^{'-1}(p_m) \quad \text{and} \quad x_0 = E - \sum_{m=1}^{M} d_m(p_m)p_m$$

Then indirect utility for any individual can be written as:

$$v(p, E) = E + \sum_{m=1}^{M} u_m(d_m(p_m)) - \sum_{m=1}^{M} d_m(p_m)p_m \tag{2.2}$$

where $\sum_{m=1}^{M} u_m(d_m(p_m)) - \sum_{m=1}^{M} d_m(p_m)p_m$ represents a consumer surplus. We denote it by the following notation $\sigma(p)$.

While equilibrium wage rate is one, the domestic price solely determines the reward to the specific factor employed in the production of good $m$. Let's denote this reward by $\pi_m(p_m)$, then it is defined as follows: $\pi_m(p_m) = max_{L_m}(p_m f(K_m, L_m) - L_m)$. By Hotelling's lemma, domestic output, or supply, of good $m$ will be the first derivative of the reward function: $y_m(p_m) = \pi_m^{'}(p_m)$.

In the economy, the incumbent government determines the trade policy. Trade taxes and subsidies are the only policy instruments that politicians can deploy. A domestic price in excess of the world price means an import tariff for imported goods and an export subsidy for exported goods, while local prices below the world prices are equivalent to import subsidies and export taxes. Furthermore, all income generated by trade policy is evenly allocated among citizens. Therefore, per capita tariff revenue from the government denoted as $r(p)$ will be:

$$r(p) = \sum_{m=1}^{M} (p_m - p_m^w)(d_m(p_m) - \frac{1}{N} y_m(p_m)) \tag{2.3}$$

In general, individuals can get income from three sources: labour income, income from owning a share of a specific factor and tariff revenues from the government. Then the indirect utility of individual $i$ holding some specific factor $m$, assuming that individuals' share in specific factor ownership is evenly distributed within groups[6], is:

$$v_i^m(p, E_m) = l_i + \frac{1}{N_m} \pi_m(p_m) + \sigma(p) + r(p) \tag{2.4}$$

The indirect utility functions will play a crucial role in the decision-making process while voting. A policy-maker cares about the votes, but at the same time, he/she needs to collect contributions from the business groups to run the election campaign. We assume that some business groups, more precisely, some of the owners of specific factors are exogenously organised and represented by a lobbyist, who decides how much to contribute for maximising the reward to the group they represent. There is an incumbent running the government, which is exogenously given and it faces a trade-off between contributions and votes. In the end, precisely this trade-off will determine the political equilibrium of the game.

The political process of trade policy determination runs through two stages:

In the first stage, lobbyists, representing organised groups, non-cooperatively and simultaneously decide on contribution functions contingent on trade policy choices of incumbent government.

In the second stage, the incumbent government observes the contribution functions offered by lobbyists, takes into consideration the possible responses from the voters, sets a trade policy and collects from each lobby the contribution associated with its policy. At the end the election takes place. The citizens cannot abstain from voting.

**2.3.2. Political Process and Decisions of Agents.** We have already defined the indirect utility function of individual $i$ owning some specific factor $m$. In order to determine how citizens vote, let's consider the indirect utility evaluated at the world

[6]The ownership structure in our model is slightly different from Grossman and Helpman (1994). There individuals might not own any specific factor.

prices:

$$v_i^m(p^w, E_m^w) = l_i + \frac{1}{N_m}\pi_m(p_m^w) + \sigma(p^w) \tag{2.5}$$

$v_i^m(p^w, E_m^w)$ shows an indirect utility of an individual $i$ owning some specific factor $m$ in the case there is not any distortion to free trade. This variable indicates what would have been the welfare of agent $i$ under the world prices.[7]

In our model the voters differ from one another in the ability/desirability to determine the relationship between the trade policies and the difference in their indirect utilities, $v_i^m(p, E^m) - v_i^m(p^w, E_w^m)$. One might think that citizens have a different level of ignorance for policies that deviate from socially optimal strategy (see footnote 7). In practice, there could be several potential reasons why citizens are ignorant with respect to policies. In some cases, it might be that the different levels of ignorance are just outcomes of rational choices of agents or citizens might have different thresholds of cognitive limitations,[8] or individuals are miscellaneously limited by the information they have.[9] The other reason could be the trade preferences of individuals, including the biases of the agents.

For our theoretical model, it does not matter what the source of ignorance is. We assume that distribution of the ignorance among the citizens is a factor (industry) specific and it is exogenously given.[10] We introduce parameter $\beta$ to capture the ignorance of the voters. In other words, $\beta_i^m$ measures the ability of a voter $i$ owning some specific factor $m$ to map the trade policy choice of the incumbent on the improvement or deterioration of his/her welfare.

[7]Note that for a small, competitive economy free trade is in general social welfare maximising policy. So far the same is true for our model as well. Based on the individual indirect utilities (equation (2.5)), aggregate welfare is $\sum_{m=1}^{M}\sum_{i=1}^{N_m} v_i^m(p, E_m) = \sum_{m=1}^{M}\sum_{i=1}^{N_m} l_i + \frac{1}{N_m}\pi_m(p_m) + \sigma(p) + r(p) = N + \sum_{m=1}^{M}\pi_m(p_m) + N(\sigma(p) + r(p))$ ($N$ stands for total labor income in the economy as the equilibrium wage is one). Now the maximisation of the aggregate welfare results in the following first order condition: $(p_k - p_k^w)(Nd_k^{'}(p_m) - y_k^{'}(p_k)) = 0$ for all $k \in M$, which is satisfied only under free trade. As the challenger is utilitarian, citizens will evaluate any distortionary trade policy relative to free trade policy.

[8]For example, Conconi et al. (2014) indirectly show that voters might have some depreciation rate for information about trade policy. There is no any difference how the members of the House (who serve 2-year terms) and the members of the Senate (who serve 6-year terms) vote for trade liberalisation when a Senator is in the last two years of his/her term. While early in their election period the members of a Senate, who serve 6-year terms more often vote for open trade. The voting pattern can be considered as an indication of the imperfect memory of voters, and they might differ in this respect from one another.

[9]Ponzeto (2011) shows that in an unbalanced panel of 162 countries from 1975 to 2003, tariffs are significantly lower the higher the rate of television ownership. The finding is both economically and statistically significant even after controlling the important economic variables.

[10]There is empirical evidence that industry of employment and education matters, for example, in the formation of trade policy preferences among voters (see Irwin (1996), Beaulieu (2002) and Blonigen (2011)). Education is a more robust determinant of trade policy preferences. If we suppose that education matters in the determination of the ignorance of the voters as well, then a factor (industry) specific distribution of ignorance seems a reasonable assumption as various industries have a different composition of the labor force by education.

For expositional clarity, we assume that $\beta_i^m$ is distributed uniformly[11] and parameters of distribution may differ over the specific factor holders. For all sectors $m \in M$, we have that:

$$\beta_i^m \sim U[0, \Phi_m] \tag{2.6}$$

The upper limit of the distribution ($\Phi_m$) determines how ignorant is the group $m$ of specific factor holders. The higher is $\Phi_m$, the lower is the density $(1/\Phi_m)$, therefore implying, on average, higher ignorance for the group $m$. If $\Phi_m$ is low (the density $(1/\Phi_m)$ is high) for a group $m$, then the voters in the group $m$ are concentrated near 0 and, on average, the group has lower ignorance.

Another variable that we introduce is the general popularity of the incumbent government, denoted by $\delta$. The incumbent cannot directly control the general popularity, but the policy-maker exploits contributions from organised groups to campaign and the campaign spending affects the general popularity. Following the literature (see Persson and Tabellini (2002)) the general popularity of a policy-maker is determined as follows:

$$\delta = \widehat{\delta} + \eta \sum_{j \in L} C_j(p) \text{ with } \eta > 0 \tag{2.7}$$

where $C_j(p)$ is a contribution from the organised group $j$, $L$ is the set of organised groups ($L \leq M$) and $\eta$ measures the effectiveness of campaign spending.[12] The incumbent government knows the distributions of ignorance in different groups, but it does not know its average popularity, as $\widehat{\delta}$ is a random uniform shock ($\widehat{\delta} \sim U[-\Psi, \Psi]$). Before the election, a positive or a negative shock may occur that together with campaign spendings determines the general popularity.

In the end, the following three elements will determine citizens' voting decisions. The first one is the difference between indirect utilities, with and without trade policy distortions. This element is group specific as individuals do not differ regarding utilities within the groups. The second element is the ignorance level of an individual, which is agent specific and the third one is the general popularity of the incumbent, common to all voters in the country. Now we formally define the voting behaviour of an individual $i$ owning the specific factor $m$:

Definition 2.1. *An individual $i$ owning the specific factor $m$ will vote for the incumbent if*

$$v_i^m(p, E_m) - v_i^m(p^w, E_m^w) + \beta_i^m + \delta \geq 0$$

[11]Persson and Tabellini (2002), p.57 conclude that in a similar set-up the usage of any unimodal distribution instead of a uniform distribution does not change the results qualitatively.

[12]In the model, organised groups give political contributions to the incumbent government only. This abstraction is not very restrictive as the financial advantage enjoyed by incumbents in general is a well-documented fact. See Fouirnais and Hall (2014) and Ansolabehere and Snyder (2002).

If we do not consider the ignorance level of individuals and the general popularity of the incumbent government, then the agent will vote for the incumbent only if the change in his/her welfare from the distortionary trade policy is positive. The ignorance becomes an essential ingredient when the net change in indirect utility from the distortionary trade policy is negative. Such a situation may occur if, for example, some groups of specific factor holders other than group $m$ are supported by the government. Then individuals in the group $m$ are harmed as consumers. In general, in the presence of ignorance, the voters with lower ignorance threshold will be more responsive to policy changes.

The next question to address is what motivates the incumbent government's trade policy choice. Grossman and Helpman (1994) assume that government maximises the weighted sum of total political contributions and aggregate welfare. We deviate from this assumption and follow the predominant view in the literature that the policy-maker conceives policy as a means to winning the election. As Downs (1957) notes "Party members have as their chief motivation the desire to obtain the intrinsic rewards of holding office; therefore they formulate policies as means to holding office rather than seeking office in order to carry out preconceived policies."

In our model, the incumbent is concerned with winning the election, so his/her expected utility depends on the probability of winning, $\Pi$ and on the return from being elected, which is considered to be some positive constant. In the case of losing the election, the return is zero. To derive the expression for the policy maker's objective function, we have to determine the probability of winning.

The incumbent knows that after implementing the trade policy, in each group $m$ of the specific factor holders there will be a voter who is indifferent between voting and not voting for the incumbent. Let's call such voter a swing voter of that group. Under the uniform distribution assumption for ignorance, every group will have the swing voter, which is determined by the condition in definition 2.1. Then every voter in the group $m$ that has a higher ignorance level than the ignorance of a swing voter (denoted as $\beta_s^m$) will vote for the incumbent. As $\beta_i^m$ is distributed uniformly, the share of voters in group $m$ who will vote for the incumbent will be:

$$\Omega^m = (\Phi_m - \beta_s^m)\frac{1}{\Phi_m} = 1 - \frac{1}{\Phi_m}(v_s^m(p^w, E_m^w) - v_s^m(p, E_m) - \delta) \tag{2.8}$$

If we consider all groups we can derive the total share of citizens who vote for the incumbent:

$$\begin{aligned} \Omega = \sum_{m=1}^{M} \alpha_m \Omega^m &= \sum_{m=1}^{M} \alpha_m (1 - \frac{1}{\Phi_m}(v_s^m(p^w, E_m^w) - v_s^m(p, E_m) - \delta)) \\ &= 1 - \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}(v_s^m(p^w, E_m^w) - v_s^m(p, E_m)) + \delta\frac{1}{\underline{\Phi}} \end{aligned} \tag{2.9}$$

where $\alpha_m = \frac{N_m}{N}$ and $\frac{1}{\underline{\Phi}} = \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}$ is the average density of ignorance in a country.

The incumbent wins the election if $\Omega > \dfrac{1}{2}$. The condition is satisfied whenever

$$\delta > \underline{\Phi} \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m} (v_s^m(p^w, E_m^w) - v_s^m(p, E_m)) - \frac{1}{2}\underline{\Phi} \tag{2.10}$$

Equation (2.7) and the requirement $\Omega > \dfrac{1}{2}$ are equivalent to the following condition

$$\widehat{\delta} > \underline{\Phi} \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m} (v_s^m(p^w, E_m^w) - v_s^m(p, E_m)) - \eta \sum_{j \in L} C_j(p) - \frac{1}{2}\underline{\Phi} \tag{2.11}$$

Let's denote the right-hand side expression of equation (2.11) by $\underline{\delta}$ and recall that $\widehat{\delta}$ is a random variable which is distributed uniformly. Then the probability of winning the election will be defined as follows:

$$\Pi[\ \widehat{\delta} > \underline{\delta}\ ] = (\Psi - \underline{\delta}) \frac{1}{2\Psi} \tag{2.12}$$

After some manipulations (see Appendix A.1.1) similar to the derivation of the equations (2.8) and (2.9), we get

$$\Pi[\ \widehat{\delta} > \underline{\delta}\ ] = \frac{\underline{\Phi}}{2\Psi} \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m} (v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{2\Psi} \sum_{j \in L} C_j(p) + \frac{\underline{\Phi}}{4\Psi} + \frac{1}{2} \tag{2.13}$$

The incumbent will maximise the probability of winning given in the equation (2.13). Note that, in the end, the rational behaviour of the policy-maker boils down to the maximisation of the following expression:

$$\sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m} (v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{\underline{\Phi}} \sum_{j \in L} C_j(p) \tag{2.14}$$

The policy-maker considers the effects of trade policy on the swing voters in each group of the specific factor holders weighted by the combination of group density of ignorance and the group's share in the total population. Besides, the incumbent government cares for political contributions. The effectiveness of campaign spending and the average density of ignorance determine the weight that the policy-maker assigns to political donations. At large, if the incumbent government is interested only in winning the election and the ignorance densities are different in the various groups of specific factor owners, then the policy-maker does not need to take into account the overall welfare of the society. It suffices to concentrate only on the swing voters in each group in addition to the political contributions.

In Grossman and Helpman (1994) when the government doesn't care about the campaign contributions, free trade is the optimal strategy as the social welfare is maximised under such a policy. Now let's assume that $\eta$ equals zero or $L$ is an empty set in our model. In the first case, policy-maker does not have any incentive to collect political donations, while in the second case, no one wants to contribute to the incumbent. In any of the two cases, contrary to Grossman and Helpman (1994), the free trade still might not be the optimal strategy.[13]

The last agent whose decision we have to present is the lobbyist who behaves on behalf of the organised group. The lobbyists, like the incumbent government, have information about the distributions of ignorance in the various groups of specific factor holders and they also possess knowledge about the distribution of the political shock that might unfold before the election. Therefore, all lobbies can correctly anticipate the policy-maker's best responses to the contribution schedules. Moreover, we assume that the lobbyists cannot influence the distribution of the ignorance among the specific factor holders. Recall that there are $M$ groups of factor holders, but only some of them are organised. The lobbyist objective is to solve the following maximisation problem:

$$\pi_j(p_j) - C_j(p) \tag{2.15}$$

where $\pi_j(p_j)$ is the reward to specific factor owned by the group $j \in L$, which depends only on its price.

**2.3.3. Equilibrium and Structure of Trade Policy.** To solve a two-stage game between the lobbies and the incumbent government, let's assume that the policy-maker chooses its policy from a bounded set of domestic price vectors denoted by $\mathcal{P}$. We retain the equilibrium notion of Grossmann and Helpman (1994) taken from Bernheim and Whinston (1986), as it directly applies to our set-up.[14] Then the sub-game perfect Nash equilibrium of a two-stage trade policy game can be defined as follows:

DEFINITION 2.2. *The collection ($\{C_j^o(p^o)\}_{j\in L}, p^o \in \mathcal{P}$) is a sub-game perfect Nash equilibrium of a two-stage game if and only if the policy vector $p^o$ is in the policy maker's best-response set to $C_j^o(p)$ and given $\{C_l^o(p^o)\}_{l\in L\setminus j}$, no lobby $j$ has any other feasible strategy $C_j(p)$ that would yield a higher payoff.*

Definition 2.2 can be further operationalised in Proposition 2.1, where we characterise the sub-game perfect Nash equilibrium of a trade policy game:

[13]Assume one of the conditions, $\eta = 0$ or $L$ is empty, holds. Then the first order conditions from government's maximisation problem are: $y_k(p_k)[\underline{\Phi}/\Phi_i - 1] + (p_k - p_k^w)m_k^{'}(p_k) = 0$ for all $k \in M$. Now it is clear that $p_k = p_k^w$ is not a sufficient requirement for the first order conditions to be satisfied. Moreover, in equilibrium we get that those groups of the specific factor owners who have higher density of ignorance (voters in the group are less ignorant) than country average will be supported in equilibrium even though there is no political contributions. The following equation demonstrates the result: $p_k - p_k^w = [\frac{1/\Phi_i}{1/\underline{\Phi}} - 1]\frac{y_k(p_k)}{(-m_k^{'}(p_k))}$ for all $k \in M$.

[14]For a general discussion of the common agency problem see Dixit et al. (1997).

Proposition 2.1. *The collection* $(\{C_j^o(p^o)\}_{j\in L}, p^o \in \mathcal{P})$ *is a sub-game perfect Nash equilibrium if and only if the following conditions are satisfied:*

1. $C_j^o(p^o)$ *is feasible for every* $j \in L$;
2. $p^o \in \mathcal{P}$ *maximises*
$$\sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}(v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{\underline{\Phi}} \sum_{j\in L} C_j^o(p);$$
3. $p^o \in \mathcal{P}$ *maximises*
$$\pi_j(p_j) - C_j^o(p) + \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}(v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{\underline{\Phi}} \sum_{j\in L} C_j^o(p);$$
4. *For every* $j \in L$ *there exists a* $p^j \in \mathcal{P}$ *that maximises*
$$\sum_{m}^{M} \frac{\alpha_m}{\Phi_m}(v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{\underline{\Phi}} \sum_{j\in L} C_j^o(p)$$
*such that* $C_j^o(p^j) = 0$.

In general there might be many contribution schedules that satisfy equilibrium conditions, but we restrict our attention to the *truthful contribution schedule* defined in Bernheim and Whinston (1986). The truthful contribution function takes the following form:

$$C_j^T(p, B_j) = max(0, \pi_j(p_j) - B_j) \tag{2.16}$$

where $B_j$ is some number determined in the equilibrium.

Bernheim and Whinston (1986) have shown that there is no cost for players to choose truthful strategies, because the set of best response functions always includes such strategies. Moreover, all equilibria supported by truthful strategies are stable as only these equilibria are Coalition Proof Nash Equilibria.[15]

With contribution functions that are differentiable, the fact that $p^o$ is the optimal price vector implies that the following first-order conditions should be satisfied in equilibrium:

$$\sum_{m=1}^{M} \triangledown \frac{\alpha_m}{\Phi_m}(v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{\underline{\Phi}} \sum_{j\in L} \triangledown C_j^o(p) = 0 \tag{2.17}$$

(2.18)
$$\triangledown\pi_j(p_j) - \triangledown C_j^o(p) + \sum_{m=1}^{M} \triangledown \frac{\alpha_m}{\Phi_m}(v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{\underline{\Phi}} \sum_{j\in L} \triangledown C_j^o(p) = 0$$

Combining equation (2.17) with equation (2.18) results in

$$\triangledown\pi_j(p_j) = \triangledown C_j^o(p) \tag{2.19}$$

[15]For thorough discussion of Truthful Contribution Schedules with detailed proofs see Grossman and Helpmann (2001), chapter 8, Dixit et al. (1997), and Rausser et al. (2011), pp 155-164.

Equation (2.19) shows that for any small change in a policy each lobby adapts its contribution schedule in the way that the difference in the contribution exactly matches the effect of policy change on the lobby's welfare.

To derive the structure of protection in equilibrium of the model, it suffices to consider the effect of a marginal policy change on swing voters' welfare and industry returns. Let's assume that the price $p_k$ changes (we provide the detailed derivations of the equilibrium conditions in Appendix A.1.2). First, consider the effect on swing voter in group $m$.

$$\frac{\partial v_s^m(p, E^m)}{\partial p_k} = (\frac{\varepsilon_{mk}}{N_m} - \frac{1}{N})y_k(p_k) + \frac{1}{N}(p_k - p_k^w)m_k^{'}(p_k) \tag{2.20}$$

where $m_k^{'}(p_k)$ is a derivative of import with respect to $p_k$ and equals $(Nd_k^{'}(p_k) - y_k^{'}(p_k))$. $\varepsilon_{mk}$ is an indicator function, that equals 1 when $m = k$ and 0, otherwise.

The effect of a marginal policy change on industry $j$'s return will be:

$$\frac{\partial \pi_j(p_j)}{\partial p_k} = \theta_{jk} y_k(p_K) \tag{2.21}$$

where $\theta_{jk}$ is an indicator function as well and it equals 1 when $j = k$ and 0, otherwise.

To exploit the equilibrium first-order conditions (2.17) and (2.19), which helps us to derive the overall effect of a marginal change in the price $p_k$, we have to sum up the effect of price change over all swing voters and sum up the effect of a marginal policy change over all organised industry returns. Note that $v_s^m(p^w, E_m^w)$ does not depend on $p_k$. Then we get

$$\frac{1}{N}(p_k - p_k^w)(-m_k^{'}(p_k)) = \eta I_k y_k(p_k) + \frac{1}{N} y_k(p_k)(\frac{\Phi}{\Phi_k} - 1) \tag{2.22}$$

where $I_k = \sum_{j \in L} \theta_{jk}$ and it represents an indicator function, which equals 1, when the industry $k$ is organised and represented by a lobbyist and 0, if industry $k$ is not organised.

Arranging terms in equation (2.22) results in the following expression, which should be satisfied in the equilibrium:

$$p_k^o - p_k^w = (\eta I_k N + \frac{1/\Phi_k}{1/\underline{\Phi}} - 1)\frac{y_k(p_k^o)}{(-m_k^{'}(p_k^o)} \tag{2.23}$$

We can reformulate the equation (2.23) using the trade taxes and subsidies and state the result as a proposition. Note that $t_k = \frac{p_k - p_k^w}{p_k^w}$.

Proposition 2.2. *Let's assume that the contribution schedules are differentiable around the equilibrium and the equilibrium lies in the interior of $\mathcal{P}$, then the government chooses the trade policy that satisfies the following condition:*

$$\frac{t_k^o}{1 + t_k^o} = (\eta I_k N + \frac{1/\Phi_k}{1/\underline{\Phi}} - 1)\frac{z_k(p_k^o)}{e_k(p_k^o)} \quad \textit{for all} \quad k \in M$$

*where* $z_k(p_k^o) = y_k(p_k^o)/m_k(p_k^o)$ *is the equilibrium ratio of domestic output to imports and* $e_k(p_k^o) = (-m_k^{'}(p_k^o))p_k^o/m_k(p_k^o)$ *is the elasticity of import demand or export supply.*

Ceteris paribus industries that have lower import or export supply elasticities (in absolute value) will have higher support from the government. Also, the local prices relative to the world prices will be higher for industries whose domestic output is more extensive. Apart from the economic variables, the density of ignorance within the group of specific factor holders relative to the overall average density of ignorance and the effectiveness of political spending matters for trade policy choices. The higher is the latter variable, the more protection the organised group can buy. For each contributed dollar, the special interest group gets higher protection. Each organised group is interested in getting more protection, but if the overall average density of ignorance in society is higher (this means that in general, the citizens are less ignorant and more responsive to trade policy) protection of any organised group will be lower. More thorough analysis of the equilibrium trade policy structure is presented in the next section.

## 2.4. Analysis

In this section, we shall further explore the implications of the proposition (2.2). The first question is when free trade is the equilibrium outcome of a trade policy game. Note that in our set-up the absence of the political contributions is not enough for a free trade regime to prevail (see footnote 13). As the incumbent government is not a social welfare maximiser even after ignoring the political donations, one should not expect the free trade to be the equilibrium outcome in such a case.

In the corollary 2.1 we present the first major implication of the proposition (2.2).

Corollary 2.1. *According to the equilibrium condition, free trade prevails only if lobbyists represent none of the groups owning the specific factors and all groups share the same ignorance density.*

Proof. In order for a free trade regime to prevail it is necessary that the following condition holds:

$$\eta I_k N + \frac{1/\Phi_k}{1/\underline{\Phi}} - 1 = 0 \quad \text{for all} \quad k \in M \tag{2.24}$$

We have to consider three cases: when lobbyists represent none of the groups of the specific factor owners, when lobbyists represent all of the groups, and when they represent only some of the groups.

First, let's consider the case when none of the groups of owners of the specific factors is organised. Under such condition $I_k = 0$ for all $k \in M$. Then from equation (2.24) we get that for a free trade regime to be an equilibrium of a trade policy game, it should happen that $1/\Phi_k = 1/\underline{\Phi}$ for any $k \in M$. This condition can only be satisfied if all groups share the same density of ignorance.

Next, let's assume that lobbyists represent all groups. Then a free trade regime to be an equilibrium, we should have that

$$\frac{1}{\Phi_k} = \frac{1}{\underline{\Phi}}(1 - \eta N) \quad \text{for every} \quad k \in M \tag{2.25}$$

Equation (2.25) implies that densities of ignorance in all groups should be the same. Recall that $\frac{1}{\underline{\Phi}} = \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}$ and $\eta > 0$, then we get contradiction.

Finally, let's assume that there are some organised and some unorganised groups. Under free trade, for any unorganised group $k \notin L$ the following condition should be satisfied $1/\Phi_k = 1/\underline{\Phi}$ and for any organised group $j \in L$, the following one $1/\Phi_j = 1/\underline{\Phi}(1 - \eta N)$. Recall that $L$ is the set of the organised groups and $L < M$, then by the definition of the average density of ignorance we should have:

$$\frac{1}{\underline{\Phi}} = \sum_{j=1}^{L} \frac{\alpha_j}{\Phi_j} + \sum_{k=L+1}^{M} \frac{\alpha_k}{\Phi_k} = \frac{1}{\underline{\Phi}}[(1 - \eta N)\sum_{j=1}^{L} \alpha_j + \sum_{k=L+1}^{M} \alpha_k] \tag{2.26}$$

We get contradiction.

So, only if none of the groups of the specific factor holders is organised and they share the same density of ignorance, we get that free trade is an equilibrium of a trade policy game. □

The intuition behind the result given in corollary 2.1 is straightforward. Since lobbyists represent no group and every group has similar densities of ignorance, the incumbent does not have incentives to support any group. The policy-maker cannot exploit the ignorance of any group; therefore, free trade prevails. Moreover, note that contrary to Grossman and Helpman (1994) we do not get a free trade as an equilibrium when all groups are organised, even in the case when they share the same density of ignorance. In such case, the incumbent government designs the trade policy based on the economic fundamentals of the model (domestic output to import or export and the elasticity of import demand or export supply) in addition to the effectiveness of the campaign spendings. In general, when all groups are organised, and they share the similar density of ignorance, all else equal, industries that have high import demand or export supply elasticities (in absolute value) will have smaller support from the government, while the sectors that have higher domestic output to import or export ratio will get higher protection.

Grossman and Helpman (1994) argue that all sectors that are represented by lobbies are protected by import tariffs or export subsidies in the equilibrium, while import subsidies and export taxes are applied to all sectors that have no organised representation. Which groups will the policy-maker support in the equilibrium in our set-up? Is it always the case that the incumbent government will support the organised groups? Or can unorganised groups get some support from the government as well? The next result of the proposition 2.2 stated as a corollary answers these questions.

COROLLARY 2.2. *Organised groups are not always supported in equilibrium. Moreover, there are cases when the incumbent government will support unorganised groups.*

Before we present the proof of corollary 2.2, notice that the policy maker's decision to protect an industry depends on whether the industry is organised or not; what is the effectiveness of the campaign spendings[16] and how the relative ignorance density of the industry relates to the average ignorance of the whole society.

PROOF. In general, the government will protect the group $k \in M$ in equilibrium if

$$\eta I_k N + \frac{1/\Phi_k}{1/\underline{\Phi}} - 1 > 0$$

First, notice that if $\frac{1}{\Phi_k} > \frac{1}{\underline{\Phi}}$, it does not matter what is the value of $I_k$ and whether $\eta N < 1$ or $\eta N > 1$, the proposition (2.2) implies that in equilibrium government will support all such groups.

Now we have to consider the cases when $\frac{1}{\Phi_k} < \frac{1}{\underline{\Phi}}$. Several situations should be assessed.

1. If $\eta N < 1$ and $I_k = 0$, the industry $k$ will not be supported.
2. If $\eta N < 1$ and $I_k = 1$, all organised sectors whose ignorance density falls in the interval $(\frac{1}{\underline{\Phi}}(1 - \eta N), \frac{1}{\underline{\Phi}})$ will be protected. But all organised groups with ignorance density satisfying the following condition $\frac{1}{\Phi_k} < \frac{1}{\underline{\Phi}}(1 - \eta N)$ will not be protected in the equilibrium.
3. If $\eta N > 1$ and $I_k = 0$, industry $k$ will not be supported in equilibrium.
4. If $\eta N > 1$ and $I_k = 1$, all such industries will be supported in the equilibrium.

□

Why might the government decide to protect the unorganised groups? Recall that trade policy affects each group of the specific factor holders through two channels. First, protection of the industry generates higher factor income, but the same industry representatives are consumers as well, therefore, supporting other sectors decreases their consumer surplus. When a group is unorganised, but it has higher ignorance density in relation to overall average ignorance density, the government might want to protect the group to compensate losses of its consumer surplus due to trade policy. In general, defending the specific factor holders, who are more responsive to policy

[16]One can interpret $\eta N$ as the effectiveness of each unit of the campaign spending in terms of the number of votes.

choices, gives the policy-maker more freedom to design the trade policy. One should remember that all these calculations happen on the margin under the clear trade-off between the political contributions and votes.

The incentives of the incumbent government to protect an industry decreases when the density of ignorance in that industry decreases. The logic applies to the organised groups as well if the effectiveness of the campaign spendings is not high enough (see the proof of corollary 2.2). The collection of the political donations from ignorant organised groups does not pay off in terms of votes for the policy-maker. But if the effectiveness of the political spendings reaches some threshold, the incumbent government will support all organised sectors.

Figure 1 summarises the results from the proof of corollary 2.2. It shows when the government will and will not support organised and unorganised sectors in the equilibrium of a trade policy game. The decision largely depends on the relative ignorance density of a sector in relation to overall ignorance density. Note that if $\eta$ is approaching zero, implying that the political spendings cannot affect the general popularity of the incumbent government, two lines in the graph will get closer to each other.

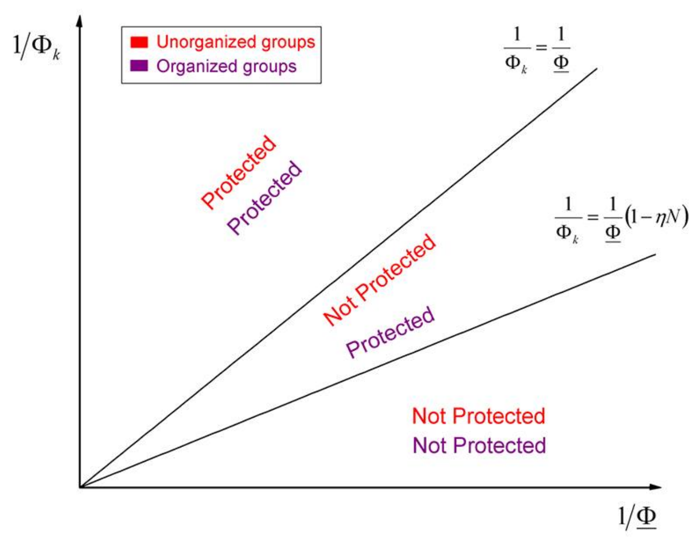



FIGURE 2.1. Relative ignorance density and policy-maker's decision

To sum up, the incumbent government will protect those groups who are relatively more responsive to the trade policy choices. Lobby representation of a sector is not a sufficient determinant for selling the protection.

## 2.5. Conclusions

Despite the consensus among economists that free trade permits the achievement of maximal social welfare in a society when a small, competitive economy is considered, in reality, we extensively observe deviations from the free trade regime through government policies. There is vast literature that aims to explain the political process behind protection. The goal of chapter 2 is to contribute to this literature.

In our work, the incumbent government faces a trade-off between the campaign contributions and votes. The policy-maker cares about winning the upcoming election. In some extent, we retain the basic structure of Grossman and Helpman's (1994) model and their notion of equilibrium, but we explicitly model the political process, assuming that voters are heterogeneously ignorant about the effects of policy choices. Based on the voting decision of citizens we derive the governmental objective function when voting is probabilistic. Adding these new features to the Grossman and Helpman model allows us to formulate an explicit micro-founded model of endogenous trade policy determination.

Examination of the structure of a political equilibrium of the model shows that groups represented by lobbies are not always supported in equilibrium. Besides, there are cases when the incumbent might support groups without any lobby representation although they do not provide any campaign contributions. On average, more ignorant groups will always experience lower (if any) protection from the policy-maker.

Our work substantiates several possible extensions of the model which are worthy of carrying out. First, in the model, we assume that the lobbyists cannot influence the ignorance density of the group they represent. In practice, the special interest groups through different information channels can affect the ignorance of voters and in this way force the policy-maker to implement some desirable policy. Intuitively the effectiveness of such informational channels will influence the incumbent's capability of collecting the campaign contributions and the structure of the equilibrium trade policy. Second, we assume that the challenger is utilitarian and the lobbyists contribute political donations to the incumbent only. Under such assumptions, we abstract in a large extent from the potentially interesting strategic interaction between the incumbent and the challenger. It is interesting to analyse the structure of the equilibrium trade policy when lobbyists can contribute to both the incumbent and the challenger, while the latter is not utilitarian as well. Lastly, the work generates interesting theoretical predictions, which in the case of finding a good proxy for voters' ignorance can be tested empirically.

CHAPTER 3

# The Farsighted Stability of Global Trade Policy Arrangements

## 3.1. Introduction

Following the General Agreement on Tariffs and Trade (GATT) of 1947, an increasing number of signatory countries liberalised their trade policies primarily via two channels: bilateral and multilateral negotiations. To the present day, there have been eight rounds of multilateral trade negotiations with the current ninth one, the Doha Round, still ongoing. At the same time, parallel to the arrangements observed on the multilateral level, the world has seen an ever-increasing number of Preferential Trade Agreements (PTAs) mainly in the wake of bilateral negotiations. Currently, about forty percent of all countries/territories are a member of more than five PTAs while about a quarter participates in more than ten.[1]

The World Trade Organization (WTO), successor of the GATT in 1995, provides the rule set for the trade liberalisation process of a significant number of countries.[2] Its Article I acts as the foundation for any multilateral trade liberalisation by formulating the so-called Most-Favoured-Nation (MFN) principle: Any concession granted to one member needs to be extended to all other members of the WTO.[3] In this chapter, trade policy arrangements that are consistent with the MFN principle are referred to as Multilateral Trade Agreements (MTAs).[4] Contrary to the core MFN principle, Article XXIV Paragraph 5 explicitly allows countries to form PTAs, specifically Customs Unions (CUs) and Free Trade Agreements (FTAs), that do not need to extend the concessions granted within the arrangement to other countries.[5] However, Article XXIV Paragraph 5 Subparagraph (a), (b), and (c) each require that these are without (negative) influence on other trade relations.

The (direction of the) influence of Article XXIV Paragraph 5 on the development of trade policy arrangements is a controversial topic and the focus of many papers.[6]

[1]Source: `http://www.wto.org`

[2]All members of the WTO account for 96.4 percent of world trade, 96.7 percent of world GDP, and 90.1 percent of world population as of 2007 (Source: `http://www.wto.org`).

[3]Article I states that 'any [...] favour [...] granted by any contracting party to any product originating in or destined for any other country shall be accorded immediately and unconditionally to the like product originating in or destined for [...] all other contracting parties' (GATT, 1947).

[4]Furthermore, we interchangeably use the terms trade policy arrangements, trade agreements, trade constellations, and trade relations.

[5]Article XXIV Paragraph 5 states that '[...] this agreement shall not prevent [...] the formation of a customs union or of a free-trade area [...]' (GATT, 1947).

[6]The next part of this chapter contains further information on the related literature.

Likewise, the primary purpose of this chapter is the analysis of the stability of different trade policy arrangements in two scenarios, that is with PTAs (current WTO rules) and without PTAs (modified WTO rules). In particular, it is our intent to examine whether PTAs act as 'building blocks' or 'stumbling blocks' on the path towards global free trade (Bhagwati (1993)).

The existing literature usually considers a limited selection of trade agreements or assumes limited farsightedness of the negotiating countries. It certainly allows for a cleaner description of the model and interpretation of its results, but ultimately raises the question about whether or not these restrictions significantly influence the analysis and to what degree these frameworks capture reality. In our opinion, certain empirical observations favour an extensive choice set and full farsightedness. During the past rounds of multilateral trade negotiations, many countries were simultaneously involved in other trade liberalisation processes.[7] Moreover, such trade negotiations are usually complicated processes with significant effect on the countries' economies and accompanied by elaborate studies about feasibility and future developments.[8] Taking these assessments into account, the contribution of our work is an answer to the question concerning the influence on the analysis.

First of all, we consider an extensive set of trade agreements, containing PTAs, i.e., CUs and FTAs, as well as MTAs. Next, endogenizing the formation of trade agreements, each country ranks them based on a three-country two-good general equilibrium model of international trade.[9] The stability of all trade agreements is then examined using these rankings together with the concept of 'consistent sets' as stable sets - a notion proposed by Chwe (1994). As a result, our work expands the set of trade agreements under consideration and also extends the farsightedness of the negotiating parties in comparison to the literature. In fact, to the best of our knowledge, no other paper considers a choice set as extensive as ours.

In the end, our analysis shows that the effect of PTAs on trade liberalisation depends on the size distribution of the countries. As long as the countries are close to symmetric, Global Free Trade (GFT) emerges as the unique stable outcome under both the existing and the hypothetical institutional arrangement. However, when two countries are considerably smaller, a modified WTO without PTAs would facilitate the formation of GFT. By contrast, if two countries are relatively larger, this modified WTO would actually obstruct the development towards GFT. Once the world is

[7] Maggi (2014) showcases the importance of an extensive set of trade constellations.

[8] Aumann and Myerson (1988) provides a (brief) description of the criticism against the use of limited farsightedness in general: 'When a player considers forming a link with another one, he does not simply ask himself whether he may expect to be better off with this link than without it, given the previously existing structure. Rather, he looks ahead and asks himself, "Suppose we form this new link, will other players be motivated to form further new links that were not worthwhile for them before? Where will it all lead? Is the end result good or bad for me?"'

[9] A model similar to that of Saggi and Yildiz (2010), which itself is a modification of the one in Bagwell and Staiger (1997). The modified one is also used in Saggi, Woodland and Yildiz (2013).

further away from symmetry, full trade liberalisation is not attainable at all and abolishing the exception for PTAs might result in the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime.

The findings of our work notably deviate from those of the existing literature. Compared to the paper of Saggi, Woodland and Yildiz (2013), the composition of the stable set of trade policy arrangements differs on a substantial part of the parameter space under consideration (while coinciding on the remainder). Beyond that, the comparison with the work of Lake (2017) yields not only a difference in terms of stability but also with respect to the driving force(s).

The remainder of this chapter is organised as follows. Section 3.2 focuses on the related literature, Section 3.3 specifies the model, Section 3.4 analyses the findings while further details are discussed in Section 3.5, and Section 3.6 concludes our work.

## 3.2. Related Literature

An ever increasing body of literature studies the different aspects of international trade agreements. It is not our goal to completely review this stream of literature.[10] The emphasis of this part of our work is on the methodology of the related papers. Further details, in particular a comparison of the model predictions, can be found in Section 3.5. In the following, the focus is on the so-called 'rules-to-make-rules' literature (Maggi (2014)) that tries to determine the role of PTAs in the global trade liberalisation process.

A number of relevant papers are the work of Saggi, Yildiz and various co-authors. Saggi and Yildiz (2010) considers a three-country trade model where the degree and nature of trade liberalisation, bilateral and multilateral, are endogenously determined. Using Coalition-Proof Nash Equilibria, the authors study the stability of FTAs and MTAs while varying the extent of asymmetry among the countries with respect to their size. In a subsequent paper Saggi, Woodland and Yildiz (2013) study the complementary case by focusing on the combination of CUs and MTAs while leaving everything else fixed (in terms of their framework). By contrast, the paper of Missions, Saggi and Yildiz (2016) analyses the effect of both forms of PTAs, i.e. CUs and FTAs, on attaining global free trade, but excludes MTAs. In a sense, this completes their '2 out of 3' pattern of trade agreements under consideration.

Another related paper (in terms of farsightedness) is the work of Lake (2017), who uses a dynamic approach to understand whether FTAs facilitate or impede the formation of GFT. The approach uses a three-country dynamic model where a fixed protocol specifies for each period the exact nature (and order) of negotiations. Then,

[10]The reader may want to consult the papers of Maggi (2014), Grossmann (2016), and Bagwell and Staiger (2016) for a detailed review of the related literature.

on the basis of Markov Perfect Equilibria in pure strategies, the author analyses the effect of country asymmetries on global trade liberalisation.

Furthermore, a variety of research focuses purely on analysing the effect of FTAs. Goyal and Joshi (2006) consider several countries with a homogeneous good in their model and study different degrees of asymmetry across countries. They employ the notion of Pairwise Stability by Jackson and Wolinsky (1996) as the solution method. Furusawa and Konishi (2007) use similar methods but introduce heterogeneity with respect to goods. In a separate section, they also briefly discuss a setting with CUs, but overall focus on FTAs. Another related paper to Goyal and Joshi (2006) is that of Zhang et al. (2013) in which the concept of Pairwise Stability is replaced with Pairwise Farsighted Stability by Herings, Mauleon and Vannetelbosch (2009), thereby comparing myopia with farsightedness in an otherwise fixed framework. Also connected to this is the paper of Zhang et al. (2014), which uses the work of Goyal and Joshi (2006) as a benchmark and analyses the evolutionary effect of the number of countries in a dynamic framework featuring random perturbations. Now, while all of the aforementioned papers employ (different) network-theoretic concepts, there is also Aghion et al. (2007), which features standard cooperative game theory. In the three-country model presented there, a single country takes on the role of negotiation leader and decides to either engage in sequential bilateral or single multilateral bargaining with the other countries.

The stability concept of our approach is that of Chwe (1994). It is (in parts) a response to the criticism of the von-Neumann-Morgenstern stable set (solution).[11] The approach aims to achieve two goals, namely to include unlimited consideration of the future by the participants while simultaneously avoiding emptiness of the stable set that plagues other (more) restrictive solution concepts.[12] It is also closely related to the stability concept found in Herings, Mauleon and Vannetelbosch (2009) and its extension (HMV (2014)). In fact, as is noted by the authors, their criterion constitutes a stricter version, but in specific cases (like our model) they coincide.

## 3.3. Model

**3.3.1. Setting.** Let $N = \{a, b, c\}$ denote the set of all (three) countries in the world. Furthermore, let $X$ denote the set of all trade agreements between these countries, see Section 3.3.3 for an explicit list. Then, the welfare function of each country induces a collection of preferences on $X$ denoted by $\{\prec_i\}_{i\in N}$, see Section 3.3.2 for a description of the employed trade model that determine the welfare functions.

[11]Consult von Neumann and Morgenstern (1944) for a description of this (solution) concept and Harsanyi (1974) for its criticism.

[12]It is also resistant to the criticism of Ray and Vohra (2015) about the sovereignty of coalitions as their main issues concerned with feasibility and distribution do not apply to our framework. Furthermore, their specific criticism about the explanatory power of Chwe's approach only applies to transferable utility games.

Moreover, the non-empty subsets $S$ of $N$ specify the coalitions of countries, i.e. the grand coalition, coalitions of two, and single coalitions. Naturally, the preferences of the individual countries induce those of the coalitions, namely for $x_1, x_2 \in X$ and $S \subseteq N$, $S \neq \emptyset$: $x_1 \prec_S x_2$ if and only if $x_1 \prec_i x_2$ for all $i \in S$. Further, the actual ability of coalitions to change the status quo of trade agreements is captured via the collection $\{\rightarrow_S\}_{S \subseteq N, S \neq \emptyset}$ of effectiveness relations defined on $X$, see Section 3.3.5 for the resulting overall network structure. In combination, the preferences together with the effectiveness relations will allow us to analyse the (potential) stability of different trade agreements, see Section 3.3.4 for a formal definition of the employed concept of stability. Finally, to determine the stable and unstable trade agreements an algorithm numerically evaluates a grid of the parameter space, see Section 3.3.6 for details.

**3.3.2. Underlying Trade Model.** In order to study the stability of different constellations of trade agreements, our framework utilises a three-country trade model with competition via exports. It will determine the welfare of each country and thereby induce preferences and rankings over all regimes. The model itself follows the one used by Saggi and Yildiz (2010).

Recall that $N = \{a, b, c\}$ denotes the set of countries. Further, let $G = \{A, B, C\}$ denote three (corresponding) non-numeraire goods. Now, each country $i$ is endowed with zero units of good $I$ (corresponding capital letter) and $e_i$ units of the others. Ultimately, it will end up importing $I$ and exporting $J$ and $K$ with $J, K \neq I$. To guarantee the 'competing exporters'-structure, a general condition needs to be applied to the degree of asymmetry with respect to the endowments of the countries. For $i$ and $j$ in $N$ with $i \neq j$, in order for the exports from $i$ to $j$ to be non-negative the condition $3e_j \leq 5e_i$ needs to be satisfied. Thus, the general condition reads:

$$\frac{3}{5}\max\{e_j, e_k\} \leq e_i \leq \frac{5}{3}\min\{e_j, e_k\} \qquad \forall\, i, j, k \in N$$

The preferences of individuals in each country are furthermore assumed to be identical. The demand for any non-numeraire good $L \in G$ in country $i \in N$ is given by the function $d(p_i^L) = \alpha - p_i^L$ with $p_i^L$ the price of good $L$ in country $i$ and the (universal) reservation price $\alpha$.[13] Each country also (possibly) imposes tariffs on the goods imported by them. Let $t_{ij}$ denote the tariff imposed by country $i$ on the import from country $j$. All prices and tariffs of a specific good $I \in G$ are connected via the following no-arbitrage condition

$$p_i^I = p_j^I + t_{ij} = p_k^I + t_{ik} \tag{3.1}$$

[13]The demand function is derived from a utility function that is additively separable and also quadratic in each non-numeraire good.

where $i, j, k \in N$ are pairwise distinct. In this model, the resulting prices together with the corresponding endowments are the only factors influencing imports and exports. In particular, the level of imports $m_i^I$ of good $I$ to country $i$ is completely determined by the demand function (depending on the price), $m_i^I = d(p_i^I) = \alpha - p_i^I$. The exports $x_j^I$ of good $I$ from country $j$ are the combination of the demand function (or prices) and the corresponding endowment, $x_j^I = e_j - d(p_j^I) = e_j + p_j^I - \alpha$. Now, a market-clearing condition for any good $I$ requires that country $i$'s import is equal to the total export of the countries $j$ and $k$ (again $i, j, k \in N$ pairwise distinct):

$$m_i^I = x_j^I + x_k^I \tag{3.2}$$

Ultimately, the objective function of country $i$ is its welfare[14], denoted $W_i$, which includes Consumer Surplus (CS), Producer Surplus (PS), and Tariff Revenue (TR):

$$W_i = \sum_{L \in G} CS_i^L + \sum_{L \in G \setminus \{I\}} PS_i^L + TR_i$$

Now, CS is composed of three parts itself, namely one for each good. The consumer surplus $CS_i^I$ with respect to the foreign good $I$ is $CS_i^I = \frac{1}{2}(\alpha - p_i^I)m_i^I$ and $CS_i^L = \frac{1}{2}(\alpha - p_i^J)(e_i - x_i^J)$ for a domestic good $L$. Also, PS splits into two. The producer surplus $PS_i^L$ for a domestic good $L$ is given by $PS_i^L = x_i^L(p_l^L - t_{li}) + (\alpha - p_i^L)p_i^L$. Finally, the tariff revenue $TR_i$ is given by $TR_i = x_j^I t_{ij} + x_k^I t_{ik}$.

3.3.2.1. *Equilibrium.* Let us start by using no-arbitrage (3.1) and market-clearing (3.2) to compute the equilibrium prices:

$$p_i^I = \frac{1}{3}\left(3\alpha - \sum_{j \neq i} e_j + \sum_{j \neq i} t_{ij}\right)$$

Using these equilibrium values, it is possible to calculate imports, exports, and also the welfare of each country up to the value of the tariffs (Appendix B.2.1). Note, that the maximisation of welfare with respect to tariffs is going to be restricted depending on the trade agreement under consideration, see Section 3.3.3. For example in the case of MFN, country $i$ maximises $W_i$ under the restriction that $t_{ij} = t_{ik}$. Therefore, country $i$ aims to maximise its welfare $W_i$ over $(t_{ij}, t_{ik}) \in T_i$ given $(t_{ji}, t_{jk}) \in T_j$ and $(t_{ki}, t_{kj}) \in T_j$, where $T_l$ is the set of possible tariff pairs for country $l$ in a fixed trade agreement.

The full equilibrium of this model is computed as follows. Fix a trade agreement and thereby the restrictions on the tariffs. Compute the best-response functions for each country (with respect to the tarifs) and determine the optimal choices. While Section 3.3.3 contains all information on the trade agreements that is necessary to

[14]In certain cases (depending on the trade agreement) the objective function of a country includes the welfare of other countries as well. See Section 3.3.3 for the details.

compute the equilibria, the actual results are presented in Appendix B.2.2. Finally, an overview of the (resulting) overall welfare can be found in Appendix B.2.3.

**3.3.3. Trade Policy Arrangements.** All trade relations in our model are one of four types: MFN, CU, FTA, and MTA. Each type, except for MFN, naturally induces different combinations of insiders and outsiders. Namely, three combinations of two members and one of three (each for CU, FTA, and MTA).[15] Additionally, the case of FTA contains the possibility of a special hub structure with two FTAs at the same time - adding another three combinations. In total, our model allows for 16 different trade constellations.[16] For each of these trade agreements the tariffs are bounded from below and above by zero and the MFN-tariff respectively, which is discussed in more detail in Appendix B.2.2. The corresponding set of tariffs for country $i$, i.e. $[0, t_i^{MFN}]$, is denoted by $T_i$. Any additional restrictions on tariffs, specific to trade agreements, are listed here:

In the baseline case, i.e. MFN, countries do not liberalise their trade relations at all, but the non-discrimination principle still applies. Each country unilaterally chooses its (optimal) tariffs accordingly. Therefore, the optimisation problem of country $i$ is $\max_{(t_{ij},t_{ik})\in T_i^{MFN}} W_i$ with $T_i^{MFN} = \{(t_{ij}, t_{ik}) \in \mathbb{R}^2_{\geq 0} \,|\, t_{ij} = t_{ik}\}$. Note, that in this reference scenario each tariff is chosen from $\mathbb{R}_{\geq 0}$ instead of $T_i$.

In case country $i$ and $j$ form CU(i,j), each of them removes any trade restriction on the other country and then jointly imposes an optimal tariff on country $k$. Thus, the optimisation problem of country $i$ and $j$ is $\max_{(t_{ij},t_{ik})\in T_i^{CU},(t_{ji},t_{jk})\in T_j^{CU}} W_i + W_j$ with $T_i^{CU} = \{(t_{ij}, t_{ik}) \in T_i^2 \,|\, t_{ij} = 0\}$ and $T_j^{CU}$ analogous. Finally, country $k$ simply follows and applies the principle of MFN (as before). However, as soon as all three countries enter a single CU together, the (common) optimisation problem is trivial, because the only possible tariff of each country towards any other country is zero, and the scenario is denoted by CUGFT.

In case country $i$ and $j$ form FTA(i,j), each of them removes any trade restriction on the other country and then unilaterally imposes an optimal tariff on country $k$. Thus, the (representative) optimisation problem of country $i$ is $\max_{(t_{ij},t_{ik})\in T_i^{FTA}} W_i$ with $T_i^{FTA} = \{(t_{ij}, t_{ik}) \in T_i^2 \,|\, t_{ij} = 0\} (= T_i^{CU})$. The optimisation problem of country $k$ is identical to that of the third country in case of a CU. Further, in case country $i$

[15]Note that in our model Global Free Trade is essentially listed in three different variations, via CUs, FTAs, and MTAs. The actual welfare is necessarily equal across all three variations, but not their position in the network (Section 3.3.5). In particular, for our concept of stability it is important which group of countries can create or destroy specific trade agreements (Appendix B.3.1). Occasionally, all three variants together are going to be referred to as 'GFT' (when applicable).

[16]The framework does not contain combinations of different classes of trade agreement due to the possibly conflicting restrictions on tariffs that the different classes entail. In order to circumvent potential conflicts one would need to fix an (arbitrary) ordering in terms of priority (or importance) of trade agreements, which would reduce the explanatory power more than the inclusion of other combinations of trade agreements would increase it (in our opinion).

forms an FTA both with $j$ and $k$, that is FTAHub(i), then both tariffs of country $i$ are set to zero by nature of its trade relation with both other countries. Each of the other two countries operates as before: Country $j$ ($k$ analogous) faces $\max_{(t_{ji},t_{jk})\in T_j^{FTA}} W_j$ where $T_j^{FTA} = \{(t_{ji}, t_{jk}) \in T_i^2 \mid t_{ji} = 0\}$. Thus, in terms of decision problem, it does not matter for a country whether its partner also forms another trade agreement with the other country. Finally, if all three countries in pairs of two countries form FTAs, then the optimisation problem is identical to the case of CUGFT, denoted FTAGFT, but the actual trade agreement is different in terms of structure and network position, see Section 3.3.5.

In case country $i$ and $j$ form MTA(i,j), then both jointly change their tariffs with respect to each other and also for the third country (at the same time). Thus, the optimisation problem of country $i$ and $j$ is $\max_{(t_{ij},t_{ik})\in T_i^{MTA},(t_{ji},t_{jk})\in T_j^{MTA}} W_i + W_j$ with $T_i^{MTA} = \{(t_{ij}, t_{ik}) \in T_i^2 \mid t_{ij} = t_{ik}\}$ and $T_j$ analogous. As seen before, the optimisation problem of country $k$ is identical to that of the third country in case of a CU. Again, as soon as all three countries enter a single MTA together, the optimisation problem is identical to the case of CUGFT, denoted MTAGFT, but also different in terms of network position, see Section 3.3.5.

**3.3.4. Stability Concept.** As concept of stability our framework makes use of the approach of Chwe (1994).[17]

Consider the tuple $\Gamma = (N, X, \{\prec_i\}_{i\in N}, \{\rightarrow_S\}_{S\subseteq N, S\neq\emptyset})$ that correspondingly describes the evolution of the status quo of trade agreements driven by the combination of preferences and effectiveness relations:

Let $x \in X$ be the status quo of trade agreements at the start. Next, each coalition $S \subseteq N$, $S \neq \emptyset$ (including individuals) is able to make $y \in X$ the new status quo as long as $x \rightarrow_S y$. Continue with such $y$ as the new status quo. If a status quo $z \in X$ is reached without any coalition moving away, then the state is actually realised and each country receives their corresponding welfare.[18] In consequence, any coalition only favours following through on their ability to move, $x \rightarrow_S y$, when preferring the final welfare over the current one, $x \prec_S z$. Formally, this comparison of states by (chains of) coalitions is captured in the definition of direct and indirect dominance:

Definition 3.1 (Dominance). *Let $x_1, x_2 \in X$. Then,*

i) *$x_1$ is directly dominated by $x_2$, write $x_1 < x_2$, if there exists $S \subseteq N$, $S \neq \emptyset$, such that $x_1 \rightarrow_S x_2$ and $x_1 \prec_S x_2$.*

[17]Consult the paper of Chwe (1994) for the proofs of the propositions that are presented here.

[18]Technically, the model is without any true sense of time. Any start (or end) as well as any sequence of actions should be interpreted as a thought-experiment. Furthermore, a path created in this fashion is generally not unique.

ii) *$x_1$ is indirectly dominated by $x_2$, write $x_1 \ll x_2$, if there exist sequences $y_0, y_1, \ldots, y_m \in X$ (with $y_0 = x_1$ and $y_m = x_2$) and $S_0, S_1, \ldots, S_{m-1} \subseteq N$, such that $S_i \neq \emptyset$, $y_i \to_{S_i} y_{i+1}$, and $y_i \prec_{S_i} y_m$ for $i = 0, 1, \ldots, m-1$.*

Note, that if $x_1 < x_2$ for some $x_1, x_2 \in X$, then automatically $x_1 \ll x_2$.

Using this definition, the concept of 'consistent set' describes a (sub-)set that exhibits internal stability in the form of a lack of incentive to deviate:

DEFINITION 3.2 (Consistent Set). *A set $Y \subseteq X$ is consistent if $y \in Y$ if and only if for all $x \in X$ and all $S \subseteq N$, $S \neq \emptyset$, with $y \to_S x$ there exists $z \in Y$ where $x = z$ or $x \ll z$ such that $y \not\prec_S z$.*

In general, a consistent set is not necessarily unique, but the following proposition allows us to talk about the unique 'largest consistent set', i.e. the (consistent) set that contains all consistent sets:

PROPOSITION 3.1. *There uniquely exists a $Y \subseteq X$ such that $Y$ is consistent and $Y' \subseteq X$ consistent implies $Y' \subseteq Y$. The set $Y$ is called the largest consistent set or simply LCS.*

*Or put differently, it is the unique fixed point of the correspondence $f\colon 2^X \to 2^X$ defined by*

$$Y \mapsto f(Y) = \{y \in X \mid \forall x \in X, \forall S \subseteq N, S \neq \emptyset, \text{ with } y \to_S x\colon \\ \exists z \in Y \text{ s.t. } (x = z \text{ or } x \ll z) \wedge y \not\prec_s z\}.$$

Now, similar to the internal stability captured in the definition of consistent sets, a form of external stability is captured via an incentive to gravitate towards the consistent set:

DEFINITION 3.3 (External Stability). *Let $Y \subseteq$ be the largest consistent set. Then, it satisfies the external stability condition if for all $x \in X \setminus Y$ there exists $y \in Y$ such that $x \ll y$.*

The following result characterises one setting in which this condition is satisfied:

PROPOSITION 3.2. *Let $X$ be finite and the underlying preferences irreflexive. Then, the LCS is non-empty and satisfies the external stability property.*

Finally, let us state a couple of comments on the application and interpretation of this stability concept with respect to our model:

3.3.4.1. *Application.* First of all, applying Proposition 3.1 to our model is trivial, because it is stated without any (additional) requirements on the involved objects. Furthermore, the application of Proposition 3.2 is straight forward as well: First, the set of outcomes $X$ is clearly finite in our setting as we are only considering a finite number of different trade agreements. Second, any strict preference is automatically irreflexive and our preferences are induced by strict welfare comparisons. Thus, while the definition of the (largest) consistent set in general only guarantees internal stability, our setting actually implies external stability as well:

COROLLARY 3.1. *In our setting, the (unique) LCS is non-empty and satisfies the external stability property (in addition to the internal stability).*

Now, the LCS is going to be the focus point of our analysis. Any trade agreement is considered to be '(potentially) stable' if it is in the LCS, 'unstable' otherwise. The nomenclature is a tribute to the fact that the LCS as a stability concept is 'weak: not so good at picking out, but ruling out with confidence', because ultimately it 'does not try to say what will happen but what can possibly happen' (Chwe (1994)).

**3.3.5. Network Structure.** The complete network structure consists of a collection of transition matrices $\{A_S\}_{S\subseteq N, S\neq\emptyset}$ induced by $\{\to_S\}_{S\subseteq N, S\neq\emptyset}$. Let $S \subseteq N, S \neq \emptyset$ be any coalition, then the entry $(A_S)_{x_i,x_j}$ is 1 if $x_i \to_S x_j$ and 0 otherwise. Thus, the matrix for $\{a, b, c\}$, the full coalition, is simply given by $(A_{\{a,b,c\}})_{x_i,x_j} = 1$ for all $x_i, x_j \in X$. Further, each of the transition matrices induces a directed graph with the trade agreements as vertices and the effectiveness relations as edges. Therefore, the corresponding directed graph of the full coalition is a complete directed graph with loops.

It is noteworthy to point out that the relation (or transition) $x \to_S x$ holds for all trade agreements $x$ and all coalitions $S$, but is ultimately irrelevant for the analysis with respect to the stability. The reason for this is the fact that our model contains no sense of time - essentially stalling negotiations serves no purpose.[19] Therefore, these transitions are ignored from now on or, put differently, the framework only considers a form of equivalence classes, namely modulo loops. Furthermore, whenever coalition $S$ is able to destroy one trade agreement, say $x_1$, and subsequently create another one, say $x_2$, then it is able to move directly, i.e. $x_1 \to_S x_2$. Finally, for the remaining coalitions (of two and one country) only the transition graphs are presented here. The corresponding transition tables can be found in Appendix B.2.4.

Let us now consider the transition graph for a single country coalition $i \in N$ with $j, k \in N \setminus \{i\}$, $j \neq k$, denoting the other two countries. In this case, MFN is connected to a number of other different elements, but not to the three variants of Global Free Trade, CUGFT, FTAGFT and MTAGFT. Now, each of those forms a separate group of connected trade agreements. Thus, the overall transition graph, see Figure 3.1 (modulo loops), consists of four sub-graphs.

[19]While staying in one trade constellation, the overall strategic situation remains the same. Specifically, for each country and each coalition the welfare of each trade agreements only depends on the parameters of the underlying trade model. Similarly, the network structure stays constant. Additionally, the number of (potential) movements in a chain of trade agreements is unlimited.

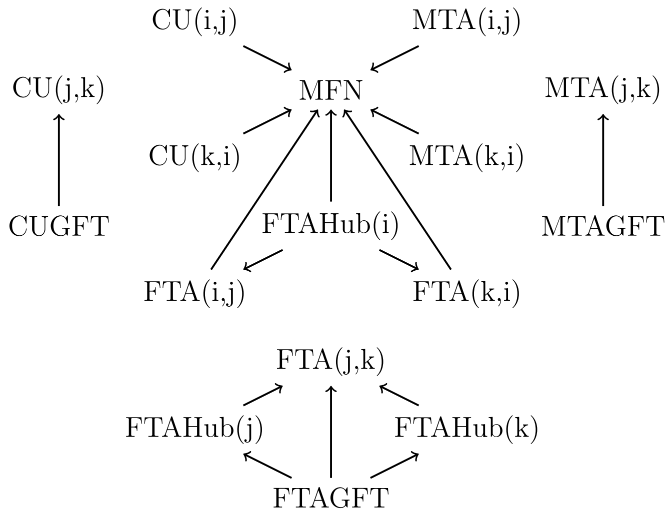



FIGURE 3.1. The transition graph for coalition $\{i\}$, $i \in N$.

Finally, consider the transition graph for a coalition of two countries $i, j \in N$, $i \neq j$ with $k \in N \setminus \{i, j\}$ denoting the other country. In this case, MFN, CU(i,j), FTA(i,j), and MTA(i,j) are all interconnected. Also, any element connected to one of these is automatically connected to all of them. Thus, in the transition graph, see Figure 3.2 (again, modulo loops), this group of four corresponds to a complete directed sub-graph pictured as one '(super) node' (dotted box).

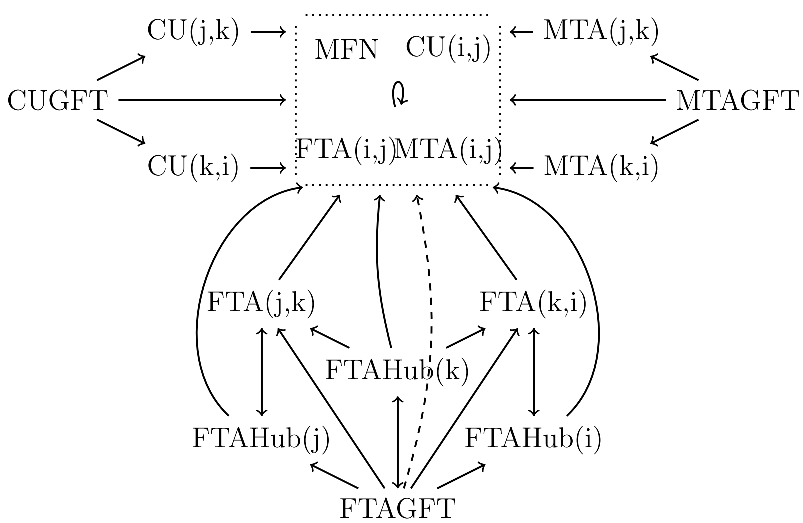



FIGURE 3.2. The transition graph for coalition $\{i, j\}$, $i, j \in N$, $i \neq j$.

**3.3.6. Algorithm and Parameters.** The (additional) explanatory power from the introduction of an extensive set of trade agreements and unlimited farsightedness comes at the cost of a complex computational problem. This problem is solved numerically with the help of an algorithm - the pseudocode of which can be found in Appendix B.1.[20] The parameter space therefore needs to be specified and discretised:

First, recall that the endowments satisfy $\frac{3}{5}\max\{e_j, e_k\} \leq e_i \leq \frac{5}{3}\min\{e_j, e_k\}$ for all $i, j, k \in N$ in order to guarantee the 'competing exporters'-structure, see Section 3.3.2. Now, without loss of generality, normalise one endowment to one, namely $e_b = 1$. Consequentially, for $i, j \in N \setminus \{b\}$: $e_{\min} := \frac{3}{5} \leq \frac{3}{5}\max\{1, e_j\} \leq e_i$ and $e_i \leq \frac{5}{3}\min\{1, e_j\} \leq \frac{5}{3} =: e_{\max}$. Furthermore, the resulting parameter space, Figure 3.3, can be split into six right-angled triangles, which are mirror images of one another (in terms of relative endowments). Thus, again without loss of generality, focus on one of them, namely the marked triangle, and then cover it with a grid for the actual computation.[21]

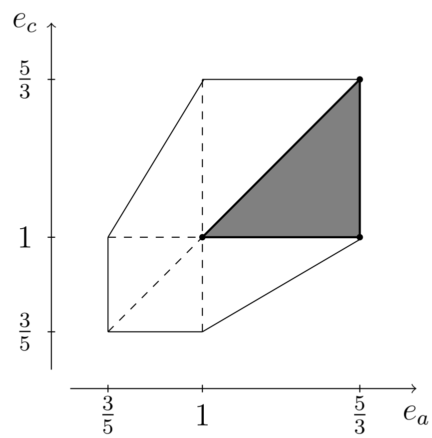



FIGURE 3.3. The parameter space of the endowments with $e_b = 1$

Additionally, to produce plausible results, e.g positive prices, the factor $\alpha$ needs to be chosen above a minimal value for each tuple of endowments, $\alpha_{\min}(e_a, e_b, e_c)$. Above this minimal value, the results remain unchanged.[22] Thus, by taking the maximum over all these minimal values, $\alpha_{\max\min} = \max_{e_a,e_b,e_c}\{\alpha_{\min}(e_a, e_b, e_c)\}$, adding an epsilon, $\alpha = \alpha_{\max\min} + \epsilon$, and using it for all endowments makes sure that all results are plausible and comparable at the same time.[23]

## 3.4. Analysis

Let us now present the resulting structure of stability among trade agreements according to our framework. Figure 3.4 depicts the parameter space of endowments under consideration for this - it is the (marked) triangle from before. The analysis starts with the three extreme points, then turns to the connecting intervals, and

[20]The authors are grateful to Michael Chwe for the provision of an exemplary algorithm.
[21]The distance is set to 0.0013360053440215 - due to 500 points per dimension of the grid.
[22]The factor $\alpha$ always enters the welfare of country $i$ as $2\alpha e_i$ (see Appendix B.2.1). Therefore, any changes above the minimal value leave the welfare levels and therefore the rankings unaffected.
[23]In our computation $\epsilon$ is simply fixed to 0.01, which yields $\alpha = 1.3988888888888888$.

finishes with the entire interior. In each of these cases, two scenarios are examined. The first scenario corresponds to the current WTO institutional arrangement while the second one assumes modified WTO rules without Article XXIV Paragraph 5, which would prevent the formation of PTAs (specifically CUs and FTAs).

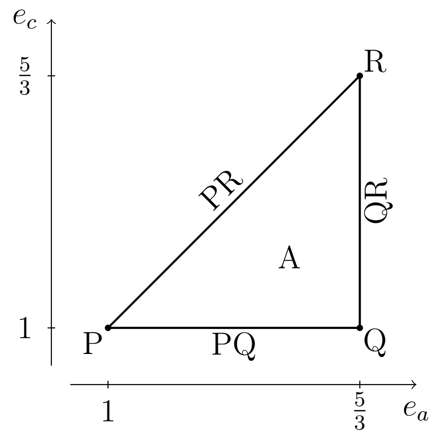



FIGURE 3.4. Overview of the different points, intervals and areas of interest depending on the (partially normalized) endowment tuple

The remainder of this analysis is structured as follows. First, Section 3.4.1 considers the symmetric case, see point P in Figure 3.4, where all countries are identical. Second, Section 3.4.2 features the two extreme asymmetric cases, points Q and R, with countries that are small, small, and large (Q) or small, large, and large (R). Next, Section 3.4.3 discusses the three related intervals, sides PQ, QR, and PR, where the countries are small, small, and varying (PQ), small, large, and varying (QR), or small with two varying equally (PR). Finally, Section 3.4.4 describes the inner area, area A, with three distinct countries.

**3.4.1. Symmetric Case.** First, let us consider the symmetric case, where symmetry refers to identical endowments for all countries, i.e. $e_a = e_b = e_c = 1 = e_{\min}$, and corresponds to point P in the triangle of Figure 3.4. As the countries do not differ from one another, the only thing that matters for welfare is whether a country is an insider or an outsider in a specific trade agreement. In the following, we present the ranking of preferences from the perspective of country $a$, which represents that of all other countries as well, for fixed $i, j \in N \setminus \{a\}$ with $i \neq j$:

$$\begin{aligned} CU(i,j) \prec_a MFN \prec_a MTA(a,i) \prec_a FTAHub(i) \prec_a FTA(i,j) \prec_a FTA(a,i) \\ \prec_a CU(a,i) \prec_a MTA(i,j) \prec_a GFT \prec_a FTAHub(a) \end{aligned}$$

The case where two countries form a CU is the least favourable trade constellation for the third country. Under such circumstances, the outsider faces the second-highest tariffs (with MFN-tariffs the highest), while the insiders cancel the tariffs among themselves. The exports of country $a$ to the other countries, $i$ and $j$, are the lowest under CU(i,j) compared to all alternative trade agreements. The same applies to the total imports. In other words, the 'trade diversion' effect is the strongest for country $a$ in case of CU(i,j). In general, the MFN regime favours country $a$ when compared to CU(i,j). The tariff revenue remains the same, while the consumer surplus is lower and the producer surplus is higher - the increase offsets the decrease. The MFN regime slackens the 'trade diversion' effect present in the case of CU(i,j) by virtue of increased export values of country $a$.

Among the group of bilateral trade agreements where the country is an insider, the MTAs result in the lowest welfare (for this country). MTA(a,i) itself generates a higher welfare for country $a$ in comparison with the MFN regime on the grounds of increased consumer and producer surplus. The FTAHub(i) constellation results in even further gains in welfare for country $a$ through higher export values and producer surplus accordingly (the tariff revenue and also the consumer surplus are lower under FTAHub(i) compared to MTA(a,i) though). However, country $a$ does not have an incentive to remain in this constellation. The unilateral deviation from FTAHub(i) to FTA(i,j) comes with a decrease of consumer and producer surplus but enough increase in tariff revenue to ultimately ensure higher welfare under the latter constellation. Nonetheless, among FTAs being an outsider is less desirable than being an insider for any country. The drop in tariff revenue is offset by an expansion of the consumer and producer surplus, resulting in higher welfare for country $a$ in case of FTA(a,i) compared to FTA(i,j). As an insider, country $a$ prefers CU(a,i) over FTA(a,i) though. More precisely, in spite of the decline in the consumer surplus, the actual welfare goes up through an expansion of tariff revenue and producer surplus.

The formation of MTA(i,j) guarantees the highest welfare for country $a$ compared to any other bilateral trade agreement. The driving factor is the MFN-principle, which implies that in case of MTA(i,j) the insiders need to apply the same tariff to both each other and the outsider - a form of free-rider problem. At the same time, country $a$ attains the highest possible tariff revenue.

Each country obtains the second-highest welfare level when the world reaches global free trade. Under full trade liberalisation, the producer surplus is also the second-highest among all trade agreements (effectively driving the ranking). It is only surpassed by that of FTAHub(a). The latter constellation brings about the highest possible welfare for country $a$. But note that such a trade agreement disproportionally favours the hub country over the other countries.

Countries' strong preference rankings are the crucial ingredient for computing the LCS. In fact, for each country all three variants of global free trade are ranked as second-best option while each first-best option, a hub structure, is ranked considerably lower for the other countries. Intuitively, global free trade seems like a stable compromise. The following proposition and its proof reinforce this:

PROPOSITION 3.3. *Under symmetry and with the current institutional arrangement of the WTO, the LCS contains three elements: CUGFT, FTAGFT, and MTAGFT. In other words, (the trinity of) global free trade is the unique stable outcome.*

PROOF. Based on the definition of indirect dominance and the transition graphs, see Section 3.3.4 and 3.3.5, the preference rankings from earlier allow us to derive the indirect dominance matrix. If the entry in the matrix is equal to one (resp. zero), then the trade arrangement corresponding to the row of the entry is (resp. isn't) indirectly dominated by the one corresponding to the column of the entry. For example, $FTAHub(a)$ is indirectly dominated by $CUGFT$ as there exists a (finite) sequence of outcomes and coalitions such that all coalitions in the sequence prefer the final outcome over the current one:

$$FTAHub(a) \rightarrow_{\{b,c\}} CU(b,c) \rightarrow_{\{a,b,c\}} CUGFT$$

Checking for all possible sequences yields the following indirect dominance matrix:[24]

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 $MFN$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 $CU(a,b)$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| 3 $CU(b,c)$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 4 $CU(c,a)$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 5 $CUGFT$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 $FTA(a,b)$ | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| 7 $FTA(b,c)$ | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 8 $FTA(c,a)$ | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 9 $FTAHub(a)$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 10 $FTAHub(b)$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 11 $FTAHub(c)$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 12 $FTAGFT$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 $MTA(a,b)$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 |
| 14 $MTA(b,c)$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| 15 $MTA(c,a)$ | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 |
| 16 $MTAGFT$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |

Note that intuitively any outcome is stable if all deviations from it are deterred. Also, a deviation from the outcome is hindered if there is a stable outcome which might be reached and some member of the deviating coalition does not prefer it over the initial outcome. In the following procedure, start with the full set and then keep removing elements that are unstable until the remaining ones are stable

[24]Appendix B.1 contains the pseudocode for this procedure.

Take $x \in \{MFN, FTA(i,j), CU(i,j), MTA(i,j)\}$, where $i,j \in N$ with $i \neq j$, and then consider the joint deviation $x \rightarrow_{\{a,b,c\}} FTAGFT$. The $FTAGFT$ regime is not indirectly dominated by any other outcome (see the matrix above) and also $x \prec_{\{a,b,c\}} FTAGFT$ for each of those $x$. Thus the deviation $x \rightarrow_{\{a,b,c\}} FTAGFT$ cannot be deterred and therefore no such $x$ can be part of the stable set.

Consider $FTAHub(i)$, $i \in N$, and the deviation $FTAHub(i) \rightarrow_{\{j,k\}} FTAGFT$, $j,k \in N \setminus \{i\}$ with $j \neq k$. Using $FTAHub(i) \prec_{\{j,k\}} FTAGFT$ together with the logic from before eliminates $FTAHub(i)$ for each $i \in N$.

Focus on the set of remaining elements $Y = \{CUGFT, FTAGFT, MTAGFT\}$. Start with any element $y$ in $Y$. If there is a deviation to any element $x \in X \setminus Y$, then there always exists an indirect dominance path (see indirect dominance matrix) $x \ll y'$ coming back to an element $y' \in Y$. In addition, for any $y_1, y_2 \in Y$, $y_1 \neq y_2$, there does not exist a coalition $S \subseteq N$, $S \neq \emptyset$, for which $y_1 \prec_S y_2$. Thus, the set $Y$ satisfies the (internal) stability condition while being maximal, i.e. $Y = LCS$. □

In the symmetric case, under the current institutional arrangement of the WTO, the global free trade variations appear as the only stable constellation according to our framework. But what would happen without Article XXIV Paragraph 5? In this case, countries would not have the option to liberalise trade through the formation of CUs or FTAs - leaving MTAs as the only possibility. The representative preference ranking of country $a$ would look as follows:

$$MFN \prec_a MTA(a,i) \prec_a MTA(i,j) \prec_a MTAGFT$$

Each country achieves the peak welfare under MTAGFT. Thus, it is reasonable to conjecture stability of MTAGFT. The following proposition proves this intuition:

PROPOSITION 3.4. *Under symmetry and with the modified institutional arrangement of the WTO (no PTAs), the LCS contains one element: MTAGFT. In other words, global free trade is the unique stable outcome.*

PROOF. The indirect dominance matrix is derived as before:

$$\begin{array}{l} \\ 1\,MFN \\ 2\,MTA(a,b) \\ 3\,MTA(b,c) \\ 4\,MTA(c,a) \\ 5\,MTAGFT \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Let us start with the full set of trade agreements again (limited to the setting). If the grand coalition moves from $MFN$ to $MTAGFT$, then the only possibility is to stay there, as $MTAGFT$ is not indirectly dominated by any other outcome. Moreover, $MFN \prec_{a,b,c} MTAGFT$. Thus, $MFN$ cannot be stable. Furthermore, if the grand coalition moves from any bilateral $MTA$ regime to $GFTMTA$, by the same argument, it is clear that no bilateral $MTA$ can be stable. Finally, any deviation from $MTAGFT$ will come back to itself due to the indirect dominance. Consequentially, the set $Y = \{MTAGFT\}$ is consistent and also the largest one. □

If symmetry among all countries holds, then Article XXIV Paragraph 5 does not change anything in terms of stability and corresponding welfare, both individual and overall. The only stable trade constellation is (the trinity of) global free trade.

**3.4.2. Asymmetric Case - Vertices of the Triangle.** It is natural to start the analysis of the asymmetric case by considering its two extreme scenarios, which correspond to the points Q and R in the triangle of Figure 3.4. In the following, Section 3.4.2.1 discusses the case of countries that are small, small, and large (Q) while Section 3.4.2.2 focuses on countries that are small, large, and large (R).

3.4.2.1. ***The case of two small and one large country.*** In this scenario, fix $e_a = e_{max}$ and $e_b = e_c = e_{min}$ (point Q). Let us start with the ranking of preferences for country $a$ and another country $i \in N \setminus \{a\}$ - representing also $j \in N \setminus \{a, i\}$:

$$CU(i,j), FTA(i,j) \prec_a FTAHub(i) \prec_a MFN, MTA(i,j) \prec_a FTA(a,i) \\ \prec_a MTA(a,i) \prec_a CU(a,i) \prec_a GFT, FTAHub(a)$$

$$MTA(a,i) \prec_i GFT, FTAHub(a) \prec_i CU(a,i) \prec_i FTA(a,i) \prec_i CU(a,j) \\ \prec_i MFN, MTA(i,j) \prec_i FTAHub(i) \prec_i FTA(a,j) \prec_i MTA(a,j) \\ \prec_i FTAHub(j) \prec_i CU(i,j), FTA(i,j)$$

One immediately notices that small and large countries have different rankings. A large country profoundly dislikes the scenarios where it is an outsider; while the small countries, by contrast, dislike any trade arrangements with the large country. Note that in certain cases countries actually do not differentiate between different trade constellations.[25] For example, CU(i,j) and FTA(i,j) result in same welfare for all countries. In this case, under the given pattern of endowments, the optimal tariffs of the small countries for CU and FTA are above the MFN-tariff. However, the Sub-paragraphs of Article XXIV Article 5 rule this out and therefore the tariffs are capped at the MFN-level. A similar argument applies to the case of FTAHub(a). Here, the optimal tariffs of the small countries would be negative. By restricting tariffs from below by zero implies that FTAHub(a) corresponds to GFT, or rather a Pseudo-GFT. Finally, the MTA between the small countries actually coincides with the MFN regime because of identical optimal tariffs for both cases.

Next, let us analyse the preferences of the large country $a$. As mentioned above, being the outsider produces the least favourable constellations for a large country.

[25]Additional details on this can be found in Appendix B.2.2 and Appendix B.2.5.1.

The worst scenarios are those where the small countries form a PTA. In such cases, the export, and hence the producer surplus, is the lowest in the large country. Now, compared to these PTAs among the small countries, both the tariff revenue and the consumer surplus are lower under FTAHub(i), but the comparably strong growth of the producer surplus produces an increase in welfare. Further increases in producer surplus are possible via the MFN regime or MTA(a,i). As soon as the large country forms an FTA, its welfare increases due to trade relations that benefit its exports within the constellation. MTA(a,i) leads to an even higher welfare, but there the driving factor are the tariff revenues. Among all bilateral trade agreements, where the large country is an insider, the optimal outcome is CU(a,i). The highest welfare for the large country occurs when trade is fully liberalised though. In that scenario, it is able to completely reap the trade benefits - the producer surplus peaks when compared to the other trade agreements.

Now, let us consider the preferences of a small country $i$, the countries $i$ and $j$ are indistinguishable from each other in this respect. Contrary to the preferences of a large country, it is in its best interest for a small country to avoid forming any trade arrangement with the large country. The smallest welfare of a small country occurs in the case of MTA(a,i), as the constellation generates one of the least desirable combinations of tariff revenue and producer surplus. Under FTAHub(a) or GFT, the small country achieves higher welfare through an increase in producer surplus and despite a decrease in tariff revenue and consumer surplus. CU(a,i) and FTA(a,i) each lead to further welfare improvements for the small country. In both cases, the driving factor is a higher consumer surplus. Next, regardless of the lower consumer and producer surplus under CU(a,j) (compared to FTA(a,i)), its higher tariff revenue actually results in a higher welfare. The tariff revenue stays at its peak under MTA(i,j) or the MFN regimes as well, but with higher welfare. Specifically, an increase in consumer surplus offsets the decrease in producer surplus. By comparison, FTAHub(i) actually increases the producer surplus and thereby also the total welfare. FTA(a,j) then generates its peak tariff revenue from before and through this a higher welfare for the small country as an outsider. Under MTA(a,j) the tariff revenue stays the same and the lower producer surplus gets (more than) compensated by the higher consumer surplus. Compared to this, FTAHub(j) decreases the tariff revenue but increases both the consumer and the producer surplus enough to increase the welfare. Finally, a small country attains the best result by forming a PTA with the other small country, either through FTA(i,j) or CU(i,j). While keeping relatively high tariff revenues, the small countries manage to have a high producer surplus as well.

Using the preference rankings to derive the LCS yields the following proposition:

PROPOSITION 3.5. *With the endowments given by $e_a = e_{max}$ and $e_b = e_c = e_{min}$, and under the current institutional arrangement of the WTO, the stable constellations are the PTAs between the two small countries, that is CU(b,c) and FTA(b,c).*

PROOF. Let us start by giving the indirect dominance matrix:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 $MFN$ | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 $CU(a,b)$ | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 3 $CU(b,c)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 $CU(c,a)$ | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 $CUGFT$ | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 6 $FTA(a,b)$ | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 $FTA(b,c)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 $FTA(c,a)$ | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 9 $FTAHub(a)$ | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 10 $FTAHub(b)$ | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 $FTAHub(c)$ | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 $FTAGFT$ | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 |
| 13 $MTA(a,b)$ | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 $MTA(b,c)$ | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 15 $MTA(c,a)$ | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 $MTAGFT$ | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

Recall that $X$ denotes the full set and let $Y = \{CU(b,c), FTA(b,c)\}$ be the candidate for the LCS. Take any element $x$ from the set $X \setminus Y$ and consider the deviation $x \rightarrow_{\{b,c\}} CU(b,c)$. Note that $CU(b,c)$ is not indirectly dominated by any other element from $X$ and furthermore $x \prec_{\{b,c\}} CU(b,c)$ for all $x \in X \setminus Y$. Thus, the deviation $x \rightarrow_{\{b,c\}} CU(b,c)$ can not be deterred for all $x \in X \setminus Y$. Therefore, no such $x$ can be part of the stable set.[26]

As each outcome in $X \setminus Y$ is indirectly dominated by $y \in Y$ (see the matrix), for any coalition and any deviation away from $y \in Y$ there always exists a path of indirect dominance back to $Y$. Moreover, no coalition is actually better off when coming back to $Y$, as $x \not\prec_S y$ for all $x, y \in Y$, $x \neq y$, and $S \subseteq N$, $S \neq \emptyset$. Therefore, the set $Y$ satisfies the (internal) stability condition while being maximal, i.e. $Y = LCS$. □

Even though global free trade is the most desirable regime for the large country, the two small countries do not have any incentive to form such a constellation and the large country can not enforce it. As a consequence, country $a$ ends up with the worst trade agreement (from its perspective). Thus, in this scenario the size advantage

[26]It might appear that this proof deviates from the general approach of eliminating element by element from the full set until the remainder forms the stable set. However, in this proof it is purely a coincidence that in one step all elements but the stable ones can be eliminated with one argument (or rather deviation).

of the large country does not translate into a favourable stable regime. Moreover, this specific case showcases the relevance of the restrictions on PTAs (remember that insiders are not allowed to raise tariffs on outsiders). The constraint makes the small countries be indifferent between the two forms of PTAs.

Now we turn to the hypothetical scenario without Article XXIV Paragraph 5. Here, the ranking of preferences for the countries, with country $a$ the large one and country $b$ and $c$ small (represented by $i$ and $j$), are as follows:

$$MTA(i,j), MFN \prec_a MTA(a,i) \prec_a GFT$$

$$MTA(a,i) \prec_i MTAGFT \prec_i MFN, MTA(i,j) \prec_i MTA(a,j)$$

As a result, the best outcome for a small country $i$ is the $MTA(i,j)$ regime, as the PTAs are not available anymore. The next proposition presents the new LCS as a consequence of these changes:

PROPOSITION 3.6. *With the endowments given by $e_a = e_{max}$ and $e_b = e_c = e_{min}$, and under a modified institutional arrangement of the WTO, the stable constellations are MFN and MTA(b,c).*

PROOF. The indirect dominance matrix is given as follows:

|  | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 $MFN$ | 0 | 0 | 0 | 0 | 0 |
| 2 $MTA(a,b)$ | 1 | 0 | 0 | 0 | 0 |
| 3 $MTA(b,c)$ | 0 | 0 | 0 | 0 | 0 |
| 4 $MTA(c,a)$ | 1 | 0 | 0 | 0 | 0 |
| 5 $MTAGFT$ | 1 | 1 | 1 | 1 | 0 |

Start with the full set again. If we consider the deviations $MTA(c,a) \to_c MFN$ and $MTA(a,b) \to_b MFN$, then no further deviations are expected as $MFN$ is not indirectly dominated by any other outcome. In addition, $MTA(c,a) \prec_c MFN$ and $MTA(a,b) \prec_b MFN$, so $MTA(c,a)$ and $MTA(a,b)$ cannot be part of the stable set. The same argument works in the case of $MTAGFT$ and the deviation $MTAGFT \to_{b,c} MFN$, as $MTAGFT \prec_{b,c} MFN$. So, the global free trade regime cannot be stable as well.

Let $Y = \{MFN, MTA(b,c)\}$. Following any deviation from the elements in $Y$, there is always an indirect dominance path coming back to $Y$ ($MFN$ in this case). In addition, for any $x, y \in Y$ with $x \neq y$ there does not exist coalition $S$ for which $x \prec_S y$. Thus, the set $Y$ is consistent and the largest one as well. □

In summary, when there are two small and one large country, the GFT regime is unstable under the current and hypothetical institutional set-up of the WTO. At best, world trade can be partially liberalised. Additionally, the small countries profit when they can form a PTA instead of an MTA, as the limiting MFN principle can be avoided that way.

3.4.2.2. ***The case of one small and two large countries.*** In this scenario, fix $e_b = e_{min}$ and $e_a = e_c = e_{max}$ (point R). Let us start with the ranking of preferences for country $b$ and another country $i \in N \setminus \{b\}$ - representing also $j \in N \setminus \{b, i\}$:

$$GFT \prec_b CU(i,b) \prec_b FTAHub(i) \prec_b FTA(i,b) \prec_b FTAHub(b) \prec_b MTA(i,b) \prec_b MFN \prec_b CU(i,j) \prec_b MTA(i,j) \prec_b FTA(i,j)$$

$$CU(j,b) \prec_i FTA(j,b) \prec_i MFN \prec_i MTA(i,b), MTA(j,b) \prec_i FTA(i,j) \prec_i MTA(i,j) \prec_i CU(i,j) \prec_i FTAHub(b) \prec_i FTAHub(j) \prec_i GFT \prec_i FTAHub(i) \prec_i FTA(i,b) \prec_i CU(i,b)$$

Under the given pattern of endowments, the preference rankings of the countries are considerably different from the previous cases. For the small country, the MFN regime generates higher welfare than any other trade agreement where it is part of. As for a large country, being an outsider is on the lower end of the ranking, while being an insider in a PTA with a small country is on the other end.

Let us take a closer look at the preference ranking of the small country. First, GFT actually generates the lowest total welfare - driven by no tariff revenue and not enough compensation via consumer and producer surplus. As mentioned before, any trade arrangement involving the small country results in lower welfare compared with other constellations (but higher welfare than GFT). The lowest among those are the CU with any of the large countries, which through increased tariff revenue (and despite a decrease in consumer surplus) yield higher welfare in comparison with the GFT regime. Even though FTAHub(i) reduces those gains in tariff revenue again, by virtue of a growing consumer surplus it still raises the total welfare. Further improvement in the welfare of the small country is possible if the world moves from FTAHub(i) to FTA(i,b); the sole reason is a higher consumer surplus. Under FTAHub(b), the export volumes to the large countries are at its peak and it generates substantially higher producer surplus. As a consequence, it results in the small country preferring to form a hub structure (as the hub node) over an FTA with one of the large countries. Replacing the FTA with an MTA with similar structure is the most desirable configuration for the small country among the constellations where it participates. Under MTA(i,b) the producer surplus is actually the smallest compared to all other alternatives, but high tariff revenue and consumer surplus determine its position in the ranking. The MFN regime surpasses all configurations mentioned above. When there are two large countries, the tariff revenue becomes an important factor in the welfare of the small country. Any further improvements with respect to the welfare of the small country depend on the large countries liberalising trade among themselves - the small country essentially free-rides in these cases (exhausting its tariff revenue to the fullest). The driving factor among these three is the export volume. Consequentially, CU(i,j) is the worst option, followed by MTA(i,j), and FTA(i,j) is the (overall) best outcome.

The following discusses the preferences of the two large countries. The least favourable scenario occurs when the other large country forms a CU together with the small country. Its position in the ranking is driven by the lowest export volumes and producer surplus. Now, FTA(j,b) produces higher welfare compared to the previous constellation due to growth in producer surplus (based on rising exports to the small country) which makes up for the drop in consumer surplus. A similar development makes the MFN regime an even better constellation (here the exports to the large country increase). All tariffs (and thus prices) are identical under both MTA(i,b) and MTA(j,b), as a consequence they generate the same welfare. On the grounds of increased exports, the welfare tops that of the MFN regime. Among the class of bilateral trade agreements between the large countries, the ranking goes as follows: FTA(i,j) followed by MTA(i,j) only surpassed by CU(i,j). In comparison with MTA(i,b) and MTA(j,b), the greater consumer and producer surplus of FTA(i,j) guarantees an increase of total welfare. An MTA between the two large countries produces more tariff revenues and actually results in a more desirable outcome. Moving from MTA(i,j) to CU(i,j) decreases tariff revenue and also consumer surplus but the gain in producer surplus through increased exports to the other large country makes more than up for this. FTAHub(b) and even more so FTAHUb(j) further improve the welfare via growth of the tariff revenue and consumer surplus (the case of FTAHub(b)), and increased exports to the other large country (for FTAHUb(j)). Now, the GFT regime allows the large country to raise the exports to the small country while retaining the same level of exports to the other large country. As a consequence, the welfare of GFT surpasses that of the previous mentioned constellations. However, when the large country is part of a hub structure as the hub node itself, then its exports to the small country increase such that the welfare exceeds that of full trade liberalisation. Furthermore, the FTA with the small country constitutes the second-best outcome for the large country on the grounds of high tariff revenue accompanied by similar consumer surplus. Finally, CU(i,b) is the most desirable constellation driven by the high exports to the small country.

Let us compute the LCS under these preference rankings in the next proposition:

Proposition 3.7. *With the endowments given by $e_b = e_{min}$ and $e_a = e_c = e_{max}$, and under the current institutional arrangement of the WTO, the stable constellation is the CU between the two large countries, that is CU(c,a).*

Proof. The indirect dominance matrix is given as follows:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 $MFN$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 2 $CU(a,b)$ | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 3 $CU(b,c)$ | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 4 $CU(c,a)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 $CUGFT$ | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 $FTA(a,b)$ | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 7 $FTA(b,c)$ | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 8 $FTA(c,a)$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 9 $FTAHub(a)$ | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 10 $FTAHub(b)$ | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| 11 $FTAHub(c)$ | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 12 $FTAGFT$ | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 13 $MTA(a,b)$ | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 14 $MTA(b,c)$ | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 15 $MTA(c,a)$ | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 16 $MTAGFT$ | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

First, take $x \in \{CU(i,b), FTA(i,b), MTA(i,b), FTAHub(i), FTAHub(b)\}$, with $i \in \{a,c\}$. Country $b$ can destroy such trade agreements and, depending on the initial constellation, either $FTA(c,a)$ or the $MFN$ regime remains. Then, further deviations are possible, namely $MTA(c,a)$ and $CU(c,a)$. However, each of the aforementioned trade agreements is indirectly dominated by $CU(c,a)$ and simultaneously country $b$ is better off compared to the initial situation. Consequently, such deviations can not be avoided and no such $x$ can be part of the stable set.

Now, consider $x \in \{MFN, FTA(c,a), MTA(c,a)\}$ for which $x \rightarrow_{\{a,c\}} CU(a,c)$ presents a deviation that can not be deterred. As in the previous paragraph, $CU(c,a)$ is not indirectly dominated any element and also $x \prec_{\{a,c\}} CU(a,c)$. Thus, no such $x$ can be the part of the stable set as well.

At last, let $x \in \{CUGFT, FTAGFT, MTAGFT\}$ and consider the deviations where country $b$ leaves the agreements. $CU(c,a)$, $FTA(c,a)$, or $MTA(c,a)$ can be the result. We have shown that the last two outcomes can not be stable. As for $CU(a,c)$, we have that for all $x$ considered $x \prec_{\{b\}} CU(a,c)$. As a result, we conclude that no such $x$ can be in the consistent set.

$CU(a,c)$ indirectly dominates each outcome, all deviations from it are deterred. So, the set containing $CU(a,c)$ is consistent and the largest one as well. □

The small country manages to block many desirable outcomes for large countries. Country $b$ can unilaterally deviate from any trade agreement with higher welfare than CU(i,j) for the large countries. Thus, the majority of countries cannot impose their will on the other country. What the large countries can achieve is the best trade agreement that they can reach without the participation of the small country, in this case one among themselves.

A similar story unfolds in the scenario without Article XXIV Paragraph 5. There, the countries' preference rankings are as follows, with country $b$ the small one and country $a$ and $c$ large (represented by $i$ and $j$):

$$MTAGFT \prec_b MTA(i,b) \prec_b MFN \prec_b MTA(i,j)$$

$$MFN \prec_i MTA(i,b), MTA(j,b) \prec_i MTA(i,j) \prec_i MTAGFT$$

As the logic of the corresponding preference rankings of the countries is similar to before, let us directly present the proposition:

PROPOSITION 3.8. *With the endowments given by $e_b = e_{min}$ and $e_a = e_c = e_{max}$, and under a modified institutional arrangement of the WTO, the stable constellation is the MTA between the two large countries, that is MTA(c,a).*

PROOF. In this case, the indirect dominance matrix has the following form:

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 $MFN$ | 0 | 0 | 0 | 1 | 0 |
| 2 $MTA(a,b)$ | 1 | 0 | 0 | 1 | 0 |
| 3 $MTA(b,c)$ | 1 | 0 | 0 | 1 | 0 |
| 4 $MTA(c,a)$ | 0 | 0 | 0 | 0 | 0 |
| 5 $MTAGFT$ | 0 | 0 | 0 | 1 | 0 |

Assume, $x \in \{MTA(a,b), MTA(b,c), MTAGFT\}$ and consider the deviations, where country $b$ dismantles any above mentioned constellation. Two possibilities: Either $MFN$ or $MTA(c,a)$ remain. From $MFN$ either no coalition moves away or, as it is indirectly dominated by $MTA(c,a)$ (see the indirect dominance matrix), the latter might be approached. In either case, $b$ is better off. Thus, no such $x$ can be part of the stable set.

Now, analyse the case of the $MFN$ regime. Take the following deviation: $MFN \rightarrow_{\{a,c\}} MTA(c,a)$. As $MTA(c,a)$ is not indirectly dominated by any other trade agreement and $MFN \prec_{\{a,c\}} MTA(c,a)$, the $MFN$ regime can not be stable as well.

As $MTA(c,a)$ indirectly dominates each trade agreement, all deviations from it are deterred. So, the set consisting of $MTA(c,a)$ is consistent and the largest one as well. □

Thus, similar to the other asymmetric case, one small and two large countries allow for partial but not full liberalisation of world trade irrespective of the actual scenario (current vs. modified WTO rules). In terms of overall welfare, the world is better off in the hypothetical scenario without Article XXIV Paragraph 5 though. Individually, the small country is in a better position in case of MTA(i,j) compared to CU(i,j), as it exploits the MFN obligation of the large countries. By contrast, the large countries are better off in the other case. Therefore, while none of the two institutional arrangement facilitate global free trade, they influence the welfare for the stable set (both overall and individual).

**3.4.3. Asymmetric Case - Edges of the Triangle.** Let us now turn to the cases where the endowments of countries vary along one dimension - corresponding to the sides PQ, QR, and PR in the triangle of Figure 3.4. Specifically, Section 3.4.3.1 presents the scenario where the countries are small, small, and varying (PQ), Section 3.4.3.2 discusses the setting where the countries are small, large, and varying (QR), and Section 3.4.3.3 describes the case of a small country with two varying equally (PR).

While in the previous cases it was still possible to solve the problems analytically, the following require the use of a numerical approach. The analysis presented here consists of graphics picturing the composition of the stable sets and accompanying descriptions that explore the underlying mechanics. The exact numerical values for these (sub-)intervals can be found in Section B.3.2.

3.4.3.1. ***The case of one small, one large, and one varying country.*** First, let us consider the case $e_b = e_{min}$, $e_a = e_{max}$, and $e_c \in (e_{min}, e_{max})$ (side QR). Under the given pattern of the endowments, a number of trade agreements can be completely ruled out (with respect to the LCS). The MFN and GFT regimes for example are never part of the stable set. Additionally, none of the PTAs between the small and the large country appear as a stable outcome. The same holds for the hub structures where either the small or the large country is the hub node. As for the actual composition of the LCS, see Figure 3.5 for a graphical representation.

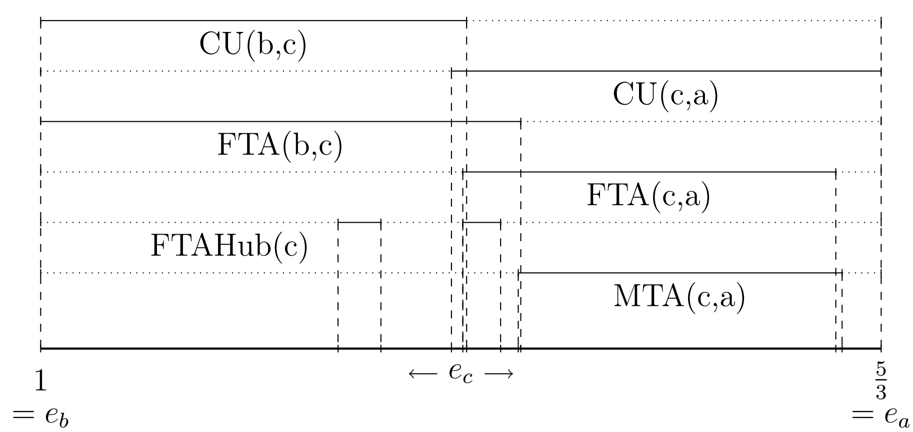



FIGURE 3.5. Characterisation of the case of small, varying, and large country

The general observation is that when the varying country is close in size to the small country, then the PTAs between these smaller countries appear as elements in the stable set. When the country becomes larger the trade constellation between the larger countries replaces these. Additionally, there are two small, separated, regions in the middle of the interval where FTAHub(c) is stable.

In order to get an intuitive understanding of the results, let us identify specific trade agreements that go from stable to unstable (or the other way around) for certain endowment tuples. Then, explore the underlying mechanics to understand why the changes happen.

Start with the PTAs between country $b$ and $c$, the small and the varying one. Interestingly, the only factor driving their stability are the preferences of country $b$ (with fixed minimal endowments). Once the MFN regime becomes more desirable than CU(b,c) for country $b$, the constellation CU(b,c) drops from the stable set. Now, an identical story holds for the case of FTA(b,c). Thus, for both constellations it only requires a single change in the preference ranking of country $b$ to influence the stable set.

The PTAs and MTAs between country $a$ and $c$ start to appear in the LCS when country $c$ is becoming relatively large and closer to country $a$ in size. At first both countries actually prefer to form a CU with country $b$, that is when country $c$ is relatively small (and CU(b,c) actually is an element in the stable set). However, once it is preferable for country $b$ to be the outsider instead of the insider in a CU, CU(c,a) emerges as a stable outcome (even though CU(b,c) still remains stable). Moreover, as soon as country $c$ prefers FTA(c,a) respectively MTA(c,a) over the MFN regime, each of them becomes part of the LCS as well. For the interval where all PTAs and MTAs between country $a$ and $c$ are stable, both countries have fixed preference relations over these outcomes:

$$FTA(c,a) \prec_a CU(c,a) \prec_a MTA(c,a)$$
$$MTA(c,a) \prec_c FTA(c,a) \prec_c CU(c,a)$$

However, as soon as country $c$ also prefers MTA(c,a) over FTA(c,a), the joint FTA drops out of the LCS. Similarly, as soon as country $a$ prefers CU(c,a) over MTA(c,a), this also applies to the joint MTA - leaving CU(c,a) as the only stable outcome.

FTAHub(c) is stable in the two small, separated, regions in (or near) the middle of the interval. In the first region, the stability is driven by the fact that country $b$ starts to value FTAHub(c) more than FTA(b,c) and gets in unison with country $a$ in this respect. Once the preferences of country $b$ over these outcomes get reversed, FTAHub(c) drops out of LCS again. In the second region, the stability of the same hub structure is largely determined by the change in the preferences of country $c$. Now, as soon as it starts to value FTA(c,a) over the MFN regime, which also puts FTA(c,a) in the LCS, both FTAs with $c$ as a partner are stable and consequentially the corresponding hub structure is stable as well. As soon as the free-riding incentives of country $b$ increase (valuing the MFN regime more than FTAHub(c)), this hub structure is not part of the stable set anymore.

The hypothetical institutional arrangement without Article XXIV Paragraph 5 does not promote the appearance of GFT as part of the stable set. GFTMTA, but also MTA(a,b) and MTA(b,c) never emerge as stable outcomes. Varying the size of country $c$ generates either the MFN regime or MTA(c,a) as the stable element. Figure 3.6 presents these findings.[27]

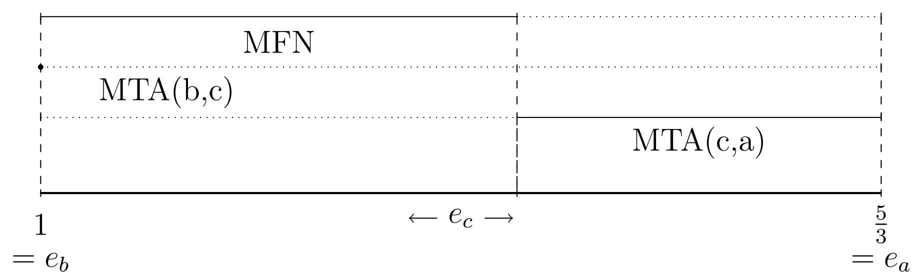



FIGURE 3.6. Characterisation of the case of small, varying, and large country

Over the whole interval, country $b$ does not have any incentive to form an MTA with any of the other countries. This is one reason why the MFN regime is stable over the specific range of the interval. The other reason is that country $c$ prefers to not have a trade agreement with country $a$ as long as its own size is not too large. Once country $c$ gets sufficiently large though, MTA(c,a) presents a better option than the MFN regime. As a consequence, MTA(c,a) replaces the MFN regime as the stable set.

As a side note, while in the first scenario (with PTAs), the LCS near and at each respective extreme point corresponded to each other (continuity), the situation is different in the second scenario (without PTAs). When country $c$ and $b$ are equal in size, MTA(b,c) appears in the LCS even though it is not there before. Here, both the MFN regime and MTA(b,c) generate the same welfares for all countries (see also the discussion on point Q in Section 3.4.2.1).

Finally, under this given pattern of endowments, the GFT regime does not appear as part of the stable set independent of the scenario (with and without PTAs). However, the choice of rules does determine whether partial trade liberalisation takes place or not. The possibility of forming PTAs reduces the incentive of the small(est) country to free ride. Otherwise, the MFN regime is the unique stable outcome when there is one small, one large, and one comparably small country.

[27]In addition to the aforementioned elements, it also pictures MTA(b,c) as a single point, see the dot, but this appears only for completion sake because that point corresponds to one of the extreme cases (point Q) discussed earlier.

3.4.3.2. ***The case of two small, and one varying country.*** Second, let us showcase the scenario with $e_b = e_c = e_{min}$ and $e_a \in (e_{min}, e_{max})$ (side PQ). In contrast to the previous case, it is not possible to rule out many of the trade agreements. Only MFN, FTAHub(a), and MTA(b,c) never appear in the LCS. The stable set is then presented in Figure 3.7.[28]

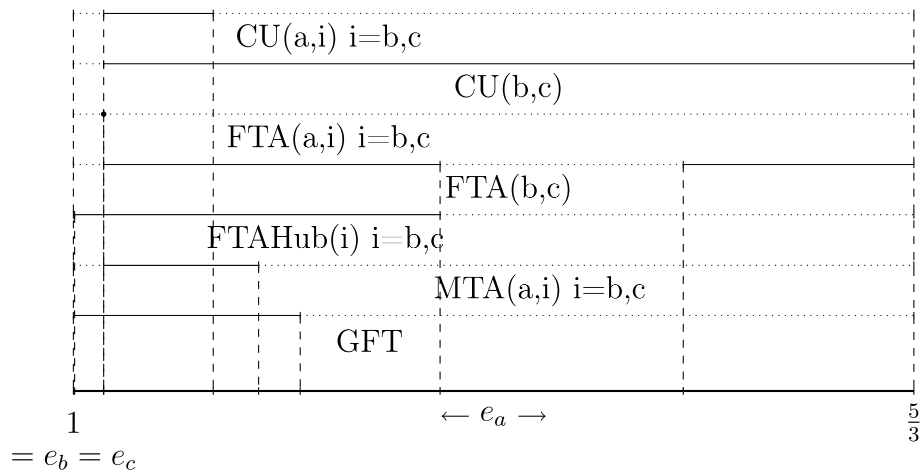



FIGURE 3.7. Characterisation of the case of small, small, and varying country

In the immediate vicinity around symmetry, the GFT regime is the only element of the LCS (or rather the group of the three variants forms the stable set), but both FTAHub(b) and FTAHub(c) emerge as stable outcomes when moving away from the extreme point. On the whole interval a number of different PTAs and MTAs, mostly between a small and the larger country, appear. Near the other point, only PTAs among the small countries are still stable.

First, the spike in the number of stable constellations close to symmetry actually follows a change in the preferences of the varying country with respect to CU(b,c) and the GFT regimes - it starts preferring the first over the latter. Furthermore, FTA(a,b) and FTA(c,a) become unstable because the small countries start to like the MFN regime more than the GFT variants (or rather these are only stable for that instance where it is not the case). When country $a$ gets sufficiently large, country $b$ prefers FTAHub(c) over CU(a,b) and $c$ prefers FTAHub(b) over CU(c,a). As a consequence, both of these CUs drop out from the LCS. Similarly, when country $b$ and $c$ start preferring FTAHub(c) and FTAHub(b) over GFT, the latter stops being stable. A similar argument also applies to the MTAs. When the size of country $a$ increases even more, both country $b$ and $c$ favour CU(b,c) over their respective hub structure, which results in FTA(b,c), FTAHub(b), and FTAHub(c) becoming unstable. When the endowment of country $a$ gets close to maximum, the small countries are constrained

[28]The dot marks a single point again.

by the MFN-tariffs and do not differentiate between CU(b,c) and FTA(b,c) anymore, which makes FTA(b,c) stable again.

In the scenario without Article XXIV Paragraph 5, the interval of GFT increases significantly. Moreover, over two-thirds of this interval the GFT regime is the unique element in the LCS. Additionally, all possible combinations of MTA appear at some point (mostly close to symmetry). Figure 3.8 demonstrates the results.[29]

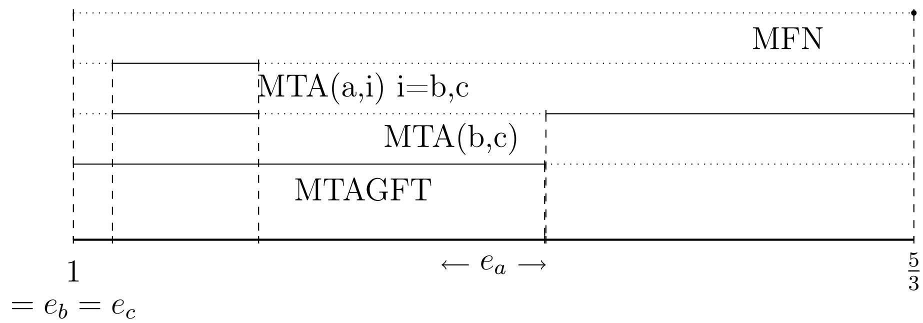



FIGURE 3.8. Characterisation of the case of small, small, and varying country

Around symmetry, GFTMTA is the only element in the stable set. As soon as the small countries start to prefer MTAs with country $a$ over GFTMTA, all three MTAs appear in the LCS. When the size of country $a$ increases, the MTAs drop out from the LCS, because the small countries rank the one with the large country as the worst trade agreements (switching last place with the MFN regime), which actually also influences the stability of the MTA among themselves. Furthermore, the GFT regime becomes unstable when the small countries start to prefer their joint MTA over GFTMTA.

Similar to the previous case, the LCS changes at one extreme point. Namely, when the endowment of country $a$ reaches the maximum, the MFN regime appears in the LCS, as MFN and MTA(b,c) generate identical welfare for all countries (again, see also the discussion on point Q in Section 3.4.2.1).

Under this pattern of endowments, the first scenario does not allow for a sharp prediction via the LCS (unlike the previous case). Especially around symmetry, where almost all trade agreements are part of the stable set. In the second scenario, the effect of the PTAs on the stability of the GFT regime is significant though - essentially the abolishment of Article XXIV Paragraph 5 would facilitate the formation of GFT as long as there are two small countries and the third country is not substantially larger.

3.4.3.3. ***The case of one small, and two varying countries***. Finally, let us turn to the case where $e_b = e_{min}$ and $e_a = e_c \in (e_{min}, e_{max})$ (side PR). In this

[29]As before, in addition to the mentioned trade agreements, the graphic also contains MFN as a single point, see the dot, at an extreme point (again point Q).

scenario, depending on the size of the larger countries, any trade agreement can be part of the stable set. The exact composition of the LCS is the basis for Figure 3.9.

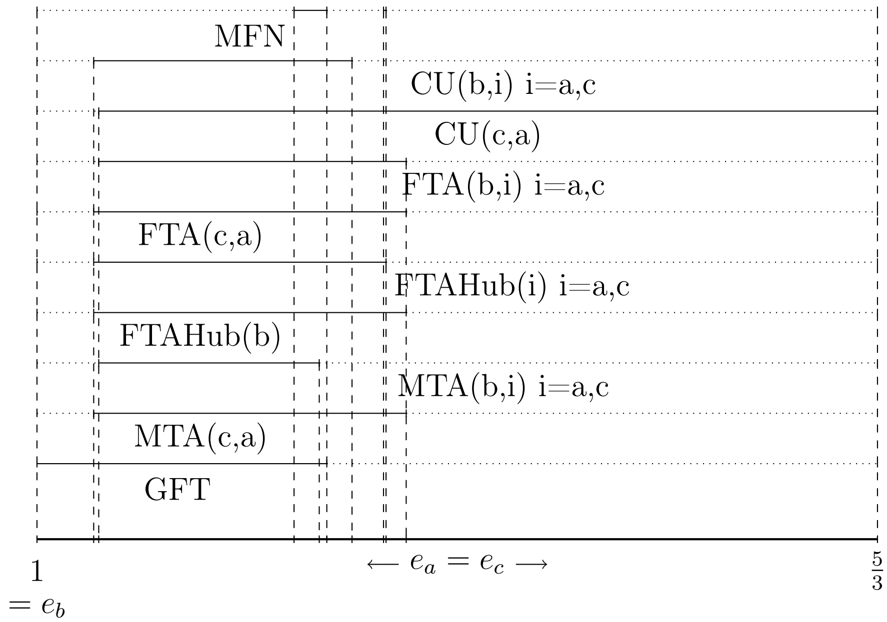



FIGURE 3.9. Characterisation of the case of small, varying, and varying country

In the interval around symmetry, the GFT variants are stable and stay the unique elements of the stable set for longer (compared to the previous case). Also, a collection of different trade agreements is stable relatively close to symmetry. However, near the other extreme point, the CUs between the varying countries is the unique stable outcome. Also, MFN is stable in two small, separated, regions.

Again, the peak in stability near symmetry comes from a shift in the preferences of the varying countries with respect to CUs. At that point, both of these start to prefer a CU with the small country over the different forms of GFT. The occurrence of the MFN regime actually follows a preference of the small country of MFN over GFT (the first region) and then FTAHub(a) and FTAHub(c) (the second region). As countries $a$ and $c$ are getting bigger, first MTA(a,b) and MTA(b,c) drop out from the LCS when they rank the lowest according to the preferences of country $b$. The three variants of GFT become unstable once the small country prefers CU(c,a). Next, CU(a,b) and CU(b,c) follow as the small country starts to prefer to be in the MFN regime over a CU with any of the larger countries. As soon as CU(c,a) becomes the more desirable trade agreement for country $b$ when comparing it to any FTA where $b$ participates or any hub structure with a large country as hub, all aforementioned constellations drop out from the LCS.

Contrary to the previous case, switching off Article XXIV Paragraph 5, actually decreases the interval where the GFT regime is part of the stable set. However, this effect is considerably smaller. A similar observation holds for the range where the GFT regime is the unique stable outcome. The exact composition can be seen in Figure 3.10.

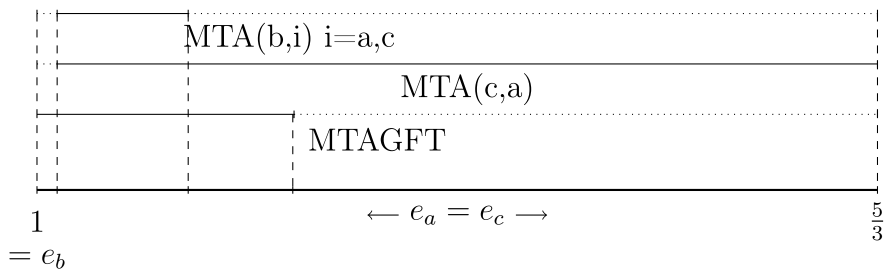



FIGURE 3.10. Characterisation of the case of small, varying, and varying country

The main driving force behind the stability are alterations in the preferences of the small country over the interval. More precisely, it is important where exactly the small country places the MFN regime in its ranking of preferences compared to the other trade agreements. As soon as country $b$ prefers MFN over another constellation, the latter drops out from the LCS. After a certain point, the MTA between the large countries remains the only stable outcome.

Similar to the previous pattern of endowments, this case makes a clear analysis in the first scenario difficult, especially around symmetry where, as before, almost all constellations are stable. The effect of Article XXIV Paragraph 5 actually works in the other direction on the stability of the GFT regime when compared to the previous case though.

**3.4.4. Asymmetric Case - Interior of the Triangle.** In the following (and final) part of the analysis, the focus lies on the interior of the triangle of Figure 3.4. Here, unlike in the previous discussions, both CU and FTA appear together under the label of PTA. However, a variation of the graphics of this analysis that actually distinguishes between the two can be found in Appendix B.3.1. For the purpose of a general overview, this level of abstraction suffices though - in fact, the members of a specific trade agreement are suppressed for clarity as well, i.e. who is insider and outsider.

First, we consider the existing institutional set-up, where PTAs are available to the countries. Figure 3.11 shows the (simplified) stable sets. In a small region close to symmetry, region one, the trinity of GFT regimes is the unique stable element. In both a neighbouring and another distant area, region two, PTAs become stable as well. The connecting area, region three, adds MTAs as another stable element. In a tiny area near the diagonal, region four, no form of trade agreement can actually be excluded from the stable set. Further along the diagonal and in the asymmetric corners, region five, PTAs are the only stable trade constellation. In between, region six, MTAs are also stable. In another tiny area, also close to the diagonal, region seven, MFN enters the stable set as well. In general, with a certain degree of asymmetry among countries at most partial trade liberalisation can be expected.

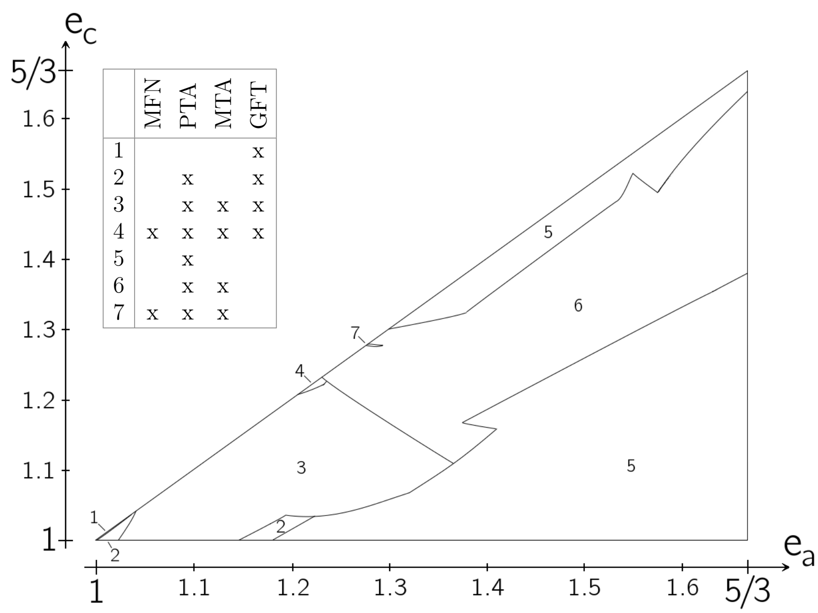



FIGURE 3.11. Simplified Overall Stability with PTAs

Next, we consider a modified institutional set-up, where PTAs are not available. Figure 3.12 depicts the corresponding stable sets. In an area near symmetry as well as in a sizeable area away from it, region one, GFT is again the unique stable element. Connected to these are two areas, region two, where MTAs become stable as well. Moving towards the asymmetric corners, region three, yields MTAs as the only stable element. In between, region four, only MFN remains in the stable set.

The comparison of the graphics allows us to deduce two compelling statements. The first noteworthy result is the extent of MFN in each scenario. In the modified institutional arrangement without PTAs the area where MFN is (uniquely) stable increases substantially (note that this effect is present away from symmetry). Under (significant) asymmetry, it seems that PTAs allow countries to move towards their international efficiency frontier (cf. Bagwell et al. (2016)).

The second interesting result is the difference in the extent of stability of GFT in the two regulatory scenarios. First, recall that once the degree of asymmetry surpasses a certain threshold, none of the GFT regimes remains in the stable set, independent of the institutional set-up. Around symmetry the opposite holds in that the GFT regimes are always stable there (in both scenarios). In between, the effect of PTAs on the stability of GFT depends on the structure of asymmetry. See Figure 3.13 for the different areas of stability depending on the regulatory scenario. Note that region one corresponds to the aforementioned stability around symmetry. In

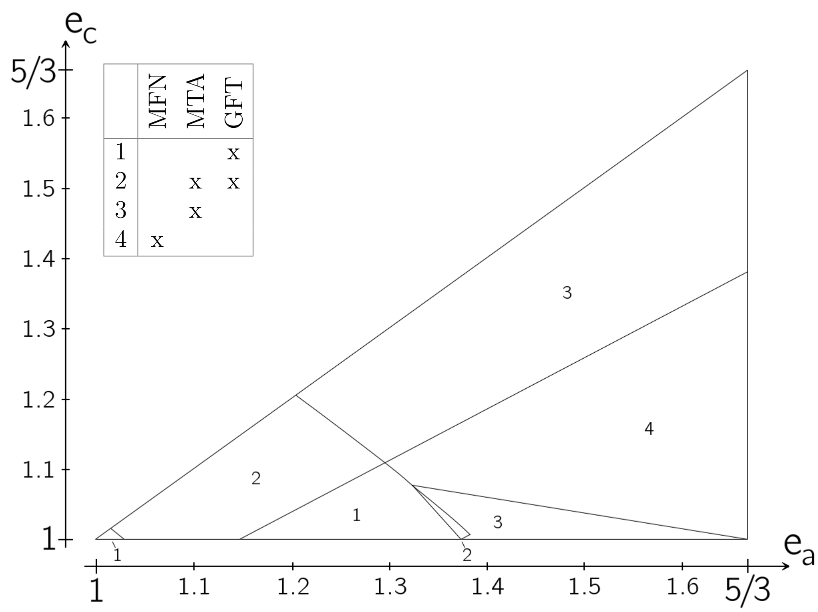



FIGURE 3.12. Simplified Overall Stability without PTAs

the case of two relatively larger countries (but not too large), the abolishment of PTAs results in a reduction of the area where GFT is stable, see region two. In this instance, PTAs act as 'building blocks' on the road to GFT. But when two countries are relatively smaller (but not too small), the same regulatory action yields the exact opposite effect, see region three. Here, PTAs are 'stumbling blocks'. Thus, whether PTAs are 'building blocks' or 'stumbling blocks' in the vicinity of symmetry depends on the relative size of the majority of the countries.

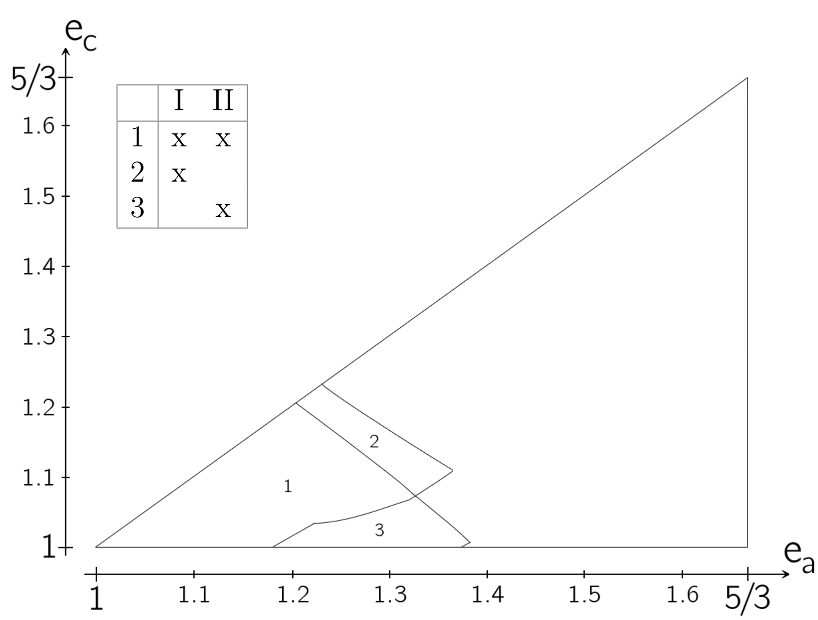



FIGURE 3.13. The different areas of stability of the GFT regime in the scenario with (I) and without (II) PTAs

In a nutshell: If the world is in the vicinity of symmetry and two out of three countries are close to identical while relatively smaller than the other one, the area where the GFT regime is stable increases when prohibiting PTAs. However, when two similar countries are relatively larger, the availability of PTAs is conducive to the stability of the GFT regime. Finally, if the world is further away from symmetry, full trade liberalisation is not attainable at all and an area where the MFN regime is stable appears in the scenario without PTAs.

## 3.5. Discussion

In this section, let us first compare the findings of our work with those of several similar studies and underline the differences in the modelling strategies, especially with respect to the explanatory power of each approach. Second, this section links our predictions to different empirical observations, thereby validating our approach.

Let us start with the paper of Saggi, Woodland and Yildiz (2013).[30] First note that the underlying trade model in our approach is similar to theirs, which allows a direct comparison of the findings in certain scenarios (found in the next paragraph). The first distinction is the set of trade agreements under consideration. While in our model countries can be involved in multilateral trade liberalisation via MTAs or they may choose to carry out their favoured form of preferential trade liberalisation through CUs or FTAs, Saggi, Woodland and Yildiz (2013) focuses on two out of these three possibilities, namely CUs and MTAs. In our opinion, the expanded set of trade arrangements in our model allows us to fully capture the trade-offs among the alternatives and make the model realistic. The second significant difference is the concept of stability. While our framework uses the notion of LCS, the paper of Saggi, Woodland and Yildiz (2013) utilises Coalition-Proof Nash Equilibria. As Bernheim, Peleg and Whinston (1987) note, their notion of self-enforceability, which is critical for Coalition-Proof Nash Equilibria, is too restrictive in one crucial aspect, mainly: 'When a deviation occurs, only members of the deviating coalition may contemplate deviations from the deviation. This rules out the possibility that some member of the deviating coalition might form a pact to deviate further with someone not included in this coalition.' Importantly, this limitation does not affect the concept of LCS. It is not a pure academic difference. The historic development of the two (disjoint) trade constellations in Europe in the 1960s, the European Economic Community (EEC) and the European Free Trade Association (EFTA), can not be captured by a model using Coalition-Proof Nash Equilibria, because it excludes those strategies

[30]As mentioned in Section 3.2, this paper analyses the case of CUs and MTA while their other papers (Saggi and Yildiz (2010) resp. Missions, Saggi and Yildiz (2016)) focus on different combinations of trade agreements (FTAs and MTAs resp. CUs and FTAs) but use a similar framework. As a consequence, the comparison of methodology applies to these papers as well.

that the UK actually followed during that time.[31] While being a member of EFTA the UK applied for EEC membership in 1961 and thereby undermined the stability of the EFTA. Furthermore, 'the more ambitious Kennedy Round between 1964 and 1967 coincided with negotiations to expand the EEC to include Britain, Ireland, Denmark, Greece, and Norway - and was motivated in part by US concerns about being excluded from an ever-broader and more unified European market.' (World Trade Report 2011). Thus, unlike the Coalition-Proof Nash Equilibrium, the LCS allows interactions among members and non-members of coalitions simultaneously, thereby accommodating these historic developments. Additionally, the (conjectured) motivation of the US reinforces the importance of the interaction among different modes and forms of trade liberalisation.

In specific cases it is actually possible to directly compare the composition of the stable sets of chapter 3 with those of Saggi, Woodland and Yildiz (2013). In fact, their 'multilateralism game' fits our scenario of a modified institutional arrangement without PTAs. Compare Figures 2 and 5 of Saggi, Woodland, and Yildiz (2013) with Figures 3.10 and 3.8 in our work correspondingly. In the case with one small country and the other two varying, both approaches predict the same stable sets near the endpoints of the interval. In our work GFT stays part of the stable set even when MTAs, either between the large countries or a small and a large country, become stable as well - which is in contrast to Saggi, Woodland, and Yildiz (2013). A similar observation follows for the case of two small and one varying country, i.e. near the endpoints of the interval results coincide while the appearance of MTAs does not prevent GFT from staying in the stable set. Furthermore, in our model there exists an interval where GFT becomes the unique stable element once more. It seems that one effect of the unlimited farsightedness is the proliferation of GFT.

Another relevant paper is that of Lake (2017). Apart from the stability concept, the approach of Lake also differs with respect to the choice set. There, the focus lies on FTAs. In this respect, a direct comparison of the findings is difficult. Moreover, compared to the previous paper, the 'multilateralism game' is further simplified in Lake (2017), as the only possible regime there is the three country constellation that results in GFT. Furthermore, as the underlying trade model, the paper employs the political economy oligopolistic model. However, according to the paper, the findings are robust with respect to various underlying trade models, including the competition via exports model. Additionally, Lake himself compares his results to those of Saggi and Yildiz (2010). Due to the similarity of the 'multilateralism game' in Saggi and Yildiz (2010) and Saggi, Woodland, and Yildiz (2013), it is only logical to compare our results with those of Lake (2017) as with the previous paper.

[31]See Baldwin and Gylfason (1995).

According to Lake (2017), specifically Figure 3, the exact role of FTAs under asymmetry depends on the nature of asymmetry (similar to our findings). However, the direction of the effect of PTAs on trade liberalisation is the opposite. There, in case of two larger and one small country, FTAs act as '(strong) stumbling blocs', and with two smaller and one larger country, as '(strong) building blocs'. Furthermore, there it seems that the determining factor are the preferences of the larger countries, while in our case the findings are driven by the preferences of the smaller countries. Thus, the aforementioned differences in the choice set and stability concept appear to shift the power to influence the negotiations among the countries, which then produces a different outcome. Specifically, in the case of the 'multilateralism game', which corresponds to our scenario without PTA, there are essentially two areas, that is one where GFT emerges as unique equilibrium and one with MFN instead. In the parameter space triangle the first makes up the upper left part of the triangle, while the second makes up the opposing lower right.[32] Therefore, for two larger countries the GFT regime remains the unique equilibrium for the whole interval, whereas in our model it only stays stable in the vicinity of symmetry (only partially unique) and then the MTA between the larger countries is the unique stable outcome, as seen in Figure 3.10. Furthermore, for two smaller countries first GFT then MFN is the unique equilibrium, while in the beginning GFT is also stable in our case (although only partially unique) the MTA between the small countries takes it place as the unique stable element near the end, depicted in Figure 3.8. Finally, in the case of three different countries, it starts with MFN and ends with GFT as the unique equilibrium, which corresponds to our findings for the first part but then in the second part the MTA between the medium and large country is the unique stable element, visible in Figure 3.6.

Another aspect of Lake (2017) necessitates a remark, namely the assumption that a once created trade agreement remains binding from then on. Lake argues that 'the binding nature of trade agreements is pervasive in the literature and realistic'. However, the latest developments in the world cast doubt on this plausibility. The USA, for example, pulled out of the negotiations for the Trans-Pacific Partnership at the final stage and currently negotiates with South Korea to amend the so-called KORUS FTA. The developments around 'Brexit' are another argument for modelling non-binding trade agreements. Using the LCS as stability concept allowed us to accommodate such deviations.

A final remark on the relation of our research with empirical observations. As the analysis has shown, a growing degree of asymmetry among countries produced a significant area of stability for the MFN regime when PTAs are prohibited, see region

[32]The first corresponds to the regions denoted WBB and SSB in Figure 3 of Lake (2017), while the second matches the regions SBB and WSB.

four in Figure 3.12. If one would interpret the expansion of the WTO rule set to an increasing number of countries as an amplification of asymmetry among its member states, then the potential of PTAs to prevent the MFN regime might be one of the driving factors of the prevalence of PTAs in recent history. However, the World Trade Report (2011) casts doubt on this motive. According to the report: 'Approximately 66 per cent of tariff lines with MFN rates above 15 percentage points have not been reduced in PTAs.' Note that a reduction of the tariff peaks for 34 percent of the tariff lines is still a significant effect considering the fact that the majority of tariff peaks occurs in agricultural and labor-intensive manufacturing sectors, which are politically sensitive and countries usually try to exclude them from the trade liberalisation via PTAs. Furthermore, according to the same report, over time more and more PTAs have included provisions regarding technical barriers to trade - a category of the Non-Tariff Barriers (NTBs). The paper of Kee, Nicita, and Olarreaga (2009) estimates that restrictiveness measures that include NTBs are on average about 87 percent higher than the measures based on tariffs alone. Thus, in order to evaluate the aforementioned motive properly, the effects of NTBs should be included in the analysis as well. Moreover, contrary to the conclusion of the report that 'preference margins are small and market access is unlikely in many cases to be an important reason for creating new PTAs', Keck and Lendle (2012) show that the preferential utilisation rates are often high even in the case of small preference margins and they increase both with the preference margin and the export volume. All in all, it is our opinion that there is (partial) evidence for the motive of avoidance of MFN via PTAs which coincides with the predictions of our model about the trade-off between trade liberalisation and the MFN regime in the asymmetric part of the parameter space.

As this overview showed, a number of core attributes of the LCS, specifically the farsightedness, the non-binding nature of agreements, and the possibility of interactions between members and non-members of coalitions, capture important mechanisms present in the world and influence the composition of the stable set significantly (when compared to other stability concepts).

## 3.6. Conclusion

Under the rules of the WTO (previously GATT), a group of countries can engage in both multilateral and preferential trade liberalisation. The formation of global trade agreements is a complex game and the rules of the game influence the nature of the exact outcomes. WTO's Article I aims at creating the global free trade system, while Article XXIV Paragraph 5 allows countries to seemingly circumvent the liberalisation process. In chapter 3, our focus lies on the stability of trade policy arrangements under two different regulatory scenarios (with and without PTAs) assuming unlimited farsightedness of the participants in the trade negotiations and considering an extensive set of trade agreements - moving our model closer to reality.

Unfortunately, the answer to the question whether PTAs are ‘building blocks’ or ‘stumbling blocks’ on the path towards global free trade is not as straightforward as one would like it to be. In the end, the results presented here are mixed and depend on the size distribution of the countries. Under symmetry, GFT is the unique stable trade constellation in both regulatory scenarios. But as soon as one moves away from symmetry, GFT might not be reached at all. In between, the effect of switching off Article XXIV Paragraph 5 depends on the exact asymmetry. In case two countries are relatively smaller, prohibiting PTAs increases the area of stability of the GFT regimes. When two countries are relatively larger, it reduces the area. Once the world is further away from symmetry, abolishing the exception for PTAs might result in the worst possible state from the perspective of overall world welfare, the non-cooperative MFN regime. Therefore, under such circumstances, PTAs act as a mechanism that prevents the MFN regime.

Our research also raises a couple of questions in need of further investigation. First, it would be interesting to study the robustness of the findings with respect to the underlying trade model. While the model of competition via exports remains popular in the related literature, economists also extensively use both oligopoly and competition via imports model. Fortunately, the framework presented here does allow for a different underlying trade model such as the ones mentioned above. Another potential area of inquiry might be an extension of the framework to increase the number of countries. Nowadays, in addition to bilateral negotiations, so-called plurilateral negotiations play an important role in the development of preferential trade liberalisation. Recent examples are the Trans-Pacific Partnership (TTP) and the Regional Comprehensive Economic Partnership (RCEP). Including more than three countries in a model would allow us to investigate the strategic interactions among countries whilst taking these negotiations into account. The introduction of political economy considerations to the underlying trade model is another area of interest[33], as it might allow us to understand the nature of tariff peaks occurring after PTAs come into effect. It is our opinion that modifications or extensions of our framework (as mentioned here) are directions worthy of further research.

As a final remark, it is perhaps important, going forward, to move the debate of ‘building blocks’ vs. ‘stumbling blocks’ to a level of detail that goes beyond this binary choice.

[33]See for example Facchini et al. (2013).

CHAPTER 4

# The Impartial Observer under Uncertainty

## 4.1. Introduction

Individuals, as members of society, continually face choices among moral rules, institutional arrangements, government policies or patterns of wealth distribution; therefore, they are repeatedly involved in value judgements about which social alternative to choose. The history of economists (and philosophers) arguing that individuals should make such decisions under sympathetic interest in the welfare of each member of society without any bias towards particular participants dates back as far as Adam Smith's 'Theory of Moral Sentiments' (1759).

Harsanyi (1953, 1955, 1977) developed a rational theory of societal judgements. According to this theory, such choices should be made based on individual's 'social' or 'moral' preferences which are derived from the concept of an 'impartial observer'. As such, you imagine a situation where you do not know your actual place in society when comparing different social arrangements. Instead, you judge the desirability of the alternatives under the personal preferences of all members of society.[1] Thus, the original premise remains that this type of theory, unlike the theory of individual rational behaviour or game theory, should be independent of selfish considerations.

The main result of this theory, now known as Harsanyi's Impartial Observer Theorem, combines Adam Smith's ideas of a sympathetic and impartial spectator with Kant's universality criterion and the utilitarian tradition of social utility maximisation using von Neumann-Morgenstern expected utility theory. In the end, Harsanyi argued that an individual facing risky prospects over social outcomes and a hypothetical lottery over identities in society should rank these according to the weighted aggregate of individuals' expected utilities.

However, Harsanyi's Impartial Observer Theorem with its implicit utilitarianism only considers scenarios where each of the involved individuals faces objective risk. It is a theory analysing societal judgements when objective probabilities over a set of social outcomes are known by each member of society. Our goal in chapter 4 is to extend Harsanyi's Impartial Observer Theorem to include Knightian Uncertainty in the model. By introducing individual belief systems about the likelihood of the social outcomes (which the impartial observer necessarily takes into consideration),

[1]The imaginary construct of impartiality is similar to John Rawls' idea of a 'veil of ignorance' in 'A Theory of Justice' (1971), as these two metaphors are attempts at capturing the same stance.

our approach allows the application of the original framework to a new area of social value judgements. In particular, it allows the analysis of scenarios where individuals agree on the ranking but not on the likelihood of social outcomes. The impartial observer in our model does not aggregate these belief systems separately though. When the impartial observer imagines herself being a particular individual, she adopts not only that individual's preferences but the belief system as well.[2]

The main result of our work is a generalised utilitarian representation of the preferences of the impartial observer under uncertainty. It is a weighted sum of Second-Order Subjective Expected Utility (SOSEU) functions, each representing the preferences of an individual. Our framework is based on the generalised version of Harsanyi (1953) by Grant et al. (2010), which accommodates common criticism of Harsanyi's approach, specifically the issue of fairness and attitude towards mixing. The introduction of individual belief systems to our framework follows Seo (2009) and as a result the SOSEU functions supersede the Expected Utility (EU) functions of both Harsanyi's and Grant et al.'s approach.

In addition to the framework, chapter 4 includes two illustrative examples. First, the moral dilemma of the 'Afghan Goatherds' (see Sandel (2010)) showcases a scenario of agreement on the ranking but disagreement on the likelihood of social outcomes. Second, an example of a simple exchange economy with endowments and different alternatives of wealth (re-)distribution demonstrates the frameworks suitability for traditional economic problems - it serves as a proof of concept in that regard.

In order to motivate the introduction of uncertainty to moral value judgements, let us preview the story of the Afghan Goatherds. In 2005 a team of four soldiers, all U.S. Navy SEALs, set out to find a Taliban leader in the Afghan mountains near the Pakistan border. Just as the team set up their base overlooking the area to fulfil their reconnaissance mission, two Afghan goatherds stumbled upon them - a young boy with them. Due to the nature of their mission and other circumstances, the team considered killing or releasing the civilians the only two viable options. Eventually, the team cast a vote, where one soldier abstained, two voted two ways and the commander of the unit made the decisive call to release them. The civilians later informed the Taliban in a nearby village about the presence of the soldiers. In the subsequent ambush three of the four soldiers died, leaving the commander as the lone survivor.

In retrospect it is easy to make the correct call for this specific scenario. However, imagine you wanted to create a guideline for commanders about how to make such moral value judgements when in the field. In that case, you would naturally assume

[2]In our opinion, this is a natural extension of Harsanyi's concept of impartiality. It differs therefore from the type of group decisions that are presented in Raiffa (1970), which features and discusses aggregation of belief systems.

the role of an impartial observer and evaluate the situation based on individual preferences. In order to truly assume a person's view however, it is necessary to also take on that persons belief system - which is not possible (or included) in the traditional setting. Our approach therefore serves as an extension of the framework to include such cases where belief systems play an important role. This work also contains a formal presentation of this specific moral dilemma and thereby connects the model with reality.

Section 2 discusses the related literature and focuses on the two relevant streams of literature, that is social choice theory and decision theory under uncertainty. Section 3 presents a minimal version of a framework based on Grant et al. (2010) and then introduces uncertainty following Seo (2009). Section 4 provides an analysis of the model, including a comparison with Grant et al. (2010). Section 5 contains the aforementioned illustrative examples as it revisits the Afghan Goatherds and presents the economic example. Section 6 summaries and concludes chapter 4. The appendix consists of one particularly extensive proof and additional details for the example of the Afghan Goatherds.

## 4.2. Related Literature

The important philosophical tradition of impartiality for moral value judgements about collective life has a long history. Vickrey (1945) and Harsanyi (1953) both independently introduced the idea to the economic literature and as Mongin (2001) formulates it: 'All in all, Harsanyi, if perhaps not Vickrey, should count as a major representative of the ethics of impartiality among 20th century writers.'

The (related) literature on Harsanyi's Impartial Observer Theorem is substantial. It is not our aim to review this vast literature as a whole, but instead to concentrate on the building blocks of our work and other closely related research.

As mentioned in the introduction, Grant et al. (2010) and Seo (2009) are the inspiration for the building blocks of our framework. In the first one, the authors revisit Harsanyi's Impartial Observer Theorem; they consider two major criticisms, concerning fairness and different risk attitudes, and derive a generalised version of the theorem that accommodates these criticisms. Furthermore, in the special case of an impartial observer that is indifferent between identity and outcome lotteries ('accidents of birth' and 'life chances') the generalised version of the theorem boils down to the standard Harsanyi doctrine. In consequence, the setting of the paper actually yields a new axiomatisation of Harsanyi's utilitarianism. The resulting generalised utilitarianism serves as inspiration for the foundation of our approach. It gives us the possibility to extend the original framework of Harsanyi from risk to uncertainty while also accommodating common criticism of it.

In the second paper mentioned above, Seo formulates a model for decision making under uncertainty using second-order beliefs, i.e. beliefs over probability measures. Existing models in this stream of literature essentially differ in the choice of the domain of preference. Seo takes the domain of Anscombe and Aumann (1963) and a similar axiomatic foundation. Klibanoff, Marinacci and Mukerji (2005) by contrast require an additonal (sub-)domain with preferences. The domain selection of Seo allows us to introduce uncertainty to the (generalised) framework without any additional modifications. The Second-Order Subjective Expected Utility (SOSEU) representation of the preferences by Seo therefore translates to a corresponding version in the context of an impartial observer.

A variety of papers already deals with Harsanyi's Impartial Observer Theorem under uncertainty. Gajdos and Kandil (2008) provide an extension of the framework where the impartial observer considers sets of identity lotteries. In their model, unlike ours, uncertainty is introduced on a societal (not individual) level. The impartial observer's preference (under additional assumptions) is then characterised by a convex combination of Harsanyi's utilitarian and Rawls' egalitarian criteria.

A work closer to ours is the one by Nascimento (2012) which presents a model of aggregating preference orderings under subjective uncertainty. A fundamental difference is the setting of each paper. Namely, the one of Nascimento is that of a group of individuals that necessarily agrees on the ranking of certain risky objects. In contrast, our setting is one where a group of individuals does not agree on the ranking of any objects (risky or ambiguous). The assumption of Nascimento fits a group of experts or specialists in a field where there is a certain consent. However, in our opinion this assumption is too restrictive for other cases, like the economic example in this chapter (see Section 4.5 for a formal discussion). Nevertheless, the results are actually closely related, compare specifically Theorem 1 by Nascimento and Theorem 4.1 in this work. In a sense, our work arrives at a similar representation but with a different axiomatic foundation and also with applications in mind that are explicitly excluded otherwise. Furthermore, as the analysis of the example of the Afghan Goatherds shows, our framework is actually also able to include scenarios with consent (see Section 4.5 and Appendix C.2).

It is worth mentioning (and repeating) that the impartial observer in our model always takes on the individual beliefs as part of the preferences thereby avoiding any aggregation of belief systems. In consequence, our model stays true to Harsanyi's thought experiment and also avoids the impossibility result of Mongin (1995).[3]

[3] A number of papers deals with decision making of societies using a mechanism of aggregating different individual beliefs, for example Cres, Gilboa and Vielle (2011), Alon and Gayer (2016), Danan, Gajdos, Hill and Talon (2016), and Qu (2017).

## 4.3. Model

Let $(X, \tau)$ be a topological space. Then, denote by $\mathcal{B}_X$ its Borel $\sigma$-algebra and by $\Delta(X)$ the set of all probability measures on $(X, \mathcal{B}_X)$. By $x$ for $x \in X$ refer both to the actual element in $X$ and to the induced one in $\Delta(X)$ - depending on context. Endow $\Delta(X)$ with the weak convergence topology. Also, endow any product of topological spaces with the product topology.

**4.3.1. General Setting.** Let $\mathcal{I} = \{i_1, \ldots, i_I\}$, $I \geq 2$, be a finite set of individuals facing a societal decision problem in the presence of individual uncertainty. Each social choice is modelled as a three-layered object (of different types of risk).[4] First, the (final) outcome space is given by $\mathcal{X}$ - a compact metrisable space with $|\mathcal{X}| \geq 2$. The outcome lotteries $p \in \Delta(\mathcal{X})$, also called one-stage lotteries, are the first layer (featuring objective risk). Further, let $\mathcal{S} = \{s_1, \ldots, s_S\}$, be the finite set of states of the world, which introduces uncertainty via individual beliefs about its probability distribution. The functions $h\colon \mathcal{S} \to \Delta(\mathcal{X})$, also called acts, are the second layer (featuring subjective risk or simply ambiguity). Denote by $\mathcal{H}$ the set of all acts. The act lotteries $P \in \Delta(\mathcal{H})$, also called two-stage lotteries, are the third layer (featuring objective risk again).

The individuals in this situation imagine themselves as an impartial observer, i.e. treating their (social) identity as an unknown component in the decision problem. As such, they face both identity lotteries $z \in \Delta(\mathcal{I})$ and act lotteries $P \in \Delta(\mathcal{H})$. Thus, the individual preferences $\succeq_i$, $i \in \mathcal{I}$, are each defined on $\Delta(\mathcal{H})$ while that of the impartial observer $\succeq$ is defined on $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$.[5] For all of these preferences, we assume a couple of 'standard' properties:

ASSUMPTION 4.1 (Individual). *For each $i$ in $\mathcal{I}$ the preference $\succeq_i$ on $\Delta(\mathcal{H})$ is complete, transitive and continuous. Its asymmetric part $\succ_i$ is non-empty.*

ASSUMPTION 4.2 (Impartial Observer). *The preference $\succeq$ on $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ is complete, transitive and continuous. Its asymmetric part $\succ$ is non-empty*

Note, that by continuous we mean that the weak upper and lower contour sets are closed with respect to the corresponding topologies. In the case of the individual this means with respect to the weak convergence topology and in the case of the impartial observer the product topology of the weak convergence topologies.

[4]The introduction of uncertainty via these three-layered objects follows Seo (2009) and by extension Anscombe and Aumann (1963).

[5]The impartiality that is presented here is based on the framework of Grant et al. (2010) which generalised the concept of Harsanyi (1953). In Grant et al. (2010) the impartial observer's preferences are defined on $\Delta(\mathcal{I}) \times \Delta(\mathcal{X})$, which naturally extends to $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ - incorporating the framework of Seo (2009). By contrast, the corresponding set in Harsanyi (1953) is $\Delta(\mathcal{I} \times \mathcal{X})$. See Grant et al. (2010) for a detailed discussion on this difference.

Axiom 4.1 (Acceptance Principle). *For all $i$ in $\mathcal{I}$ and all $P, Q$ in $\Delta(\mathcal{H})$:*

$$P \succeq_i Q \Leftrightarrow (i, P) \succeq (i, Q)$$

The acceptance principle establishes the intuitive link between the preferences of the individuals and that of the impartial observer. The intuition is that when the impartial observer imagines herself to be a particular individual that she in fact takes on the preferences of that individual (including the belief system).

Axiom 4.2 (Independence over Identity Lotteries). *Suppose elements $(z, P), (z', Q)$ in $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ are such that $(z, P) \sim (z', Q)$. Then, for all $\tilde{z}, \tilde{z}'$ in $\Delta(\mathcal{I})$ and all $\alpha$ in $(0, 1]$:*

$$(\tilde{z}, P) \succeq (\tilde{z}', Q) \Leftrightarrow (\alpha\tilde{z} + (1-\alpha)z, P) \succeq (\alpha\tilde{z}' + (1-\alpha)z', Q)$$

The independence over identity lotteries as well as the acceptance principle are each concerned with the nature of the impartial observer's preferences with respect to identities. As our approach considers uncertainty on the level of outcomes and not identities, these two axioms naturally carry over from the traditional setting.

Assumption 4.3 (Absence of Unanimity). *For all $P, Q$ in $\Delta(\mathcal{H})$:*

$$\exists i \in \mathcal{I} : P \succ_i Q \Rightarrow \exists j \in \mathcal{I} : Q \succ_j P$$

The absence of unanimity can be interpreted as a required heterogeneity on the social alternatives and (preferences of) individuals. It is also not a new addition, but controversial enough to require an additional comment. First of all, normative decision-making is clearly trivial when all individuals agree on all rankings. Thus, it is possible to exclude this extreme case without losing any explanatory power. However, in our opinion it is too restrictive to completely leave out the opposite where everyone disagrees about everything - like it is done in Nascimento (2012) with the requirement of agreement on risky prospects. In general, our aim is to focus on scenarios that exhibit substantial heterogeneity in terms of (dis-)agreement.[6]

Next, let us state a lemma (which is going to be useful later on) about the representation of the preferences of the impartial observer and the individuals. Now, the structure of the results and the proof itself (see Appendix C.1) follow the ideas of Grant et al. (2010):

Lemma 4.1. *Suppose absence of unanimity applies. Then, the impartial observer satisfies the acceptance principle and independence over identity lotteries if and only if there exists a continuous function $V\colon \Delta(\mathcal{I}) \times \Delta(\mathcal{H}) \to \mathbb{R}$ that represents $\succeq$ and for*

[6]It might seem that absence of unanimity is too restrictive as well (just in the other direction). Yet, adding a dummy individual that provides the (technically) required heterogeneity allows us to relax the restriction while still staying in our framework. See Section 4.5 and Appendix C.2 for the formal presentation of the story of the Afghan Goatherds with a demonstration of a dummy.

*each individual $i$ in $\mathcal{I}$ a function $V_i\colon \Delta(\mathcal{H}) \to \mathbb{R}$ that represents $\succeq_i$ such that for all $(z, P)$ in $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$:*

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i V_i(P)$$

*Moreover, the functions $V$ and $V_i$, $i \in \mathcal{I}$, are unique up to common positive affine transformation.*

Note that the main arguments of the proof actually work for a general compact set $\mathcal{H}$ as well, i.e. the proof requires no specific structure of $\mathcal{H}$. Thus, a modified Lemma 4.1 could potentially serve as a foundation for conceptually similar approaches to ours that only differ in terms of additional structure, specifically with respect to individual utilities.

Now, up to this point all assumptions and axioms follow Grant et al. (2010). Specifically, their axiom of independence over outcome lotteries (for individuals) is the only missing axiom. However, in the next part the axioms follow Seo (2009) instead and introduce uncertainty to the framework.

**4.3.2. Introducing Uncertainty.** In order to formulate the remaining axioms and introduce uncertainty to the model, it is necessary to define additional objects. Namely, let us define what we mean when talking about mixing two acts or lotteries (of acts). In the end, the two different kind of mixtures depend on the timing of the resolution of uncertainty (or of the mixing - depending on how you look at it).

First, consider the case where, when combining two (pure) acts, the uncertainty is resolved first and then the mixing takes place:

Definition 4.1. *For $f, g$ in $\mathcal{H}$ and $\alpha$ in $[0, 1]$ and for $s \in S$, $B \in \mathcal{B}_{\mathcal{X}}$ we set*

$$(\alpha f \oplus (1-\alpha) g)(s)(B) = \alpha f(s)(B) + (1-\alpha) g(s)(B).$$

*This operation is called a second-stage mixture.*

Now, with this in mind, we introduce a 'standard' independence axiom with respect to second-stage mixtures:

Axiom 4.3 (Second-Stage Independence). *For all $i$ in $\mathcal{I}$, all $\alpha$ in $(0, 1]$ and lotteries $p, q, r$ in $\Delta(\mathcal{X})$:*

$$\alpha p \oplus (1-\alpha) r \succeq_i \alpha q \oplus (1-\alpha) r \Leftrightarrow p \succeq_i q$$

Second, consider the case where, when combining two lotteries of acts, the mixing takes place first and then the uncertainty is resolved:

Definition 4.2. *For $P, Q$ in $\Delta(\mathcal{H})$ and $\alpha$ in $[0, 1]$ and for $B \in \mathcal{B}_{\mathcal{H}}$ we set*

$$(\alpha P + (1-\alpha) Q)(B) = \alpha P(B) + (1-\alpha) Q(B).$$

*This operation is called a first-stage mixture.*

Again, with this in mind, we introduce a 'standard' independence axiom with respect to first-stage mixtures:

AXIOM 4.4 (First-Stage Independence). *For all $i$ in $\mathcal{I}$, all $\alpha$ in $(0,1]$ and lotteries $P, Q, R$ in $\Delta(\mathcal{H})$:*

$$\alpha P + (1-\alpha)R \succeq_i \alpha Q + (1-\alpha)R \Leftrightarrow P \succeq_i Q$$

Finally, it is necessary to introduce an additional technical object that essentially serves as a tool for scenario analysis (for the individuals):

DEFINITION 4.3. *Each $f \in \mathcal{H}$ and $\mu \in \Delta(\mathcal{S})$ induce a one-stage lottery*

$$\Psi(f,\mu) := \bigoplus_{s\in\mathcal{S}} \mu(s) f(s),$$

*each $P \in \Delta(\mathcal{H})$ and $\mu \in \Delta(\mathcal{S})$ induce a two-stage lottery*

$$\Psi(P,\mu)(B) := P(\{f \in \mathcal{H} \colon \Psi(f,\mu) \in B\})$$

*for $B \in \mathcal{B}_{\mathcal{H}}$.*

In other words, the element $\Psi(P,\mu)$ is the (induced) lottery that corresponds to the lottery $P$ in the scenario where $\mu$ is the probability distribution over the states.

AXIOM 4.5 (Dominance). *For all $i$ in $\mathcal{I}$, all $P, Q$ in $\Delta(\mathcal{H})$:*

$$\Psi(P,\mu) \succeq_i \Psi(Q,\mu)\ \forall \mu \in \Delta(\mathcal{S}) \Rightarrow P \succeq_i Q$$

Imagine an individual only knows there exists a 'true' probability distribution but does not know which one it is. Then, the axiom of dominance captures the intuition that if the individual prefers one induced lottery over another one - substituting all possible probability distributions as the true one, then this individual should prefer that one lottery over the other (and vice-versa).

Finally, using the additional axioms with Lemma 4.1 it is possible to formulate the main result of our work:

THEOREM 4.1. *Suppose absence of unanimity applies. Then, the impartial observer satisfies the acceptance principle as well as independence over identity lotteries, and each individual satisfies first-stage and second-stage independence, and dominance if and only if the impartial observer's preference admits a representation $\langle\{U_i, \phi_i\}_{i\in\mathcal{I}}\rangle$ of the form*

$$V(z,P) = \sum_{i\in\mathcal{I}} z_i \phi_i(U_i(P))$$

*where each*

- $\phi_i \colon \mathbb{R} \to \mathbb{R}$ *is an increasing continuous function*
- $U_i \colon \Delta(\mathcal{H}) \to \mathbb{R}$ *is a SOSEU representation of* $\succeq_i$

*i.e. the impartial observer is a generalised (weighted) utilitarian under uncertainty.*

*In addition, the* $U_i$ *are unique up to uniqueness of the SOSEU representation. Further, the functions* $V$ *and* $\phi_i \circ U_i$ *are unique up to a common positive affine transformation.*

Proof. Let absence of unanimity apply.

Part 1 ('⇐'):

First, the representation of the impartial observer is affine in identity lotteries and therefore satisfies the acceptance principle and independence over identity lotteries. Note that alternatively this specific step also follows by application of Lemma 4.1. Second, the representation of each individual is of SOSEU form and thus satisfies its axioms, that is first-stage and second-stage independence, and dominance, using Theorem 4.2 of Seo (2009).

Part 2 ('⇒'):

In this part, the result of Lemma 4.1 is required. Namely, as the impartial observer satisfies the acceptance principle and independence over identity lotteries, it gives us a continuous function $V \colon \Delta(\mathcal{I}) \times \Delta(\mathcal{H}) \to \mathbb{R}$ representing $\succeq$ and for each $i \in \mathcal{I}$ functions $V_i \colon \Delta(\mathcal{H}) \to \mathbb{R}$ representing $\succeq_i$ such that for all $(z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$:

$$V(z, P) = \sum_{i \in \mathcal{I}} z_i V_i(P)$$

The preferences of each individual satisfy first-stage and second-stage independence, and dominance, and so by Theorem 4.2 of Seo (2009) each $V_i$ is a SOSEU function. However, this only holds up to transformation via an increasing continuous function. Thus, for each $i \in \mathcal{I}$ it follows that $V_i = \phi_i \circ U_i$ where $U_i$ is a SOSEU function and $\phi_i$ is a transformation.

The uniqueness of the $U_i$ follows directly from Lemma C.1 in Seo (2009), while the uniqueness of the $V$ and $\phi_i \circ U_i$ follows from Lemma 4.1. □

In the theorem and its proof the actual SOSEU representation remained hidden. The following remark (re-)states the formal definition of a SOSEU representation and its uniqueness properties (see Seo (2009)). It will be helpful in the analysis and for the applications later on.

Remark 4.1. *A SOSEU representation is generally characterised by a triple* $(u, v, m)$ *and of the form*

$$U(P) = \int_{\mathcal{H}} \int_{\Delta(\mathcal{S})} v\left(\int_{\mathcal{S}} u(f) d\mu\right) dm(\mu) dP(f)$$

*for* $P \in \Delta(\mathcal{H})$*, where*

- *$u$ is a bounded continuous mixture linear function, $u\colon \Delta(\mathcal{X}) \to \mathbb{R}$,*
- *$v$ is a bounded continuous strictly increasing function, $v\colon u(\Delta(\mathcal{X})) \to \mathbb{R}$,*
- *$m$ is a probability measure, $m \in \Delta(\Delta(\mathcal{S}))$.*

*The uniqueness of the SOSEU representation implies that for any another triple $(u', v', m')$*

i) *$u$ and $u'$ as well as $v \circ u$ and $v' \circ u'$ are each identical up to positive affine transformation,*

ii) *$\int_{\Delta(\mathcal{S})} \varphi dm = \int_{\Delta(\mathcal{S})} \varphi dm'$ for all continuous functions $\varphi$ on $\Delta(\mathcal{S})$ for which there exists a Borel signed measure $\lambda$ on $T := [u(\Delta(\mathcal{X}))]^S$ with bounded variation such that for all $\mu \in \Delta(\mathcal{S})$ it holds that $\varphi(\mu) = \int_T v(\mu \cdot t) d\lambda(t)$.*

## 4.4. Analysis

Naturally, the starting point of our analysis is the connection between our result and that of Grant et al. (2010). As expected, completely eliminating uncertainty by reducing it to risk produces their result as a special case of ours (see 4.4.1).

Furthermore, one (significant) indeterminacy in Theorem 4.1 is the specific form of the $\phi_i$, $i \in \mathcal{I}$. In the current setting, the only restriction is that each of them is an increasing continuous function. Let us therefore analyse the specific form of these functions in relation with different (additional) properties of the preferences. In particular, let us compare this to the findings presented in Grant et al. (2010). Fortunately, their results and proofs actually do not rely on the underlying structure of the outcome space and therefore all of their findings translate into our setting without any modifications.

**4.4.1. The Special Case of Risk.** In the special case of (only) risk, i.e. $\mathcal{S} = \{s\}$, all belief systems are trivial. In consequence, the three-layer objects containing uncertainty in the middle reduce to two-layer objects with only risk present. Further, in order to identify this with the setting of Grant et al. (2010) assume that each $v_i$, $i \in \mathcal{I}$, is actually a linear function (corresponding to indifference to uncertainty) or equivalently assume reversal of order or reduction of compound lotteries.[7] Thus, the two-layer objects with risk collapse to single-layer objects with risk.

In this (special case of a) setting, first/second-stage independence simply reduces to independence over outcome lotteries, dominance is now an empty statement, and

[7]Take $i \in \mathcal{I}$. Then, reversal of order and reduction of compound lotteries each describe a property of the preferences of individual $i$ with respect to first-stage and second-stage mixtures.
Namely, reversal of order is satisfied if for all $f, g \in \mathcal{H}$ and $\alpha \in [0,1]$

$$\alpha f \oplus (1-\alpha) g \sim_i \alpha f + (1-\alpha) g.$$

Similarly, reduction of compound lotteries is satisfied if for all $p, q \in \Delta(\mathcal{X})$ and $\alpha \in [0,1]$

$$\alpha p \oplus (1-\alpha) q \sim_i \alpha p + (1-\alpha) q.$$

The axiom of dominance implies the equivalence of these two properties (see Seo (2009)).

the representation of the impartial observer takes on the form of the generalised (weighted) utilitarian of Grant et al. (2010).

**4.4.2. Fairness.** One common criticism of Harsanyi's utilitarianism is concerned with fairness (see for example Diamond (1967)). It is one of the two issues that Grant et al. (2010) solved by generalising the theory. The notion of fairness in this context refers to a preference of the impartial observer for mixing act lotteries over mixing identity lotteries.[8]

Consider from now on those tuples of identity and act lotteries between which the impartial observer is indifferent, that is $(z, P')$ and $(z', P)$ with $(z, P') \sim (z', P)$. Following the arguments in favour of fairness, the impartial observer always prefers mixing these pairs on the level of acts over mixing on the level of identities, i.e. $(z, \alpha P + (1-\alpha)P') \succeq (\alpha z + (1-\alpha)z', P)$ for all $\alpha \in (0,1)$, which is also referred to as a preference for life chances (compared to accidents of birth). In the case of risk, Grant et al. (2010) show that this holds if and only if each $\phi_i$, $i \in \mathcal{I}$, is concave, which actually translates one-to-one into our setting.

Conversely, consider a scenario where the impartial observer is indifferent between life chances and accidents of birth, that is $(z, \alpha P + (1-\alpha)P') \sim (\alpha z + (1-\alpha)z', P)$ for all $\alpha \in (0,1)$. In the case of risk, Grant et al. (2010) show that this holds if and only if each $\phi_i$, $i \in \mathcal{I}$, is affine, which again translates one-to-one into our setting.[9]

**4.4.3. Mixtures.** In the previous part, the focus was on the relation between mixing either identity or act lotteries. Another criticism of Harsanyi's utilitarianism is concerned with different attitudes among the individuals towards mixing.[10]

Fix two individuals, say $i$ and $j$, and consider from now on those act lotteries which the impartial observer ranks equally from each perspective, that is $P, \tilde{P}, Q, \tilde{Q}$ with $(i, P) \sim (j, Q)$ and $(i, \tilde{P}) \sim (j, \tilde{Q})$. Imagine that the impartial observer prefers facing the (first-stage) mixtures of each of those two pairings of act lotteries as $i$ rather than as $j$, i.e. $(i, \alpha P + (1-\alpha)\tilde{P}) \succeq (j, \alpha Q + (1-\alpha)\tilde{Q})$ for all $\alpha \in (0,1)$. Now, Grant et al. (2010) show that this holds if and only if the composite function $\phi_i^{-1} \circ \phi_j$ is convex on the domain $\mathcal{U}_{ji} := \{u \in \mathbb{R} \mid \exists P, Q \in \Delta(\mathcal{H})\colon (i, P) \sim (j, Q) \wedge U_j(Q) = u\}$ for the case of risk but it actually also applies to our setting under uncertainty.

Alternatively, imagine that the impartial observer is indifferent when comparing to face these mixtures as different individuals: $(i, \alpha\tilde{P} + (1-\alpha)P) \sim (j, \alpha\tilde{Q} + (1-\alpha)Q)$ for all $\alpha \in (0,1)$ and all $i, j \in \mathcal{I}$. Following Grant et al. (2010) this holds if and only if $\phi_i = \phi$, $i \in \mathcal{I}$, both for risk and under uncertainty. Further, let $i_1$ and $i_2$ be

[8]The paper of Grant et al. (2010) also provides an example justifying this notion of fairness.

[9]Note that in this case the resulting representation is actually equivalent to a framework with a single probability distribution over states for every individual and simultaneously a modified probability distribution over individuals that includes belief systems.

[10]In Grant et al. (2010) this issue is actually referred to as 'different attitude towards risk'.

a pair of individuals such that there exists a sequence of individuals $j_1, \dots, j_N$ with $j_1 = i_1$ and $j_N = i_2$ where each $\mathcal{U}_{j_n, j_{n-1}}$ has non-empty interior. Then, the functions $U_i$, $i \in \mathcal{I}$, are unique up to a common positive affine transformation.

## 4.5. Applications

In the following, two applications (or examples) for our approach are presented. The first example picks up the story of the Afghan Goatherds from the introduction; the second one is a simple economic example.

As mentioned in the introduction, the moral dilemma of the Afghan Goatherds features a scenario where individuals agree on the ranking of social outcomes, but disagree on the likelihood of these. It showcases the effect of the introduction of uncertainty, as its results are purely driven by the nature of different belief systems.

The economic example on the other hand essentially serves as a proof of concept. It illustrates that our framework is able to accommodate not only pure philosophical but also economic problems. As a bonus, the example allows us to demonstrate the effect of the degree of fairness on the level of the impartial observer.

**4.5.1. Afghan Goatherds.** First, recall the moral dilemma of the Afghan Goatherds described in the introduction. As mentioned there, our claim is actually not that the commander of the unit necessarily made the decision using anything related to our approach (even though it could very well be the case). However, our approach allows a normative analysis of this (and similar) situations. You could for example be developing a moral guideline for the military in the vein of the United States Army Field Manuals. Naturally, you would then be in the position of a neutral or impartial observer and evaluate the situation using the point of view of each involved individual including their perception of the situation's uncertainty.

Let us define the formal structure of the decision problem now. It is deliberately kept simplistic compared to the actual events. Hence, it might not be a perfect fit for every aspect of the original story, but should still serve as a proper demonstration of an application of our theory.

First, let $\mathcal{X} = \{0,1\}^2 \backslash \{(0,0)\}$ be the set of outcomes. Each element $x = (x_1, x_2)$ corresponds to the survival of the soldiers, $x_1$, and that of the afghan goatherds, $x_2$. Each entry then indicates either 'alive', 1, or 'dead', 0. Furthermore, let $\mathcal{S} = \{t, u\}$ be the states of the world, where $t$ and $u$ correspond to talking to the Taliban about the soldiers and keeping quiet about them, respectively. Finally, the two available moral choices are killing or sparing the civilians, denoted by $K$ and $L$, respectively.

Thus, using the previous notation, they are given by:

$$K = \begin{cases} (1,0) & s = t, u \end{cases}$$

$$L = \begin{cases} (0,1) & s = t \\ (1,1) & s = u \end{cases}$$

Note that these elements only contain subjective risk (with respect to the states). Together with the remaining specifications, this is actually going to guarantee that the example is purely driven by individual belief systems.

In addition, let $I = \{1, \ldots, 3\}$ be the set of individuals - corresponding to the team of soldiers.[11] Following the traditional setting, the identity lottery that is part of the choice problem is going to be fixed to $(1/3, 1/3, 1/3)$. Note that in this setting, i.e. with this set of individuals and this (fixed) identity lottery, the specifications of the next part actually conflict with our initial assumption of absence of unanimity. Appendix C.2 analyses and corrects this problem by introduction of a dummy variable without any changes to the preferences. In the rest of this analysis therefore consider the assumption of absence of unanimity to be fulfilled.

4.5.1.1. *Specifying the Moral Dilemma.* In order for us to conduct a detailed analysis, we need to specify more about the preferences of the individuals and in particular about the the nature of their belief systems. First of all, to ensure that our results are purely driven by the beliefs of the soldiers, we assume that their preferences are otherwise completely identical, that is $u_i = u$ and $v_i = v$ for each $i \in \mathcal{I}$. Similarly, assume that the impartial observer treats the soldiers identical with respect to facing similar mixtures. Consequently, $\phi_i = \phi$, $i \in \mathcal{I}$, by Section 4.4.3.

Furthermore, let us specify the ranking of all of the possible survival outcomes. Naturally, you would assume that the soldiers rank the survival of all the highest. Additionally, our assumption is going to be that the soldiers, when confronted with the exclusive survival outcomes, prefer their own survival over that of the civilians. This essentially captures the idea of universal self-preservation instincts. Therefore, after using a positive affine transformation (to specify the lower and upper bound):

$$0 = u(0,1) < u(1,0) < u(1,1) = 1$$

Moreover, the individuals are assumed to be uncertainty-averse, which then implies a concave function $v$. In particular, it is going to be of the form $z^q$ for $q \in (0,1)$. Assume further a preference for life chances (of the impartial observer). Thus, Section 4.4.2 yields that $\phi$ is also a concave function. As before, take $z^r$ for $r \in (0,1)$.

[11]In the original story, the team consists of four soldiers including the commander of the unit. Each of the three regular soldiers exhibited different individual preferences, i.e. $K \succ_1 L$, $L \succ_2 K$ and $K \sim_3 L$, which is already enough to construct a simple (yet interesting) example. Therefore, excluding the commander as an individual is of no (significant) consequence to our analysis.

In general, the belief systems $m_i$ are probability distributions over $\Delta(\mathcal{S})$, which in this specific scenario is equivalent to probability distributions over $[0,1]$ by simply identifying $\mu \in \Delta(\mathcal{S})$ with $p \in [0,1]$ via $p = \mu(u)$ (alternatively $p = \mu(t)$). Now, the actual belief systems are going to be truncated normal distributions on $[0,1]$.[12] Let $(\mu_i, \sigma_i)$ denote a pair of parameters of the initial normal distributions. Then, assume that $\mu_i = \mu = 0.5$ for $i \in \mathcal{I}$ and furthermore $0 < \sigma_1 < \sigma_2 < \sigma_3 < +\infty$. Hence, the individual belief systems are mean-preserving spreads of each other, which allows us to showcase the effect of introducing uncertainty to the framework. Additionally, the centered mean captures the idea of unbiased individuals. Finally, combining everything together yields the following utility of the impartial observer for the two moral choices:

$$\begin{aligned}
V(z,K) &= \sum_{i=1}^{3} \frac{1}{3} \phi(v(u(1,0))) \\
&= (u(1,0)^q)^r \\
V(z,L) &= \sum_{i=1}^{3} \frac{1}{3} \phi\left( \int_0^1 v(pu(1,1) + (1-p)u(0,1)) dm_i(p) \right) \\
&= \sum_{i=1}^{3} \frac{1}{3} \left( \int_0^1 p^q dm(\mu, \sigma_i)(p) \right)^r
\end{aligned}$$

4.5.1.2. *Numerical Analysis.* In addition to calculating the results for specific values, the aim is to demonstrate the effect of uncertainty in our framework. Therefore, consider a modified version of the framework where each individual only uses a single (subjective) probability distribution over states but on a societal level there is an additional probability distribution over these subjective probability distributions.[13] In other words, instead of uncertainty, the model features subjective risk. Denote by $\tilde{V}(z,P)$ the evaluation of $(z,P)$ corresponding to the modified model. Ideally, the comparison between our model and this modified one produces a difference and thus justifies to a certain degree the use of the concept of uncertainty in our model.

Finally, fix $q = 0.75$, $r = 0.25$, $\sigma_1 = 0.01$, $\sigma_2 = 0.1$, $\sigma_3 = 1$, and $u(1,0) = 0.48$. Now, the parameter choices here (and also the previous functional choices) should be understood as part of an example. Our (numerical) analysis uses reasonable but still debatable choices to enable us to showcase interesting phenomena for one of potentially many formal interpretations of the story within our framework. Anyhow, using these parameters produces the following values when rounded to the fourth decimal:

---

[12]Alternatively, take beta distributions $Beta(\alpha, \beta)$ with varying $\alpha = \beta$.

[13]Formally, this corresponds to the use of indicator functions as belief systems, i.e. $m_i = \mathbb{1}_{\mu_i}$ for $\mu_i \in \Delta(\mathcal{S})$ and all $i \in \mathcal{I}$, and an extended set of individuals that includes beliefs, i.e. $\mathcal{I} \times \Delta(\mathcal{S})$ where the density function is given by $f_z(i,\mu) = z_i \cdot m_i(\mu)$ for $(i,\mu) \in \mathcal{I} \times \Delta(\mathcal{S})$, instead of $\mathcal{I}$ and the corresponding $z \in \Delta(\mathcal{I})$.

| | $V_1$ | $V_2$ | $V_3$ | $V$ | $\tilde{V}$ |
|---|---|---|---|---|---|
| $K$ | 0.5767 | 0.5767 | 0.5767 | 0.8714 | 0.8714 |
| $L$ | 0.5946 | 0.5923 | 0.5723 | 0.8751 | 0.8657 |

As a consequence, the values produce the following rankings:

$$V_3(L) < V_{1/2/3}(K) < V_2(L) < V_1(L)$$
$$V(z, K) < V(z, L)$$
$$\tilde{V}(z, L) < \tilde{V}(z, K)$$

Therefore, the higher the (perceived) uncertainty on the level of the individuals the lower the evaluation of $L$ compared to that of $K$, which stays constant. In fact, the dynamic actually produces different individual rankings of the two alternatives. Additionally, the comparison of our model and the modified one actually produces a different ranking on the level of the impartial observer. Thus, the introduction of uncertainty influences (and to a certain extent drives) the final ranking. Further, the impartial observer actually agrees with the result of a simple majority vote in this specific case (which in general is not guaranteed).

**4.5.2. Exchange Economy.** Consider a simple exchange economy with two goods and two individuals, with each of them receiving endowments. In this setting, compare two alternative re-distributions rules, namely the Walrasian auctioneer and the Egalitarian rule.[14] Uncertainty enters the model via a possible bias in the distribution of the endowments. This specific example serves as a proof-of-concept in the sense that it shows an interpretation of a traditional economic problem within our framework. It is based on an example by Eichberger and Pethig (1994).

Let $\mathcal{I} = \{1, 2\}$ be the set of individuals and let $x$ and $y$ denote the two goods. In order to keep it simple, let us assume that the total endowment for each good is set to 3 and restricted to positive integers. Thus, the possible initial endowments are given by the two by two matrix

$$\begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} = \begin{pmatrix} ((1,1),(2,2)) & ((1,2),(2,1)) \\ ((2,1),(1,2)) & ((2,2),(1,1)) \end{pmatrix}$$

where $e_{i,j}$ in row $i$ and column $j$ corresponds to individual 1 receiving $(i, j)$ and individual 2 receiving $(3 - i, 3 - j)$ of the pair of goods $(e_{x,k}, e_{y,k})$, $k = 1, 2$.

In the following, our analysis focuses on two possible re-distributions of these initial endowments, namely the Walrasian auctioneer and the Egalitarian rule. Let the utility function of an individual $i$ for the (re-distributed) goods $x_i$ and $y_i$ be given

[14]In general there is also a combination of the two, namely the Walras rule from Equal Division. However, in this scenario, it coincides with the Egalitarian rule. See Nagahisa and Suh (1995) for a characterisation of the Walras rules.

by the Cobb-Douglas form $(x_i y_i)^{\alpha_i}$, where $\alpha_i \in \mathbb{R}_{>0}$. Then, the individuals evaluate the results of the re-distributions as follows (depending on endowments):

In case of the Egalitarian rule, any endowment vector $((e_{x,1}, e_{y,1}), (e_{x,2}, e_{y,2}))$ yields the consumption bundles $((3/2, 3/2), (3/2, 3/2))$. In consequence, the utility for an individual $i$ is always given by $(9/4)^{\alpha_i}$, independent of initial endowments. Now, in case of the Walrasian auctioneer rule, any fixed endowment vector induces a corresponding unique Walrasian equilibrium. The resulting utility of individual $i$ is then given by $((e_{x,i} + e_{y,i})^2/4)^{\alpha_i}$, where $e_{x,i}$ and $e_{y,i}$ are the initial endowments. Using our matrix notation, the following then characterises the re-distribution rules with respect to the individual utilities for all possible initial endowments:

$$\tilde{E}\colon \begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} \mapsto \begin{pmatrix} ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \\ ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \end{pmatrix}$$

$$\tilde{W}\colon \begin{pmatrix} e_{1,1} & e_{1,2} \\ e_{2,1} & e_{2,2} \end{pmatrix} \mapsto \begin{pmatrix} (1, 4^{\alpha_2}) & ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) \\ ((9/4)^{\alpha_1}, (9/4)^{\alpha_2}) & (4^{\alpha_1}, 1) \end{pmatrix}$$

The notation with the tilde is deliberate in order to distinguish these descriptions from their counterparts that take the uncertainty into consideration.

As mentioned earlier, the uncertainty is about the probability distribution over the initial endowments. Assume that there are two states, i.e. $\mathcal{S} = \{s_1, s_2\}$ where $s_1$ corresponds to a bias towards individual 1 and analogously $s_2$ towards 2. Thus, these are described via the following probability distributions ($\pi_1$ for $s_1$, $\pi_2$ for $s_2$):

$$\pi_1(e_{i,j}) = \begin{cases} \frac{1}{2} & i = j = 2 \\ \frac{1}{4} & i = 1, j = 2 \text{ or } i = 2, j = 1 \\ 0 & i = j = 1 \end{cases}$$

$$\pi_2(e_{i,j}) = \begin{cases} \frac{1}{2} & i = j = 1 \\ \frac{1}{4} & i = 1, j = 2 \text{ or } i = 2, j = 1 \\ 0 & i = j = 2 \end{cases}$$

In a sense, our example exhibits uncertainty about (the state of) the economy instead of a fixed (state of the) economy with inherent uncertainty. Consequently, society chooses the re-distribution rule before knowing the initial distributions.

Finally, to formally state the two rules taking uncertainty into consideration, i.e. to formulate the act lotteries corresponding to them, combine the aforementioned functions and probability distributions as follows:

$$E\colon s_k \mapsto \left(\tilde{E}(e_{i,j}) \text{ with probability } \pi_k(e_{i,j})\right)$$

$$W\colon s_k \mapsto \left(\tilde{W}(e_{i,j}) \text{ with probability } \pi_k(e_{i,j})\right)$$

Furthermore, using the inherent symmetries and other similarities simplifies it to:

$$E\colon s \mapsto \left((9/4)^{\alpha_1}, (9/4)^{\alpha_2}\right)$$

$$W\colon s \mapsto \begin{cases} \left(4^{\alpha_1 \mathbb{1}_{\{s_1\}}(s)}, 4^{\alpha_2 \mathbb{1}_{\{s_2\}}(s)}\right) & \text{with probability } 1/2 \\ \left((9/4)^{\alpha_1}, (9/4)^{\alpha_2}\right) & \text{with probability } 1/2 \end{cases}$$

Following the traditional setting, take again the 'fair' uniform identity lottery, that is $z = (1/2, 1/2)$. As in the previous example assume moreover that the impartial observer treats individuals identical with respect to similar mixtures, thus $\phi_i = \phi$, $i \in \mathcal{I}$. In addition, set $\alpha_i = 1/2$ for both $i$ (assuming similar risk-aversion) and fix $v_i$ to be the identity function for both $i$ (assuming uncertainty-neutrality).[15] Also, take $\phi = z^r$ again but with $r \in (0, +\infty)$ this time.

As individual belief systems $m_i$ take truncated normal distributions on $[0, 1]$ again[16], where $p \in [0, 1]$ corresponds to the probability of state $s_i$ realizing. Further, let $(\mu_i, \sigma_i)$ denote a pair of parameters of the initial normal distributions. Then assume that $\mu_1 = 0.75$, $\mu_2 = 0.25$, and $\sigma_i = \sigma = 0.25$, i.e. each of the individuals suspects a bias towards individual 1 and the same level of volatility.

Note that due to the exclusive nature of the game, essentially a zero-sum game, and the additional assumptions, absence of unanimity requires no modifications. Finally, combine everything together for the following evaluations:

$$\begin{aligned} V(z, E) &= \sum_{i=1}^{2} \frac{1}{2} \phi_i \left( v_i \left( \left( \frac{9}{4} \right)^{\alpha_i} \right) \right) \\ &= \left( \frac{3}{2} \right)^r \\ V(z, W) &= \sum_{i=1}^{2} \frac{1}{2} \phi_i \left( \int_0^1 v_i \left( \frac{1}{2} \left( \frac{9}{4} \right)^{\alpha_i} + \frac{1}{2} \left( 4^{\alpha_i} p + 1 \, (1 - p) \right) \right) dm_i(p) \right) \\ &= \sum_{i=1}^{2} \frac{1}{2} \left( \frac{5}{4} + \frac{1}{2} \int_0^1 p dm(\mu_i, \sigma)(p) \right)^r \end{aligned}$$

Consider different values for $r \in (0, +\infty)$ now, which captures different degrees of fairness on the level of the impartial observer. It results in the following values when rounded to the fourth decimal:

| | $V_1$ | $V_2$ | $V_{\vert r=1.5}$ | $V_{\vert r=1}$ | $V_{\vert r=0.5}$ |
|---|---|---|---|---|---|
| $W$ | 1.5897 | 1.4103 | 1.8396 | 1.5000 | 1.2242 |
| $E$ | 1.5000 | 1.5000 | 1.8371 | 1.5000 | 1.2247 |

[15]It seems completely counter-intuitive to assume uncertainty-neutrality in our framework as the introduction of uncertainty is our main contribution. However, in this example and specifically this (numerical) analysis our focus is on the effect of different transformations $\phi$.

[16]Alternatively, take beta distributions $Beta(\alpha, \beta)$ with varying $\alpha = \beta$.

Evidently, the ranking of the impartial observer depends on the exact value of $r$ and the rankings of the individuals are diametrically opposed:

$$V_2(W) < V_2(E) = V_1(E) < V_1(W)$$
$$V(z,E)_{|r=1.5} < V(z,W)_{|r=1.5}$$
$$V(z,E)_{|r=1} = V(z,W)_{|r=1}$$
$$V(z,E)_{|r=0.5} > V(z,W)_{|r=0.5}$$

In other words, a preference for life chances of the impartial observer actually leads to a preference of the Egalitarian rule over that of the Walrasian auctioneer in a setting with a clear bias towards one specific individual. It essentially provides another example for the discussion on the issue of fairness.

## 4.6. Conclusion

The focus of this chapter is the normative decision-making of an individual when considering all other individuals in society and under the influence of uncertainty. Based on the works of Grant et al. (2010) and Seo (2009) we provide an axiomatic foundation for an extension of Harsanyi's Impartial Observer Theorem that includes Knightian uncertainty while also accommodating certain common criticism of the traditional result. The main result shows that the impartial observer's preferences admit a representation in form of a weighted average of the individual (transformed) second-order subjective expected utilities. This representation allows for a tractable analysis of the normative choice problems under consideration. Furthermore, the framework re-establishes links between additional properties of the preferences of the impartial observer on one hand (e.g. the issues of fairness and mixtures) and the specific form of the individual transformations on the other.

The main appeal of our model is the extension to normative decision-making in situations where a group of individuals faces subjective instead of objective risk. The story of the Afghan Goatherds is an example for such moral value judgements. It is also purely driven by individual belief systems and thus provides justification for the introduction of uncertainty to the framework. Moreover, the example shows that our model in fact extends beyond the limitations of absence of unanimity via the use of a dummy individual. Finally, the economic example is an application of our theory to a scenario that demonstrates the effect of a preference for life chances - compared to accidents of birth - of the impartial observer. It also serves as a proof-of-concept for the application of our theory to economic problems in general.

## APPENDIX A

# Appendix to Chapter 2

### A.1. Derivations

**A.1.1. Election winning probability.** In equation (2.11) we denoted the right-hand side of the inequality by $\underline{\delta}$.

$$\underline{\delta} = \underline{\Phi} \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}(v_s^m(p^w, E_m) - v_s^m(p, E_m^w)) - \frac{1}{2}\underline{\Phi}$$

Recall that $\widehat{\delta}$ is distributed uniformly ($\widehat{\delta} \sim U[-\Psi, \Psi]$). Then using equation (2.12)

$$\Pi[\ \widehat{\delta} > \underline{\delta}\ ] = (\Psi - \underline{\delta})\frac{1}{2\Psi}$$

we have

$$\Pi[\ \widehat{\delta} > \underline{\delta}\ ] = \frac{1}{2} - \frac{1}{2\Psi}[\frac{1}{\underline{\Phi}} \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}(v_s^m(p^w, E_m^w) - v_s^m(p, E_m)) - \eta \sum_{j \in L} C_j(p) - \frac{1}{2}\underline{\Phi}] \tag{A.1}$$

After the simple manipulation equation (A.1) results in equation (2.13).

**A.1.2. Equilibrium Conditions.** Recall that the following formulas give the effects of a marginal trade policy change on a specific swing voter and an industry:

$$\frac{\partial v_s^m(p, E^m)}{\partial p_k} = (\frac{\varepsilon_{mk}}{N_m} - \frac{1}{N})y_k(p_k) + \frac{1}{N}(p_k - p_k^w)m_k^{'}(p_k)$$

where $\varepsilon_{mk}$ is an indicator function, that equals 1 when $m = k$ and 0, otherwise.

$$\frac{\partial \pi_j(p_j)}{\partial p_k} = \theta_{jk} y_k(p_k)$$

where $\theta_{jk}$ is an indicator function as well and it equals 1 when $j = k$ and 0, otherwise.

To exploit the equilibrium first-order conditions, we have to sum up the effects over all swing voters and industries respectively. First let's consider the swing voters:

$$\begin{aligned} \sum_{m=1}^{M} \frac{\partial v_s^m(p, E^m)}{\partial p_k} &= \sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}((\frac{\varepsilon_{mk}}{N_m} - \frac{1}{N})y_k(p_k) + \frac{1}{N}(p_k - p_k^w)m_k^{'}(p_k)) \\ &= \frac{1}{N}\frac{1}{\underline{\Phi}}(p_k - p_k^w)m_k^{'}(p_k) + y_k(p_k)\sum_{m=1}^{M} \frac{\alpha_m}{\Phi_m}(\frac{\varepsilon_{mk}}{N_m} - \frac{1}{N}) \\ &= \frac{1}{N}\frac{1}{\underline{\Phi}}(p_k - p_k^w)m_k^{'}(p_k) + y_k(p_k)(\frac{1}{N_k}\frac{\alpha_k}{\Phi_k} - \frac{1}{N}\frac{1}{\underline{\Phi}}) \end{aligned} \tag{A.2}$$

Now derive the effect of a marginal policy change on all organized industries:

$$\begin{aligned} \sum_{j \in L} \frac{\partial \pi_j(p_j)}{\partial p_k} &= \sum_{j \in L} \theta_{jk} y_k(p_k) \\ &= I_k y_k(p_k) \end{aligned} \tag{A.3}$$

where

$$I_k = \begin{cases} 1 & \text{if industry } k \text{ is organized;} \\ 0 & \text{if industry } k \text{ is unorganized.} \end{cases}$$

Based on equation (2.19) the first-order condition in equation (2.17) can be written as follows:

$$\sum_{m=1}^{M} \bigtriangledown \frac{\alpha_m}{\Phi_m}(v_s^m(p, E_m) - v_s^m(p^w, E_m^w)) + \frac{\eta}{\underline{\Phi}} \sum_{j \in L} \bigtriangledown \pi_j(p_j) = 0 \tag{A.4}$$

Based on equation (A.4), considering only the change in one dimension, in the price $p_k$, the relevant first-order condition accordingly will be:

$$\frac{1}{N}\frac{1}{\underline{\Phi}}(p_k - p_k^w)m_k^{'}(p_k) + y_k(p_k)(\frac{1}{N_k}\frac{\alpha_k}{\Phi_k} - \frac{1}{N}\frac{1}{\underline{\Phi}}) + I_k y_k(p_k) = 0 \tag{A.5}$$

After some simple manipulations, equation (2.22) follows directly.

APPENDIX B

# Appendix to Chapter 3

## B.1. Pseudocode

Note, that a couple of functions and variables are directly baked into the program without any further explanation in the pseudocode below - for example the matrix that determines the general network structure (for each player and all coalitions). The origin and characterisation of these can be found in their respective parts in the main paper. The network structure $A$ and the preference relations $B$ both enter as a collection of $|X| \times |X|$-matrices, $\{A_S\}_{S \subseteq N, S \neq \emptyset}$ and $\{B_S\}_{S \subseteq N, S \neq \emptyset}$ resp., where $(A_S)_{i,j} = \mathbb{1}_{\{i \to_S j\}}(i, j)$ and $(B_S)_{i,j} = \mathbb{1}_{\{i \prec_S j\}}(i, j)$ for $(i, j) \in X \times X$. The detailed code is available at: `https://pub.uni-bielefeld.de/librecat/record/preview/2931412`

---

**Algorithm** Largest Consistent Set

---

**Input:** Countries $N$, Outcomes $X$, Network Structure $A$, Preference Relations $B$
**Output:** Largest Consistent Set $\{Y\}$

**procedure** PARAMETERSPACELCS$(N, X, A, B)$
  $E = eMaxArea$ ▷ See Section 3.3.6
  $\alpha = \alpha MinValue(E)$ ▷ See Section 3.3.6
  **for** $e \in E$ **do**
    $Y = GeneralLCS(N, X, A, B)$
  **return** $\{Y\}$

**function** GENERALLCS$(N, X, A, B)$
  **for** $S \subseteq N$ **do**
    $C_S = \min\{A_S, B_S\}$
  $D^0 = \max_{S \subseteq N}\{C_S\}$ ▷ : Direct Dominance
  $n = 0$
  **repeat**
    $n = n + 1$
    **for** $S \subseteq N$ **do**
      $A_S^n = (\mathbb{1}_{\{(A_S \cdot D^{n-1})_{i,j} \neq 0\}}(i, j))_{(i,j) \in X \times X}$
      $D_S^n = \min\{A_S^n, B_S\}$
    $D^n = \max_{S \subseteq N}\{D_S^n\}$ ▷ : Indirect Dominance
  **until** $D^n = D^{n-1}$
  $D = \mathbb{1}_X + D^n$
  $Y^0 = (1)_{x \in X}$
  $m = 0$
  **repeat**
    $m = m + 1$
    **for** $x \in X$ **do**
      **if** $Y_x^{m-1} = 0$ **then**
        $Y_x^m = 0$
      **else**
        $y = \max_{k \in X, S \subseteq N} \left\{ (A_S)_{x,k} \left( 1 - \max_{z \in X} \{ Y_z^{m-1} (D)_{k,z} \left(1 - (B_S)_{x,z}\right) \} \right) \right\}$
        $Y_x^m = Y_x^{m-1} - y$
  **until** $Y^m = Y^{m-1}$
  $Y = Y^m$
  **return** $Y$

## B.2. Model

**B.2.1. Individual Welfare.** The following table lists the individual welfare for each (representative) trade agreement, depending on endowments and tariffs, multiplied with the factor 18. Note that for MFN, CUGFT, FTAGFT, and MTAGFT the welfare $W_i$ resembles $W_j$ and $W_k$. In case of CU(i,j), FTA(i,j), and MTA(i,j) the welfare $W_i$ is similar to $W_j$. For FTAHub(i) the welfare $W_j$ resembles $W_k$.

| Trade Agreement | Individual Welfare |
|---|---|
| MFN | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 - 8t_i^2 + t_j^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k - t_j - t_k)$ |
| | $+2e_j(e_k + t_i + t_k) + 2e_k(t_i + t_j)$ |
| CU(i,j) | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 - 11t_{ik}^2 + t_{jk}^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} - t_k)$ |
| | $+2e_j(e_k - 4t_{ik} + t_k) + 2e_k(5t_{ik} - t_{jk})$ |
| ↳ $W_k$ | $2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{ik}^2 + 4t_{jk}^2 - 8t_k^2 + 2e_i(e_j - 2e_k + 2t_{jk} + t_k)$ |
| | $+2e_j(-2e_k + 2t_{ik} + t_k) + 4e_k(9\alpha - 2t_{ik} - 2t_{jk})$ |
| CUGFT | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_je_k$ |
| FTA(i,j) | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 - 11t_{ik}^2 + t_{jk}^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} - t_k)$ |
| | $+2e_j(e_k - 4t_{ik} + t_k) + 2e_k(5t_{ik} - t_{jk})$ |
| ↳ $W_k$ | $2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{ik}^2 + 4t_{jk}^2 - 8t_k2 + 2e_i(e_j - 2e_k + 2t_{jk} + t_k)$ |
| | $+2e_j(-2e_k + 2t_{ik} + t_k) + 4e_k(9\alpha - 2t_{ik} - 2t_{jk})$ |
| FTAHub(i) | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 + t_{jk}^2 + t_{kj}^2 + 4e_i(9\alpha - e_j - e_k + t_{jk} + t_{kj})$ |
| | $+2e_j(e_k - t_{kj}) - 2e_kt_{jk}$ |
| ↳ $W_j$ | $2e_i^2 + 2e_j^2 - 10e_k^2 + 4t_{jk}^2 - 11t_{kj}^2 + 2e_i(e_j - 2e_k + 2t_{jk} - 4t_{kj})$ |
| | $+e_j(-4e_k + 10t_{kj}) + 4e_k(9\alpha - 2t_{jk})$ |
| FTAGFT | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_je_k$ |
| MTA(i,j) | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 - 8t_i^2 + t_j^2 + t_k^2 + 4e_i(9\alpha - e_j - e_k - t_j - t_k)$ |
| | $+2e_j(e_k + t_i + t_k) + 2e_k(t_i + t_j)$ |
| ↳ $W_k$ | $2e_i^2 + 2e_j^2 - 10e_k^2 + t_i^2 + t_j^2 - 8t_k^2 + 2e_i(e_j - 2e_k + t_j + t_k)$ |
| | $+2e_j(-2e_k + t_i + t_k) + 4e_k(9\alpha - t_i - t_j)$ |
| MTAGFT | |
| ↳ $W_i$ | $-10e_i^2 + 2e_j^2 + 2e_k^2 + 4e_i(9\alpha - e_j - e_k) + 2e_je_k$ |

TABLE B.1. The individual welfare for each trade agreement depending on endowments and tariffs

**B.2.2. Tariffs.** The following describes the tariffs that the countries choose for each trade agreement. In addition to the specific restrictions mentioned in Section 3.3.3, all tariffs are bounded both from below and above by zero and the MFN-tariff respectively. As per WTO rule, the formation of any PTA does not allow additional tariffs towards others - which results in the upper bound of the MFN-tariff. Also, any form of subsidies is excluded here - which results in the lower bound of zero. Now, the following determines and describes the optimal tariffs for each scenario and the cases where capping occurs:

B.2.2.1. *MFN.* In this case, the optimal tariff of country $i$, given by $t_i^* = \frac{1}{8}(e_j + e_k)$, is always greater than zero as the endowments themselves are greater than zero. Additionally, $t_i^*$ is going to play the role of the maximal tariff for country $i$ for all the other agreements, then denoted $t_i^{MFN}$.

B.2.2.2. *CU.* Consider the scenario CU(i,j), then the optimal tariff of country $i$ towards country $k$, given by $t_{ik}^* = \frac{1}{5}(2e_k - e_j)$, is always greater than zero but not always less than the MFN-tariff (and the one towards country $j$, $t_{ij}^*$, is always zero):

i) Lower Bound. By assumption on the endowments $e_k \geq \frac{3}{5}e_j$ and thus $e_k > \frac{1}{2}e_j$, which guarantees $t_{ik}^* > 0$.

ii) Upper Bound. By assumption on the endowments $e_k \leq \frac{5}{3}e_j$ however $t_{ik}^* \leq t_i^{MFN}$ requires $e_k \leq \frac{13}{11}e_j$, which leaves the interval $\frac{13}{11}e_j < e_k \leq \frac{5}{3}e_j$ to require capping. For this interval, the (maximal) MFN-tariff is optimal as the derivative of the joint welfare with respect to $t_{ik}$ is always greater than zero on the interval $[0, t_i^{MFN}]$:

$$\frac{\partial(W_i + W_j)}{t_{ik}} = \frac{1}{9}\left(-10t_{ik} - 2e_j + 4e_k\right) \geq \frac{1}{36}\left(-13e_j + 11e_k\right) > 0$$

B.2.2.3. *FTA.* Consider the scenario FTA(i,j), then the optimal tariff of country $i$ towards country $k$, given by $t_{ik}^* = \frac{1}{11}(5e_k - 4e_j)$, is neither always greater than zero nor always less than the MFN-tariff (but the one towards country $j$, $t_{ij}^*$, is zero):

i) Lower Bound. By assumption on the endowments $e_k \geq \frac{3}{5}e_j$ however $t_{ik}^* \geq 0$ requires $e_k \geq \frac{4}{5}e_j$, which leaves the interval $\frac{3}{5}e_j \leq e_k < \frac{4}{5}e_j$ to require capping. For this interval, the (minmal) zero-tariff is optimal as the derivative of the welfare with respect to $t_{ik}$ is always lesser than zero on the interval $[0, t_i^{MFN}]$:

$$\frac{\partial W_i}{\partial t_{ik}} = \frac{1}{9}\left(-11t_{ik} - 4e_j + 5e_k\right) \leq \frac{1}{9}\left(5e_k - 4e_j\right) < 0$$

ii) Upper Bound. By assumption on the endowments $e_k \leq \frac{5}{3}e_j$ however $t_{ik}^* \leq t_i^{MFN}$ requires $e_k \leq \frac{43}{29}e_j$, which leaves the interval $\frac{43}{29}e_j < e_k \leq \frac{5}{3}e_j$ to require capping. For this interval, the (maximal) MFN-tariff is optimal as the derivative of the welfare with respect to $t_{ik}$ is always greater than zero on the interval $[0, t_i^{MFN}]$:

$$\frac{\partial W_i}{\partial t_{ik}} = \frac{1}{9}\left(-11t_{ik} - 4e_j + 5e_k\right) \geq \frac{1}{72}\left(-43e_j + 29e_k\right) > 0$$

B.2.2.4. *MTA.* Consider the scenario MTA(i,j), then the optimal tariff of country $i$, given by $t_i^* = \frac{1}{7}(2e_k - e_j)$, is greater than zero and less or equal to the MFN-tariff as per assumption on the endowments $\frac{3}{5}e_j \leq e_k \leq \frac{5}{3}e_j$.

B.2.2.5. *Notes.* The analysis considered country $i$ and an agreement with country $j$, but it naturally extends to all other combinations. Also, the perspective of the third country needs no further analysis as it always chooses the MFN-tariff. Furthermore, the case of FTAHub(i) is simply a combination of FTA(i,j) and FTA(i,k). Finally, the three variants of GFT require no additional analysis as every country always chooses the zero-tariff. Information on another form of GFT, Pseudo-GFT, that technically exists but turns out to be negligible, can be found in Appendix B.2.5.1.

**B.2.3. Overall Welfare.** The following table lists the overall welfare for each (representative) trade agreement, depending purely on endowments, computed modulo $2\alpha\left(\sum_{n\in N} e_n\right)$, which is the common term associated with the factor $\alpha$. Also, the notation $l_c$ and $l^c$ is used to indicate that country $l$ is capped in terms of tariffs from below or above respectively. Note that one specific comparison of trade agreements is presented in more detail in Appendix B.2.5.2.

| Trade Agreement | Overall Welfare |
|---|---|
| MFN | |
| ↳ no cap | $\frac{11}{32}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$ |
| CU(i,j) | |
| ↳ no cap | $\frac{1}{1600}(-563e_i^2 - 550e_ie_j - 448e_ie_k - 563e_j^2 - 448e_je_k - 704e_k^2)$ |
| ↳ $i^c$ | $\frac{1}{1600}(-563e_i^2 - 550e_ie_j - 448e_ie_k - 550e_j^2 - 550e_je_k - 627e_k^2)$ |
| ↳ $i^c, j^c$ | $\frac{11}{32}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$ |
| CUGFT | |
| ↳ no cap | $\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$ |
| FTA(i,j) | |
| ↳ no cap | $\frac{1}{7744}(-2963e_i^2 - 2662e_ie_j - 1728e_ie_k - 2963e_j^2 - 1728e_je_k - 3648e_k^2)$ |
| ↳ $i_c$ | $\frac{1}{23232}(-8889e_i^2 - 7986e_ie_j - 5184e_ie_k - 7865e_j^2 - 7744e_je_k - 9344e_k^2)$ |
| ↳ $i_c, i_c$ | $\frac{1}{192}(-65e_i^2 - 66e_ie_j - 64e_ie_k - 65e_j^2 - 64e_je_k - 64e_k^2)$ |
| ↳ $i^c$ | $\frac{1}{7744}(-2963e_i^2 - 2662e_ie_j - 1728e_ie_k - 2662e_j^2 - 2662e_je_k - 3155e_k^2)$ |
| ↳ $i^c, j^c$ | $\frac{11}{32}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$ |
| FTAHub(i) | |
| ↳ no cap | $\frac{1}{363}(-153e_i^2 - 81e_ie_j - 81e_ie_k - 146e_j^2 - 121e_je_k - 146e_k^2)$ |
| ↳ $j_c$ | $\frac{1}{363}(-137e_i^2 - 81e_ie_j - 121e_ie_k - 146e_j^2 - 121e_je_k - 121e_k^2)$ |
| ↳ $j_c, k_c$ | $\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$ |
| ↳ $j^c$ | $\frac{1}{23232}(-8889e_i^2 - 5184e_ie_j - 7986e_ie_k - 9344e_j^2 - 7744e_je_k - 7865e_k^2)$ |
| ↳ $j^c, k^c$ | $\frac{1}{192}(-66e_i^2 - 66e_ie_j - 66e_ie_k - 65e_j^2 - 64e_je_k - 65e_k^2)$ |
| FTAGFT | |
| ↳ no cap | $\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$ |
| MTA(i,j) | |
| ↳ no cap | $\frac{1}{3136}(-1083e_i^2 - 1078e_ie_j - 960e_ie_k - 1083e_j^2 - 960e_je_k - 1216e_k^2)$ |
| MTAGFT | |
| ↳ no cap | $\frac{1}{3}(-e_i^2 - e_ie_j - e_ie_k - e_j^2 - e_je_k - e_k^2)$ |

TABLE B.2. The overall welfare for each trade agreement depending on endowments

**B.2.4. Transition Tables.** The following lists the network structure of Section 3.3.5, specifically Figure 3.1 and 3.2, in the form of transition tables:

| $x_1 \in X$ | $x_2 \in X \setminus \{x_1\}$ with $x_1 \to_{\{i\}} x_2$ |
|---|---|
| MFN | - |
| CU(i,j) | MFN |
| CU(j,k) | - |
| CU(k,i) | MFN |
| CUGFT | CU(j,k) |
| FTA(i,j) | MFN |
| FTA(j,k) | - |
| FTA(k,i) | MFN |
| FTAHub(i) | MFN, FTA(i,j), FTA(k,i) |
| FTAHub(j) | FTA(j,k) |
| FTAHub(k) | FTA(j,k) |
| FTAGFT | FTA(j,k), FTAHub(j), FTAHub(k) |
| MTA(i,j) | MFN |
| MTA(j,k) | - |
| MTA(k,i) | MFN |
| MTAGFT | MTA(j,k) |

(A) The transition table for coalition $\{i\}$, $i \in N$.

| $x_1 \in X$ | $x_2 \in X \setminus \{x_1\}$ with $x_1 \to_{\{i,j\}} x_2$ |
|---|---|
| MFN | CU(i,j), FTA(i,j), MTA(i,j) |
| CU(i,j) | MFN, FTA(i,j), MTA(i,j) |
| CU(j,k) | MFN, CU(i,j), FTA(i,j), MTA(i,j) |
| CU(k,i) | MFN, CU(i,j), FTA(i,j), MTA(i,j) |
| CUGFT | MFN, CU(i,j), CU(j,k), CU(k,i), FTA(i,j), MTA(i,j) |
| FTA(i,j) | MFN, CU(i,j), MTA(i,j) |
| FTA(j,k) | MFN, CU(i,j), FTA(i,j), FTAHub(j), MTA(i,j) |
| FTA(k,i) | MFN, CU(i,j), FTA(i,j), FTAHub(i), MTA(i,j) |
| FTAHub(i) | MFN, CU(i,j), FTA(i,j), FTA(k,i), MTA(i,j) |
| FTAHub(j) | MFN, CU(i,j), FTA(i,j), FTA(j,k), MTA(i,j) |
| FTAHub(k) | MFN, CU(i,j), FTA(i,j), FTA(j,k), FTA(k,i), FTAGFT, MTA(i,j) |
| FTAGFT | MFN, CU(i,j), FTA(i,j), FTA(j,k), FTA(k,i), FTAHub(i), FTAHub(j), FTAHub(k), MTA(i,j) |
| MTA(i,j) | MFN, CU(i,j), FTA(i,j) |
| MTA(j,k) | MFN, CU(i,j), FTA(i,j), MTA(i,j) |
| MTA(k,i) | MFN, CU(i,j), FTA(i,j), MTA(i,j) |
| MTAGFT | MFN, CU(i,j), FTA(i,j), MTA(i,j), MTA(j,k), MTA(k,i) |

(B) The transition table for coalition $\{i, j\}$, $i, j \in N$, $i \neq j$.

TABLE B.3. The network structure as transition tables

**B.2.5. Additional Remarks.**

B.2.5.1. *Pseudo-GFT.* In Appendix B.2.2 a special case of 'Pseudo-GFT' is a possibility. Namely, in the case of a hub structure with both non-hub nodes capping at zero the trade agreement amounts to the same tariff structure (and welfare) of a GFT. If it were ever part of the stable set, then it would necessarily need to be considered de facto GFT even though it is not de jure GFT. However, in our analysis this case never occurred and it is therefore a negligible oddity.

B.2.5.2. *A Special Case.* As can be seen in Table B.2 the overall welfare is equal in case of MFN, CU($i^c, j^c$), and FTA($i^c, j^c$) even though the tariff structure is different. The following explores this equivalence in order to provide an insight into the underlying mechanics. In terms of tariff structure both CU($i^c, j^c$) and FTA($i^c, j^c$) are the same and therefore it is sufficient to compare MFN with CU($i^c, j^c$) when only interested in (effects on) welfare. Now, Table B.4 shows us the differences in the welfare (components) both on the individual as well as on the joint/overall level, which are computed from the expressions in Table B.5 and B.6.

| | $\Delta(MFN, CU(i^c, j^c))$ |
|---|---|
| $TR_i$ | $1/24(e_j + e_k)(2e_j - e_k)$ |
| $CS_i$ | $(32e_i^2 + 64e_ie_k - 13e_j^2 - 26e_je_k + 19e_k^2)/1152$ |
| $PS_i$ | $1/12e_i(-e_i - e_k)$ |
| $W_i$ | $(-64e_i^2 - 32e_ie_k + 83e_j^2 + 22e_je_k - 29e_k^2)/1152$ |
| $TR_j$ | $1/24(e_i + e_k)(2e_i - e_k)$ |
| $CS_j$ | $(-13e_i^2 - 26e_ie_k + 32e_j^2 + 64e_je_k + 19e_k^2)/1152$ |
| $PS_j$ | $1/12e_j(-e_j - e_k)$ |
| $W_j$ | $(83e_i^2 + 22e_ie_k - 64e_j^2 - 32e_je_k - 29e_k^2)/1152$ |
| $TR_k$ | 0 |
| $CS_k$ | $19(-e_i^2 - 2e_ie_k - e_j^2 - 2e_je_k - 2e_k^2)/1152$ |
| $PS_k$ | $1/24e_k(e_i + e_j + 2e_k)$ |
| $W_k$ | $(-19e_i^2 + 10e_ie_k - 19e_j^2 + 10e_je_k + 58e_k^2)/1152$ |

(A) The difference in the individual welfare (components) depending on endowments

| | $\Delta(MFN, CU(i^c, j^c))$ |
|---|---|
| $TR_i + TR_j$ | $1/24(2e_i^2 + e_ie_k + 2e_j^2 + e_je_k - 2e_k^2)$ |
| $CS_i + CS_j$ | $19(e_i^2 + 2e_ie_k + e_j^2 + 2e_je_k + 2e_k^2)/1152$ |
| $PS_i + PS_j$ | $1/12(-e_i^2 - e_ie_k - e_j^2 - e_je_k)$ |
| $W_i + W_j$ | $(19e_i^2 - 10e_ie_k + 19e_j^2 - 10e_je_k - 58e_k^2)/1152$ |
| $\sum_{n\in N} TR_n$ | $1/24(2e_i^2 + e_ie_k + 2e_j^2 + e_je_k - 2e_k^2)$ |
| $\sum_{n\in N} CS_n$ | 0 |
| $\sum_{n\in N} PS_n$ | $1/24(-2e_i^2 - e_ie_k - 2e_j^2 - e_je_k + 2e_k^2)$ |
| $\sum_{n\in N} W_n$ | 0 |

(B) The difference in the joint/overall welfare (components) depending on endowments

TABLE B.4. The difference in the welfare (components) depending on endowments

| Welfare Components | |
|---|---|
| **MFN** | |
| $TR_i$ | $(e_j+e_k)^2/32$ |
| $CS_i$ | $(18e_i^2+13e_j^2+13e_k^2+8e_je_k+18e_i(e_j+e_k))/128$ |
| $PS_i$ | $-e_i(-16\alpha+6e_i+3e_j+3e_k)/8$ |
| $TR_j$ | $(e_i+e_k)^2/32$ |
| $CS_j$ | $(13e_i^2+18e_j^2+13e_k^2+8e_ie_k+18e_j(e_i+e_k))/128$ |
| $PS_j$ | $-e_j(-16\alpha+3e_i+6e_j+3e_k)/8$ |
| $TR_k$ | $(e_i+e_j)^2/32$ |
| $CS_k$ | $(13e_i^2+13e_j^2+18e_k^2+8e_ie_j+18e_k(e_i+e_j))/128$ |
| $PS_k$ | $-e_k(-16\alpha+3e_i+3e_j+6e_k)/8$ |
| **CU($i^c, j^c$)** | |
| $TR_i$ | $-(5e_j-7e_k)(e_j+e_k)/96$ |
| $CS_i$ | $(65e_i^2+65e_j^2+49e_k^2+49e_je_k+e_i(81e_j+49e_k))/576$ |
| $PS_i$ | $-e_i(-48\alpha+16e_i+9e_j+7e_k)/24$ |
| $TR_j$ | $-(5e_i-7e_k)(e_i+e_k)/96$ |
| $CS_j$ | $(65e_i^2+65e_j^2+49e_k^2+49e_ie_k+e_j(81e_i+49e_k))/576$ |
| $PS_j$ | $-e_j(-48\alpha+9e_i+16e_j+7e_k)/24$ |
| $TR_k$ | $(e_i+e_j)^2/32$ |
| $CS_k$ | $(17e_i^2+17e_j^2+25e_k^2+25e_ie_k+25e_je_k+9e_ie_j)/144$ |
| $PS_k$ | $-e_k(-24\alpha+5e_i+5e_j+10e_k)/12$ |

(A) The individual welfare (components) depending on endowments

| Welfare Components | |
|---|---|
| **MFN** | |
| $TR_i+TR_j$ | $1/32(e_i^2+2e_ie_k+e_j^2+2e_je_k+2e_k^2)$ |
| $CS_i+CS_j$ | $1/128(31e_i^2+36e_ie_j+26e_ie_k+31e_j^2+26e_je_k+26e_k^2)$ |
| $PS_i+PS_j$ | $1/8(-6e_i^2-6e_ie_j-3e_ie_k-6e_j^2-3e_je_k)+2\alpha(e_i+e_j)$ |
| $\sum_{n\in N} TR_n$ | $1/16(e_i^2+e_ie_j+e_ie_k+e_j^2+e_je_k+e_k^2)$ |
| $\sum_{n\in N} CS_n$ | $11/32(e_i^2+e_ie_j+e_ie_k+e_j^2+e_je_k+e_k^2)$ |
| $\sum_{n\in N} PS_n$ | $1/4(-3e_i^2-3e_ie_j-3e_ie_k-3e_j^2-3e_je_k-3e_k^2)+2\alpha\left(\sum_{n\in N} e_n\right)$ |
| **CU($i^c, j^c$)** | |
| $TR_i+TR_j$ | $1/96(-5e_i^2+2e_ie_k-5e_j^2+2e_je_k+14e_k^2)$ |
| $CS_i+CS_j$ | $1/288(65e_i^2+81e_ie_j+49e_ie_k+65e_j^2+49e_je_k+49e_k^2)$ |
| $PS_i+PS_j$ | $1/24(-16e_i^2-18e_ie_j-7e_ie_k-16e_j^2-7e_je_k)+2\alpha(e_i+e_j)$ |
| $\sum_{n\in N} TR_n$ | $1/48(-e_i^2+3e_ie_j+e_ie_k-e_j^2+e_je_k+7e_k^2)$ |
| $\sum_{n\in N} CS_n$ | $11/32(e_i^2+e_ie_j+e_ie_k+e_j^2+e_je_k+e_k^2)$ |
| $\sum_{n\in N} PS_n$ | $1/24(-16e_i^2-18e_ie_j-17e_ie_k-16e_j^2-17e_je_k-20e_k^2)+2\alpha\left(\sum_{n\in N} e_n\right)$ |

(B) The joint/overall welfare (components) depending on endowments

TABLE B.5. The welfare (components) depending on endowments

| Trade Agreement | Individual/Joint/Overall Welfare |
|---|---|
| MFN | |
| ↳ $W_i$ | $1/128(-78e_i^2 + 256\alpha e_i - 30e_ie_j - 30e_ie_k + 17e_j^2 + 16e_je_k + 17e_k^2)$ |
| ↳ $W_j$ | $1/128(17e_i^2 - 30e_ie_j + 16e_ie_k - 78e_j^2 + 256\alpha e_j - 30e_je_k + 17e_k^2)$ |
| ↳ $W_k$ | $1/128(17e_i^2 + 16e_ie_j - 30e_ie_k + 17e_j^2 - 30e_je_k - 78e_k^2 + 256\alpha e_k)$ |
| ↳ $W_i + W_j$ | $1/128(-61e_i^2 - 60e_ie_j - 14e_ie_k - 61e_j^2 - 14e_je_k + 34e_k^2) + 2\alpha(e_i + e_j)$ |
| ↳ $\sum_{n\in N} W_n$ | $1/32(-11e_i^2 - 11e_ie_j - 11e_ie_k - 11e_j^2 - 11e_je_k - 11e_k^2) + 2\alpha\left(\sum_{n\in N} e_n\right)$ |
| CU$(i^c, j^c)$ | |
| ↳ $W_i$ | $1/576(-319e_i^2 + 1152\alpha e_i - 135e_ie_j - 119e_ie_k + 35e_j^2 + 61e_je_k + 91e_k^2)$ |
| ↳ $W_j$ | $1/576(35e_i^2 - 135e_ie_j + 61e_ie_k - 319e_j^2 + 1152\alpha e_j - 119e_je_k + 91e_k^2)$ |
| ↳ $W_k$ | $1/288(43e_i^2 + 36e_ie_j - 70e_ie_k + 43e_j^2 - 70e_je_k - 190e_k^2 + 576\alpha e_k)$ |
| ↳ $W_i + W_j$ | $1/288(-142e_i^2 - 135e_ie_j - 29e_ie_k - 142e_j^2 - 29e_je_k + 91e_k^2) + 2\alpha(e_i + e_j)$ |
| ↳ $\sum_{n\in N} W_n$ | $1/32(-11e_i^2 - 11e_ie_j - 11e_ie_k - 11e_j^2 - 11e_je_k - 11e_k^2) + 2\alpha\left(\sum_{n\in N} e_n\right)$ |

TABLE B.6. The welfare depending on endowments

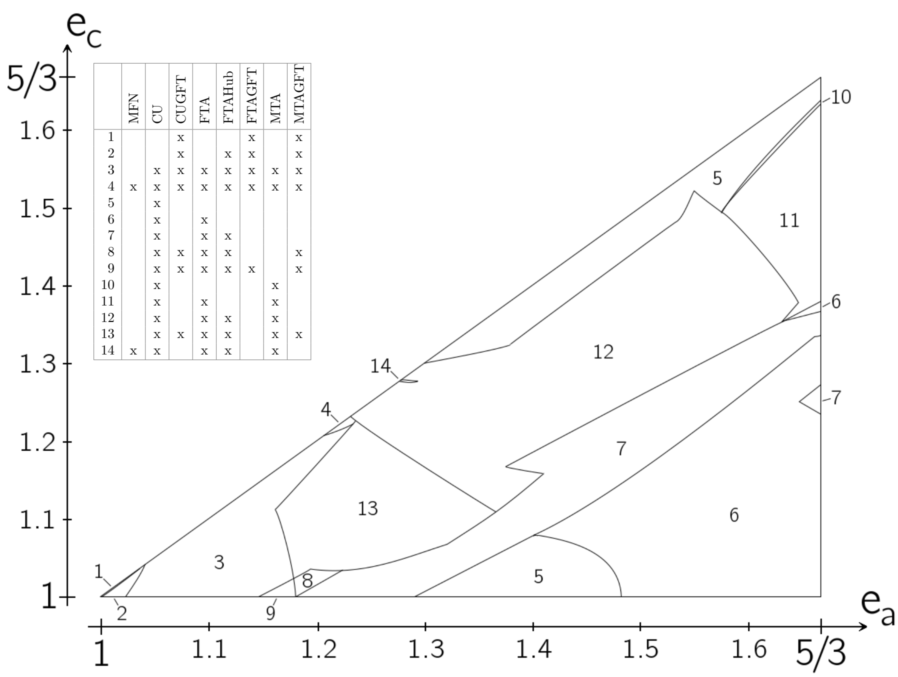
Now, recall that capping at the MFN-tariff for both members of a customs union, in this case CU$(i^c, j^c)$, occurs when the endowment of the non-member is above a minimal value determined by the endowments of the members, $\max\{e_i, e_j\} < \frac{11}{13}e_k$ (Appendix B.2.2). Using this together with the general assumptions on the relation of endowments, the following effects on welfare (components) take place.

| $\Delta(MFN, CU(i^c, j^c))$ | | |
|---|---|---|
| Country $i$ | Country $j$ | Country $k$ |
| $TR$ | | |
| $+\tau_i$ | $+\tau_j$ | 0 |
| $+\tau_{ij}$ | | 0 |
| $+\tau$ | | |
| $CS$ | | |
| $+\gamma_i$ | $+\gamma_j$ | $-\gamma_k$ |
| $+\gamma_{ij}$ | | $-\gamma_k$ |
| 0 | | |
| $PS$ | | |
| $-\rho_i$ | $-\rho_j$ | $+\rho_k$ |
| $-\rho_{ij}$ | | $+\rho_k$ |
| $-\rho$ | | |
| $W$ | | |
| $\pm\omega_i$ | $\pm\omega_j$ | $+\omega_k$ |
| $-\omega_{ij}$ | | $+\omega_k$ |
| 0 | | |

TABLE B.7. The effect on the welfare (components)

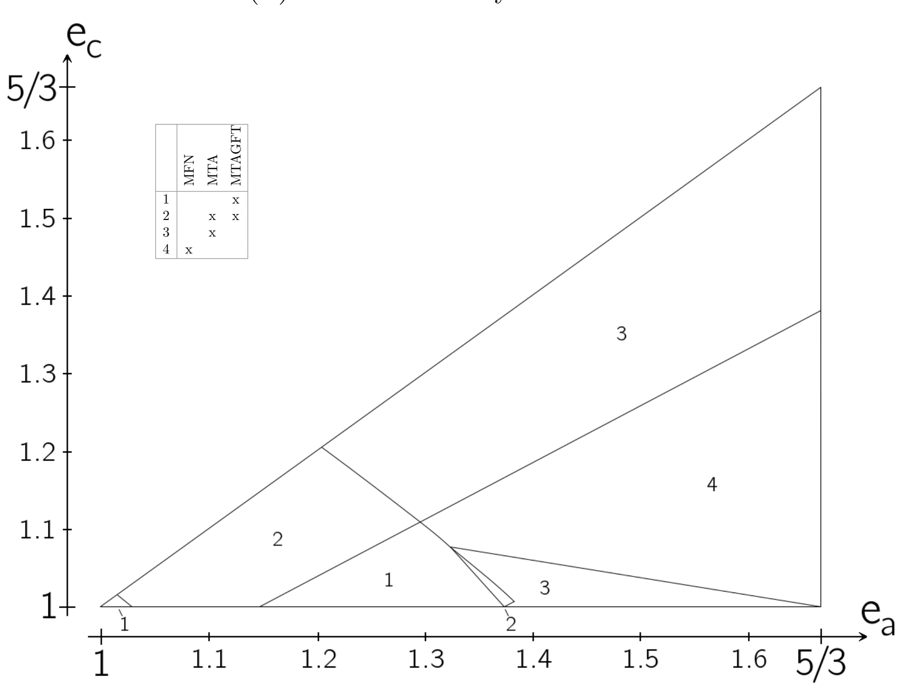
## B.3. Analysis

**B.3.1. Additional Graphics.** The following provides detailed figures:




| | MFN | CU | CUGFT | FTA | FTAHub | FTAGFT | MTA | MTAGFT |
|---|---|---|---|---|---|---|---|---|
| 1 | | | x | | | x | | x |
| 2 | | | x | | x | x | | x |
| 3 | | x | x | x | x | x | x | x |
| 4 | x | x | x | x | x | x | x | x |
| 5 | | x | | | | | | |
| 6 | | x | | x | | | | |
| 7 | | x | | x | x | | | |
| 8 | | x | x | x | x | | | x |
| 9 | | x | x | x | x | x | | x |
| 10 | | x | | | | | x | |
| 11 | | x | | x | | | x | |
| 12 | | x | | x | x | | x | |
| 13 | | x | x | x | x | | x | x |
| 14 | x | x | | x | x | | x | |

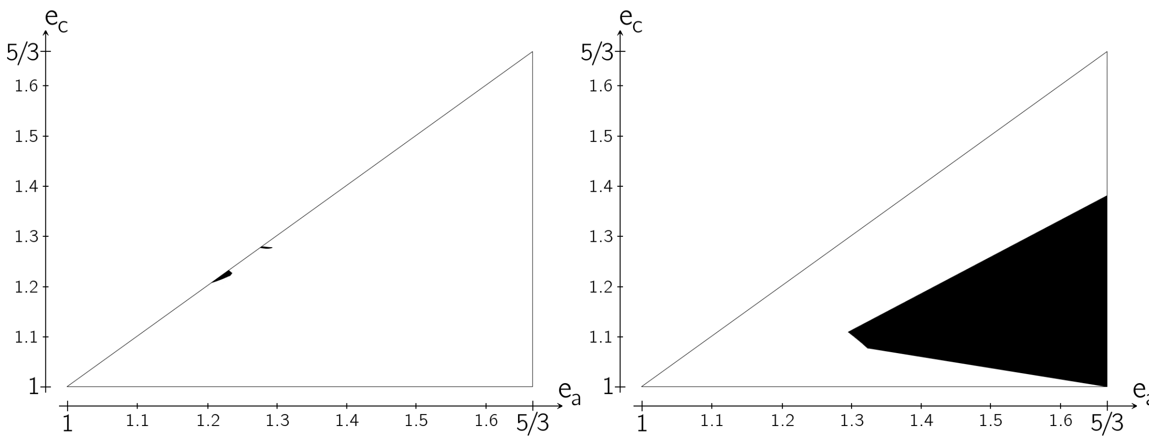
(A) Overall Stability with PTAs




| | MFN | MTA | MTAGFT |
|---|---|---|---|
| 1 | | | x |
| 2 | | x | x |
| 3 | | x | |
| 4 | x | | |

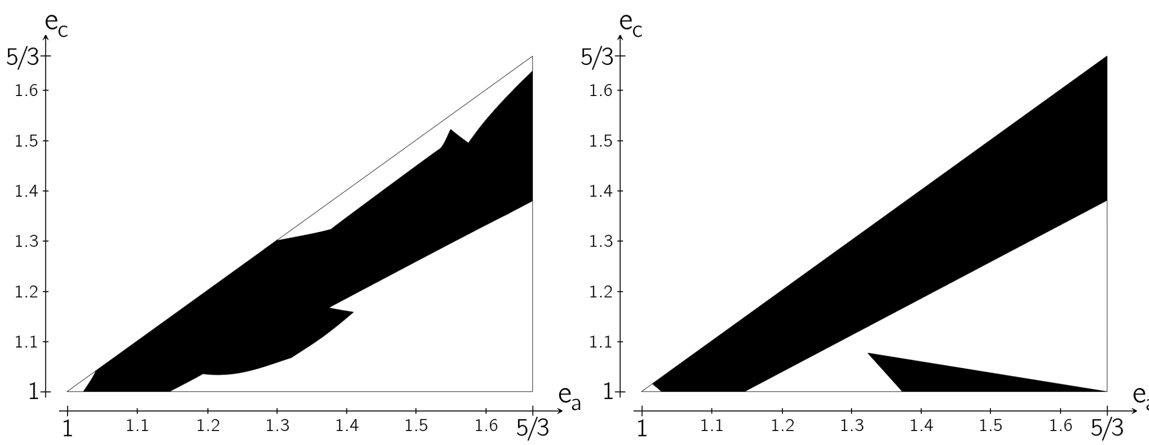
(B) Overall Stability without PTAs

FIGURE B.1. Overall Stability with and without PTAs



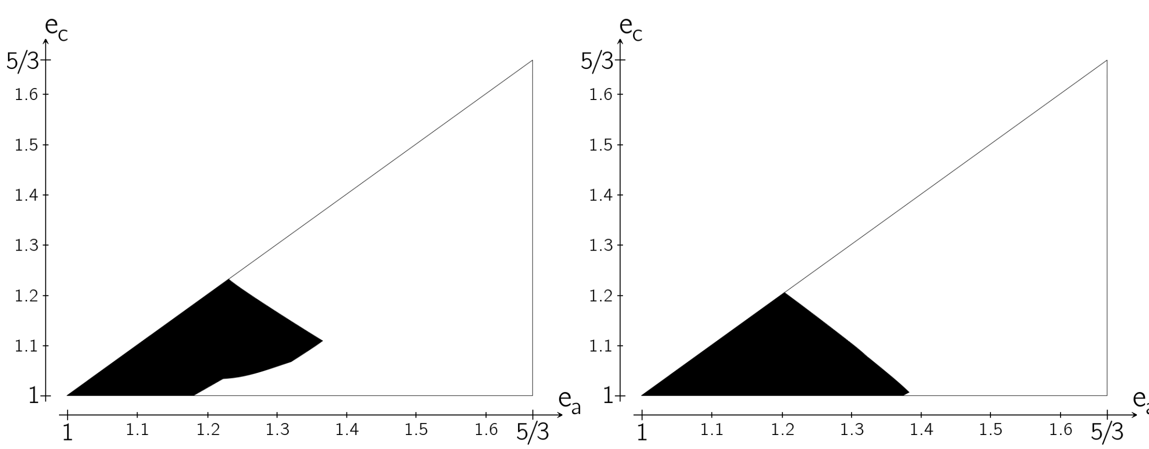

(A) with PTAs (B) without PTAs

Figure B.2. Stability of MFN



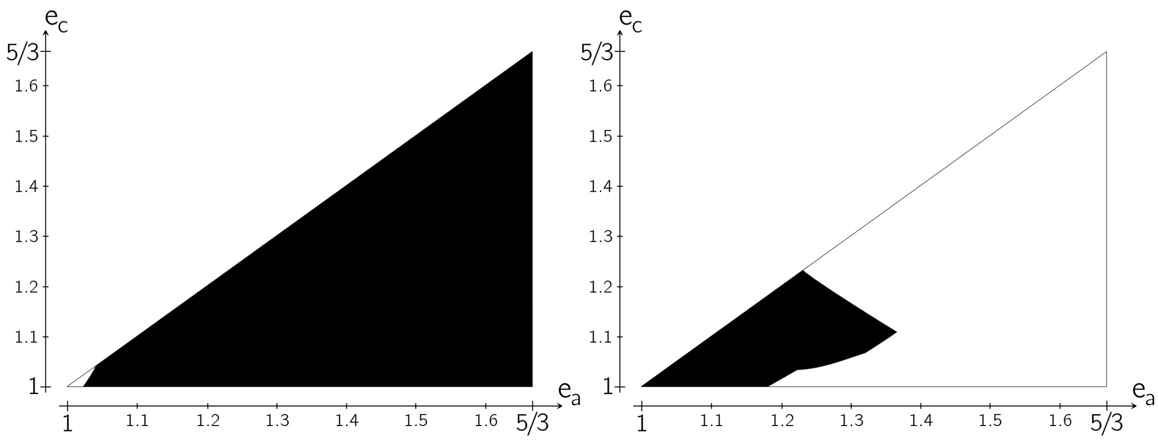

(A) with PTAs (B) without PTAs

Figure B.3. Stability of MTA



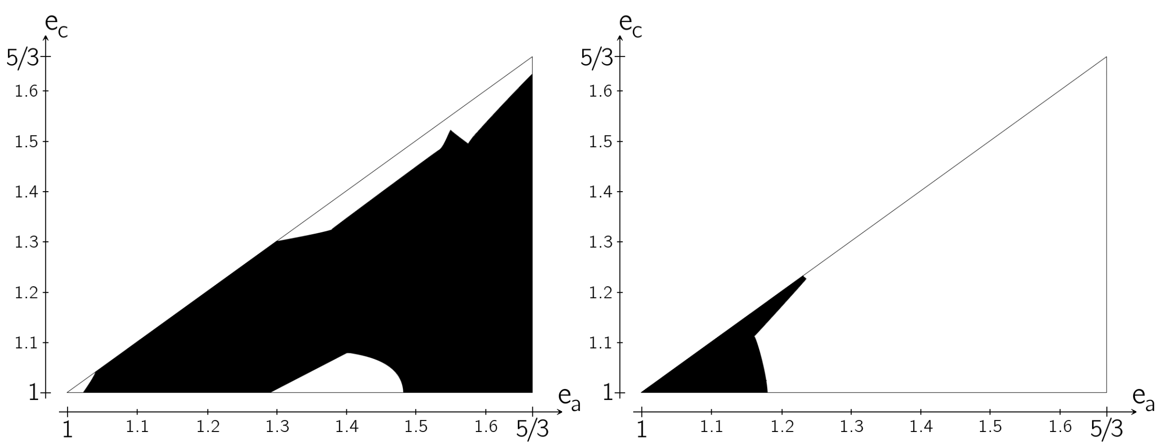

(A) with PTAs (B) without PTAs

Figure B.4. Stability of MTAGFT

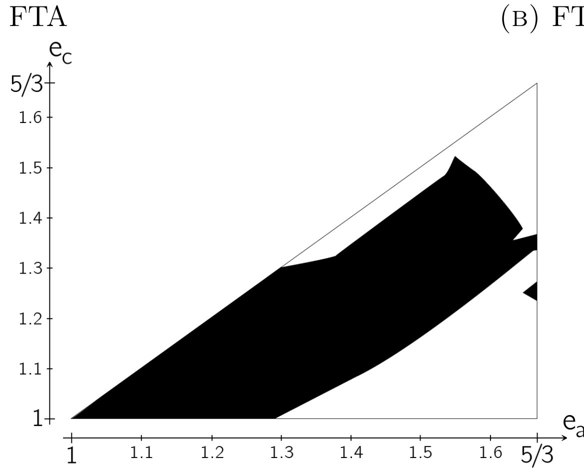



(A) CU (B) CUGFT

Figure B.5. Stability of CU




(A) FTA (B) FTAGFT




(C) FTAHub

Figure B.6. Stability of FTA

**B.3.2. Exact Intervals.** The table here lists the exact intervals where each specific trade agreement is part of the stable set (for the border of the parameter space):

| Trade Agreement | Exact Interval(s) |
|---|---|
| | $e_b = e_{\min} \leq e_c \leq e_{\max} = e_a$ |
| CU(b,c) | [1.0000000000000000,1.3380093520374081] |
| CU(c,a) | [1.3259853039412157,1.6666666666666667] |
| FTA(b,c) | [1.0000000000000000,1.3807615230460921] |
| FTA(c,a) | [1.3353373413493654,1.6305945223780896] |
| FTAHub(c) | [1.2364729458917836,1.2698730794923179] |
| | [1.3353373413493654,1.3647294589178356] |
| MTA(c,a) | [1.379425517702071,1.635938543754175] |
| | $e_b = e_c = e_{\min} \leq e_a \leq e_{\max}$ |
| CU(a,b) | [1.0240480961923848,1.1108884435537743] |
| CU(b,c) | [1.0240480961923848,1.6666666666666667] |
| CU(c,a) | [1.0240480961923848,1.1108884435537743] |
| CUGFT | [1.0000000000000000,1.1803607214428857] |
| FTA(a,b) | [1.0240480961923848,1.2404809619238477] |
| FTA(b,c) | [1.0240480961923848,1.29124916499666] |
| | [1.483633934535738,1.6666666666666667] |
| FTA(c,a) | [1.0240480961923848,1.2404809619238477] |
| FTAHub(b) | [1.0013360053440215,1.29124916499666] |
| FTAHub(c) | [1.0013360053440215,1.29124916499666] |
| FTAGFT | [1.0000000000000000,1.1803607214428857] |
| MTA(a,b) | [1.0240480961923848,1.1469605878423514] |
| MTA(c,a) | [1.0240480961923848,1.1469605878423514] |
| MTAGFT | [1.0000000000000000,1.1803607214428857] |

| $e_b = e_{\min} \leq e_a = e_c \leq e_{\max}$ | |
|---|---|
| MFN | [1.2044088176352705,1.2297929191716768] |
| | [1.2752171008684035,1.276553106212425] |
| CU(a,b) | [1.0454241816967267,1.2498329993319974] |
| CU(b,c) | [1.0454241816967267,1.2498329993319974] |
| CU(c,a) | [1.0494321977287908,1.6666666666666667] |
| CUGFT | [1.0000000000000000,1.2297929191716768] |
| FTA(a,b) | [1.0494321977287908,1.2925851703406814] |
| FTA(b,c) | [1.0494321977287908,1.2925851703406814] |
| FTA(c,a) | [1.0454241816967267,1.2925851703406814] |
| FTAHub(a) | [1.0454241816967267,1.276553106212425] |
| FTAHub(b) | [1.0454241816967267,1.2925851703406814] |
| FTAHub(c) | [1.0454241816967267,1.276553106212425] |
| FTAGFT | [1.0000000000000000,1.2297929191716768] |
| MTA(a,b) | [1.0494321977287908,1.2244488977955912] |
| MTA(b,c) | [1.0494321977287908,1.2244488977955912] |
| MTA(c,a) | [1.0454241816967267,1.2925851703406814] |
| MTAGFT | [1.0000000000000000,1.2297929191716768] |

TABLE B.8. The exact intervals of stability with PTAs

| Trade Agreement | Exact Interval(s) |
|---|---|
| $e_b = e_{\min} \leq e_c \leq e_{\max} = e_a$ | |
| MFN | [1.0000000000000000,1.3780895123580494] |
| MTA(b,c) | [1.0000000000000000,1.0000000000000000] |
| MTA(c,a) | [1.379425517702071,1.6666666666666667] |
| $e_b = e_c = e_{\min} \leq e_a \leq e_{\max}$ | |
| MFN | [1.6666666666666667,1.6666666666666667] |
| MTA(a,b) | [1.0307281229124916,1.1469605878423514] |
| MTA(b,c) | [1.0307281229124916,1.1469605878423514] |
| | [1.3754175016700068,1.6666666666666667] |
| MTA(c,a) | [1.0307281229124916,1.1469605878423514] |
| MTAGFT | [1.0000000000000000,1.3740814963259853] |
| $e_b = e_{\min} \leq e_a = e_c \leq e_{\max}$ | |
| MTA(a,b) | [1.0160320641282565,1.1202404809619237] |
| MTA(b,c) | [1.0160320641282565,1.1202404809619237] |
| MTA(c,a) | [1.0160320641282565,1.6666666666666667] |
| MTAGFT | [1.0000000000000000,1.203072812291249] |

TABLE B.9. The exact intervals of stability without PTAs

APPENDIX C

# Appendix to Chapter 4

## C.1. Proof of Lemma 4.1

Proof. Let us start the proof with a general remark on the compactness of the involved sets: For any compact topological space $(X, \tau)$ the set $\Delta(X)$ is always compact with the weak convergence topology. Now, $\mathcal{X}$ is compact by assumption and thus $\Delta(\mathcal{X})$ is compact. Furthermore, $\mathcal{H} = \Delta(\mathcal{X})^S$ as a finite Cartesian product of compact spaces is compact and thus $\Delta(\mathcal{H})$ is compact.[1] In addition, $\mathcal{I}$ is compact and therefore $\Delta(\mathcal{I})$ is compact as well. Consequently, $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ is compact. Next, let absence of unanimity apply and split the proof into its two directions.

Assume absence of unanimity from now on. If there exist functions $V$ and $V_i$, $i \in \mathcal{I}$, that represent the corresponding preferences, then this representation of the impartial observer's preferences is affine in identity lotteries. Therefore, it satisfies the acceptance principle and independence over identity lotteries.

Conversely, let the preferences of the impartial observer now satisfy both the acceptance principle and independence over identity lotteries. Then, let us first prove that there exist lotteries $\overline{P}$ and $\underline{P}$ in $\Delta(\mathcal{H})$ and lotteries $\overline{z_1}$, $\underline{z_1}$, $\overline{z_2}$, $\underline{z_2}$ in $\Delta(\mathcal{I})$ with

1) $(\overline{z_1}, \overline{P}) \succ (\underline{z_2}, \underline{P})$

2) $(\overline{z_1}, \overline{P}) \succeq (z, P) \succeq (\underline{z_2}, \underline{P}) \qquad \forall (z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$

3) $(\overline{z_1}, \overline{P}) \succeq (z, \overline{P}) \succeq (\underline{z_1}, \overline{P}) \qquad \forall z \in \Delta(\mathcal{I})$

4) $(\overline{z_2}, \underline{P}) \succeq (z, \underline{P}) \succeq (\underline{z_2}, \underline{P}) \qquad \forall z \in \Delta(\mathcal{I})$

Note that these lotteries do not necessarily have to be distinct. The existence itself follows from continuity of $\succeq$, non-emptiness of $\succ$, and compactness of $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$:

First of all, consider the collection of weak upper (resp. lower) contour sets, i.e. $\{(z', P') \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H}) | (z', P') \succeq (z, P)\}$ (resp. $\preceq$) for $(z, P)$ in $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$. Each of them is a closed set because of the continuity of $\succeq$. Also, this collection satisfies the finite intersection property as the intersection of a finite collection of contour set is simply the contour set of the 'maximal' or 'minimal' element. Together with the fact that $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ is compact, this yields that the infinite intersection over all elements is non-empty. In other words, there exists a maximal and minimal element (or pair) satisfying the second relation(s) and by using the non-emptiness of $\succ$ also the first. In addition, the remaining relations of three and four follow either directly or by using similar arguments for closed subsets of the contour sets.

Moreover, by independence over identity lotteries, we can take $\overline{z_1}$, $\underline{z_1}$, $\overline{z_2}$, $\underline{z_2}$ to be degenerate identity lotteries $\overline{i_1}$, $\underline{i_1}$, $\overline{i_2}$, $\underline{i_2}$. Namely, take the relation $(\overline{z_1}, \overline{P}) \succeq (z, P)$ and assume (for contradiction) that there exists no $i \in \mathcal{I}$ such that $(i, \overline{P}) \succeq (z, P)$ for all $(z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$. In other words, for all $i \in \mathcal{I}$ there exists a corresponding element $(z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ such that $(z, P) \succ (i, \overline{P})$. However, $(\overline{z_1}, \overline{P}) \succeq (z, P)$ and thus $(\overline{z_1}, \overline{P}) \succ (i, \overline{P})$ for all $i \in \mathcal{I}$. Take now one $i \in \mathcal{I}$ such that $\overline{z_1}_i \neq 0$ and define $\alpha := (\overline{z_1}_i)$ and $z' := \frac{1}{(1-\alpha)} (\overline{z_1}_1, \ldots, \overline{z_1}_{i-1}, 0, \overline{z_1}_{i+1}, \ldots, \overline{z_1}_I)$. Then, using independence over identity

[1] A modified proof that requires no specific structure of $\mathcal{H}$ but compactness could replace the last one and a half sentences with 'The set $\mathcal{H}$ is compact by assumption.'.

lotteries, $(\overline{z_1}, \overline{P}) \succ (i, \overline{P})$ implies by mixing with $(z', \overline{P})$ that $(\alpha\overline{z_1} + (1-\alpha)z', \overline{P}) \succ (\alpha i + (1-\alpha)z', \overline{P}) = (\overline{z_1}, \overline{P})$ - a contradiction to the maximality of $(\overline{z_1}, \overline{P})$. Similarly, the same chain of arguments work for $\underline{z_2}$ as well as $\underline{z_1}$ and $\overline{z_2}$.

Next, let us show the following two (related) statements for these lotteries:

- A) At least one of the following three relations always holds:
  - A1) $(\underline{i_1}, \overline{P}) \sim (\underline{i_2}, \underline{P})$;
  - A2) $(\overline{i_2}, \underline{P}) \sim (\overline{i_1}, \overline{P})$;
  - A3) $(\overline{i_2}, \underline{P}) \succ (\underline{i_1}, \overline{P})$.
- B) Take $(z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$, then at least one of two holds:
  - B1) $(\overline{i_1}, \overline{P}) \succeq (z, P) \succeq (\underline{i_1}, \overline{P})$;
  - B2) $(\overline{i_2}, \underline{P}) \succeq (z, P) \succeq (\underline{i_2}, \underline{P})$.

Note, that B) follows from A) with 2)-4). Thus, let us show A). First, if $\underline{P} = \overline{P}$, then clearly both A1) and A2) hold by property of $\overline{i_1}, \underline{i_1}, \overline{i_2}, \underline{i_2}$. Second, if $\underline{P} \neq \overline{P}$, then assume that A1) and A2) do not hold, i.e. $(\underline{i_1}, \overline{P}) \succ (\underline{i_2}, \underline{P})$ and $(\overline{i_1}, \overline{P}) \succ (\overline{i_2}, \underline{P})$. By definition of $\underline{i_1}$, it holds that $(\underline{i_2}, \overline{P}) \succeq (\underline{i_1}, \overline{P})$ and hence $(\underline{i_2}, \overline{P}) \succ (\underline{i_2}, \underline{P})$. Using absence of unanimity and the acceptance principle, there must exist another individual $\hat{i} \neq \underline{i_2}$ such that $(\hat{i}, \underline{P}) \succ (\hat{i}, \overline{P})$. Again, by definition of $\underline{i_1}$ it follows that $(\hat{i}, \overline{P}) \succeq (\underline{i_1}, \overline{P})$ and hence $(\hat{i}, \underline{P}) \succ (\underline{i_1}, \overline{P})$. On the other hand, by definition of $\overline{i_2}$, it follows that $(\overline{i_2}, \underline{P}) \succeq (\hat{i}, \underline{P})$ and hence $(\overline{i_2}, \underline{P}) \succ (\underline{i_1}, \overline{P})$ as desired.[2]

Using B) together with continuity, it follows that for any pair of lotteries $(z, P)$ there exists $z'$ with $(z, P) \sim (z', \overline{P})$ or there exists $z''$ with $(z, P) \sim (z'', \underline{P})$. Moreover, $z'$ and $z''$ can actually be chosen such that they are convex combinations of $\overline{i_1}$ and $\underline{i_1}$ (for $z'$) and $\overline{i_2}$ and $\underline{i_2}$ (for $z''$).

Finally, let us construct the (claimed) representation of the impartial observer's preferences. First, assume again that $\underline{P} = \overline{P}$. Then, as before $(\overline{i_1}, \overline{P}) \succ (\underline{i_1}, \overline{P})$ and for all $(z, P)$ it holds that $(\overline{i_1}, \overline{P}) \succeq (z, P) \succeq (\underline{i_1}, \overline{P})$. Define for each $(z, P)$ the expression $V(z, P)$ by

$$(z, P) \sim \big(V(z, P)\overline{i_1} + (1 - V(z, P))\underline{i_1}, \overline{P}\big)$$

whose existence and uniqueness follows by continuity of $\succeq$ (see previous paragraph) and independence over identity lotteries. In addition, the function $V$ is continuous and represents $\succeq$ on $\Delta(\mathcal{I}) \times \Delta(\mathcal{H})$. In order to show that this representation is also affine in identity lotteries, take $(z, P)$ and $(z', P)$ with their representations given by

$$(z, P) \sim \big(V(z, P)\overline{i_1} + (1 - V(z, P))\underline{i_1}, \overline{P}\big)$$
$$(z', P) \sim \big(V(z', P)\overline{i_1} + (1 - V(z', P))\underline{i_1}, \overline{P}\big)$$

and note that independence over identity lotteries implies that the following mixture

$$\big([\alpha V(z, P) + (1-\alpha)V(z', P)]\overline{i_1} + [1 - \alpha V(z, P) - (1-\alpha)V(z', P)]\underline{i_1}, \overline{P}\big)$$

is on the same indifference curve as $(\alpha z + (1-\alpha)z', P)$. Thus, by definition and uniqueness of $V(\cdot, \cdot)$ it holds that $V(\alpha z + (1-\alpha)z', P) = \alpha V(z, P) + (1-\alpha)V(z', P)$. Now, any identity lottery $z$ in $\Delta(\mathcal{I})$ can be written as $z = \sum_{i \in \mathcal{I}} z_i i$. Therefore, affinity implies $V(z, P) = \sum_{i \in \mathcal{I}} z_i V(i, P)$. The acceptance principle then yields that $V_i \colon \Delta(\mathcal{H}) \to \mathbb{R}$ defined by $V_i(P) := V(i, P)$ corresponds to $\succeq_i$ on $\Delta(\mathcal{H})$.

Uniqueness follows by standard arguments: Assume that $W$, $W_i$ is another such representation of $\succeq$, $\succeq_i$ with $W(z, P) = \sum_{i \in \mathcal{I}} z_i W_i(P)$. Let $(z, P) \in \Delta(\mathcal{I}) \times \Delta(\mathcal{H})$ and assume that without loss of generality B1) holds, i.e. $(\overline{i_1}, \overline{P}) \succeq (z, P) \succeq (\underline{i_1}, \overline{P})$, (the case of B2), i.e. $(\overline{i_2}, \underline{P}) \succeq (z, P) \succeq (\underline{i_2}, \underline{P})$, follows similarly). Next, define $a := W(\underline{i_1}, \overline{P})$ and $b := W(\overline{i_1}, \overline{P}) - W(\underline{i_1}, \overline{P}) > 0$. Now, by definition of $V$ it holds that $(z, P) \sim$

[2] Even though our approach only uses ¬A1), it also works with ¬A2) and similar arguments.

$(V(z,P)\overline{i_1} + (1 - V(z,P))\underline{i_1}, \overline{P})$ and therefore, using that $W$ is affine,

$$\begin{aligned} W(z,P) &= V(z,P)W(\overline{i_1}, \overline{P}) + (1 - V(z,P))W(\underline{i_1}, \overline{P}) \\ &= bV(z,P) + a \end{aligned}$$

with the definitions of $a$ and $b$. Also, $W_i(P) = W(i,P) = bV(i,P) + a = bV_i(P) + a$.

Second, assume now that $\underline{P} \neq \overline{P}$. If $(\underline{i_1}, \overline{P}) \sim (\underline{i_2}, \underline{P})$ or $(\overline{i_2}, \underline{P}) \sim (\overline{i_1}, \overline{P})$, then for all $(z,P)$ we have $(\overline{i_1}, \overline{P}) \succeq (z,P) \succeq (\underline{i_1}, \overline{P})$ or $(\overline{i_2}, \underline{P}) \succeq (z,P) \succeq (\underline{i_2}, \underline{P})$. Thus, in both cases, the arguments from the previous part still hold. Otherwise, by A3), it follows that $(\overline{i_1}, \overline{P}) \succ (\overline{i_2}, \underline{P}) \succ (\underline{i_1}, \overline{P}) \succ (\underline{i_2}, \underline{P})$. In other words, two overlapping intervals span the entire range of the impartial observer's preferences. Construct for each of those intervals $V^1$ and $V^2$ as before. In order to merge these into a single $V$ on the whole range, define $\alpha_1 := V^1(\overline{i_2}, \underline{P})$, $\alpha_2 := V^2(\underline{i_1}, \overline{P})$ and then the (affine) transformed

$$\begin{aligned} \tilde{V}^1(z,P) &:= \frac{1 - \alpha_2}{1 - (1 - \alpha_1)\alpha_2} V^1(z,P) + \frac{\alpha_1 \alpha_2}{1 - (1 - \alpha_1)\alpha_2} \\ \tilde{V}^2(z,P) &:= \frac{\alpha_1}{1 - (1 - \alpha_1)\alpha_2} V^2(z,P) \end{aligned}$$

to get agreement on the overlap $(\overline{i_2}, \underline{P}) \succeq (z,P) \succeq (\underline{i_1}, \overline{P})$. Define $V$ to be the merged representation on the whole range, which is affine as both $V^1$ and $V^2$ as well as $\tilde{V}^1$ and $\tilde{V}^2$ are affine. Then, with the same arguments as before we get that $V(z,P) = \sum_{i\mathcal{I}} z_i V_i(P)$, where $V_i$ represent $\succeq_i$ for $i \in \mathcal{I}$ and $V$ represents $\succeq$, and the representation exhibits the claimed uniqueness. □

## C.2. Construction of a Dummy

As mentioned in chapter 4, in order to actually satisfy the assumption of absence of unanimity, without distorting any preferences, the introduction of a dummy individual $d$ is necessary. The following explains this necessity:

Assuming sufficiently heterogeneous beliefs, it is certainly possible to imagine a ranking of the form $K \succ_1 L$, $L \succ_2 K$ and $K \sim_3 L$ - essentially mirroring reality. However, the assumption of absence of unanimity also applies to all degenerate outcome lotteries, like the one always yielding the outcome $(1,1)$ irrespective of the state of the world and also the one always yielding $(0,1)$. Certainly, it is counter-intuitive to assume that in reality one of the soldiers would prefer the second over the first one, i.e. preferring being dead over being alive with everything else fixed. Now, the dummy individual takes care of this problem by preferring $(0,1)$ over $(1,1)$ and therefore maintaining absence of unanimity. At the same time, the probability of imagining yourself as the dummy individual is set to zero (for both options) to prevent any distortion on the level of preferences of the impartial observer.

A dummy individual seems artificial, especially one that is necessary because of an assumption that is imposed by us on the model. Yet, it is not an actual restriction or invalidates the assumption. It is merely a technical solution to a technical problem. The two presented (degenerate) lotteries that conflict with absence of unanimity otherwise are (or were) not part of the set of feasible options in reality anyway. A dummy individual in this example is necessary due to the homogeneity of the individuals (and their preferences), which is the result of a simplistic structure. Thus, a dummy individual allows us to apply our theory to examples where the size of the choice set collides with absence of unanimity otherwise. Essentially, this weakens absence of unanimity while still remaining in the framework of our theory.

Formally: Consider $\mathcal{I}' = \mathcal{I} \cup \{d\}$ with the (fixed) identity lottery $(1/3, 1/3, 1/3, 0)$ and set $u_d(x_1, x_2) = u(x_1, x_2)$ and $v_d(y) = 1 - v(y)$ with $m_d = m_1$ (or $m_d = m_{2/3}$). It results in the following (ultimately irrelevant) utilities for the two moral choices: $V_d(K) = 0.4233$ and $V_d(L) = 0.4054$; Consequently: $V_d(L) < V_d(K)$.

# Bibliography


1. *General agreement on tariffs and trade*, 1947, Source: `http://www.wto.org`.
2. *World trade report 2011*, 2011, Source: `http://www.wto.org`.
3. Philippe Aghion, Pol Antràs, and Elhanan Helpman, *Negotiating free trade*, Journal of International Economics **73** (2007), no. 1, 1–30.
4. Shiri Alon and Gabi Gayer, *Utilitarian preferences with multiple priors*, Econometrica **84** (2016), no. 3, 1181–1201.
5. Gary M. Anderson and Robert D. Tollison, *Ideology, interest groups, and the repeal of the corn laws*, Journal of Institutional and Theoretical Economics **141** (1985), no. 2, 197–212.
6. F.J. Anscombe and R.J. Aumann, *A definition of subjective probability*, Annals of Mathematical Statistics **34** (1963), 199–205.
7. Stephen Ansolabehere and James M. Snyder, *The incumbency advantage in U.S. elections: An analysis of state and federal offices, 1942–2000*, Election Law Journal **1** (2002), no. 3, 315–338.
8. Robert Aumann and Roger Myerson, *Endogenous formation of links between players and coalitions: an application of the shapley value*, The Shapley Value (1988), 175–191.
9. David Austen-Smith, *Rational consumers and irretional voters: A review essay on black hole tariffs and endogenous policy theory*, Economics & Politics **3** (1991), no. 1, 73–92.
10. Kyle Bagwell, Chad P. Bown, and Robert W. Staiger, *Is the wto passé?*, Journal of Economic Literature **54** (2016), no. 4, 1125–1231.
11. Kyle Bagwell and Robert W. Staiger, *Regionalism and multilateral tariff cooperation*, Working Paper 5921, National Bureau of Economic Research, February 1997.
12. Kyle Bagwell and Robert W Staiger, *The design of trade agreements*, Handbook of Commercial Policy **1** (2016), 435–529.
13. Richard E. Baldwin and Thorvaldur Gylfason, *A domino theory of regionalism*, Expanding Membership of the European Union, Cambridge University Press, 1995, pp. 25–53.
14. David Baron, *Electoral competition with informed and uniformed voters*, The American Political Science Review **88** (1994), no. 1, 33–47.
15. Eugene Beaulieu, *Factor or industry cleavages in trade policy? an empirical analysis of the stolper–samuelson theorem*, Economics & Politics **14** (2002), no. 2, 99–131.
16. B. Douglas Bernheim, Bezalel Peleg, and Michael Whinston, *Coalition-proof nash equilibria i. concepts*, Journal of Economic Theory **42** (1987), no. 1, 1–12.
17. Douglas B. Bernheim and Michael D. Whinston, *Menu auctions, resource allocation, and economic influence*, The Quarterly Journal of Economics **101** (1986), no. 1, 1–31.
18. Jagdish Bhagwati, *Protectionism*, The MIT Press, 1988.
19. Jagdish Bhagwati, *Regionalism and multilateralism: an overview*, New dimensions in regional integration **22** (1993), 51.
20. Bruce A. Blonigen, *Revisiting the evidence on trade policy preferences*, Journal of International Economics **85** (2011), no. 1, 129 – 135.

21. Matilde Bombardini and Francesco Trebbi, *Votes or money? theory and evidence from the us congress*, Journal of Public Economics **95** (2011), no. 7, 587 – 611.
22. ______, *Competition and political organization: Together or alone in lobbying for trade policy?*, Journal of International Economics **87** (2012), no. 1, 18 – 26, Symposium on the Global Dimensions of the Financial Crisis.
23. Michael Suk-Young Chwe, *Farsighted coalitional stability*, Journal of Economic Theory **63** (1994), no. 2, 299 – 325.
24. Paola Conconi, Giovanni Facchini, and Maurizio Zanardi, *Policymakers' horizon and trade reforms: The protectionist effect of elections*, Journal of International Economics **94** (2014), no. 1, 102 – 118.
25. Hervé Crès, Itzhak Gilboa, and Nicolas Vieille, *Aggregation of multiple prior opinions*, Journal of Economic Theory **146** (2011), no. 6, 2563–2582.
26. Eric Danan, Thibault Gajdos, Brian Hill, and Jean-Marc Tallon, *Robust social decisions*, American Economic Review **106** (2016), no. 9, 2407–25.
27. Peter A. Diamond, *Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment*, Journal of Political Economy **75** (1967), no. 5, 765–766.
28. Avinash Dixit, Gene M. Grossman, and Elhanan Helpman, *Common agency and coordination: General theory and application to government policy making*, Journal of Political Economy **105** (1997), no. 4, 752–769.
29. Anthony Downs, *An economic theory of democracy*, 1st ed., Harper and Row, 1957.
30. Juergen Eichberger and Ruediger Pethig, *Constitutional choice of rules*, European Journal of Political Economy **10** (1994), no. 2, 311 – 337.
31. Haluk Ergin and Faruk Gul, *A theory of subjective compound lotteries*, Journal of Economic Theory **144** (2009), no. 3, 899–929.
32. Giovanni Facchini, Peri Silva, and Gerald Willmann, *The customs union issue: Why do we observe so few of them?*, Journal of International Economics **90** (2013), no. 1, 136 – 147.
33. Alexander Fouirnaies and Andrew B. Hall, *The financial incumbency advantage: Causes and consequences*, The Journal of Politics **76** (2014), no. 3, 711–724.
34. Caroline Freund and Christine McDaniel, *How long does it take to conclude a trade agreement with the us?*, Tech. report, Peterson Institute for International Economics, 2016.
35. Taiji Furusawa and Hideo Konishi, *Free trade networks*, Journal of International Economics **72** (2007), no. 2, 310–335.
36. Thibault Gajdos and Feriel Kandil, *The ignorant observer*, Social Choice and Welafre **31** (2008), 193–232.
37. Kishore Gawande and Usree Bandyopadhyay, *Is protection for sale? evidence on the grossman-helpman theory of endogenous protection*, The Review of Economics and Statistics **82** (2000), no. 1, 139–152.
38. Pinelopi Koujianou Goldberg and Giovanni Maggi, *Protection for sale: An empirical investigation*, American Economic Review **89** (1999), no. 5, 1135–1155.
39. Sanjeev Goyal and Sumit Joshi, *Bilateralism and free trade*, International Economic Review **47** (2006), no. 3, 749–778.
40. Simon Grant, Atsushi Kajii, Ben Polak, and Zvi Safra, *Generalized utilitarianism and harsanyi's impartial observer theorem*, Econometrica **78** (2010), no. 6, 1939–1971.
41. Gene M Grossman, *The purpose of trade agreements*, Handbook of Commercial Policy **1** (2016), 379–434.

42. Gene M. Grossman and Elhanan Helpman, *Protection for sale*, The American Economic Review **84** (1994), no. 4, 833–850.
43. ______, *Electoral competition and special interest politics*, The Review of Economic Studies **63** (1996), no. 2, 265–286.
44. ______, *Special interest politics*, 1st ed., MIT University Press, 2001.
45. John Harsanyi, *An equilibrium-point interpretation of stable sets and a proposed alternative definition*, Management Science **20** (1974), no. 11, 1472–1495.
46. John C. Harsanyi, *Cardinal utility in welfare economics and in the theory of risk-taking*, Journal of Political Economy **61** (1953), no. 5, 434–435.
47. ______, *Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility*, Journal of Political Economy **63** (1955), no. 4, 309–321.
48. ______, *Rational behaviour and bargaining equilibrium in games and social situations*, Campbirdge University Press, 1977.
49. Elhanan Helpman, *Politics and trade policy*, Working Paper 5309, National Bureau of Economic Research, October 1995.
50. P. Jean-Jacques Herings, Ana Mauleon, and Vincent Vannetelbosch, *Farsightedly stable networks*, Games and Economic Behavior **67** (2009), no. 2, 526–541.
51. P.J.J. Herings, A. Mauleon, and V. Vannetelbosch, *Stability of networks under level-k farsightedness*, Research Memorandum 030, Maastricht University, Graduate School of Business and Economics (GSBE), January 2014.
52. Arye Hillman, *Declining industries and political-support protectionist motives*, The American Economic Review **72** (1982), no. 5, 1180–1187.
53. Douglas A. Irwin, *Multilateral and bilateral trade policies in the world trading system: An historical perspective*, New Dimensions in Regional Integration (J. De Melo and A. Panagariya, eds.), Cambridge University Press, 1993, pp. 90–119.
54. ______, *Industry or class cleavages over trade policy? evidence from the British general election of 1923*, The political economy of trade policy: Papers in honor of Jagdish Bhagwati (Robert C. Feenstra, Gene M. Grossman, and Douglas A. Irwin, eds.), MIT Press, 1996, pp. 53–76.
55. ______, *Clashing over commerce: A histroy of US trade policy*, The University of Chicago Press, 2017.
56. Matthew O Jackson and Asher Wolinsky, *A strategic model of social and economic networks*, Journal of economic theory **71** (1996), no. 1, 44–74.
57. Alexander Keck and Andreas Lendle, *New evidence on preference utilization*, WTO Staff Working Papers ERSD-2012-12, World Trade Organization (WTO), Economic Research and Statistics Division, September 2012.
58. Hiau Looi Kee, Alessandro Nicita, and Marcelo Olarreaga, *Estimating trade restrictiveness indices*, The Economic Journal **119** (2009), no. 534, 172–199.
59. Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji, *A smooth model of decision making under ambiguity*, Econometrica **73** (2005), no. 6, 1849–1892.
60. James Lake, *Free trade agreements as dynamic farsighted networks*, Economic Inquiry **55** (2017), no. 1, 31–50.
61. Stephen S. Magee, William A. Brock, and Leslie Young, *Black hole tariffs and endogenous policy theory: Political economy in general equilibrium*, 1st ed., Cambridge University Press, 1989.
62. Giovanni Maggi, *International trade agreements*, Handbook of International Economics, vol. 4, 2014, pp. 317–390.

63. Wofgang Mayer, *Endogenous tariff formation*, The American Economic Review **74** (1984), no. 5, 970–985.

64. Wolfgang Mayer and Jun Li, *Interest groups, electoral competition, and probabilistic voting for trade policies*, Economics & Politics **6** (1994), no. 1, 59–77.

65. John McLaren, *Chapter 2 - the political economy of commercial policy*, Handbook of Commercial Policy, vol. 1, North-Holland, 2016, pp. 109 – 159.

66. Paul Missios, Kamal Saggi, and Halis Murat Yildiz, *External trade diversion, exclusion incentives and the nature of preferential trade agreements*, Journal of International Economics **99** (2016), 105–119.

67. Devashish Mitra, Dimitrios D. Thomakos, and Mehmet A. Ulubaşoğlu, *Protection for sale in a developing country: Democracy vs. dictatorship*, The Review of Economics and Statistics **84** (2002), no. 3, 497–508.

68. Florian Mölders, *On the path to trade liberalisation: Political regimes in trade negotiations*, The World Economy **39** (2015), 890–924.

69. Philippe Mongin, *Consistent bayesian aggregation*, Journal of Economic Theory **66** (1995), 313–351.

70. ______, *The impartial observer theorem of social ethics*, Economics and Philosophy **17** (2001), 147–179.

71. Ryo Ichi Nagahisa and Sang Chul Suh, *A characterization of the walras rule*, Social Choice and Welfare **12** (1995), no. 4, 335–352.

72. Leandro Nascimento, *The ex ante aggregation of opinions under uncertainty*, Theoretical Economics **7** (2012), 535–570.

73. Robert F. Nau, *Uncertainty aversion with second-order utilities and probabilities*, Management Science **52** (2006), no. 1, 136–145.

74. Sam Peltzman, *Toward a more general theory of regulation*, The Journal of Law and Economics **19** (1976), no. 2, 211–240.

75. Torsten Persson and Guido Tabellini, *Political economics: Explaining economic policy*, 1st ed., The MIT Press, 2002.

76. Giacomo A. M. Ponzetto, *Heterogenous information and trade policy*, Barcelona GSE Working Paper Series (2011), no. 596.

77. Xiangyu Qu, *Separate aggregation of beliefs and values under ambiguity*, Economic Theory **63** (2017), no. 2, 503–519.

78. H. Raiffa, *Decision analysis - introductory lectures on choices under uncertainty*, Addison Wesley, Reading, MA, 1970, traduction française: Analyse de la décision : introduction aux choix en avenir incertain, Dunod, 1973.

79. Gordon C. Rausser, Johan Swinnen, and Pinhas Zusman, *Political power and economic policy: Theory, analysis, and empirical applications*, 1st ed., Cambridge University Press, 2011.

80. John Rawls, *A theory of justice*, Belknap, 1971.

81. Debraj Ray and Rajiv Vohra, *The farsighted stable set*, Econometrica **83** (2015), no. 3, 977–1011.

82. Bryan Riley, *U.S. trade policy gouges american sugar consumers*, Tech. report, The Heritage Foundation, 2014.

83. Dani Rodrik, *Political economy of trade policy*, Handbook of International Economics, vol. 3, Elsevier, 1995, pp. 1457 – 1494.

84. Kamal Saggi, Alan Woodland, and Halis Murat Yildiz, *On the relationship between preferential and multilateral trade liberalization: the case of customs unions*, American Economic Journal: Microeconomics **5** (2013), no. 1, 63–99.

85. Kamal Saggi and Halis Murat Yildiz, *Bilateralism, multilateralism, and the quest for global free trade*, Journal of International Economics **81** (2010), no. 1, 26–37.
86. Michael Sandel, *Justice: What's the right thing to do?*, Farrar, Straus and Giroux, 2010.
87. Kyoungwon Seo, *Ambiguity and second-order belief*, Econometrica **77** (2009), no. 5, 1575–1605.
88. Adam Smith, *The theory of moral sentiments*, London: A. Millar, 1759.
89. George Stigler, *The theory of economic regulation*, The Bell Journal of Economics and Management Science **2** (1971), no. 1, 3–21.
90. Patricia Tovar, *The effects of loss aversion on trade policy: Theory and evidence*, Journal of International Economics **78** (2009), no. 1, 154 – 167.
91. Frans van Winden, *Interest group behavior and influence*, The Encyclopedia of Public Choice (Charles K. Rowley and Friedrich Schneider, eds.), Springer US, Boston, MA, 2004, pp. 118–129.
92. William S. Vickrey, *Measuring marginal utility by reaction to risk*, Econometrica **13** (1945), no. 4, 319–333.
93. John von Neumann and Oskar Morgenstern, *Theory of games and economic behavior*, Princeton University Press, 1944.
94. Tim Wegenast, *Uninformed voters for sale: Electoral competition, information and interest groups in the US congress*, Kyklos **63** (2010), no. 2, 271–300.
95. Jin Zhang, Zhiwei Cui, and Lei Zu, *The evolution of free trade networks*, Journal of Economic Dynamics and Control **38** (2014), 72–86.
96. Jin Zhang, Licun Xue, and Lei Zu, *Farsighted free trade networks*, International Journal of Game Theory **42** (2013), no. 2, 375–398.