Mediation, Tax-Compliance and Gerrymandering

Three Essays in Public Economics

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2022

vorgelegt von
Raphaela Hennigs

Referent: Prof. Dr. Kai A. Konrad
Korreferent: Prof. Dr. Florian Englmaier
Promotionsabschlussberatung: 01.02.2023

Acknowledgements

Without the guidance, help and moral support of a number of people, I would not have completed this dissertation. First and foremost, I would like to thank my supervisor, Professor Kai A. Konrad, who gave me the opportunity to pursue my PhD at the Max-Planck-Institute for Tax Law and Public Finance. During the past years, he constantly gave me constructive advice and helpful feedback. In particular, I owe the research ideas for all the three chapters of my dissertation to Professor Kai A. Konrad and the third chapter of this dissertation benefited largely from the discussions with Professor Kai A. Konrad. I would like to express my special thanks for this.

Further, I would like to thank my second supervisor Professor Florian Englmaier for very fruitful discussions and valuable comments on the first two chapters of this dissertation. He was always open for a talk and each time I used this opportunity, I left with ideas to think about and, not to be neglected, more optimistic.

I also thank Professor Simeon Schudy for being part of my examination committee.

On the first chapter of this dissertation, I received helpful comments from current and former colleagues, being Laura Arnemann, Jana Cahlíkova, Jonas Send, Marco Serena and Bharat Goel, and a guest at the institute, Professor Rabah Armir.

My colleagues at the institute made the past years more enjoyable. I thank each of those already mentioned and further Anwesha Banerjee, Mariana Lopes de Fonseca, Andrea Martinangeli, Biljana Meiske, Raisa Sherif, Sven Arne Simon, Yixuan Shi, Carmen Sainz Villalba and Lisa Windsteiger.

Finally, I thank Sandra Sundt-Johannesen for helping me with any administrative matters and I thank Hans Müller and Andreas Kraus for always being very supportive with IT matters.

Contents

List of Figures

Preface

The public sphere is marked by opposing interests between individuals, groups or organisations, whose self-interested engagement in conflict given these opposing interests often leads to inefficient or undesirable outcomes. Understanding the mechanisms behind these conflicts is of key interest for public economics.

A conflict between two parties typically occurs given some framework, consisting of a set of rules and an informational environment, which limits, but also influences the parties behaviour. Understanding how rules and information feed into the parties behaviour is important, if we want to choose those rules and control the available information appropriately to induce certain outcomes of the conflict or to prevent the conflict from taking an inefficient outcome.

In this dissertation, I look at three such situations and show how information (Chapters 1 and 2) or rules (Chapter 3) can induce (more) desirable outcomes. In the first Chapter, I consider a situation in which two conflicting parties escalate a conflict to costly fighting due to underlying uncertainty about each other and show how the strategic provision of information can prevent the costly escalation of the conflict. The Chapter is phrased in the context of international relations, but the key result about that information, which feeds into conflicting parties behaviour, can be used strategically, is well applicable to other contexts. The second Chapter considers the conflict of interest between a tax payer and a tax authority, with the former trying to evade taxes. Also in this Chapter, I show how information can be used strategically to induce certain behaviour, in this case that of the tax payer. The third Chapter considers the problem of gerrymandering, where two political parties manipulate electoral district boarders, each aiming at increasing its representation in an elected house of representatives. Here, I show that choosing the process of setting district boarders appropriately induces a desirable outcome of this process, being that each party is represented in the elected house of representative according to its representation in voters' preferences.

All three papers address use game-theoretic methods to address the given quesitons. As the first two chapters apply Bayesian persuasion to situations with incomplete information, I will briefly introduce this framework. The paper titled "Bayesian Persuasion" by Kamenica and Gentzkow (2011) was one among the first in the growing literature on information design. Bergemann and Morris (2019) provide an overview of this literature. Information design asks how a sender optimally provides information to one or multiple receivers to induce a certain equilibrium behaviour.[1] Kamenica and Gentzkow (2011) focus on the case of one sender and one receiver with symmetric information. The payoffs of the sender and the receiver depend on an action taken by the receiver and an unknown state of the world. Sender and receiver share a common prior belief about the state of the world. The sender chooses an arbitrarily informative signal about the state of the world and commits to reveal the signal realisation to the receiver. A signal consists of a finite set of signal realisations and a mapping from the true state of the world to a distribution over the set of signal realisations. The receiver perfectly understands the signal, meaning that he knows how the observed signal realisation was generated. The receiver uses the observed signal realisation to form a posterior belief about the state of the world, and takes an action to maximise his expected payoff. The sender uses the signal to induce the receiver to take certain actions in equilibrium.

A key feature of Bayesian persuasion is that sender and receiver are ex-ante symmetrically informed and that the sender commits to reveal the signal's realisation. This feature distinguishes the problem from sender-receiver games, in which the sender chooses a communication rule after the true state has been realised. The interaction between sender and receiver becomes non-strategic, meaning that the signal chosen by the sender does not have to be incentive compatible. (This is generally the key feature of the information design literature.)

Kamenica and Gentzkow (2011) simplify this problem using a 'concavification' approach, which rests on two observations: first, the sender's expected

[1] The use of terminology in this literature is not consistent. For instance, the 'sender' is also referred to as 'information designer', 'mediator' or 'principal', and the 'receiver' is also referred to as 'agent' or 'player'.

payoff depends only on the receiver's posterior belief (which equals the sender's posterior belief). This allows to formulate the sender's expected payoff as a value function over posterior beliefs. Second, any possible distribution of posterior beliefs can be induced by an appropriately chosen signal subject to the constraint that the expectation over the posterior beliefs equals the prior belief. Taken together, these results allow to find the optimal signal by constructing the smallest concave function which takes everywhere values at least as high as the value function (i.e. the value function's concavification). Finding the optimal signal is trivial if there are only two or three states of the world so that the value function and its concavification can be depicted graphically. In this case, the posterior beliefs induced by the optimal signal and the sender's expected payoff achieved by the use of the optimal signal can be read off the graph, and the optimal signal can be backed up. However, if the state space is large, the problem is generally not trivial.

In the first Chapter, I model the information provision of a mediator in a situation of conflict as Bayesian persuasion. In the second Chapter, I model the choice of an income estimation technology of the tax authority as Bayesian persuasion. I explain in the introduction of each of both chapters how I use Bayesian persuasion in the respective framework. The model in Chapter 3, on the other hand, assumes complete information. I will now briefly outline the three Chapters of this dissertation.

In Chapter 1 titled "Conflict Prevention by Bayesian Persuasion", I analyse a model of mediation in a situation of conflict. The chapter is framed in the context of international relations, where the escalation of conflicts can be very costly, making it important to understand how escalation can be prevented. One important tool for conflict prevention is mediation and one form of mediation uses information to prevent conflicting parties from fighting each other. In Chapter 1, I ask how a mediator can generate information strategically to prevent war in a situation where conflicting parties have private information about their military strength. The conflict situation takes the following form: there is a dispute that needs to be settled and a default resolution makes both conflicting parties equally well off. However, parties can trigger a costly war by fighting each other. The outcome of a war is not clear ex-ante, as it depends on the military strength of both conflicting parties. Each party is privately informed about it strength but does not know the opponent's strength. A strong conflicting party has an incentive to wage costly war because it wins the war if it encounters with a weak opponent. Therefore, without mediation, a strong conflicting party always fights and escalates the conflict.

I show that a mediator can decrease the ex-ante probability of war by providing conflicting parties strategically with information. Thereby, strong conflicting parties fight less often. The conflicting parties benefit from mediation, as the ex-ante war probability is reduced. The benefit is taken up by weak conflicting parties. This benefit is larger when war is costlier and when the war probability absent mediation is higher.

In Chapter 2, titled "Estimating Income in a Tax Compliance Game - A Bayesian Persuasion Approach", I analyse the interaction between a tax-payer and the tax-authority. The tax-payer chooses whether to report his income and if he does not report his income, the tax-authority can choose whether to conduct a costly tax. However, without any information, the tax authority has no incentive to audit the tax payer, because in expectation, collected taxes do not pay off the audit cost. I show how the tax-authority can solve this problem by conditioning its choice of whether to audit on an observed estimate about the tax-payer's income. The technology which estimates the tax-payer's income is chosen at a prior state and I model this estimation technology using Bayesian persuasion. Although the tax-authority could choose an arbitrarily precise estimation technology, the optimal technology overestimates income. This allows the tax authority to commit on auditing the tax payer in case he does not report his income. In equilibrium, the tax-payer always reports his income and the tax-authority never needs to conduct an audit.

In Chapter 3, titled "Sequential Gerrymandering", I use a version of a Colonel Blotto game to analyse a gerrymandering process, which involves two political parties. Gerrymandering, the self-interested manipulation of electoral boarder districts, has a negative connotation and is criticised for leading to a misrepresentation of voter preferences in the elected house of representatives. In my analysis, I show that such misrepresentation does not occur in a process in which both parties choose electoral districts in alternating order. In equilibrium, the party which would win in a popular vote wins the majority of districts.

Chapter 1

Conflict Prevention by Bayesian Persuasion

This chapter is based on a published paper.[1]

1.1 Introduction

Understanding conflict resolution and prevention is a central concern of international relations. Mediation is one of the most widely used techniques in international crises management. Mediation occurred in 37 out of the 84 conflicts registered by the International Crisis Behaviour Project in the period from 1990 to 2015 and for 27 out of these 37 cases, mediation was named as either an "important" or the "most important" factor for easing tensions.[2] This Chapter contributes to the theoretical understanding of effective mediation by showing how a mediator can use research and intelligence to reduce the ex-ante war probability. Specifically, a mediator can convince conflicting parties not to fight each other by strategically providing them with information about their respective opponent.

Illustrative examples for this idea are aerial reconnaissance missions. Mediators conduct observational flights or take satellite pictures to obtain information about a conflicting party's military power and activity in a conflict and can provide the opponent with this information. For instance, the United

[1] Hennigs (2021)

[2] See Brecher and Wilkenfeld (1997) and Brecher et al. (2017).

States conducted observational flights to monitor the compliance with treaties between Israel and Egypt, and Israel and Syria in the Israel-Arab conflict during the 1970s and 1980s. The obtained pictures were reported to both sides in each case (Laipson et al., 1995). During the Falklands War, the United States provided Argentina with photographical coverage of the Falklands Islands and open sea areas Argentina had asked for based on an agreement with NASA (Freedman, 2007). Potentially, these pictures could have shown the movements of the British fleet. How can the provision of such information change conflicting parties' behaviour? If the information indicates that a conflicting party is strong, the opponent has less incentive to fight. On the downside, if the information indicates that the conflicting party is weak, this may increase the opponent's incentive to fight. The result of observational flights or satellite pictures might, therefore, prevent or trigger war. I show that a mediator can provide conflicting parties strategically with information to decrease the ex-ante war probability. I do so in a game-theoretic setting under full rationality, meaning that the conflicting parties are aware of the mediator's incentive to provide information strategically, so as to lower the ex-ante war probability.

This idea is modelled in a conflict game between two conflicting parties with mediation as information provision using a Bayesian persuasion framework (see Kamenica and Gentzkow (2011)). Two conflicting parties divide a pie of unit size with the default division being an equal split. The conflicting parties simultaneously decide whether to fight their opponent. Each conflicting party is strong or weak and does not know whether its opponent is strong or weak. The probability with which a conflicting party is strong is given by its militarisation level, which is common knowledge. Absent mediation, this uncertainty can lead to war. As a strong conflicting party always wins against a weak opponent, it fights if the probability to encounter a weak opponent is sufficiently high. A mediator can decrease the ex-ante war probability by decreasing conflicting parties' uncertainty and providing them with information indicating whether each opponent is strong or weak. In the context of international relations, such information could be for instance: whether or not a conflicting party has recently made large investments into military capabilities; whether or not a conflicting party has taken measures in preparation for a war, by, for example, moving troops close to the border; whether or not a conflicting party has a certain type of military equipment, such as specific missiles. The mediator can decide how precise the provided information is.

I derive how a mediator provides information optimally to minimise the ex-ante war probability. Comparative statics show that mediation decreases the ex-ante war probability the more, the more costly the war is and the more likely it is that conflicting parties fight when the mediator does not provide any information. The conflicting parties benefit from mediation as a costly war is less likely. Using Bayesian persuasion to model mediation stresses a strategic aspect of the information-collecting process during mediation.

1.2 Literature Review

The modelling approach in this chapter draws on Bayesian persuasion as developed by Kamenica and Gentzkow (2011). I refer the reader to the Preface for a short explanation of the work of Kamenica and Gentzkow (2011) and related literature in the field on information design. Kamenica and Gentzkow (2011) consider the case of one receiver and one sender.

Further work has extended Bayesian persuasion to games with multiple receivers.[3] Few papers address these games in a general form. (See Bergemann and Morris (2019), Mathevet et al. (2020) and Taneva (2019).) More often, specific games with multiple receivers have been analysed, a large group being voting games / games of collective decision making. Alonso and Câmara (2016) study a voting game between a sender and a group of voter with heterogeneous preferences, in which the sender can use a public signal to influence voters' choice over different policy options. They show that the sender can exploit the heterogeneity of voters' preferences by choosing a signal, which targets different winning coalitions. Whereas Alonso and Câmara (2016) focus on a public signal, Chan et al. (2019) allow for private signals in a voting game, in which voters are heterogeneous with respect to voting costs. They show that the sender benefits from using private and correlated signals. In both papers, voters share a common prior belief about the state of the world, being a single random variable. In contrast, I analyse a setting with uncertainty about two

[3]Kamenica (2019) provides a recent survey on the literature of Bayesian persuasion, covering some of the extensions.

random variables, being the conflicting parties' strengths, and each of both conflicting parties being privately informed about its own strength.[4]

Further related to the present chapter are applications of Bayesian persuasion to contests. For instance, Zhang and Zhou (2016) consider a Tullock contest, in which one conflicting party is privately informed about its valuation. The sender chooses a signal to inform the uninformed party about the private valuation of its opponent. Chen (2019) compares public and private signals in an all-pay auction with two parties with private binary valuations. The sender's signal space is restricted to unbiased signals. Feng and Lu (2016) consider a contest with random entry of parties and derives how the sender optimally informs the entered parties about the number of entered parties. In these models, the sender's objective is to maximise total expected effort (Zhang and Zhou (2016), Chen (2019) and Feng and Lu (2016)) or conflicting parties' expected payoff (Chen (2019)), whereas in the present chapter, the mediator's objective is conflict prevention.

To the best of my knowledge, only Balzer and Schneider (2019) have studied conflict prevention from an information design perspective. They consider a general class of conflicts between two conflicting parties with private information about strengths. Conflict can either be solved peacefully or escalates to war. In the case of war, payoffs are determined endogenously by actions taken by the conflicting parties. Balzer and Schneider (2019) show the equivalence between formulating these conflicts as arbitration problems and as information design problems. Finding the optimal arbitration is a mechanism design problem: Once agreed upon by the conflicting parties, arbitration prescribes payoffs in a peaceful settlement, a probability of war and an information structure for the case of war. The information design problem consists of finding the optimal information structure for the case of war. Thus, solving the information design problem is necessary for solving the arbitration problem and the authors show that it is moreover sufficient for solving the arbitration problem. While Balzer and Schneider (2019) look at a class of conflicts, I focus on a specific conflict game. This allows for an intuitive understanding of the information design problem and its solution. Moreover, I provide comparative statics which

[4]Further applications of Bayesian persuasion which are related to the present chapter look at contest models (see Zhang and Zhou (2016), Feng and Lu (2016), Chen (2019)).

illustrate the effect of mediation.

Arbitration with exogenous conflict escalation has been analysed extensively. With exogenous escalation payoffs are determined exogenously once war occurs, implying that information revealed during arbitration does not affect payoffs in case of war. Bester and Warneryd (2006) consider the conflict between two parties with asymmetric information about their strengths, where strength determines the winning probability in case of war. The authors focus on the existence of peaceful arbitration, which reduces the war probability to zero and is ex-post efficient. Bester and Warneryd (2006) show that peaceful arbitration cannot be reached if the cost of war is sufficiently low. If the cost of war is too low, strong parties have sufficiently high incentives to go to war such that no peaceful settlement can be found.

Fey and Ramsay (2009) build upon Bester and Warneryd (2006), but focus on self-enforcing mechanisms by adding an additional ex-post participation constraint. This constraint requires each suggested settlement to be agreed to by both parties. The authors analyse the case of parties with interdependent types (parties are privately informed about their strength) and that of parties with independent, but correlated types (parties are privately informed about their payoffs in case of war, which are correlated). In the case of interdependent types, a peaceful mechanism exists only if the cost of war is sufficiently high. In the case of correlated types, a peaceful mechanism does not exist.[5] Hörner et al. (2015) compare the effectiveness of self-enforcing mechanisms to mechanisms with enforcement power. They focus on a symmetric two type model with asymmetric information about parties' strengths, and find that the self-enforcing mechanism is as effective in avoiding war as a mechanism with enforcement power.

As the literature on arbitration is also interested in peaceful conflict resolution the arbitrator can be interpreted as a mediator. Modelling mediation as arbitration assumes that the mediator has means to control the payoffs scheme of a conflict. Formulating mediation as an information design problem does not rest upon this assumption, but assumes that the mediator can control the information the parties in a conflict hold. The different assumption distinguished the present chapter from the arbitration literature.

[5]See also Fey and Ramsay (2011).

The present chapter also adds to the literature addressing mediation in sender-receiver games. These games typically model the mediator as a self-interested player in a multi-stage game, who is equipped with some private information. Potentially, the use of this information allows a peaceful conflict resolution. The central question in this setting is if or under which conditions the mediator can credibly communicate with the conflicting parties. The models by Kydd (2003), Rauchhaus (2006) and Kydd (2006) stress the question of whether bias or impartiality enhance mediation success. Kydd (2003) formulates a bargaining model of war, in which one party can make a take-it-or-leave-it offer to its opponent, which is privately informed about its private cost of war. War occurs upon rejection of the offer. Absent mediation, the offer-making party makes an offer which is accepted by high cost opponents, but not by low cost opponent. Consequently, war occurs with a strictly positive probability. Kydd (2003) argues that an informed mediator needs to be biased towards the offer-making party to credibly communicate information about the opponent's cost of war. If the mediator is unbiased and only interested in preventing war, he has incentives to misreport his private information such that the offer-making party would make offers which are accepted by any opponent.

Rauchhaus (2006) partially weakens the argumentation in Kydd (2003) by showing that credible communication is also possible if the mediator is impartial. The different result stems from a different definition of the mediator's preferences. Whereas Kydd (2003) assumes that the mediator either shares the ideal point of one of both conflicting parties or is indifferent between any peaceful conflict resolution, Rauchhaus (2006) allows the mediator to have an ideal point which lies between those of the two conflicting parties.

Searching for different justifications for mediation success, Crescenzi et al. (2011) add a cost function for false reports to the mediator's payoff. The authors argue that democratic mediators incur a cost if being caught to misreport private information and that this cost can induce credible communication.

Using a multi-period model of war, Smith and Stam (2003) argue that biased mediators cannot communicate credibly. In their setting, conflicting parties disagree about their relative strengths and fight until their beliefs have sufficiently converged such that an agreement can be reached. The mediator is assumed to have some private signal about the parties' relative strengths.

While a true and honest communication of this information would shorten the war duration, a mediator who is either biased towards one party or has an interest in terminating the war cannot credibly communicate the private information to the parties.

Fey and Ramsay (2010) elaborate on the cheap-talk framework by asking how a mediator can obtain private information in the first place. They assume that the conflicting parties can reveal their private information to the mediator, who can subsequently report back to the conflicting parties. The authors argue that this form of mediation cannot outperform any direct communication between both parties. Either the conflicting parties have incentives to misreport their private information, or the mediator has incentives to misreport the information he obtains, depending on whether the mediator is biased towards one party or only interested in preventing war.

Mediation in the present chapter differs from mediation in sender-receiver games. Whereas the above models assume cheap-talk communication, I assume that the mediator commits to a signal which provides information to the conflicting parties. The commitment assumption guarantees the credibility of the provided information. The present analysis therefore focuses on the question how to optimally obtain information about conflicting parties instead of how to distribution given information.

1.3 Model Setup

Conflict Stage The game consists of a mediation stage and a subsequent conflict stage. The conflict stage follows Hörner et al. (2015). The three players of the game are the two conflicting parties $i \in \{1, 2\}$ and a mediator. Throughout, I refer to conflicting party j as i's opponent, assuming $i \neq j$. All players are risk neutral. The two conflicting parties are heterogeneous with respect to *strength* $t_i \in T = \{H, L\}$, a conflicting party with $t_i = H$ is referred to as *strong* and a conflicting party with $t_i = L$ is referred to as *weak*. The probability with which conflicting party i is strong is given by its *militarisation level* q_i.

Militarisation levels q_1 and q_2 are common knowledge among all players. Each conflicting party is privately informed about its own strength. Additionally, each conflicting party receives a private signal about its opponent's strength as described below.

The two conflicting parties dispute over a pie of unit size. The conflicting parties simultaneously take an action $a_i \in \{f, nf\}$, where f denotes *fight* and nf denotes *not fight*. If at least one conflicting party fights, war occurs. If both conflicting parties do not fight, no war occurs.

If no war occurs, each conflicting party receives a payoff of $1/2$. If war occurs, the unit pie shrinks to θ, with $\theta \in (1/2, 1)$. The smaller θ, the costlier war is. If both conflicting parties are of equal strength, each receives a payoff of $\theta/2$. If a strong conflicting party encounters a weak conflicting party, the strong one receives θ, whereas the weak one walks away empty-handed.[6] Let u_i denote the payoff, conflicting party i receives.

Note that the cost of war $(1 - \theta)$ is the same regardless of whether only one or both conflicting parties fight. Further, the conflicting parties do not incur any private cost upon fighting. Hence, if party i fights, its payoff does only depend on the opponent's type (along with its own type), but not on the opponent's action a_j.

Mediation Stage Before the conflicting parties' strengths realise, the mediator shapes the informational environment of the conflict stage by choosing a signal profile $\boldsymbol{\pi} = \{\boldsymbol{\pi_1}, \boldsymbol{\pi_2}\}$ publicly and at no cost. Signal $\boldsymbol{\pi_i}$ is addressed at conflicting party i and can provide information about j's strength. It consists of a binary realisation space $S = \{h, l\}$, with elements denoted as s_i, and a pair of conditional probability distributions $\{\pi_i(\cdot \mid t_j)\}_{t_j=H,L}$ over S. I refer to $s_i = h$ as a *high signal realisation* and to $s_i = l$ as a *low signal realisation*. Signal $\boldsymbol{\pi_i}$ prescribes, for example, that conflicting party i receives a high signal realisation with probability $\pi_i(h \mid H)$ when its opponent j is strong. Throughout and without loss of generality, I restrict attention to signals for which $\pi_i(h \mid H) \geq \pi_i(l \mid H)$ holds. Further, I restrict the analysis to a profile of uncorrelated signals, which is at no cost as will become clear at a later point. The mediator receives a payoff of (-1) if war occurs and a payoff of 0 if no war

[6] I simplify the model used in Hörner et al. (2015) by assuming that a a strong conflicting party receives θ with probability 1 when fighting against a weak conflicting party, while Hörner et al. (2015) allow this probability to take any interior value grater than $1/2$. Note that for $\theta \leq 1/2$, each conflicting party would always prefer to receive $1/2$ and never fight.

occurs. Let v denote the mediator's payoff.

Upon receiving signal realisation s_i of signal $\boldsymbol{\pi_i}$, conflicting party i forms the posterior belief $q_j^{s_i}$ about its opponent's strength. The posterior belief q_j^h is the probability with which j is strong given that i receives a high signal realisation, and the posterior belief q_j^l is the probability with which j is strong given that i receives a low signal realisation:

$$q_j^h = \text{Prob}(t_j = H \mid s_i = h) = \frac{\pi_i(h \mid H)q_j}{\pi_i(h \mid H)q_j + \pi_i(h \mid L)(1 - q_j)} \tag{1.1}$$

$$q_j^l = \text{Prob}(t_j = H \mid s_i = l) = \frac{\pi_i(l \mid H)q_j}{\pi_i(l \mid H)q_j + \pi_i(l \mid L)(1 - q_j)} \tag{1.2}$$

The posterior beliefs are formed by Bayesian updating. In equation (1.1), the denominator states the total probability with which i receives a high signal realisation. The numerator states the joint probability of i receiving a high signal realisation and j being strong, given as the prior probability of j being strong multiplied by the probability of a high signal realisation conditional on j being strong. The ratio is, thus, the posterior probability of j being strong given that i receives a high signal realisation.

Two special cases of a signal are the fully informative signal and an uninformative signal. The fully informative signal has $\pi_i(h \mid H) = 1$ and $\pi_i(h \mid L) = 0$. Upon receiving a realisation of the fully informative signal, conflicting party i perfectly learns its opponent's strength. An uninformative signal has $\pi_i(h \mid H) = \pi_i(h \mid L) \in [0, 1]$. Upon receiving the realisation of an uninformative signal, conflicting party i does not learn anything about its opponent's strength. Denote the profile of fully informative signals as $\boldsymbol{\pi^{FI}}$, and a profile of uninformative signals as $\boldsymbol{\pi^{NI}}$. A signal with $\pi_i(h \mid H) > \pi_i(h \mid L)$ is informative, meaning that conflicting party i's belief about its opponent's strength strictly increases after receiving a high signal realisation and strictly decreases after receiving a low signal realisation. The precision of signal $\boldsymbol{\pi_i}$ strictly increases in the difference between $\pi_i(h \mid H)$ and $\pi_i(h \mid L)$. The more precise signal $\boldsymbol{\pi_i}$ is, the larger the difference between the posterior beliefs q_j^h and q_j^l is.

Timing The timing is as follows:

Stage 1 The mediator publicly chooses the signal profile $\boldsymbol{\pi}$.

Stage 2 The conflicting parties' strengths t_1 and t_2 realise.

Stage 3 Signal realisations s_1 and s_2 realise according to $\boldsymbol{\pi}$, each conflicting party i privately observes the signal realisation s_i and forms the posterior belief $q_j^{s_i}$.

Stage 4 The conflicting parties take actions a_1 and a_2 and payoffs realise.

Militarisation Levels Consider the expected payoff of conflicting party i, assuming i is strong and assigns the probability q_j to its opponent being strong. Then i prefers war over no war if

$$(1 - q_j)\theta + q_j{}^{\theta}\!/\!_2 \geq {}^1\!/\!_2.$$

This condition can be reformulated to

$$\frac{q_j}{1 - q_j} \leq \frac{\theta - {}^1\!/\!_2}{{}^1\!/\!_2 - {}^{\theta}\!/\!_2}. \tag{1.3}$$

The right-hand side of inequality (1.3) states the ratio of the benefit when fighting a weak opponent to the loss when fighting a strong opponent. The left-hand side states the probability ratio of encountering a strong opponent to encountering a weak opponent. Let $\bar{q}(\theta)$ be defined such that $q_j = \bar{q}(\theta)$ solves inequality (1.3) with equality. If $q_j < \bar{q}(\theta)$, i strictly prefers war over no war. If $q_j > \bar{q}(\theta)$, i strictly prefers no war over war. I refer to militarisation level q_i as *deterrent* if $q_i \geq \bar{q}(\theta)$. If both parties have deterrent militarisation levels, war can always be avoided and mediation is needless. Therefore, I restrict the further analysis:

Assumption 1.1 (Parameter Region)**.** *At least one conflicting party does not have a deterrent militarisation level.*

1.4 Analysis

1.4.1 Strategies and Expected Payoffs

A pure strategy of conflicting party i is defined as $\phi_i : T \times S \to \{f, nf\}$. For a given signal profile $\boldsymbol{\pi}$, a Bayesian Nash equilibrium in pure strategies in the conflict stage consists of a strategy profile $\{\phi_1^*, \phi_2^*\}$ such that for all i,

$a_i = \phi_i^*(t_i, s_i)$ maximises the expected payoff $E_i[u_i(a_i, a_j, t_i, t_j) \mid t_i, s_i]$ for all t_i and s_i given ϕ_j^*.

The expected payoff $E_i[u_i(a_i, a_j, t_i, t_j) \mid t_i, s_i]$ is defined at the end of stage 3 given the strength t_i of conflicting party i and the received signal realisation s_i. Expectations are taken over the opponent's strength t_j and the opponent's equilibrium action $a_j = \phi_j^*(t_j, s_j)$ using the posterior belief $q_j^{s_i}$. If a pure strategy equilibrium $\{\phi_1^*, \phi_2^*\}$ exists in the conflict stage for a given signal profile $\boldsymbol{\pi}$, the ex-ante expected payoff of a conflicting party i can be defined at the end of stage 1 as $E[u_i(a_i, a_j, t_i, t_j) \mid \boldsymbol{\pi}, \phi_i^*, \phi_j^*]$. Ex-ante here refers to the realisation of the conflicting parties' strengths and the signals realisations and the ex-ante expected payoff is derived by taking expectations over the realisations of conflicting parties' strengths and the signal realisations. I abbreviate the notation by writing $E[u_i \mid \boldsymbol{\pi}, \phi_i^*, \phi_j^*] \equiv E[u_i(a_i, a_j, t_i, t_j) \mid \boldsymbol{\pi}, \phi_i^*, \phi_j^*]$. Additionally, if the equilibrium $\{\phi_1^*, \phi_2^*\}$ is unique given signal profile $\boldsymbol{\pi}$, I further abbreviate the notation to $E[u_i \mid \boldsymbol{\pi}]$.

The expected payoff of the mediator is denoted as $E[v(a_1, a_2) \mid \boldsymbol{\pi}, \phi_1^*, \phi_2^*]$ and equals the negative of the ex-ante war probability. Analogously, I abbreviate the notation to $E[v \mid \boldsymbol{\pi}, \phi_1^*, \phi_2^*]$, and further to $E[v \mid \boldsymbol{\pi}]$ if the equilibrium $\{\phi_1^*, \phi_2^*\}$ is unique given signal profile $\boldsymbol{\pi}$.

Two observations regarding the conflicting parties' strategies simplify the analysis.

Remark 1.1. *A weak conflicting party never strictly prefers to fight in the conflict stage, independently of the belief it holds about its opponent's strength.*

Remark 1.2. *No signal exists that induces a strong conflicting party to never fight unless its opponent has a deterrent militarisation level.*

To understand Remark 1.2, remember that a strong conflicting party strictly prefers war over no war unless the opponent has a deterrent militarisation level and that any informative signal $\boldsymbol{\pi_i}$ induces a posterior belief $q_j^l < q_j$.

I begin the analysis by considering the two benchmark cases, being a profile of uninformative signals $\boldsymbol{\pi^{NI}}$ (Section 1.4.2) and the profile of fully informative signals $\boldsymbol{\pi^{FI}}$ (Section 1.4.3). In Section 1.4.4, I derive the signal profile which maximises the mediator's ex-ante expected payoff (i.e. minimises the ex-ante war probability).

1.4.2 No Information

Assume the mediator chooses a profile of uninformative signals $\boldsymbol{\pi}^{\boldsymbol{NI}}$. As conflicting parties do not learn anything by receiving a signal realisation, $q_j^h = q_j^l = q_j$. A unique equilibrium exists in the conflict stage:

Proposition 1.1 (No Information)**.** *If the mediator does not provide any information to the conflicting parties about their opponents' strengths, the following symmetric equilibrium always exists in the conflict stage: a strong conflicting party fights and a weak conflicting party does not fight. Call this the* ***Aggressive Equilibrium****. Under Assumption 1.1 (Parameter Region), the Aggressive Equilibrium is the unique equilibrium in the no information case.*

Proof. Assume $\boldsymbol{\pi} = \boldsymbol{\pi}^{\boldsymbol{NI}}$, such that $q_j^h = q_j^l = q_j$ for $j = 1, 2$. In the Aggressive Equilibrium, the strategy of both conflicting parties is given by

$$\phi^*(t_i, s_i) = \begin{cases} f & \text{if } t_i = H \\ nf & \text{if } t_i = L, \end{cases} \tag{1.4}$$

for $s_i = l, h$. I prove (a) existence of the Aggressive Equilibrium, and (b) uniqueness of the Aggressive Equilibrium under Assumption 1.1.
(a) I show that no conflicting party has an incentive to deviate from (1.4). Assume $\phi_j = \phi^*$. First, if conflicting party i is weak, its expected payoff is given by

$$E_i[u_i(a_i, a_j, L, t_j) \mid s_i] = \begin{cases} (1 - q_j)\theta/2 & \text{for } a_i = f \\ (1 - q_j)1/2 & \text{for } a_i = nf \end{cases}$$

for $s_i = l, h$. As $E_i[u_i(nf, a_j, L, t_j) \mid s_i] > E_i[u_i(f, a_j, L, t_j) \mid s_i]$ given $\theta < 1$, choosing $a_i = nf$ is optimal. Second, if conflicting party i is strong, its expected payoff is given by

$$E_i[u_i(a_i, a_j, H, t_j) \mid s_i] = \begin{cases} (1 - q_j)\theta + q_j\theta/2 & \text{for } a_i = f \\ (1 - q_j)1/2 + q_j\theta/2 & \text{for } a_i = nf \end{cases}$$

for $s_i = l, h$. As $E_i[u_i(f, a_j, H, t_j) \mid s_i] > E_i[u_i(nf, a_j, H, t_j) \mid s_i]$ given $\theta > 1/2$, choosing $a_i = f$ is optimal. Hence, $\phi_i = \phi^*$ is optimal and no conflicting party has an incentive to deviate from strategy (1.4). As this holds independently of

the militarisation levels q_1 and q_2, the Aggressive Equilibrium always exists.
(b) I show that no further payoff-undominated equilibrium exists under Assumption 1.1
(b.1) An asymmetric equilibrium in pure strategies does not exist. Consider an asymmetric strategy profile with $\phi_i = \phi^*$ and j either never fights or always fights. It follows from part (a) of the proof that this is not an equilibrium. Further, assume i never fights and j always fights. This is not an equilibrium, as j has an incentive to deviate to nf if j is weak.
(b.2) An equilibrium in which both conflicting parties never fight does not exist under Assumption 1.1. Assume j sticks to strategy

$$\phi(t_j, s_j) = \begin{cases} nf & \text{if } t_j = H \\ nf & \text{if } t_j = L, \end{cases} \tag{1.5}$$

for $s_i = l, h$. Then, if i is strong, its the expected payoff is given by

$$E_i[u_i(a_i, a_j, H, t_j) \mid s_i] = \begin{cases} (1 - q_j)\theta + q_j \theta/2 & \text{for } a_i = f \\ 1/2 & \text{for } a_i = nf \end{cases}$$

for $s_i = l, h$. It is optimal for i to choose $a_i = nf$ if and only if $q_j \geq \bar{q}(\theta)$. By Assumption 1.1, there is at least one opponent j with $q_j < \bar{q}_j$, such that there is at least one conflicting party i for which strategy (1.5) is not optimal.[7]
(b.3) An equilibrium in which both conflicting parties always fight is payoff dominated by the Aggressive Equilibrium. The same applies to any equilibrium in non-degenerate mixed strategies. □

To understand the intuition for the Aggressive Equilibrium, consider a strong conflicting party i which expects a strong opponent j to fight and a weak opponent not to fight. Then, i can trigger war only if the opponent j is weak. If a strong conflicting party i did not fight, it would forego the potential benefit from fighting a weak opponent. Hence, a strong conflicting party i replies optimally by fighting. This reasoning holds independently of the militarisation levels. Moreover, the Aggressive Equilibrium is unique if at least

[7] If both conflicting parties have deterrent militarisation levels, an equilibrium in which both conflicting parties use strategy (1.5) exists and is moreover the payoff-dominant equilibrium.

one conflicting party strictly prefers war over no war if it is strong. This holds under Assumption 1.1.

In the Aggressive Equilibrium, war occurs if at least one conflicting party is strong. The mediator's ex-ante expected payoff for choosing an uninformative signal profile $\boldsymbol{\pi}^{\boldsymbol{NI}}$ is given by

$$E\left[v \mid \boldsymbol{\pi}^{\boldsymbol{NI}}\right] = -\left[q_1 q_2 + q_1(1-q_2) + (1-q_1)q_2\right]. \tag{1.6}$$

1.4.3 Full information

Assume the mediator chooses $\boldsymbol{\pi}^{\boldsymbol{FI}}$, the profile of fully informative signals. Each conflicting party perfectly learns its opponent's strength after receiving a signal realisation and the conflict stage is changed to a game of full information.

Proposition 1.2 (Full Information)**.** *If the mediator provides full information to the conflicting parties, the following equilibrium always exists: A strong conflicting party fights upon receiving a low signal realisation and does not fight upon receiving a high signal realisation. A weak conflicting party never fights. Call this the* ***Mediated Equilibrium****. The Mediated Equilibrium is the unique payoff undominated equilibrium in the full information case.*

A formal proof is omitted. In the full information case, a strong conflicting party conditions its action on the received signal realisation. The ex-ante war probability is reduced compared to the no information case, as war between a pair of strong conflicting parties is avoided. The mediator's ex-ante expected payoff for choosing the profile of fully informative signals $\boldsymbol{\pi}^{\boldsymbol{FI}}$ is given by

$$E\left[v \mid \boldsymbol{\pi}^{\boldsymbol{FI}}\right] = -\left[q_1(1-q_2) + (1-q_1)q_2\right]. \tag{1.7}$$

1.4.4 Optimal Information

The question emerges whether the mediator can choose a signal profile $\boldsymbol{\pi}$ to improve upon the full information case and if he can do so, which signal profile minimises the ex-ante war probability. Remark 1.1 and Remark 1.2, stated in the beginning of the analysis, help to simplify the mediator's problem: The mediator aims at inducing strong conflicting parties to choose not to fight as often as possible. The following Proposition characterises the optimal signal profile to do so together with the equilibrium it induces in the conflict stage.

Proposition 1.3 (Optimal Information). *To minimise the ex-ante war probability, the mediator provides each conflicting party strategically with information about its opponent's strength. Signals of the optimal signal profile* $\boldsymbol{\pi^*} = \{\boldsymbol{\pi_1^*}, \boldsymbol{\pi_2^*}\}$ *take the following form:* [8]

$$\pi_i^*(h \mid H) = 1 \qquad \textit{and} \qquad \pi_i^*(h \mid L) = \frac{q_j}{1-q_j}\frac{{}^1\!/_2 - {}^\theta\!/_2}{\theta - {}^1\!/_2} \qquad \textit{if } q_j < \bar{q}(\theta) \tag{1.8}$$

and

$$\pi_i^*(h \mid H) = 1 \qquad \textit{and} \qquad \pi_i^*(h \mid L) = 1 \qquad \textit{if } q_j \geq \bar{q}(\theta) \tag{1.9}$$

for $i = 1, 2$. *Given that the mediator chooses the optimal signal profile, the Mediated Equilibrium is the unique payoff undominated equilibrium in the conflict stage: a strong conflicting party fights upon receiving a low signal realisation and does not fight upon receiving a high signal realisation. A weak conflicting party never fights.*

Proof. It follows from Remark 1.2 that no equilibrium can be induced in which war never occurs and it follows from Remark 1.1 that it comes at no cost to induce weak conflicting parties to never fight. Hence, it is optimal to induce the Mediated Equilibrium and to minimise the probability with which strong conflicting parties fight under the Mediated Equilibrium. n the Mediated Equilibrium, the strategy of both conflicting parties is given by

$$\phi^*(t_i, s_i) = \begin{cases} f & \text{if } t_i = H \text{ and } s_i = l \\ nf & \text{otherwise.} \end{cases} \tag{1.10}$$

I derive (a) conditions under which a signal profile $\boldsymbol{\pi}$ induces the Mediated Equilibrium, and (b) the signal profile which minimises the ex-ante war probability in the Mediated Equilibrium subject to these conditions being satisfied. (a) It is sufficient to concentrate on the behaviour of a strong conflicting party.

[8] To see that $\pi_i^*(h \mid L) \in [0, 1]$ holds for any $q_i < \bar{q}(\theta)$, it helps to recall that $\bar{q}(\theta)$ increases with θ and $\lim_{\theta \to 1/2} \bar{q}(\theta) = 0$.

Assume $\phi_j = \phi^*$ and i is strong. First, if i receives a high signal realisation, its expected payoff for given $\boldsymbol{\pi}$ is given by

$$E_i[u_i(a_i, a_j, H, t_j)] = \begin{cases} (1 - q_j^h)\theta/2) + q_j^h \theta/2 & \text{for } a_i = f \\ 1/2 - q_j^h \pi_j(l \mid H)(1/2 - \theta/2) & \text{for } a_i = nf. \end{cases}$$

If the posterior belief q_j^h satisfies

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h \mid H)} \leq \frac{q_j^h}{1 - q_j^h}, \tag{C1$_\text{i}$}$$

$E_i[u_i(nf, a_j, H, t_j)] \geq E_i[u_i(f, a_j, H, t_j)]$ and it is optimal to choose $a_i = nf$. Second, if i receives a low signal realisation, its expected payoff for given $\boldsymbol{\pi}$ is given by

$$E_i[u_i(a_i, a_j, H, t_j)] = \begin{cases} (1 - q_j^l)\theta/2) + q_j^l \theta/2 & \text{for } a_i = f \\ 1/2 - q_j^l \pi_j(l \mid H)(1/2 - \theta/2) & \text{for } a_i = nf. \end{cases}$$

If the posterior belief q_j^l satisfies

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h \mid H)} > \frac{q_j^l}{1 - q_j^l}, \tag{C2$_\text{i}$}$$

$E_i[u_i(f, a_j, H, t_j)] > E_i[u_i(nf, a_j, H, t_j)]$ and it is optimal to choose $a_i = f$. Thus, if the posterior beliefs q_j^h and q_j^l satisfy (C1$_\text{i}$) and (C2$_\text{i}$), $\phi_i = \phi^*$ is the optimal reply to $\phi_j = \phi^*$. A signal profile $\boldsymbol{\pi}$ induces the Mediated Equilibrium if signal $\boldsymbol{\pi_i}$ induces posterior beliefs q_j^h and q_j^l satisfying (C1$_\text{i}$) and (C2$_\text{i}$) for $i = 1, 2$. Substituting for q_j^h and q_j^l, the constraints (C1$_\text{i}$) and (C2$_\text{i}$) can be restated as

$$\frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h \mid H)} \leq \frac{q_j}{1 - q_j} \frac{\pi_i(h \mid H)}{\pi_i(h \mid L)} \tag{C1$_\text{i}$}$$

$$\text{and } \frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h \mid H)} > \frac{q_j}{1 - q_j} \frac{\pi_i(l \mid H)}{\pi_i(l \mid L)}. \tag{C2$_\text{i}$}$$

(b) The mediator chooses the signal profile $\boldsymbol{\pi}$ to maximise the expected payoff $E[v \mid \boldsymbol{\pi}]$ under the Mediated Equilibrium, subject to (C1$_\text{i}$) and (C2$_\text{i}$) being satisfied for $i = 1, 2$. The choice of a signal profile $\boldsymbol{\pi}$ corresponds to that of the four probabilities $\left\{\{\pi_i(h \mid t_j)\}_{t_j=H,L}\right\}_{i=1,2}$. The mediator's objective function can be equivalently formulated as the ex-ante probability of no war:

$$\begin{aligned} E[v \mid \boldsymbol{\pi}] &= q_1 q_2 \pi_1(h|H)\pi_2(h|H) + q_1(1-q_2)\pi_1(h|L) \\ &+ (1-q_1)q_2\pi_2(h|L) + (1-q_1)(1-q_2) \end{aligned}$$

The first term corresponds to the case that both conflicting parties are strong and receive a high signal realisation each. The second and the third term each corresponds to the case that exactly one conflicting party is strong and receives a high signal realisation. The fourth term corresponds to the case that both conflicting parties are weak and is constant in $\boldsymbol{\pi}$. The mediator's maximisation problem can be written as

$$\begin{aligned} \underset{\left\{\{\pi_i(h|t_j)\}_{t_j=H,L}\right\}_{i=1,2}}{Max} \quad & E[v \mid \boldsymbol{\pi}] \\ \text{s.t.} \quad & \frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h \mid H)} \leq \frac{q_j}{1-q_j}\frac{\pi_i(h \mid H)}{\pi_i(h \mid L)} && (\mathrm{C1_i}) \\ & \frac{\theta - 1/2}{(1/2 - \theta/2)\pi_j(h \mid H)} > \frac{q_j}{1-q_j}\frac{\pi_i(l \mid H)}{\pi_i(l \mid L)} && (\mathrm{C2_i}) \\ & \{\pi_i(h \mid t_j)\}_{t_j=H,L} \in [0,1] \\ & \text{for } i = 1,2. \end{aligned}$$

The objective function increases in $\pi_i(h \mid H)$ and $\pi_i(h \mid L)$ for $i = 1, 2$. Assume $\pi_i(h \mid H) = 1$ for $i = 1, 2$. The constraints reduce to

$$\begin{aligned} \frac{\theta - 1/2}{(1/2 - \theta/2)} &\leq \frac{q_j}{1-q_j}\frac{1}{\pi_i(h \mid L)} && (\mathrm{C1_i'}) \\ \frac{\theta - 1/2}{(1/2 - \theta/2)} &> 0. && (\mathrm{C2_i'}) \end{aligned}$$

($\mathrm{C2_i'}$) holds trivially. If $q_j \geq \bar{q}(\theta)$, ($\mathrm{C1_i'}$) holds for any $\pi_i(h \mid L)$ and it is optimal to choose $\pi_i(h \mid L) = 1$. If $q_j < \bar{q}(\theta)$, it is optimal to choose $\pi_i(h \mid L)$ such that ($\mathrm{C1_i'}$) binds:

$$\pi_i(h \mid L) = \frac{q_j}{1-q_j}\frac{1/2 - \theta/2}{\theta - 1/2}.$$

Now consider whether $\pi_i(h \mid H) = 1$ is optimal. $\pi_i(h \mid H) = 1$ is optimal unless choosing $\pi_i(h \mid H) < 1$ allows to increase $\pi_i(h \mid L)$ or $\pi_j(h \mid L)$. However, decreasing $\pi_i(h \mid H)$ decreases the right hand side of ($\mathrm{C1_i}$) and increases the left hand side of ($\mathrm{C1_j}$). This implies that if $\pi_i(h \mid H)$ was reduced, $\pi_i(h \mid L)$ and $\pi_j(h \mid L)$ would need to decrease to satisfy ($\mathrm{C1_i}$) and ($\mathrm{C1_j}$). Thus, $\pi_i(h \mid H) = 1$ for $i = 1, 2$ is optimal. To conclude, optimal signals are given by (1.8) and (1.9). □

As in the full information case, a strong conflicting party conditions its action on the received signal realisation in the optimal information case. Different to the full information case, a high signal realisation is at most indicative of a strong opponent. The mediator improves upon the full information case by choosing a signal which sometimes has a high realisation although the opponent is weak.

The optimal signal takes one of two forms, depending on the opponent's militarisation level. If the opponent has a deterrent militarisation level, the signal does not need to provide any information to convince a strong conflicting party not to fight. In this case, the uninformative signal given by (1.9) is optimal such that the conflicting party always receives a high signal realisation and never fights.

If the opponent does not have a deterrent militarisation level, the signal does need to provide information to convince a strong conflicting party not to fight. In this case, the signal given by (1.8) is optimal. The signal has always a high realisation if the opponent is strong and sometimes does so if the opponent is weak. To maximise the occurrence of high signal realisations, the probability with which the signal has a high realisation if the opponent is weak is chosen such that a strong conflicting party is indifferent between fighting and not fighting upon receiving a high signal realisation.

The more incentive a strong conflicting party has to fight without receiving any further information about the opponent's strength, the more information is necessary to convince it not to fight. For this reason, the probability $\pi_i(h \mid L)$ decreases from 1 to 0 and the signal $\boldsymbol{\pi_i}$ becomes more precise as the opponent's militarisation level decreases from $\bar{q}(\theta)$ to 0. When the opponent's militarisation level q_j is close to the deterrent threshold $\bar{q}(\theta)$, a strong conflicting party has little incentive to fight. Little information is therefore necessary to convince it not to fight and an imprecise signal with a high probability $\pi_i(h \mid L)$ is optimal. As q_j decreases, a strong conflicting party has more incentive to fight and more information is necessary to convince it not to fight. The probability $\pi_i(h \mid L)$ decreases such that the signal becomes more precise.

Parallel reasoning applies regarding θ. The probability $\pi_i(h \mid L)$ decreases from 1 to 0 and the signal $\boldsymbol{\pi_i}$ becomes more precise as θ increases from $1/2$ to 1. For θ close to $1/2$, a strong conflicting party i has little incentive to fight and little information is necessary to convince it not to fight. As θ increases, a strong conflicting party has more incentive to fight and a more precise signal is necessary to convince it not to fight.

Note that the mediator could not improve upon the obtained results by choosing correlated signals. The reason why the two signals can be chosen independently is twofold: As noted above, the payoff of a strong conflicting party who chooses to fight is independent of its opponent's action. Second, by choosing the two uncorrelated signals optimally, the mediator can reduce the probability of war when both conflicting parties are strong to zero.

Given the mediator chooses the optimal signal profile and war occurs in the Mediated Equilibrium if exactly one conflicting party is strong and receives a low signal realisation. The mediator's expected payoff for choosing the optimal signal profile $\boldsymbol{\pi}^*$ is given by

$$E[v \mid \boldsymbol{\pi}^*] = \begin{cases} -[q_1(1-q_2)\pi_1^*(l \mid L) + q_2(1-q_1)\pi_2^*(l \mid L)] & \text{if } q_1, q_2 < \bar{q}(\theta) \\ -[q_1(1-q_2)\pi_1^*(l \mid L)] & \text{if } q_2 < \bar{q}(\theta) \leq q_1 \\ -[q_2(1-q_1)\pi_2^*(l \mid L)] & \text{if } q_1 < \bar{q}(\theta) \leq q_2. \end{cases} \tag{1.11}$$

1.5 Effect of Mediation

1.5.1 War Probability

I discuss how optimal information affects the ex-ante war probability (Sections 1.5.1 and 1.5.2) and the conflicting parties' ex-ante expected payoff (Section 1.5.3). For ease of exposition, I restrict the discussion by assuming that *no* conflicting party has a deterrent militarisation level.

Optimal information reduces the ex-ante war probability compared to the no information case and the full information case: No war occurs between two strong conflicting parties and the probability of war between a strong and a weak conflicting party is strictly less than one. Figure 1.1 depicts the ex-ante war probability in the no information case, the full information case and in the optimal information case for $\theta = 0.8$ assuming a symmetric militarisation level q. In the no information case, the ex-ante war probability strictly increases in the militarisation level q up to the deterrent threshold $\bar{q}(0.8)$. As q increases, so does the probability of at least one conflicting party being strong which fights in the no information case. Once the militarisation levels reach $\bar{q}(0.8)$, war can always be avoided in the no information case. In the full information case, war occurs if exactly one conflicting party is strong. The probability of this being the case is concave in q, reaching its maximum at $q = 1/2$.[9]

Optimal information improves further upon full information. In the optimal information case, war occurs if exactly one conflicting party is strong, but does so with a probability strictly less than one. Figure 1.1 shows that the difference between the ex-ante war probability in the full information case and in the optimal information case increases in q. As the opponent's militarisation level increases, less information is necessary to convince a strong conflicting party not to fight. The optimal signal is less precise, implying a higher probability of high signal realisations. When q is close to $\bar{q}(0.8)$, the probability with which the optimal signals have a high realisation is close to one and war can almost always be avoided. Once militarisation levels reach $\bar{q}(\theta)$, the mediator can use uninformative signals, such that both conflicting parties always receive high signal realisations and never fight.

1.5.2 Mediation Success

Let $\Delta E[v \mid \boldsymbol{\pi}^*]$ be defined as the increase in the mediator's expected payoff over the no information case achieved by choosing the optimal signal profile:

$$\Delta E[v \mid \boldsymbol{\pi}^*] \equiv E[v \mid \boldsymbol{\pi}^*] - E[v \mid \boldsymbol{\pi}^{NI}]$$

[9] It can be seen that full information provokes war when there is no war in the no information case if both conflicting parties have deterrent militarisation levels. However, this case is not the focus of this chapter and was ruled out by Assumption 1.1.

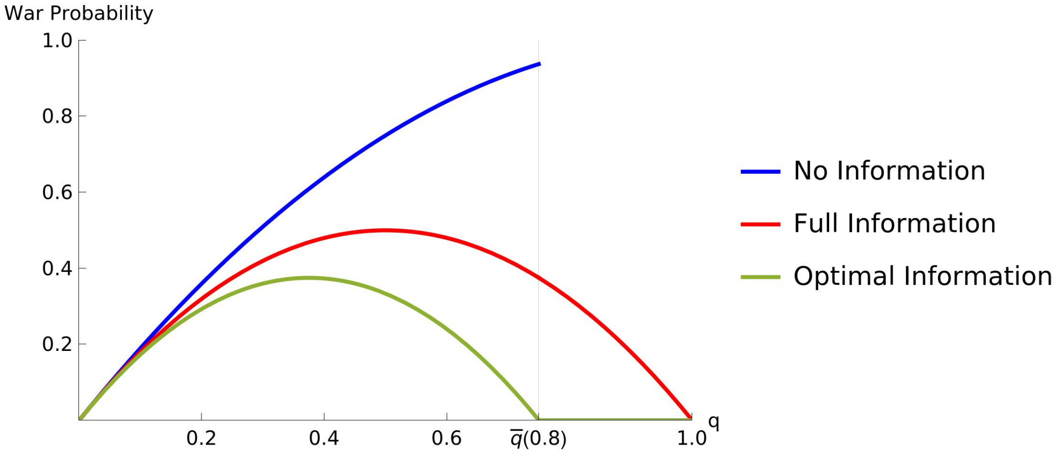


Figure 1.1: The ex-ante war probability for no information, full information and optimal information depicted as a function of a symmetric militarisation level q for $\theta = 0.8$. In the no information case, the ex-ante war probability strictly increases in the militarisation level q. In the optimal information case, the ex-ante war probability increases for low militarisation levels and decreases for high militarisation levels. For $q \geq \bar{q}(0.8)$, the ex-ante war probability is zero in the no information case and in the optimal information case.

I refer to $\Delta E[v \mid \boldsymbol{\pi}^*]$ as *mediation success*, meaning that the higher $\Delta E[v \mid \boldsymbol{\pi}^*]$, the more successful mediation is. Assuming $q_i < \bar{q}(\theta)$ for $i = 1, 2$,

$$\Delta E[v \mid \boldsymbol{\pi}^*] = q_1 q_2 \frac{1/2}{\theta - 1/2}. \tag{1.12}$$

Mediation success increases in militarisation levels q_1 and q_2 and decreases in θ. The relationship between militarisation levels and mediation success follows directly from the discussion in Section 1.5.1. The negative relationship between θ and mediation success stems from the effect an increase in θ has on the optimal signals. As θ increases, more information about the opponent's strength is needed to convince a strong conflicting party not to fight. The optimal signals get more precise and the probability of high signal realisations decreases, implying a higher probability of war between a strong and a weak conflicting party. Thus, mediation success increases faster in the militarisation level q the lower θ is, as can be seen in Figure 1.2, which also shows that the parameter range of q, for which mediation is relevant, is the smaller the smaller θ is.

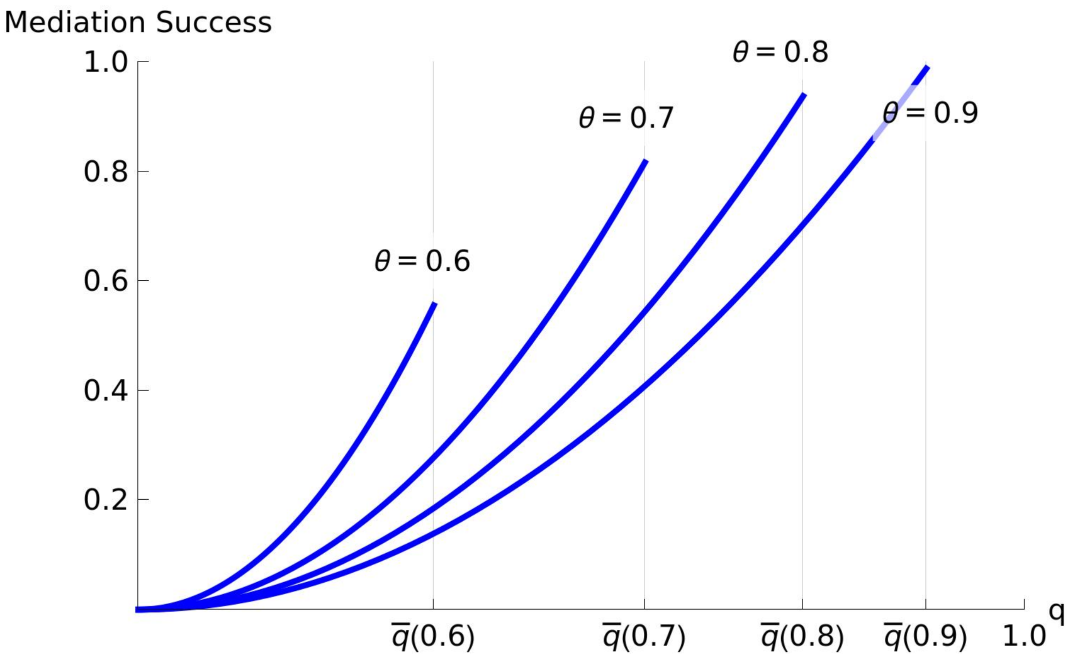


Figure 1.2: Mediation success depicted as a function of the militarisation level q for different values of θ, assuming a symmetric militarisation level q. Mediation success increases with q and is the steeper, the costlier war is.

1.5.3 Conflicting Parties' Benefit

Each conflicting party benefits from optimal information. Let $\Delta E[u_i \mid \boldsymbol{\pi}^*]$ be defined as the increase in conflicting party i's ex-ante expected payoff in the optimal information case compared to the no information case:

$$\Delta E[u_i \mid \boldsymbol{\pi}^*] \equiv E[u_i \mid \boldsymbol{\pi}^*] - E[u_i \mid \boldsymbol{\pi}^{NI}].$$

I refer to $\Delta E[u_i \mid \boldsymbol{\pi}^*]$ as conflicting party i's *mediation benefit.* Assuming $q_i < \bar{q}(\theta)$ for $i = 1, 2$,

$$\Delta E[u_i \mid \boldsymbol{\pi}^*] = q_j q_i \frac{1/2 - \theta/2}{\theta - 1/2} 1/2. \tag{1.13}$$

The mediation benefit of conflicting party i increases in i's militarisation level q_i, in the opponent's militarisation level q_j and decreases in θ. Conflicting party i's own militarisation level has two countervailing effects on i's mediation benefit. The probability with which i is weak and benefits from optimal information decreases in q_i. This effect is negative. On the other hand, the extent to which a weak conflicting party i benefits from optimal information increases in q_i, as the higher q_i, the less precise the signal $\boldsymbol{\pi}_j^*$ is. The second effect dominates, implying an overall positive relationship between i's militarisation level q_i and i's mediation benefit.

The relationship between the opponent's militarisation level and i's mediation benefit is more straightforward. As q_j increases, so does the probability with which i faces a strong opponent and can, therefore, benefit from optimal information.

Lastly, the relationship between θ and the mediation benefit is negative, because the greater θ is, the more precise optimal signals are and the less often the opponent receives high signal realisations.

Decomposing the mediation benefit by the realisation of conflicting party i's strength reveals that the whole benefit is taken up by weak conflicting parties:

$$\Delta E[u_i \mid \boldsymbol{\pi}^*] = (1-q_i)\left[q_j \underbrace{\frac{q_i}{1-q_i}\frac{1/2-\theta/2}{\theta-1/2}}_{\pi_j^*(h|L)} 1/2\right] + q_i 0. \tag{1.14}$$

Strong conflicting parties are as good off in the optimal information case as in the no information case. This is intuitive: if a strong conflicting party benefited from optimal information, the ex-ante war probability could be reduced further.

1.6 Conclusion

Uncertainty and informational asymmetries are important factors to understand conflict and conflict resolution. In this chapter, I use Bayesian persuasion (see Kamenica and Gentzkow (2011)) to show how a mediator can reduce the ex-ante war probability by providing conflicting parties strategically with information about the respective opponent's strength. When a conflicting party receives information indicating that it faces a strong opponent, it refrains from fighting. On the downside, if a conflicting party receives information indicating a weak opponent, it fights. I derive conditions under which the provided information has this effect and how the mediator uses this effect optimally. The effective use of information to prevent war rests on the assumption that the mediator obtains the information independently and commits on sharing this information with the conflicting parties. Mediation in this form is the more successful, the higher the ex-ante war probability is absent mediation and the costlier war is.

Chapter 2

Estimating Income in a Tax Compliance Game: A Bayesian Persuasion Approach

2.1 Introduction

Tax compliance is a major concern for tax authorities for at least two reasons: first, if tax payers comply with the tax schedule, tax revenues are high. Second, tax compliance is necessary so that the tax schedule implements its intended net income distribution. To induce tax compliance, tax authorities use tax audits, whereby a fraction of tax reports is audited to check whether tax payers comply with the tax law or try to evade taxes. A tax payer caught under-reporting his income needs to pay the evaded tax and an additional fine, which intends to deter tax payers from evading taxes: tax payers should prefer to state correct tax reports to avoid the risk of being audited and fined.

The choice of *which* tax reports to audit is important so that tax audits achieve their intended goals while being cost efficient. As tax audits are costly, auditing all tax reports is generally not cost efficient. Usually, the tax authority has information about tax payers which allows to differentiate between tax reports with higher and lower risk of fraud. Tax authorities have employed computerised technologies to make use of this information and to decide whether or not to audit a specific tax report. For instance, the United States' Internal Revenue Service (IRS) uses the Discriminant Index Function (DIF) to score tax reports according to their audit potential. A high DIF score indicates a high audit potential, meaning that an audit is likely to generate high additional tax revenue. Tax collectors conduct audits among reports with a high DIF score. While this practice is publicly known, it remains a black box for tax payer how that score was generated. The DIF itself is designed at a prior stage. More recently, the German tax authority adopted a similar procedure. An automated risk management system (RMS) classifies tax reports according to the risk of tax fraud. Only reports within the highest risk class are later checked manually by tax collectors.[1]

While computerised technologies, such as the DIF or the German RMS, facilitate tax collectors' choice of whether to audit a specific tax report, they shift the problem to a prior stage: How to design the technology which later categorises tax reports? As this technology is decisive for whether the tax audits are targeted wisely and increase tax compliance, its design is not a trivial task. This Chapter addresses this design question, by assuming that the tax authority can choose a technology to estimate income. Importantly, the technology is chosen prior to tax reports being made, similar as the DIF or the German RMS are chosen before tax payers report their income. At a later stage, the technology is used to decide whether or not to audit a given tax report. The analysis shows that the tax authority benefits from choosing the technology strategically. Even if income can be estimated arbitrarily precisely, a biased technology, which overestimates low income, can be optimal.

The analysis differs from previous literature, which analyses how limited and given information about tax payers' income can be used to select tax reports for audits. This literature typically assumes that the tax authority observes an exogenous signal about tax payers' income, and that it commits to an audit strategy before tax payers make their tax reports. I deviate from this literature, as I ask how the tax authority should estimate income in the first place. I do so by endogenising the signal which provides information about a tax payer's income. The tax authority's choice of the estimation technology is that of a signal. Further, I do not assume that the tax authority can commit to an audit strategy before the tax payer makes his tax report.

[1]A description of the DIF can be found in the Internal Revenue Manual, Chapter 4.1.2 of the IRS. Information, in German, about the RMS is provided here: `https://www.steuer-it-konsens.de/auf-einen-blick-die-verfahren/`.

The underlying assumption of this chapter is that the tax authority can in principle estimate income perfectly. As this assumption appears to be stark, it desires some justification. Canonical sources of information about tax payers are their tax reports and mandatory third party reports, such as employers' reports about their employees' salaries. Further information is provided by observable characteristics, such as a tax payer's address or his profession. New and abundant sources of information are becoming available with the enhancing digitalisation. For instance, credit card transactions allow to infer the tax payer's consumption behaviour; online adverts for flats allow to observe a landlord's rental income; and social media activity signals the user's living standards.[2] Thus, a scenario in which sufficient information is available to estimate income perfectly becomes more and more realistic.

Against this background, a shift of attention is interesting: instead of asking how a limited amount of information about tax payers is used optimally (as the literature has done so far), it is worthwhile to ask how much of the abundant information should be used and how this should be done. My analysis shows that it might not be optimally to use all available information.

I set up the analysis in a three-stage game between the tax authority and the tax payer. The tax payer is privately informed about his income, which is high or low. In the first stage, the tax authority chooses a technology to estimate income. This choice corresponds to the IRS' choice of the DIF or the German tax authority's choice of the RMS. In the second stage, the tax payer makes a tax report. He either reports his true income or he makes no tax report. A false report is not possible. If the tax payer makes no tax report, the tax authority observes an income estimate generated by the previously chosen estimation technology. Lastly, the tax authority chooses whether to audit the tax payer at a fix audit cost. If the tax payer does not report his income, but is audited by the tax authority, he pays a fine additional to the tax duty. Fine and taxes collected pay off the audit cost if and only if the tax payer's income is high. The estimation technology is modelled using the Bayesian persuasion framework developed by Kamenica and Gentzkow (2011). Having chosen the

[2]It has been recognised that the digitalisation increases tax authorities' informational resources (See for instance Hatfield, 2015). Tax authorities have started to use online data ore are planning to do so, for instance in France (BBC News, 2019), in the UK (Houlder, 2017), and in the United States (Houser and Sanders, 2016).

estimation technology, the tax authority uses the observed income estimate as an indication of whether the tax payer's income is high or low.

Absent the estimation technology, the tax authority suffers a commitment problem in this setting. If the tax authority could commit to audit the tax payer once he does not report his income, the tax payer would always prefer to report his income to avoid being audited and fined. However, a commitment is not credible. The taxes and fines collected in an audit pay off the audit cost only if the tax payer's income is high, and on average, they do not pay it off. I show how the tax authority can use the estimation technology to solve this commitment problem. The tax authority can choose an estimation technology, which overestimates low income: the optimal estimation technology always indicates income to be high when it is high and it does so with a strictly positive probability when income is low. The probability with which low income is indicated to be high is chosen, such that the tax authority prefers to audit the tax payer when the estimation technology indicates high income. As a result, the tax authority would always audit a high income tax payer and with some probability it would audit a low income tax payer. On the other hand, the tax authority learns income to be low after observing a low estimate, in which case it does not audit the tax payer. The estimation technology solves the commitment problem, as it allows the tax authority to commit to audit a tax payer with low income with a strictly positive probability. Anticipating this audit behaviour, the tax payer prefers to report his income independently of whether it is high or low. In equilibrium, the tax payer always reports his income and the tax authority never needs to audit the tax payer.

2.2 Literature Review

This chapter adds to the game-theoretic literature on tax compliance, which studies the strategic interaction between tax payers and the tax authority. This literature[3] typically studies variations of the following setting: tax payers are privately informed about their income and choose how much income to report to the tax authority. The tax authority chooses an audit strategy which

[3] Andreoni et al. (1998), Sandmo (2005) and Alm (2019) provide informative reviews on the literature on tax compliance.

consists of an audit probability contingent on the tax report. In some cases, the tax authority further chooses tax and fine rates. Two approaches can be distinguished within this literature, depending on whether the tax authority can commit to an audit strategy before tax payers report their income.[4]

Important contributions assuming commitment on side of the tax authority are Reinganum and Wilde (1985), Border and Sobel (1987), Mookherjee and Png (1989) and Sánchez and Sobel (1993). Reinganum and Wilde (1985) set up a mechanism design problem in which the tax authority chooses lump-sum taxes and fines and an audit strategy. The authors show that the use of a cut-off rule for audits weakly dominates random audits which are not contingent on a tax report. Using a cut-off rule, the tax authority audits all tax payers who report an income below the cut-off, but does not audit tax payers who report an income above the cut-off. In equilibrium, tax payers with low income report honestly, while those with high income report an income equal to the cut-off. Thus, only truth telling, low income tax payers are audited. Border and Sobel (1987) generalise this model by allowing for varying tax and fine rates and find that it is optimal to have taxes increasing in reported income and the audit probability decreasing in reported income. Mookherjee and Png (1989) show that an audit strategy using random audits always dominates one using deterministic audits if the tax payer is risk averse. Sánchez and Sobel (1993) separate the problem of finding an optimal audit strategy from that of setting taxes and fines. In a first stage, the government decides on taxes, fines and a budget for tax enforcement. Subsequently, the tax authority chooses and commits to an audit strategy.

When the tax authority cannot commit to an audit strategy, a tax payer's tax report serves as a signal about his true income. The signalling structure can lead to many equilibria, as reported by Reinganum and Wilde (1986a). Reinganum and Wilde (1986a) focus their analysis on a separating equilibrium, in which the tax authority perfectly learns a tax payer's true income by observing the tax report. In this equilibrium, absolute under-reporting declines with income and the audit probability decreases with reported income. The latter incentivises high income tax payers to make a higher tax report. In this

[4]Earlier approaches investigate how tax payers reporting behaviour is affected by the audit probability and tax and fine rates assuming these to be exogenous(Allingham and Sandmo, 1972; Yitzhaki, 1974).

separating equilibrium, the tax authority does not audit to learn a tax payer's true income, but to enforce the payment of the total amount of due taxes and fines.[5]

Regarding the commitment assumption, the present chapter undertakes the second approach and assumes that the tax authority cannot commit to an audit strategy. This chapter differs from all of the contributions mentioned so far, as it assumes that the tax authority observes an exogenous signal about a tax payer's income.

The use of exogenous signals about tax payers' income has been considered by Scotchmer (1987) and Macho-Stadler and Perez-Castrillo (2002). In Scotchmer (1987), the tax authority observes an exogenous signal, which is correlated with tax payers' income. The signal is used to sort tax payers into different income classes. Within each income class, the optimal audit strategy takes the form of a cut-off rule. The focus of this analysis is to determine the effect of an exogenous signal about tax payers' income on the distributional characteristics of the effective tax scheme. Within each income class, the use of a cut-off rule leads to a regressive bias in the effective tax scheme, as high income tax payers pay proportionally less taxes. Sorting tax payers into income classes countervails this bias and renders the effective tax scheme more progressive. Macho-Stadler and Perez-Castrillo (2002) differ from Scotchmer (1987) by considering a setting with discrete income levels and by assuming that the tax payer knows which signal is observed by the tax authority. The results do not change qualitatively.

Kuchumova (2017) adds to a model similar to Scotchmer (1987) a first stage in which the tax authority can costly increase the precision of the signal about tax payers' income. Thereby, the author models a trade-off between costly information acquisition and costly audits: a fix budget can be used to increase the signal's precision or to increase audit probabilities. Kuchumova (2017) shows that the value of an increase in the signal's precision depends on the audit probability. This leads to a non-monotonic relationship between the budget and the optimal level of signal precision. The optimal level of precision initially rises with the budget, but decreases with the budget when this is high,

[5] See also Reinganum and Wilde (1986b) for a related model. Graetz et al. (1986) and Erard and Feinstein (1994) incorporate a fraction of honest tax payers into models in which the tax authority cannot commit to an audit strategy.

because a high budget allows for high audit probabilities.

In a different framework, Sansing (1993) also looks at the optimal choice of a signal's precision by the tax authority. In this model, tax payers' income is constant and common knowledge, but tax payers have private information regarding whether they are entitled to claim a tax deduction. The tax authority observes a signal about whether the tax payer is entitled to claim the deduction. Sansing (1993) shows that an increase in the signal's precision can have a negative effect on tax compliance and on the tax authority's audit costs. Allowing the tax authority to costly invest in the signal's precision, Sansing (1993) further shows that the optimal level of investment varies non-monotonically with the model's parameters.

The present chapter differs from Scotchmer (1987) and Macho-Stadler and Perez-Castrillo (2002) by focusing on optimal information acquisition. It further differs from Kuchumova (2017) and Sansing (1993) by focusing on the strategic use of information and by assuming information acquisition at not cost.

Loosely related to the present chapter, is work which takes into account that the tax authority can use a tax payer's consumption decision to infer his income. This idea was taken up by Yaniv (2013) and Bronsert (2016). In Yaniv (2013), the audit probability is conditioned on a tax payer's income report and on his consumption of an observable, conspicuous good. Yaniv (2013) does not derive the optimal audit strategy, but analyses effects of changes in the exogenously given audit strategy on tax evasion and consumption behaviour. Bronsert (2016) derives the optimal audit strategy given that the tax authority can infer a tax payer's income from his consumption of a conspicuous good. Both papers take into account that the tax authority can use information beyond a tax payer's income report to target audits. As consumption is chosen by the tax payer, the information is endogenous in both papers, whereas I consider an exogenous signal about income.

In the present chapter, an exogenous signal allows the tax authority to commit to audit tax payers who do not report their income. Alternative commitment strategies have been suggested: Melumad and Mookherjee (1989) show that the government can delegate audits to the tax authority to achieve commitment. Finkle and Shin (2007) argue that the tax authority can choose audits to be inaccurate to set ex-post incentives to conduct audits. Using inaccurate audits, the tax authority collects a fine if it correctly detects an under-reporting, high income tax payer or if the audit wrongly suggests that a truth-telling, low income tax payer is under-reporting.

I model information acquisition using the Bayesian persuasion framework developed in Kamenica and Gentzkow (2011). Bayesian persuasion belongs to the growing literature on information design which addressed the question, how one sender or multiple senders can generate information strategically to induce certain equilibrium behaviour among a set of receivers. Kamenica and Gentzkow (2011) address this question in a setting with one sender and one receiver. For a short explanation of this framework, I refer the reader to the Preface. Bayesian persuasion has been extended into several directions and applied to different contexts.[6] Applications typically consider settings in which sender and receiver are distinct from each other. In this Chapter, sender and receiver of the signal are identical: the tax authority chooses the signal in a prior stage and observes its realisation in a later stage. To the best of my knowledge, none of the applications of Bayesian persuasion features a similar structure.

2.3 Model

2.3.1 Setup

The game consists of three stages and has two players, being the tax authority and the tax payer. In Stage 0, the tax authority chooses an estimation technology which allows to estimate the tax payer's income at a later stage. In Stages 1 and 2, the tax payer and the tax authority interact in a tax compliance game. I begin by describing the tax compliance game abstracting from the estimation technology.

Tax Compliance Game The tax payer has income $\theta \in \{H, L\}$, where H denotes high income and L denotes low income with $0 < L < H$. Income θ is

[6] Kamenica (2019) provides a recent survey on the literature of Bayesian persuasion, covering some of the extensions.

high with probability $q \in (0,1)$ and private information to the tax payer. The probability q is common knowledge.

The tax payer either reports his income θ or he reports no income. This choice is denoted by $r \in \{0,1\}$, where $r = 1$ refers to the choice to report income θ. If the tax payer reports his income, the tax authority observes θ. If the tax payer does not report his income, the tax authority does not observe θ unless it conducts an audit. The tax authority chooses whether to audit the tax payer by taking action $a \in \{0,1\}$, where $a = 1$ refers to the choice to conduct an audit. If the tax authority audits the tax payer, it observes θ.

The tax payer owes taxes $t\theta$ to the tax authority, where the linear tax rate t satisfies $0 < t < 1$. If the tax payer reports his income, he pays taxes $t\theta$ to the tax authority and is left with a payoff equal to his net income $(1-t)\theta$, while the tax authority receives the tax revenue $t\theta$ as payoff. If the tax payer does not report his income, but the tax authority conducts an audit, taxes owed are multiplied by the fine rate f and the tax payer's payoff equals $(1-ft)\theta$, where $f > 1$. The tax authority pays an audit cost c and receives a payoff of $ft\theta - c$, where $c > 0$. If neither the tax payer reports his income nor the tax authority conducts an audit, the tax payer is left with income θ, while the tax authority has zero tax revenue.[7] The tax rate t, the fine rate f, and the audit cost c are exogenously given. The tax payer and the tax authority are both risk neutral.

Estimation Technology Before the tax payer's income θ realises, the tax authority chooses the estimation technology π which allows to estimate income θ. The estimation technology π consists of a binary realisation space $S = \{h, l\}$ and a pair of conditional probability distributions $\{\pi(\cdot \mid \theta)\}_{\theta=H,L}$ over S. Elements of the realisation space S are denoted by s and referred to as high $(s = h)$ and low $(s = l)$ estimates. The pair of conditional probability distributions describes with which probability the estimation technology has high and low estimates given that income θ is high or low. For instance, estimation technology π has a high estimate with probability $\pi(h \mid H)$ when income θ is high. The tax authority chooses the estimation technology π publicly and at no cost. Throughout and without loss of generality, I restrict attention to the case $\pi(h \mid L) \leq \pi(h \mid H)$.

[7] There is no need to consider the case that the tax authority conducts an audit after the tax payer reports his income, as the tax authority does not have any incentive to do so.

Timing The timing of the complete game is as follows:

Stage 0 The tax authority chooses estimation technology π publicly and at no cost.

Stage 1 The tax payer's income θ realises. The tax payer takes action $r \in \{0, 1\}$. If $r = 1$, the tax payer pays taxes $t\theta$ to the tax authority and the game ends. If $r = 0$, the game proceeds to Stage 2.

Stage 2 Estimate s realises. The tax authority forms a posterior belief $\hat{q}^s$ about the tax payer's income and takes action $a \in \{0, 1\}$. If $a = 1$, the tax payer pays taxes and fines $ft\theta$ to the tax authority, the tax authority pays audit cost c and the game ends. If $a = 0$, no payments are made and the game ends.

2.3.2 Cost Parameter

I make two assumptions regarding the relative size of the audit cost c to exclude trivial cases:

Assumption 2.1. $ftL < c < ftH$

Assumption 2.2. $qftH + (1 - q)ftL < c$

Under Assumption 2.1, the net tax revenue after conducting an audit is positive if and only if the tax payer's income is high.

Under Assumption 2.2, the net tax revenue after conducting an audit is negative in expectation when the tax authority assigns probability q to income being high. Given Assumption 2.1, Assumption 2.2 poses a restrictions on the income distribution: the probability q with which income is high needs to be sufficiently low to satisfy Assumption 2.

To see that Assumptions 2.1 and 2.2 exclude trivial cases, consider the tax compliance game abstracting from the estimation technology π. If the second inequality of Assumption 2.1 was violated, an audit would never pay off the audit cost and the tax authority would optimally never audit the tax payer. If Assumption 2.2 was violated, the tax authority could credibly commit to always audit the tax payer.

2.4 Analysis

2.4.1 Definitions

A pure strategy of the tax payer is defined as $\rho : \{H, L\} \to \{0, 1\}$. A pure strategy of the tax authority is defined as $\alpha : \{h, l\} \to \{0, 1\}$, taking the estimate s as its argument. I restrict the analysis to pure strategies. The tax authority's belief about ρ is denoted by $\hat{\rho}$. The tax payer's belief about α is denoted by $\hat{\alpha}$. The tax authority's posterior belief about q after observing action $r = 0$ is denoted by $\hat{q}$ and that after observing action $r = 0$ and the estimate s is denoted by $\hat{q}^s$. Given estimation technology π, a perfect Bayesian Nash equilibrium of the game consists of a strategy profile (ρ, α) and a set of beliefs $\{\hat{\rho}, \hat{\alpha}, \hat{q}^h, \hat{q}^l\}$ such that

a) $\rho(\theta)$ maximises the tax payer's expected payoff given $\hat{\alpha}$ for any θ,

b) $\alpha(s)$ maximises the tax authority's expected payoff given $\hat{q}^s$ for any s,

c) the posterior belief $\hat{q}^s$ is correct given $\hat{\alpha}$ for any s,

d) and the beliefs $\hat{\rho}$ and $\hat{\alpha}$ are correct.

2.4.2 Belief Updating

If the tax payer does not report his income, the tax authority observes the estimate s at the beginning of Stage 2. The tax authority uses observation $r = 0$ and the estimate s to form a posterior belief $\hat{q}^s$ about income θ in a two step procedure:

First Step The tax authority forms the posterior belief $\hat{q}$, which is the conditional probability of income being high given action $r = 0$ and the tax authority's belief $\hat{\rho}$. Restricting the analysis to pure strategies drastically simplifies this step and in equilibrium, $\hat{q} \in \{0, q, 1\}$. However, an out of equilibrium belief has to be defined for the case that the tax payer always reports income in equilibrium, such that the tax authority would never observe $r = 0$ in equilibrium. I define the out of equilibrium belief for a deviation to $r = 0$ in this case to be $\hat{q} = q$.

Second Step The tax authority forms the posterior belief $\hat{q}^s$ which is the conditional probability of income being high given action $r = 0$ and estimate s of estimation technology π:

$$\hat{q}^s = \frac{\hat{q}\pi(s \mid H)}{\hat{q}\pi(s \mid H) + (1 - \hat{q})\pi(s \mid L)} \text{ for } s = h, l \tag{2.1}$$

This conditional probability equals the joint probability of observing the estimate s and income θ being high proportional to the total probability of observing the estimate s. If the tax authority perfectly learns income in the first step after observing $r = 0$, equation (2.1) reduces to $\hat{q}^s = 1$ or $\hat{q}^s = 0$ for $s = h, l$, respectively. For the more interesting case $\hat{q} = q$, equation (2.1) becomes

$$\hat{q}^s = \frac{q\pi(s \mid H)}{q\pi(s \mid H) + (1 - q)\pi(s \mid L)} \text{ for } s = h, l,$$

where $\hat{q}^s$ depends on the estimation technology π.

In the extreme, the estimation technology π is either fully informative or uninformative about income θ. The fully informative estimation technology π^{FI} is given by

$$\pi^{FI}(h \mid H) = 1 \qquad\qquad \pi^{FI}(h \mid L) = 0.$$

Any estimate of π^{FI} fully reveals income θ, such that $\hat{q}^h = 1$ and $\hat{q}^l = 0$. An uninformative estimation technology π^{NI} is given by

$$\pi^{NI}(h \mid H) = x \qquad\qquad \pi^{NI}(h \mid L) = x,$$

with $x \in [0, 1]$. No estimate of an uninformative estimation technology contains any information about income θ, such that $\hat{q}^h = \hat{q}^l = q$. If estimation technology π is of neither extreme case, the posterior belief about income θ increases after observing a high estimate and decreases after observing a low estimate: $\hat{q}^l \leq q \leq \hat{q}^h$. The larger the difference $\pi(h \mid H) - \pi(h \mid L)$ is, the more precise the estimation technology π is and the more the tax authority learns by observing an estimate of it.

In equilibrium, equation (2.1) reaches the correct posterior belief. However, out of equilibrium, equation (2.1) cannot be defined in two cases. The first being that the tax authority holds the belief $\hat{q} = 0$, but observes the estimate $s = h$ of the fully informative technology π^{FI}. The second being that the tax authority holds belief $\hat{q} = 1$, but observes the estimate $s = l$ of the fully informative technology π^{FI}. For both cases, I assume that the tax authority uses *only* the observed estimate to reach a posterior belief about the tax payer's income. That is, if $\hat{q} = 1$ and $s = l$ given $\pi = \pi^{FI}$, then $\hat{q}^l = 0$ and if $\hat{q} = 0$ and $s = h$ given $\pi = \pi^{FI}$, then $\hat{q}^h = 1$.

I begin the analysis with two benchmark cases, being that the tax authority chooses the estimation technology to be fully informative or uninformative. Subsequently, I show how and under which conditions the tax authority can choose the estimation technology π strategically to improve upon the full information case.

2.4.3 No Information

If the estimation technology is uninformative, the tax compliance game in Stages 1 and 2 is identical to one without any estimation technology:

Proposition 2.1 (No Information). *If the tax authority chooses an uninformative estimation technology in Stage 0, the subsequent tax compliance game has as a unique perfect Bayesian equilibrium. In Stage 1, the tax payer never reports his income. In Stage 2, the tax authority conducts no audit. In equilibrium, the tax payer's payoff equals his income θ and the tax authority receives zero tax revenue.*

Proof. I proof that Proposition 2.1 describes an equilibrium when $\pi = \pi^{NI}$. A proof of uniqueness is delegated to the Appendix. Suppose the tax payer sticks to the proposed strategy in Stage 1:

$$\rho(\theta) = \begin{cases} 0 & if \;\; \theta = L \\ 0 & if \;\; \theta = H \end{cases} \tag{2.2}$$

I show that a) Proposition 2.1 describes the tax authority's optimal reply to strategy (2.2), and b) the tax payer has no incentive to deviate from strategy (2.2).

a) In Stage 2, the tax authority forms the posterior belief $\hat{q}^s = q$ after observing $r = 0$ and s for $s = l, h$. Thus, the tax authority's expected payoff for conducting an audit is

$$qftH + (1-q)ftL - c.$$

This term is negative under Assumption 2, such that the tax authority's optimal strategy is

$$\alpha(s) = \begin{cases} 0 & if \quad s = l \\ 0 & if \quad s = h. \end{cases} \tag{2.3}$$

b) The tax payer's expected payoff for $r = 0$ is θ, while that of deviating to $r = 1$ is $(1-t)\theta$. As $(1-t)\theta < \theta$ for $\theta = L, H$, the tax payer has no incentive to deviate from strategy (2.2) and Proposition 2.1 describes an equilibrium when $\pi = \pi^{NI}$. □

2.4.4 Full Information

If the estimation technology is fully informative, the subsequent tax compliance game is identical to one with complete information:

Proposition 2.2 (Full Information)**.** *If the tax authority chooses the fully informative estimation technology in Stage 0, the subsequent tax compliance game has a unique subgame perfect equilibrium. If the tax payer's income is high, the tax payer reports his income in Stage 1, pays taxes tH to the tax authority and the game ends after Stage 1. If the tax payer's income is low, the tax payer does not report his income and the game proceeds to Stage 2. In Stage 2, the tax authority learns that the tax payer's income is low and conducts no audit.*

Proof. I proof that Proposition 2.2 describes an equilibrium when $\pi = \pi^{FI}$. A proof of uniqueness is delegated to the Appendix. Suppose the tax payer sticks to the proposed strategy in Stage 1:

$$\rho(\theta) = \begin{cases} 0 & if \quad \theta = L \\ 1 & if \quad \theta = H \end{cases} \tag{2.4}$$

I show that a) Proposition 2.2 describes the tax authority's optimal reply to strategy (2.4), and b) the tax payer has no incentive to deviate from strategy (2.4).

a) In Stage 2, the tax authority forms the posterior belief $\hat{q} = 0$ after observing $r = 0$. Estimate s realises according to π^{FI} and the tax authority forms the posterior belief

$$\hat{q}^s = \begin{cases} 0 & if \;\; s = l \\ 1 & if \;\; s = h, \end{cases} \tag{2.5}$$

where I make use of the assumption that out of equilibrium $\hat{q}^h = 1$ when $\hat{q} = 0$ and $s = h$ with $\pi = \pi^{FI}$. After observing $r = 0$ and $s = l$, the tax authority's expected payoff for conducting an audit is $ftL - c$, which is negative under Assumption 1, such that the tax authority has no incentive to audit the tax payer.

b) When $\theta = L$, the tax payer's expected payoff for $r = 0$ is L and that of deviating to $r = 1$ is $(1 - t)L$, such that he has no incentive to deviate from strategy (2.4). To check whether the tax payer has an incentive to deviate from strategy (2.4) when $\theta = H$, consider the out of equilibrium behaviour by the tax authority after observing $r = 0$ and $s = h$. With $\hat{q}^h = 1$, $a = 1$ is optimal for the tax authority, such that the tax payer's expected payoff for deviating to $r = 0$ is $(1 - ft)H$. Hence, the tax payer has no incentive to deviate from strategy (2.4) and Proposition 2.2 describes an equilibrium when $\pi = \pi^{FI}$. □

In the no information case, the tax authority suffers a commitment problem: if the tax authority could commit to audit the taxpayer once he does not report his income, the taxpayer would always report his income and the tax authority would never need to audit the tax payer. However, lacking any commitment device, the tax authority cannot credibly commit to audit the tax payer. If the tax payer never reports his income, the tax authority assigns probability q to income being high. In expectation, an audit results in a loss. Facing the decision whether to audit the tax payer, the tax authority prefers not to audit the tax payer.

The commitment problem is partially solved in the full information case. As the estimation technology perfectly identifies high and low income, the tax authority can commit to audit the tax payer when his income is high. Anticipating this, the tax payer reports high income to avoid being fined. On the other hand, the tax payer does not report his income when it is low, because he will not be audited.

2.4.5 Strategic Information

The tax authority can use the estimation technology π to solve the commitment problem completely. To do so, it chooses the information technology to be biased towards high estimates.

Proposition 2.3 (Strategic Information)**.** *If the audit cost c satisfies $c \leq \bar{c}$, the tax authority can induce the tax payer to always report his income by strategically overestimating income. The optimal estimation technology π^* to do so is given by*

$$\pi^*(h \mid H) = 1 \qquad \textit{and} \qquad \pi^*(h \mid L) = \frac{ftH - c}{c - ftL}\frac{q}{1-q}. \tag{2.6}$$

When the tax authority chooses π^* in Stage 0 and $c \leq \bar{c}$, the subsequent tax compliance game has a unique perfect Bayesian equilibrium. In Stage 1, the tax payer always reports his income, pays taxes $t\theta$ to the tax authority and the game ends after Stage 1. The threshold $\bar{c}$ is given by

$$\bar{c} = ft\frac{qfH + (1-q)L}{1 + qf - q}. \tag{2.7}$$

Proof. I proof that Proposition 2.3 describes an equilibrium when $\pi = \pi^*$ and $c \leq \bar{c}$. A proof of uniqueness is delegated to the Appendix. Suppose the tax payer sticks to the proposed strategy in Stage 1:

$$\rho^*(\theta) = \begin{cases} 1 & if \;\; \theta = L \\ 1 & if \;\; \theta = H. \end{cases} \tag{2.8}$$

I a) derive out of equilibrium beliefs and behaviour for the tax authority for $r = 0$, and b) show that the tax payer has no incentive to deviate from strategy (2.8).

a)Suppose the tax payer deviates to $r = 0$ in Stage 1. The tax authority forms the posterior belief $\hat{q} = q$ after observing $r = 0$, where I make use of the assumption that the tax authority sticks to the prior belief q when $r = 0$ is out of equilibrium. Estimate s realises according to π^* and the tax authority forms the posterior belief

$$\hat{q}^{s*} = \begin{cases} 0 & if \ \ s = l \\ \frac{1}{H-L}\left[\frac{c}{ft} - L\right] & if \ \ s = h. \end{cases} \tag{2.9}$$

After observing $r = 0$ and $s = l$, the tax authority's expected payoff for conducting an audit is

$$ftL - c,$$

which is negative under Assumption 1. Further, after observing $r = 0$ and $s = h$, the tax authority's expected payoff for conducting an audit is

$$\hat{q}^{h*} ftH + \left(1 - \hat{q}^{h*}\right) ftL - c. \tag{2.10}$$

Reformulation (2.10) shows

$$0 \leq \hat{q}^{h*} ftH + \left(1 - \hat{q}^{h*}\right) ftL - c$$

to hold with equality. The tax authority's optimal reply to the tax payer's deviation to $r = 0$ is therefore

$$\alpha^*(s) = \begin{cases} 0 & if \ \ s = l \\ 1 & if \ \ s = h. \end{cases} \tag{2.11}$$

b) Now consider the tax payer's choice in Stage 1 given π^* and α^*. first, consider a tax payer with $\theta = H$. The expected payoff for $r = 0$ is $(1 - ft)H$, while that for is $r = 1$ is $(1 - t)H$, and the tax payer has no incentive to deviate from (2.8) when $\theta = H$.

Second, consider a tax payer with $\theta = L$. The expected payoff for deviating to $r = 0$ is

$$\pi^*(h \mid L)(1 - ft)L + (1 - \pi^*(h \mid L))L,$$

while that for choosing $r = 1$ is $(1 - t)L$. The tax payer has an incentive to deviate to $r = 0$ if and only if

$$\pi^*(h \mid L)(1 - ft)L + (1 - \pi^*(h \mid L))L > (1 - t)L, \tag{2.12}$$

which reduces to

$$\pi^*(h \mid L) < \frac{1}{f}.$$

Reformulating (2.4.5) in terms of c yields $c > \bar{c}$ with $\bar{c}$ as defined in (2.7). Thus, if and only if $c \leq \bar{c}$, the tax payer has no incentive to deviate from strategy ρ^* in Stage 1 and Proposition 2.3 describes an equilibrium when $\pi = \pi^*$.[8] □

The estimation technology π^* is biased towards high estimates. The probability of it having a high estimate exceeds that of income being high as a high estimate realises with probability $\pi^*(h \mid L) > 0$ when income is low. Therefore, a high estimate indicates high income, but does not reveal high income. On the other hand, a low estimate reveals low income. Consequently, the tax authority's posterior belief about the tax payer's income increases after a high estimate and decreases to 0 after a low estimate:

$$0 = \hat{q}^{l*} < q < \hat{q}^{h*}$$

The probability $\pi^*(h \mid L)$ is chosen such that the tax authority has an incentive to audit the tax payer after observing a high estimate. Holding the posterior belief $\hat{q}^{h*}$, taxes and fines collected by an audit exactly pay off the audit cost in expectation. Thereby, the estimation technology works as a commitment device. If the tax payer does not report his income and the tax authority observes a high estimate, is has an incentive to audit the tax payer. On the downside, if the tax payer does not report his income and the tax authority observes a low estimate, it will not audit the tax payer.

In Stage 1, the tax payer anticipates estimates realising according to π^* and the tax authority to audit after observing a high estimate. If his income is high, the tax payer expects a high estimate to realise with certainty and therefore clearly prefers to report income to avoid being audited and fined.

[8] In the Appendix, I show that for any parameters (L, H, q, t, f) with $0 < L < H, 0 < q < 1, 1 < f$, and $0 < t$, we can find a cost parameter c which satisfies Assumption 2.1, Assumption 2.2 and the constraint $c \leq \bar{c}$.

Further, if his income is low, the tax payer expects a high estimate to realise with probability $\pi^*(h \mid L)$. In this case, the tax payer prefers to report income provided the probability $\pi^*(h \mid L)$ is sufficiently high. To maximise this probability, $\pi^*(h \mid L)$ is such that the tax authority is exactly indifferent between auditing and not auditing the tax payer after receiving a high estimate.

Therefore, the probability $\pi^*(h \mid L)$ needs to be sufficiently high for the tax payer to prefer to report low income. As the probability $\pi^*(h \mid L)$ decreases in the cost parameter c, this requirement puts an upper constraint on c. Only if $c < \bar{c}$, the tax payer prefers to report low income given that he expects to be audited with probability $\pi^*(h \mid L)$. Thus, provided $c < \bar{c}$, the tax authority can use the estimation technology π^* as a commitment device, such that the tax payer always prefers to report his income. Intuitively, a low audit cost allows the tax authority to choose a less informative estimation technology, as a less precise signal is sufficient to provide incentives to audit the tax payer after observing a high estimate. And the less informative the estimation technology, the higher the probability with which the tax payer expects to be audited when his income is low. This explains why the tax authority can use the biased estimation technology provided the audit cost is sufficiently low.

If the tax authority can and does use the biased estimation technology π^*, the tax payer always reports his income in equilibrium and is left with net income $(1-t)\theta$. The tax authority never needs to audit the tax payer and the expected tax revenue increases to $qtfH + (1-q)tfL$, compared to qtH in the full information case or 0 in the no information case.

2.5 Conclusion

Tax authorities can use a large amount of information about tax payers to decide whether to audit tax reports. In this Chapter, I study the tax authority's choice of how to use this information to estimate the tax payer's income. I use the Bayesian persuasion framework (Kamenica and Gentzkow, 2011) to model the tax authority's choice of a technology to estimate income. This technology is chosen in a prior stage before the tax payer reports his income and later used to decide whether to audit the tax payer. Although the technology could be chosen to estimate income perfectly, I show that it can be optimal to strategically overestimate income.

By overestimating income, the tax authority solves a commitment problem, which it suffers when income cannot be estimated. Because in expectation, the taxes and fines collected in an audit do not pay off the audit cost, the tax authority has no ex-post incentive to audit the tax payer. Consequently, the tax payer anticipates not to be audited and does never report his income. By overestimating low income strategically, the tax authority can fully solve the commitment problem. The optimal estimation technology always indicates high income to be high and sometimes indicates low income to be high. Further, it is chosen such that the tax authority has an incentive to audit the tax payer if income is indicated to be high. Anticipating to be audited with a certain probability, the tax payer prefers to report not only high, but also low income. I derive that the tax authority benefits from strategically overestimating income in this way provided the audit cost is sufficiently low.

Appendices

2.A Proof of Uniqueness - Proposition 2.1

I proof that Proposition 2.1 describes the unique sequential equilibrium in pure strategies when $\pi = \pi^{NI}$, by showing by contradiction that no alternative pure strategy of the tax payer can be sustained in an equilibrium. Three alternative candidates for a pure strategy exist.

First, suppose the tax payer takes action always reports income in Stage 1, such that the game would ends after Stage 1, with the tax payer's payoff being $(1-t)\theta$. This cannot be sustained in equilibrium, as the tax payer has an incentive to deviate. If the tax payer deviates to $r = 0$, the tax authority forms the posterior belief $\hat{q}^s = q$ after observing $r = 0$ and s for $s = h, l$, and takes action $a = 0$. Therefore, the tax payer's expected payoff for taking action $r = 0$ is θ for $\theta = H, L$, such that the tax payer has an incentive to deviate from the proposed strategy.

Second, suppose the tax payer only reports high income, such that the tax authority forms the posterior belief $\hat{q}^s = 0$ after observing $r = 0$ and s for $s = h, l$ and takes action $a = 0$. However, the the tax payer has an incentive to deviate from the proposed strategy when $\theta = H$, anticipating not to be audited.

Third, suppose the tax payer only reports low income, such that the tax authority forms the posterior belief $\hat{q}^s = 1$ after observing $r = 0$ and s for $s = h, l$ and takes action $a = 1$. However, the the tax payer would have an incentive to deviate from the proposed strategy in Stage 1 when $\theta = H$ to avoid being audited and fined.

2.B Proof of Uniqueness - Proposition 2.2

I proof that Proposition 2.2 describes the unique sequential equilibrium in pure strategies when $\pi = \pi^{FI}$, by showing that no alternative pure strategy of the tax payer can be sustained in an equilibrium.

First, suppose the tax payer always reports income, such that the game ends after Stage 1, with the tax payer's payoff being $(1-t)\theta$. This cannot be sustained in equilibrium, as the tax payer has an incentive to deviate from the proposed strategy when $\theta = L$: if the tax payer deviates to $r = 0$ in Stage 1, the tax authority forms the posterior belief $\hat{q} = q$ after observing $r = 0$ (making use of the assumption that the tax authority sticks to the prior belief q when $r = 0$ is not observed in equilibrium) and the posterior belief $\hat{q}^l = 0$ after observing $s = l$. Given $\hat{q}^l = 0$, the tax authority takes action $a = 0$, such that the tax payer's expected payoff for deviating to $r = 0$ when $\theta = L$ is L, which is larger than $(1-t)\theta$.

Second, suppose the tax payer never reports income, such that the tax authority forms the posterior belief $\hat{q} = q$ after observing $r = 0$ and posterior beliefs $\hat{q}^h = 1$ and $\hat{q}^l = 0$ after observing s. Given $\hat{q}^h = 1$, the tax authority takes action $a = 1$ and the tax payer's payoff is $(1-ft)H$ when $\theta = H$. However, as the tax payer's payoff for taking acting action $r = 1$ when $\theta = H$ is $(1-t)H$, the tax payer would have an incentive to deviate from the proposed strategy.

Third, suppose the tax payer takes action only reports low income. It follows from above that this strategy cannot be optimal.

2.C Proof of Uniqueness - Proposition 2.3

I proof that Proposition 2.3 describes the unique sequential equilibrium in pure strategies when $\pi = \pi^{SI}$, by showing that no alternative pure strategy of the tax payer can be sustained in an equilibrium.

First, suppose the tax payer never reports income, such that the tax authority forms the posterior belief $\hat{q} = q$ after observing $r = 0$ and the posterior belief $\hat{q}^{s*}$ according to (2.9) as described in the proof of Proposition 2.3 after observing estimate s. Then, after observing $s = h$, the tax authority takes action $a = 1$, which is weakly optimal given $\hat{q}^{h*}$, and the tax payer's payoff is

$(1 - ft)H$ when $\theta = H$. However, as the tax payer's payoff for taking acting action $r = 1$ when $\theta = H$ is $(1 - t)H$, the tax payer has an incentive to deviate from the proposed strategy.

Second, suppose the tax payer only reports high income, such that the tax authority forms the posterior belief $\hat{q} = 0$ after observing $r = 0$ and the posterior belief $\hat{q}^s = 0$ for $s = h, l$ after observing estimate s. The tax authority takes action action $a = 0$, which is optimal given $\hat{q}^s = 0$ for $s = h, l$. However, anticipating action $a = 0$, the expected payoff for the tax payer for taking action $r = 0$ when $\theta = H$ is H, such that the tax payer has an incentive to deviate to $r = 0$ when $\theta = H$.

Lastly, suppose the tax payer only reports low income, such that the tax authority forms the posterior belief $\hat{q} = 1$ after observing $r = 0$ and the posterior belief $\hat{q}^s = 1$ after observing s for $s = h, l$. With $\hat{q}^s = 1$, action $a = 1$ would be optimal and the tax payer's payoff would be $(1 - ft)H$ when $\theta = H$. However, the tax payer would have an incentive to deviate from the proposed strategy when $\theta = H$ to avoid being fined. Thus, the proposed reporting strategy cannot be part of an equilibrium.

2.D Permissible Range of the Cost Parameter

I show that for any parameters L, H, q, t, f with $0 < L < H, 0 < q < 1, 0 < t$ and $1 > f$, we can find a cost parameter c such that

$$ftL < c, \tag{2.13}$$

$$c < ftH, \tag{2.14}$$

$$ft[qH + (1 - q)L] < c \tag{2.15}$$

$$\text{and } c < \frac{ft[fqH + (1 - q)L]}{1 + q(f - 1)}. \tag{2.16}$$

Assume $0 < L < H, 0 < q < 1, 0 < t$ and $1 > f$. First, observe that $ftL < ft[qH + (1 - q)L]$ and hence, inequality (2.15) implies inequality (2.13). Further, observe that

$$\frac{ft[fqH + (1 - q)L]}{1 + q(f - 1)} < ftH,$$

as

$$\begin{aligned}\frac{ft[fqH+(1-q)L]}{1+q(f-1)} &< ftH \leftrightarrow \\ fqH+L-qL &< H+fqH-qH \leftrightarrow \\ 0 &< (H-L)(1-q),\end{aligned}$$

and hence, inequality (2.16) implies inequality (2.14). Finally, observe that

$$ft[qH+(1-q)L] < \frac{ft[fqH+(1-q)L]}{1+q(f-1)},$$

as

$$\begin{aligned}&ft[qH+(1-q)L] < \frac{ft[fqH+(1-q)L]}{1+q(f-1)} \leftrightarrow \\ &qH+L-qL+fqqH+fqL-fqqL-qqH-qL+qqL < fqH+L-qL \\ &\leftrightarrow q(H-L)(1-f)(1-q) < 0.\end{aligned}$$

Hence, a parameter c always exists satisfying inequalities (2.15) and (2.16). And inequalities (2.13) and (2.14) are implied.

Chapter 3

Sequential Gerrymandering

This chapter benefited heavily from discussions with Kai A. Konrad.

3.1 Introduction

Dynamic contests in which the player who wins a given minimum number of battles wins the overall contest have attracted considerable scholarly interest in recent years. Starting with Konrad and Kovenock (2009) and Klumpp and Polborn (2006), various practically relevant structures and rules of such best-of-n contests have been studied.[1] I build on this literature and analyse a best-of-n contest that captures the strategic elements of a gerrymandering process, in which two political parties choose electoral district boarders in alternating order.

The specific elements of this problem are: two parties compete in an election across a given number of single-member districts. The total electorate is divided between two fractions of partisan voters, each fraction being partisan of one of both parties. A district map needs to be chosen to allocate the total electorate across the districts. An equal number of voters needs to be allocated to each district. In each district, a partisan representative is elected by simple majority voting. As voters vote according to their partisanship, the elected representative of a district belongs to the party with a majority of partisan voters in this district. The college of representatives then decides on policies

[1]See, for instance, Sela (2011), Gelder (2014), Konrad (2018) and Barbieri and Serena (2020).

by majority vote.

If one of both parties itself (or both of them) is (are) responsible for choosing a new district map, this choice is strategic: if one voter is excluded from district i, this voter must be allocated to a different district j, and another voter must replace this voter in district i. A self-interested party choosing a district map will optimise the allocation of voters to pursue a specific objective. When the party's objective is to increase its representation in the college of representatives, this is referred to as 'partisan gerrymandering'. I focus in this Chapter on partisan gerrymandering, however, a party can also pursue different objectives when choosing electoral districts, such as the protection of an incumbent or discrimination against minority groups.

Although gerrymandering occurs across a wide range of electoral systems, the vast majority of literature on gerrymandering focuses on the United States.[2] In the United States, new electoral district maps have to be defined timely after each decennial census to ensure the equal representation of voters and to prevent malapportionment. Typically, the current state legislator, being the Republican or Democratic party, decides on a new district map. The two basic constraints thereby are that districts have to be contiguous and that districts need to be equal in population. While the second requirement is practically binding, the first is not. Examples exist of electoral districts consisting of two disjoint areas connected only by a thin line.[3]

(Partisan) gerrymandering is widely criticised for harming democratic elections. A main concern is that the elected representatives might no longer represent the overall distribution of voter preferences: one party might be supported by only a minority of voters, while having a majority in the college of representatives, such that decision outcomes differ from those of a popular vote.

In the model I analyse, misrepresentation of voter preferences does not occur in equilibrium. Specifically, I analyse a sequential process, in which both parties allocate voters to one district at a time in alternating order. Each party aims at winning a majority of districts and there is no uncertainty about voters'

[2]See Bickerstaff et al. (2020) for an overview.

[3]In fact, the term 'gerrymandering' dates back to a very oddly shaped district in Massachusetts. This district was signed off by the then-governor Elbridge Gerry and later compared to a salamander, which gave birth to the word 'gerry-mander'.

partisanship. In equilibrium, the party with a majority in the popular vote wins a majority of the districts. Neither party has a first-mover or a second-mover advantage, because the two parties' strategic redistricting choices neutralise each other. In an extension to the model, I assume that parties maximise the number of districts they win and show that for this case, a first-mover advantage does emerge.

Seen as a dynamic contest, the problem is a variant of a sequential Blotto contest. Both players choose sequentially how much of a given amount of resources to allocate to a battle. The total amount of resources allocated to a battle is given and constant (the number of voters in a district) and a player chooses how this amount is composed of own resources (own partisan voters) and the other player's resources (the other party's partisan voters). The analysis contributes to the large literature on dynamic contests on one side, and the large literature on Blotto contests[4] on the other side. A Blotto contest similar to my model was analysed by Klumpp et al. (2019): two players enter into a best-of-n contest with given resources. In each round, players move simultaneously and allocate resources to the given battle. My analysis differs from theirs by restrictions that are motivated by the gerrymandering problem: I assume that the total amount of resources is necessarily used up in the whole sequence of battles and that the same total amount of resources is allocated to each battle. Further, I assume that players choose sequentially and can dispose of their own and the other party's resources. This structure generates a trade-off that has not been studied: resources that are not allocated to a given battle must be allocated to one of the subsequent battles, and resources allocated to a given battle cannot be allocated to a subsequent battle.

This chapter contributes to the economic literature studying partisan gerrymandering.[5] This literature was started by Owen and Grofman (1988). They look at the problem of a party in power to choose a district map optimally to either maximise the number of districts won or to maximise the probability of

[4]This literature dates back to Borel (1921)and Borel and Ville (1938). For recent developments see Roberson (2006), Kvasov (2007), Kovenock and Roberson (2012) and Kovenock and Roberson (2021).

[5]A connected line of research studies redistricting from a welfare perspective, with the focus on identifying welfare maximising district maps (see Coate and Knight (2007), Besley and Preston (2007) and Bracco (2013)) or district maps satisfying certain fairness criteria (see for instance Puppe and Tasnádi (2015)).

winning a majority of the districts. Two types of districts make up the optimal solution: 'cracking' districts, which are won by the smallest possible winning margin; and 'packing' districts, which are lost by the highest possible margin. Friedman and Holden (2008) further developed on this problem by varying the underlying assumptions about the distribution of voters' preferences and uncertainty.[6] Kolotilin and Wolitzky (2020) summarise a large part of this literature in a very general framework with a continuum of voter types and uncertainty about the aggregate vote share for the party being in charge of the redistricting process. They identify a connection to the information design literature (see Kamenica and Gentzkow (2011)).[7]

Building upon the literature with one party choosing the district map, Gul and Pesendorfer (2010) and Friedman and Holden (2020) account for the fact that, on a national level, two parties are involved in the redistricting process, each in a predefined region (i.e. each party chooses the district map within those states it currently controls). Both papers model this interaction as a simultaneous move game, in which both parties choose independently the district map within the region they control. The setting is strategic insofar as the district map chosen by the respective other party influences a party's expected payoff and thus its maximisation problem.

Most closely related to this chapter is the analysis by Bierbrauer and Polborn (2022) of a sequential redistricting process.[8] They take a design perspective on the probelm and develop a redistricting protocol that ensures the party with a majority in the popular vote can always obtain a majority of the districts. I obtain a similar result, however in a different model: whereas in the model by Bierbrauer and Polborn (2022), a constant fraction of voters is allocated to each of the districts in each stage, in my model, one district is chosen at each stage. Further, I abstract from any uncertainty about voters' voting behaviour, which allows to solve for a concrete equilibrium, which is not possible in the general framework of Bierbrauer and Polborn (2022).

[6]See also Sherstyuk (1998), Shotts (2001) and Gilligan and Matsusaka (2006).

[7]This connection is also explored by Lagarde and Tomala (2021).

[8]Pegden et al. (2017) also consider a sequential redistricting process. In each stage, one player chooses a redistricting plan and the other player 'freezes' one of the districts of this plan. In the subsequent round, a new plan for the remaining districts is proposed. Ely (2019) considers a two-stage game in a geometric setting.

3.2 Baseline Model

3.2.1 Setup

I consider the following dynamic contest. Two players, A and B, compete in an odd number $N = 2n + 1$ of consecutive battles. The player who wins $n + 1$ battles, wins the contest.[9] This defines a grid of possible states: the initial state is $(0, 0)$. The subsequent state is either $(1, 0)$, if A wins the first battle, or $(0, 1)$, if B wins the first battle. A generic state (i, j) describes with i and j the number of previous battles won by A and B, respectively. If $i, j < n + 1$, the contest winner is not yet determined. If either $i \geq n + 1$ or $j \geq n + 1$, the contest winner was determined prior to state (i, j). The number of remaining battles, including the battle at state (i, j), is given by $N - i - j$.

The players allocate resources to each of the N battles in alternating order. The initial amount of resources has size $2N$ and consists of two types. Resources of 'type a', support player A, resources of 'type b' support player B. The initial amounts of type a and type b resources are denoted by $a^{0,0} \geq 0$ and $b^{0,0} \geq 0$, respectively, with $a^{0,0} + b^{0,0} = 2N$. At a given state (i, j), the resources left from the initial amount are denoted by $a^{i,j}$ and $b^{i,j}$, with

$$b^{i,j} + a^{i,j} = 2(N - i - j). \tag{3.1}$$

The total initial amount of resources is used up in the N consecutive battles.

I call states with even $i + j$ 'even states', and states with odd $i + j$ 'odd states'. A chooses at the first state and at all even states, B chooses at all odd states. The player in charge at state (i, j) chooses a resource allocation $(a_{i,j}, b_{i,j})$, with $a^{i,j} \geq a_{i,j} \geq 0$, $b^{i,j} \geq b_{i,j} \geq 0$ and

$$a_{i,j} + b_{i,j} = 2. \tag{3.2}$$

By (3.2), the choice at a given state is fully characterised by $a_{i,j}$.

The choice of $a_{i,j}$ has two consequences. First, it determines who wins

[9] The structure of the contest I look at resembles best-of-n-contests, as in Konrad and Kovenock (2009) or Klumpp, Konrad, and Solomon (2019), differences being that in the present contest, at each stage, one player chooses the composition of a predetermined amount of the overall resources to be allocated to one battle.

battle (i, j). If A chooses at state (i, j), A wins this battle if $a_{i,j} \geq 1$ and loses it otherwise. If B chooses at state (i, j), B wins this battle if $b_{i,j} = 2 - a_{i,j} \geq 1$, and loses it otherwise.[10] Accordingly, the next state is $(i + 1, j)$ or $(i, j + 1)$. Second, the resource amounts adjust from $a^{i,j}$ to $a^{i,j} - a_{i,j}$ and from $b^{i,j}$ to $b^{i,j} - b_{i,j}$.

Initial resources, resource allocations in previous states (and the induced battle outcomes) are observed by both players. The player who wins a majority of battles wins the contest. The payoff of A is $\pi_A = V$ for all sequences of battles for which A wins at least $n + 1$ battles and zero otherwise. Similarly, the payoff of B is $V - \pi_A$.

3.2.2 Strategies and Equilibrium Concept

Generally, players can condition a choice at state (i, j) on the full history of resource allocations up to this state. However, in line with the reasoning in Klumpp et al. (2019), I focus on Markov strategies, meaning that the choice at state (i, j) depends only on the directly payoff-relevant information, being the state itself and the amounts of remaining resources $a^{i,j}$ and $b^{i,j}$. The allocation function

$$a_{i,j} : \; [0, 2(N - i - j)] \to [\max\{0, 2 - b^{i,j}\}, \min\{2, a^{i,j}\}],$$

with $b^{i,j} = 2(N - i - j) - a^{i,j}$, describes the allocation choice $a_{i,j}(a^{i,j})$ for any amount of left resources $a^{i,j}$ at state (i, j).

A continuation strategy for A at an even state (i, j) is a collection of allocation functions $a_{i',j'}$ for the current and all possible future even states that can be reached from (i, j). Analogously, a continuation strategy for B at a given odd state (i, j) is a collection of allocation functions $a_{i',j'}$ for the current and all possible future odd states that can be reached from (i, j). A Markov strategy for A is then a continuation strategy at state $(0, 0)$ and a Markov strategy for B consists of continuation strategies at states $(1, 0)$ and $(0, 1)$.

A Markov perfect equilibrium is described by a strategy profile such that, for all non-terminal states (i, j) the continuation strategies described in the

[10] The assumed tie-breaking rule is reasonable: the player in charge could allocate $1 + \epsilon$ of his type's resource to win the battle if the tie-breaking rule was different.

strategy profile are Nash equilibria at the respective subgames (analogous to Klumpp et al. (2019)).

3.2.3 Equilibrium

The number of possible subgames in the sequential contest just described is typically infinite[11] given that the resource allocation at each state is chosen from a continuous space, as long as there are resources of both types left. For large N, describing the complete set of Markov perfect equilibria would be cumbersome. (For the simple case $N = 3$, I describe the full range of equilibria in Section 3.3.) In the following, I describe a specific Markov perfect equilibrium that exists for any N and any amount of initial resources. I will use this result to show that in any subgame perfect equilibrium, the player with the larger amount of initial resources wins the contest.

Proposition 3.1. *The following describes a Markov perfect equilibrium for initial resources* $(a^{0,0}, b^{0,0})$. *The resource allocation A chooses at each even state (i, j), and the resource allocation B chooses at each odd state (i, j) is given by:*

$$a_{i,j}(a^{i,j}) = \begin{cases} 1 & \textit{if } a^{i,j}, b^{i,j} \geq 1 \\ a^{i,j} & \textit{if } a^{i,j} < 1 \\ 2 - b^{i,j} & \textit{if } b^{i,j} < 1, \end{cases} \tag{3.3}$$

where $b^{i,j} = 2(N - i - j) - a^{i,j}$. *In this Markov perfect equilibrium, resource allocations at state (i, j) take the form $(1, 1)$ as long as the remaining resources $a^{i,j}$ and $b^{i,j}$ allow this. In each of these states, the player allocating resources wins the respective battle. Once the remaining resources of one player fall below 1, the remaining battles are won by the respective other player. In this equilibrium, the player with the larger initial amount of resources wins the contest.*

Proof of Proposition 3.1. Using backward induction, I show that allocating resources according to (3.3) is optimal for each player at each state at which he needs to make a choice.

[11] The exception being the trivial case that the initial amount of resources consists of only one type.

State (i,j) with $i+j=N-1$: In the last state, A's choice is predetermined by

$$a_{i,j}(a^{i,j}) = a^{i,j} = 2 - b^{i,j}. \tag{3.4}$$

Observe that this choice coincides with (3.3).

State (i,j) with $i+j=N-2$: I show that:

Lemma 3.1. *In the last but one state, choosing $a_{i,j}$ according to* (3.3) *is optimal for B.*

Proof of Lemma 3.1. In state (i,j) with $i+j=N-2$, either $i \geq n+1$, or $j \geq n+1$, or $(i,j)=(n,n-1)$, or $(i,j)=(n-1,n)$. If $i \geq n+1$ or $j \geq n+1$, the contest outcome was determined at a prior state and B's choice of $a_{i,j}$ is not payoff relevant, so that choosing $a_{i,j}$ according to (3.3) is (weakly) optimal for B. If $(i,j)=(n-1,n)$, B wins the contest if he wins at least one of the remaining battles which is possible if $a^{i,j} \leq 3$. If $(i,j)=(n,n-1)$, B wins the contest if he wins both remaining battles, which is possible if $a^{i,j} < 2$. Choosing $a_{i,j}$ according to (3.3) guarantees B to win the contest for both cases whenever possible. □

State (i,j) with $i+j=N-k, k \in \{3,4,\ldots N\}$: Assume that for any subsequent state (i',j') with $i'+j' > N-k$, both players allocate resources according to (3.3). Given these resource allocations, I show that:

Lemma 3.2. *If state (i,j) is even, choosing resources according to* (3.3) *is optimal for A and if state (i,j) is odd, choosing resources according to* (3.3) *is optimal for B.*

Proof of Lemma 3.2. Given that resource allocations in states (i',j') with $i'+j' > N-k$ are given by (3.3), the total number of battles A wins in the states subsequent to state (i,j) depends on the resources $a^{i,j}-a_{i,j}$, remaining at the next state, only. Denote this number by $X(a^{i,j}-a_{i,j})$. I make two observations regarding $X(a^{i,j}-a_{i,j})$:

$$X(a^{i,j}-a_{i,j}) \geq X(a^{i,j}-a'_{i,j}) \qquad \text{if} \qquad a_{i,j} \leq a'_{i,j} \tag{3.5}$$

$$X(a^{i,j}-a_{i,j}) - X(a^{i,j}-a'_{i,j}) \leq 1 \qquad \text{if} \qquad a_{i,j} = a'_{i,j}-1 \tag{3.6}$$

Inequality (3.5) states that the total number of subsequent battles A wins is weakly increasing in the amount of type a resources remaining at the beginning of the next state. Inequality (3.6) states that the total number of battles A wins increases at most by 1 if the amount of type a resources remaining at the beginning of the next state increases exactly by 1. (Reversely, inequality (3.6) implies that the number of battles won by player B increases at most by 1 if the amount of type b resources remaining at the beginning of the next state increase exactly by 1.) Both observations follow directly from the assumption I made regarding resource allocations in states subsequent to state (i, j).

I am now in a position to prove Lemma 3.2.
First, assume $a^{i,j}, b^{i,j} \geq 1$ and state (i, j) is even. Suppose A chooses $a_{i,j} = 1$, implying A wins this battle and the next state is $(i+1, j)$ with remaining type a resources $a^{i+1,j} = a^{i,j} - 1$. A wins the contest if and only if $i+1+X(a^{i,j}-1) \geq n+1$. To show that the resource choice according to (3.3) is optimal, I need to show that A has no profitable deviation for the case $i + 1 + X(a^{i,j} - 1) < n+1$. Consider possible deviations for this case: if A deviates to $\hat{a}_{i,j} \in [0, 1)$, he looses battle (i, j) and the next stage is $(i, j + 1)$ with remaining type a resources $a^{i,j-1} = a^{i,j} - \hat{a}_{i,j} \in (a^{i,j} - 1, a^{i,j}]$. Here, inequality (3.6) implies that with remaining resources $a^{i,j-1} = a^{i,j}$, the total number of battles A wins in subsequent states increases by at most 1. Hence, a deviation to $\hat{a}_{i,j} \in [0, 1)$ is not profitable. If A deviates to $\hat{a}_{i,j} \in (1, 2]$, he still wins battle (i, j), but the resources at the beginning of the next state $(i + 1, j)$ reduce compared to A choosing $a_{i,j} = 1$. This, by inequality (3.5), implies that the total number of battles A wins in subsequent states does not increase. Hence, this deviation is also not profitable.
Second, assume $a^{i,j}, b^{i,j} \geq 1$ and state (i, j) is odd. By parallel reasoning, suppose B chooses $a_{i,j} = 1$, implying B wins this battle and the next state is $(i, j + 1)$ with resources $a^{i,j+1} = a^{i,j} - 1$. B wins the contest if and only if $i + X(a^{i,j} - 1) < n + 1$. To show that the resource choice according to (3.3) is optimal, I need to show that B has no profitable deviation for the case $i + X(a^{i,j} - 1) \geq n + 1$. Consider a deviation to $\hat{a}_{i,j} \in (1, 2]$. Then A wins the battle at stage (i, j) and the next stage is $(i + 1, j)$ with resources $a^{i+1,j} = a^{i,j} - \hat{a}_{i,j} \in [a^{i,j} - 2, a^{i,j} - 1)$. Here, inequality (3.6) implies that with remaining type a resources $a^{i+1,j} = a^{i,j} - 2$, the total number of battles A wins

in subsequent states reduces by at most 1. Hence, a deviation to $\hat{a}_{i,j} \in (1, 2]$ is not profitable. Now consider a deviation to $\hat{a}_{i,j} \in [0, 1)$. While B still wins the battle at state (i, j), the type a resources at the beginning of the next state $(i, j + 1)$ increase compared to B choosing $a_{i,j} = 1$. This, by inequality (3.5), implies that the total number of battles A wins in subsequent states does not decrease. Hence, this deviation is also not profitable.
Finally, observe that if $a^{i,j} < 1$ or $b^{i,j} < 1$, any choice at state (i, j) results in the same player winning the battle at state (i, j) and all remaining battles (namely the player who still has a positive amount of resources left). Hence, choosing $a_{i,j}$ according to (3.3) is (weakly) optimal. □

I have shown that choosing resources according to (3.3) is optimal in the last but one state (i, j) with $i + j = N - 2$ and that, under the assumption that for any subsequent state (i', j') with $i' + j' > N - k$, both players allocate resources according to (3.3), choosing resources according to (3.3) at state (i, j) with $i' + j' = N - k$ is optimal for player A if (i, j) is even and is optimal for player B if (i, j) is odd. This allows to conclude the Proof of Proposition 3.1 by induction. □

In this equilibrium, each player allocates the minimal amount of resources necessary to win a battle, each time he makes a choice and as long as this is possible. Particularly, neither player has a first- or second mover advantage. To see this, consider the case $|a^{0,0} - b^{0,0}| = 2x, x \in (0, 1]$, for which the resources of the player with the smaller amount of initial resources fall below 1 in the last state on equilibrium path. In this case, during the first $N - 1$ states, both players win battles in alternating sequence, with each player winning a total number of n battles. The last battle is won by the player with the larger amount of initial resources, independently whether this is A or B.

The described equilibrium is not unique, as noted before. However, I can use this equilibrium to draw a general conclusion about the contest outcome.

Proposition 3.2. *If two subgame perfect equilibria exist in the contest described above, they result in the same contest outcome.*

Proof of Proposition 3.2. I proof Proposition 3.2 by contradiction. Suppose two subgame perfect equilibria exist with A's payoff in equilibrium being $\pi^*_A = 1$ and $\pi^{**}_A = 0$. Denote the resource allocations chosen on equilibrium path in these equilibria by $a^*_{i,j}$ and $a^{**}_{i,j}$, respectively. Define a state (i,j) with $i+j=k$ such that, for any state (i',j') with $i'+j'<k$, $a^*_{i',j'}=a^{**}_{i',j'}$ and for state (i,j), $a^*_{i,j}\neq a^{**}_{i,j}$. If k is even or $k=0$, the allocation $a^{**}_{i,j}$ can not be optimal for A. If k is odd, the allocation $a^*_{i,j}$ can not be optimal for B. □

Corollary 3.1. *In any subgame perfect equilibrium, the player with the larger amount of initial resources wins the contest.*

Corollary 3.1 follows directly from Propositions 3.1 and 3.2.

3.3 The Simple Case with N=3

When $N=3$, A allocates resources to the first battle, B allocates resources to the second battle and the remaining resources are allocated to the last battle. The player who wins at least two battles wins the contest. It is illustrative to describe the full range of equilibria for this simple case. If $a^{0,0}>b^{0,0}$, two types of equilibria exist. In the first type, A chooses

$$a_{0,0}\in[max\{1,2-b^{0,0}\},min\{a^{0,0}-2,2\}]. \tag{3.7}$$

Thereby, A wins the first battle and the next state is $(1,0)$ with $a^{1,0}\geq 2$. Independent of B's choice at state $(1,0)$, A wins at least one further battle. This makes B indifferent at state $(1,0)$. A wins the contest.

In the second type, A chooses

$$a_{0,0}\in[max\{0,2-b^{0,0}\},min\{a^{0,0}-3,2\}]. \tag{3.8}$$

In this case, the second state (i,j) can be $(1,0)$ or $(0,1)$. In any of both cases, resources at the second state satisfy $a^{i,j}\geq 3$, such that independent of B's choice at the second state, A wins both remaining battles. Again, B is indifferent at the second state and A wins the contest. Note that for $a^{0,0}>5$, A wins for sure all three battles and (3.7) coincides with (3.8).

If $b^{0,0}>a^{0,0}$, A is indifferent at state $(0,0)$ because independent of its choice $a_{0,0}$, B win at least two of the three battles. If A chooses $a_{0,0}\in[1,2]$, A wins the first battle, but as $b^{1,0}>2$, B can win both remaining battles. If A chooses

$a_{0,0} \in [0,1)$, B wins the first battle and as $b^{0,1} > 1$, B also wins one further battle.

The range of all equilibria is depicted in Figure 3.1. Each of the subfigures depicts optimal resource allocations for B (Subfigures 3.1a and 3.1b) and A (Subfigure 3.1c) for any amount of resources left. The horizontal axis corresponds to the amount of available resources of type a (top axis) and of type b (bottom axis). The vertical axis corresponds to the allocation of resources of type a (left axis) and of type b (right axis) between 0 and 2. The coloured areas in Subfigures 3.1a and 3.1b depict optimal allocations for B in states $(0,1)$ and $(1,0)$, respectively, with which B wins the contest. For $a^{0,1} \geq 3$ in state $(0,1)$ (Subfigure 3.1a) and for $a^{1,0} \geq 2$ in state $(1,0)$ (Subfigure 3.1b), B is indifferent between any resource allocation. The coloured areas in Subfigure 3.1c depicts optimal allocations for A in the first state with which A wins the contest. If $a^{0,0} < 3$, A is indifferent between any resource allocation.

3.4 Linear Payoff Function

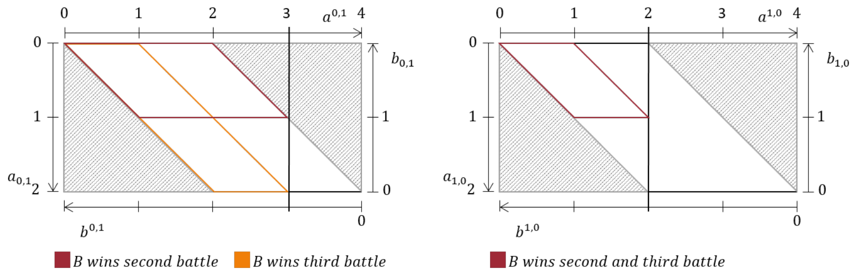
I now return to the general case $N \geq 3$ and consider a different payoff function: suppose players' payoff is linear in the number of battles won, with each player receiving a payoff of 1 for each battle won. Let the number of battles A wins be denoted by X, the total payoff A receives is then given by $\pi_A = X$ and the total payoff B receives is given by $N - X$. I show that the equilibrium described in Proposition 3.1 extends to this payoff function. However, A has a first-mover advantage.

Proposition 3.3. *The resource allocations described in equation* (3.3) *describe a Markov perfect equilibrium for any initial resources* $(a^{0,0}, b^{0,0})$ *if the players' payoff is linear in the number of battles won. The payoff A receives in equilibrium is given by*

$$\pi_A = \begin{cases} \lceil \frac{1}{2} \cdot a^{0,0} \rceil & \text{if } a^{0,0} > b^{0,0} \\ \lceil \frac{1}{2} \cdot \lfloor a^{0,0} \rfloor \rceil & \text{if } a^{0,0} < b^{0,0} \end{cases} \tag{3.9}$$

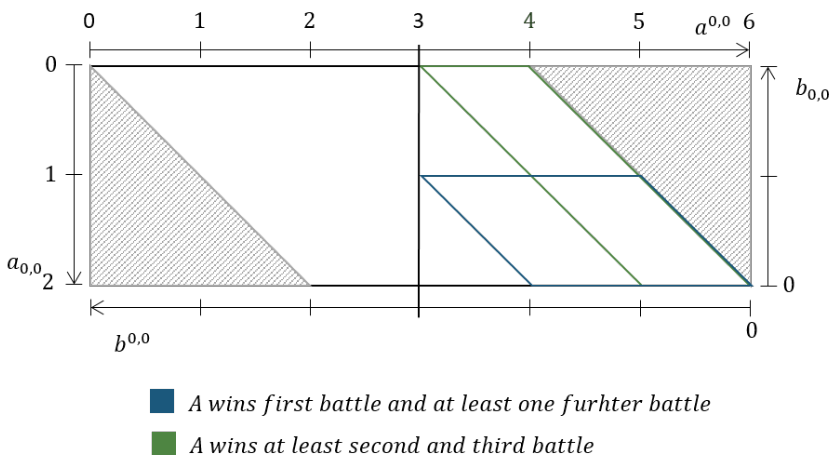
In this equilibrium, A has a first-mover advantage.

The Proof of Proposition 3.3 follows the reasoning of the Proof of Proposition 3.1 (see Appendix).



(a) Optimal allocations for B in state $(0, 1)$

(b) Optimal allocations for B in state $(1, 0)$



(c) Optimal allocations for A in state $(0, 0)$

Figure 3.1: Range of Equilibria for $N = 3$

The following explains the number of battles A wins in equilibrium, as described in Proposition 3.3: the sequence of battles on equilibrium path can be divided into three parts. During an initial sequence of an *even* number of battles, resource allocations take the form $a_{i,j} = 1$, such that each player wins 1 battle for a total amount of 2 resources allocated to a pair of battles. During an ending sequence of battles, resource allocations take the form $a_{i,j} = 0$ or $a_{i,j} = 2$, such that the player with the larger amount of initial resources wins 1 battle for a total amount of 2 resources allocated to each of these battles. Thus, for the battles during the initial and during the ending sequence, an amount of 2 of initial resources corresponds to 1 battle won on equilibrium path.

In between, in a short sequence of one or two battles, one battle won can correspond to *less* than a total amount of 2 resources. Equation (3.9) controls for this by rounding up (and down) the initial amount of type a resource or its halve. At state (i, j), at which the resources of one of both players fall below 1, the respective other player wins this battle using an amount of resources between 1 and 2. If this state (i, j) is odd, additionally, the prior battle $(i - 1, j)$ is won by A with an amount of 1 resources. Here, the two battles won by A correspond to amount of less than a total amount of 4 initial resources. Additionally, a first-mover advantage for A emerges: to see this, consider the following two examples:

Example 1 $N = 11, a^{0,0} = 9.5, b^{0,0} = 12.5$. On equilibrium path, $a_{i,j} = 1$ for the first 9 battles, of which 5 are won by A and 4 are won by B. In state $(5, 4)$, $a^{5,4} = 0.5$, such that B wins the last two battles.

Example 2 $N = 11, a^{0,0} = 12.5, b^{0,0} = 9.5$. On equilibrium path, $a_{i,j} = 1$ for the first 9 battles, of which 5 are won by A and 4 are won by B. In state $(5, 4)$, $b^{5,4} = 0.5$, such that A wins the last two battles.

In both examples, A wins 5 of the first 9 battles and thus 1 battle more than B. While in Example 1, B wins both remaining battles and thus 6 battles in total, in Example 2, A wins both remaining battles and thus 7 battles total. Hence, for the reverse allocation of initial resources, A wins 1 battle more if he has the larger amount of initial resources. Therefore, A has a first-mover advantage.

As for the main result, the equilibrium described in Proposition 3.3 is not unique. However, as before, Proposition 3.2 allows to conclude:

Corollary 3.2. *In any subgame perfect equilibrium for the described contest with linear payoffs, the payoff of A is given as described in Proposition 3.3.*

Corollary 3.2 follows directly from Propositions 3.3 and 3.2.

3.5 Conclusion

This chapter studied a sequential contest with fixed resource amounts. In this contest, players allocate resources to each battle in alternating order and the player with a majority of resources allocated to a battle wins the respective battle. The structure of the contest was chosen to resemble the dynamics of a sequential gerrymandering process, in which two parties select the boarders of electoral districts, each party one district at a time, in alternating order. The model was solved for a Markov perfect equilibrium. In this equilibrium, each player allocates equal amounts of own resources and opposing resources to each battles. As long as this is feasible, battles are won by the player allocating resources to the respective battle. Once the resources of one player are used up, the remaining battles are won by the respective other player. With this result, I show that in a sequential gerrymandering process, the party with a majority of voters in the total electorate wins a majority of electoral districts.

Whereas the baseline model assumes that players aim at winning a majority of battles, I consider a linear payoff function in an extension. I showed that the same equilibrium can be obtained with the modified payoff function. However, whereas there is no first- or second mover advantage in the baseline model, the first player to move has an advantage when players aim at maximising the number of battles won.

Interpreting the theoretic results in the context of gerrymandering relies on the stark assumption that voter preferences are perfectly known by political parties and that there is no uncertainty in the voting process. An interesting question to be explored in further research would be, in how far the obtained result can be generalised to settings that include uncertainty.

Appendices

3.A Proof of Proposition 3.3

Following the structure of the Proof of Proposition 3.1, I show that resource allocations according to (3.3) also constitute an equilibrium if the payoff is linear in the number of battles won.

State (i, j) **with** $i + j = N - 1$**:** See Proof of Proposition 3.1.

State (i, j) **with** $i + j = N - 2$**:**

Lemma 3.3. *In the last but one state, B maximises the number of battles won in the last two stages by choosing* $a_{i,j}$ *according to* (3.3)*.*

A formal proof is omitted.

State (i, j) **with** $i + j = N - k, k \in \{3, 4, \ldots N\}$**:** Assume that for any subsequent state (i', j') with $i' + j' > N - k$, both players allocate resources according to (3.3). Given these resource allocations, I show that:

Lemma 3.4. *If state* (i, j) *is even, choosing resources according to* (3.3) *is optimal for A and if state* (i, j) *is odd, choosing resources according to* (3.3) *is optimal for* B*.*

Proof. I use the same observation made in the Proof of Proposition 3.1: the total number of battles A wins in the states subsequent to state (i, j) is denoted by $X(a^{i,j} - a_{i,j})$, with properties described in equations (3.5) and (3.6). Now, consider the following cases:
First, assume $a^{i,j}, b^{i,j} \geq 1$ and state (i, j) is odd. Suppose A chooses $a_{i,j} = 1$,

implying A wins this battle and the next state is $(i+1, j)$ with remaining type a resources $a^{i+1,j} = a^{i,j} - 1$. A's total payoff is given by $i + 1 + X(a^{i,j} - 1)$. To show that the resource choice according to (3.3) is optimal, I need to show that A's payoff does not increase if A chooses a different resource allocation. If A deviates to $\hat{a}_{i,j} \in [0, 1)$, he looses battle (i, j) and the next stage is $(i, j+1)$ with remaining type a resources $a^{i,j-1} = a^{i,j} - \hat{a}_{i,j} \in (a^{i,j} - 1, a^{i,j}]$. Here, inequality (3.6) implies that with remaining resources $a^{i,j-1} = a^{i,j}$, the total number of battles A wins in subsequent states increases by at most 1. Hence, a deviation to $\hat{a}_{i,j} \in [0, 1)$ is not profitable. If A deviates to $\hat{a}_{i,j} \in (1, 2]$, he still wins battle (i, j), but the resources at the beginning of the next state $(i + 1, j)$ reduce compared to A choosing $a_{i,j} = 1$. This, by inequality (3.5), implies that the total number of battles A wins in subsequent states does not increase. Hence, this deviation is also not profitable.

Second, assume $a^{i,j}, b^{i,j} \geq 1$ and state (i, j) is even. By parallel reasoning, suppose B chooses $a_{i,j} = 1$, implying B wins this battle and the next state is $(i, j + 1)$ with resources $a^{i,j+1} = a^{i,j} - 1$. B's total payoff is given by $N - i - X(a^{i,j} - 1)$. To show that the resource choice according to (3.3) is optimal, I need to show that B's payoff does not increase if B chooses a different resource allocation: consider a deviation to $\hat{a}_{i,j} \in (1, 2]$. Then A wins the battle at stage (i, j) and the next stage is $(i + 1, j)$ with resources $a^{i+1,j} = a^{i,j} - \hat{a}_{i,j} \in [a^{i,j} - 2, a^{i,j} - 1)$. Here, inequality (3.6) implies that with remaining type a resources $a^{i+1,j} = a^{i,j} - 2$, the total number of battles A wins in subsequent states reduces by at most 1. Hence, a deviation to $\hat{a}_{i,j} \in (1, 2]$ is not profitable. Now consider a deviation to $\hat{a}_{i,j} \in [0, 1)$. While B still wins the battle at state (i, j), the type a resources at the beginning of the next state $(i, j + 1)$ increase compared to B choosing $a_{i,j} = 1$, which, by inequality (3.5) implies that the total number of battles A wins in subsequent states does not decrease. Hence, this deviation is also not profitable.

Finally, observe that if $a^{i,j} < 1$ or $b^{i,j} < 1$, any choice at state (i, j) results in the same player winning the battle at state (i, j) and all remaining battles (namely the player who still has a positive amount of resources left). Hence, choosing $a_{i,j}$ according to (3.3) is (weakly) optimal. □

I have shown that choosing resources according to (3.3) is optimal in the last but one state (i, j) with $i + j = N - 2$ and that, under the assumption that for any subsequent state (i', j') with $i' + j' > N - k$, both players allocate resources according to (3.3), choosing resources according to (3.3) at state (i, j) with $i' + j' = N - k$ is optimal for player A if (i, j) is even and is optimal for player B if (i, j) is odd. This allows to conclude the proof by induction.

References

Allingham, M. G. and Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, 1(3-4):323–338.

Alm, J. (2019). What motivates tax compliance? *Journal of Economic Surveys*, 33(2):353–388.

Alonso, R. and Câmara, O. (2016). Persuading voters. *American Economic Review*, 106(11):3590–3605.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of economic literature*, 36(2):818–860.

Balzer, B. and Schneider, J. (2019). Belief management and optimal arbitration. Working Paper. Uc3m, UTS.

Barbieri, S. and Serena, M. (2020). Fair representation in primaries: Heterogeneity and the new hampshire effect. Working Paper of the Max Planck Institute for Tax Law and Public Finance No. 2020-07.

BBC News (2019). French government to scan social media for tax cheats. Available at: `https://www.bbc.com/news/world-europe-50930094` (accessed 08.09.2022).

Bergemann, D. and Morris, S. (2019). Information design: A unified perspective. *Journal of economic literature*, 57(1):44–95.

Besley, T. and Preston, I. (2007). Electoral bias and policy choice: Theory and evidence. *The Quarterly Journal of Economics*, 122(4):1473–1510.

Bester, H. and Warneryd, K. (2006). Conflict and the social contract. *The Scandinavian Journal of Economics*, 108(2):231–249.

Bickerstaff, S. et al. (2020). *Election Systems and Gerrymandering Worldwide.* Springer.

Bierbrauer, F. and Polborn, M. (2022). Competitive fair redistricting. Working Paper. Vanderbilt University and University of Cologne.

Border, K. C. and Sobel, J. (1987). Samurai accountant: A theory of auditing and plunder. *The Review of Economic Studies*, 54(4):525–540.

Borel, E. (1921). La théorie du jeu les équations intégrales à noyau symétrique, Comptes Rendus l'Académie 173; English translation by Savage, L. (1953), The theory of play and integral equations with skew symmetric kernels. *Econometrica*, 21:97–100.

Borel, E. and Ville, J. (1938). Application de la théorie des probabilitiés aux jeux de hasard, Gauthier-Villars; reprinted in Borel, E., & Chéron, A.(1991). *Théorie mathematique du bridgea la portée de tous, Editions Jacques Gabay.*

Bracco, E. (2013). Optimal districting with endogenous party platforms. *Journal of Public Economics*, 104:1–13.

Brecher, Michael, Wilkenfeld, J., Beardsley, K., James, P., and Quinn, D. (2017). International crisis behavior data codebook, version 12. Available at: `http://sites.duke.edu/icbdata/data-collections/` (accessed 08.09.2022).

Brecher, M. and Wilkenfeld, J. (1997). *A Study of Crisis.* University of Michigan Press.

Bronsert, A.-K. (2016). *Marriage markets and tax compliance games.* PhD thesis, Ludwig-Maximilians-Universität München.

Chan, J., Gupta, S., Li, F., and Wang, Y. (2019). Pivotal persuasion. *Journal of Economic Theory*, 180:178–202.

Chen, Z. (2019). Information disclosure in contests: Private versus public signals. Working Paper. SSRN Electronic Journal.

Coate, S. and Knight, B. (2007). Socially optimal districting: A theoretical and empirical exploration. *The Quarterly Journal of Economics*, 122(4):1409–1471.

Crescenzi, M. J., Kadera, K. M., Mitchell, S. M., and Thyne, C. L. (2011). A supply side theory of mediation1. *International Studies Quarterly*, 55(4):1069–1094.

Ely, J. (2019). A cake-cutting solution to gerrymandering. Working Paper. Northwestern Universtiy.

Erard, B. and Feinstein, J. S. (1994). Honesty and evasion in the tax compliance game. *The RAND Journal of Economics*, 25(1):1–20.

Feng, X. and Lu, J. (2016). The optimal disclosure policy in contests with stochastic entry: A bayesian persuasion perspective. *Economics Letters*, 147:103–107.

Fey, M. and Ramsay, K. W. (2009). Mechanism design goes to war: Peaceful outcomes with interdependent and correlated types. *Review of Economic Design*, 13(3):233–250.

Fey, M. and Ramsay, K. W. (2010). When is shuttle diplomacy worth the commute? information sharing through mediation. *World Politics*, 62(04):529–560.

Fey, M. and Ramsay, K. W. (2011). Uncertainty and incentives in crisis bargaining: Game-free analysis of international conflict. *American Journal of Political Science*, 55(1):149–169.

Finkle, A. and Shin, D. (2007). Conducting inaccurate audits to commit to the audit policy. *International Journal of Industrial Organization*, 25(2):379–389.

Freedman, L. (2007). *The Official History of the Falklands Campaign: Volume 2: War and Diplomacy*. Routledge, London, revised. edition.

Friedman, J. N. and Holden, R. (2020). Optimal gerrymandering in a competitive environment. *Economic Theory Bulletin*, 8(2):347–367.

Friedman, J. N. and Holden, R. T. (2008). Optimal gerrymandering: Sometimes pack, but never crack. *American Economic Review*, 98(1):113–144.

Gelder, A. (2014). From custer to thermopylae: Last stand behavior in multistage contests. *Games and Economic Behavior*, 87:442–466.

Gilligan, T. W. and Matsusaka, J. G. (2006). Public choice principles of redistricting. *Public Choice*, 129(3):381–398.

Graetz, M. J., Reinganum, J. F., and Wilde, L. L. (1986). The tax compliance game: Toward an interactive theory of law enforcement. *Journal of Law, Economics & Organization*, 2(1).

Gul, F. and Pesendorfer, W. (2010). Strategic redistricting. *American Economic Review*, 100(4):1616–1641.

Hatfield, M. (2015). Taxation and surveillance: An agenda. *Yale Journal of Law & Technology*, 17:319.

Hennigs, R. (2021). Conflict prevention by bayesian persuasion. *Journal of Public Economic Theory*, 23(4):710–731.

Hörner, J., Morelli, M., and Squintani, F. (2015). Mediation and peace. *The Review of Economic Studies*, 82(4):1483–1501.

Houlder, V. (2017). Ten ways hrms can tell if you're a tax cheat: The authoritiy's latest battlefront in the war on evasion. Financial Times, Available at: `https://www.ft.com/content/0640f6ac-5ce9-11e7-9bc8-8055f264aa8b` (accessed 08.09.2022).

Houser, K. A. and Sanders, D. (2016). The use of big data analytics by the irs: Efficient solutions or the end of privacy as we know it. *Vanderbilt Journal of Entertainment & Technology Law*, 19(4):817.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics*, 11(1):249–272.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Klumpp, T., Konrad, K. A., and Solomon, A. (2019). The dynamics of majoritarian blotto games. *Games and Economic Behavior*, 117:402–419.

Klumpp, T. and Polborn, M. K. (2006). Primaries and the new hampshire effect. *Journal of Public Economics*, 90(6-7):1073–1114.

Kolotilin, A. and Wolitzky, A. (2020). The economics of partisan gerrymandering. UNSW Economics Working Paper No. 2020-12.

Konrad, K. A. (2018). Budget and effort choice in sequential colonel blotto campaigns. *CESifo Economic Studies*, 64(4):555–576.

Konrad, K. A. and Kovenock, D. (2009). Multi-battle contests. *Games and Economic Behavior*, 66(1):256–274.

Kovenock, D. and Roberson, B. (2012). Conflicts with multiple battlefields. In Garfinkel, M.R., S. S., editor, *The Oxford Handbook of the Economics of Peace and Conflict*. Oxford University Press, Oxford.

Kovenock, D. and Roberson, B. (2021). Generalizations of the general lotto and colonel blotto games. *Economic Theory*, 71(3):997–1032.

Kuchumova, Y. (2017). The optimal deterrence of tax evasion: The trade-off between information reporting and audits. *Journal of Public Economics*, 145:162–180.

Kvasov, D. (2007). Contests with limited resources. *Journal of Economic Theory*, 136(1):738–748.

Kydd, A. (2003). Which side are you on? bias, credibility, and mediation. *American Journal of Political Science*, 47(4):597–611.

Kydd, A. H. (2006). When can mediators build trust? *American Political Science Review*, 100(03):449–462.

Lagarde, A. and Tomala, T. (2021). Optimality and fairness of partisan gerrymandering. *Mathematical Programming*, pages 1–37.

Laipson, E., Kurtzer, D., and Moore, J. L. (1995). *Intelligence and the Middle East: What Do We Need To Know?* Washington Institute for Near East Policy.

Macho-Stadler, I. and Perez-Castrillo, J. D. (2002). Auditing with signals. *Economica*, 69(273):1–20.

Mathevet, L., Perego, J., and Taneva, I. (2020). On information design in games. *Journal of Political Economy*, 128(4):1370–1404.

Melumad, N. D. and Mookherjee, D. (1989). Delegation as commitment: The case of income tax audits. *The RAND Journal of Economics*, 20(2):139–163.

Mookherjee, D. and Png, I. (1989). Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics*, 104(2):399–415.

Owen, G. and Grofman, B. (1988). Optimal partisan gerrymandering. *Political Geography Quarterly*, 7(1):5–22.

Pegden, W., Procaccia, A. D., and Yu, D. (2017). A partisan districting protocol with provably nonpartisan outcomes. *arXiv preprint arXiv:1710.08781*.

Puppe, C. and Tasnádi, A. (2015). Axiomatic districting. *Social Choice and Welfare*, 44(1):31–50.

Rauchhaus, R. W. (2006). Asymmetric information, mediation, and conflict management. *World Politics*, 58(02):207–241.

Reinganum, J. F. and Wilde, L. L. (1985). Income tax compliance in a principal-agent framework. *Journal of Public Economics*, 26(1):1–18.

Reinganum, J. F. and Wilde, L. L. (1986a). Equilibrium verification and reporting policies in a model of tax compliance. *International Economic Review*, 27(3):739–760.

Reinganum, J. F. and Wilde, L. L. (1986b). Settlement, litigation, and the allocation of litigation costs. *The RAND Journal of Economics*, 17(4):557–566.

Roberson, B. (2006). The colonel blotto game. *Economic Theory*, 29(1):1–24.

Sánchez, I. and Sobel, J. (1993). Hierarchical design and enforcement of income tax policies. *Journal of Public Economics*, 50(3):345–369.

Sandmo, A. (2005). The theory of tax evasion: A retrospective view. *National tax journal*, 58(4):643–663.

Sansing, R. C. (1993). Information acquisition in a tax compliance game. *Accounting Review*, pages 874–884.

Scotchmer, S. (1987). Audit classes and tax enforcement policy. *The American Economic Review*, 77(2):229–233.

Sela, A. (2011). Best-of-three all-pay auctions. *Economics Letters*, 112(1):67–70.

Sherstyuk, K. (1998). How to gerrymander: A formal analysis. *Public Choice*, 95(1):27–49.

Shotts, K. W. (2001). The effect of majority-minority mandates on partisan gerrymandering. *American Journal of Political Science*, 45(1):120–135.

Smith, A. and Stam, A. (2003). Mediation and peacekeeping in a random walk model of civil and interstate war. *International Studies Review*, 5(4):115–135.

Taneva, I. (2019). Information design. *American Economic Journal: Microeconomics*, 11(4):151–85.

Yaniv, G. (2013). Tax evasion, conspicuous consumption, and the income tax rate. *Public Finance Review*, 41(3):302–316.

Yitzhaki, S. (1974). A note on income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3:201–202.

Zhang, J. and Zhou, J. (2016). Information disclosure in contests: A bayesian persuasion approach. *The Economic Journal*, 126(597):2197–2217.